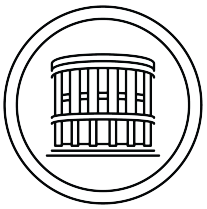
UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Vyhodnocovanie účinnosti nástrojov aktívnej politiky trhu práce na Slovensku

DIPLOMOVÁ PRÁCA

2025 Bc. Kristína Ivánová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Vyhodnocovanie účinnosti nástrojov aktívnej politiky trhu práce na Slovensku

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	Matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	Mgr. Miroslav Štefánik, PhD.

Bratislava 2025 **Bc. Kristína Ivánová**





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Kristína Ivánová
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Vyhodnocovanie účinnosti nástrojov aktívnej politiky trhu práce na Slovensku
Impact evaluation of Slovak active labour market policy measures

Anotácia: Slovenský trh práce čoraz intenzívnejšie pociťuje nedostatok pracovnej sily. Investovanie do aktivácie a rekvalifikácie disponibilných uchádzačov o zamestnanie, tak nadobúda na význame. Aj prostriedky smerujúce do aktívnej politiky trhu práce (APTP) sú obmedzeným zdrojom, ich efektívna alokácia je preto výzvou ekonomického výskumu. Vďaka dostupným administratívnym dátam je možné merať účinnosť jednotlivých nástrojov APTP relatívne spoľahlivými metódami. Empirická mikroekonómia v tejto oblasti sa pomerne dynamicky rozvíja; či už ide o metódy založené na rekonštrukcií vedeckého experimentu párovaním účastníkov s oprávnenými uchádzačmi, metódy spracovania panelových dát, či ďalšie typy kauzálnej analýzy. Úlohou diplomanta/diplomantky bude popísať intervenčnú logiku poskytovania nástrojov APTP na Slovensku a vyhodnotiť účinnosť vybraného nástroja APTP na Slovensku aplikovaním najnovších techník kontrafaktuálneho vyhodnotenia dopadov.

Vedúci: Mgr. Miroslav Štefánik, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: doc. Mgr. Igor Melicherčík, PhD.

Dátum zadania: 11.01.2024

Dátum schválenia: 25.01.2024

prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Poďakovanie Touto cestou by som sa chcela veľmi pekne poďakovať svojmu vedúcemu diplomovej práce pánovi Mgr. Miroslavovi Štefánikovi, PhD. za veľmi seriózne, vysokoodborné vedenie mojej práce, odborné rady a hodnotné pripomienky pri riešení mojej diplomovej práce. Poskytol mi expertný pohľad na problematiku a množstvo praktických skúseností, čo mi pomohlo pri realizovaní práce ale aj osobnom raste.

Abstrakt v štátnom jazyku

IVÁNOVÁ, Kristína: Vyhodnocovanie účinnosti nástrojov aktívnej politiky trhu práce na Slovensku [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Miroslav Štefánik, PhD., Bratislava, 2025, 72 strán

Moja diplomová práca sa zaoberá matematickými, štatistickými a analytickými postupmi v súvislostiach s problematikou trhu práce v kontexte implementovaných nástrojov aktívnych politík. Zameriava sa na vyhodnocovanie účinnosti nástrojov politiky trhu práce prostredníctvom kontrafaktuálnej analýzy. Hlavným cieľom diplomovej práce bolo vykonanie analýzy kauzálnych účinkov účasti v programe APTP na zamestnanosť na trhu práce na Slovensku. Analýza bola uskutočnená na dátovej vzorke vygenerovaných dát založených na skutočných anonymizovaných administratívnych dátach z Úradu práce, sociálnych vecí a rodiny Slovenskej republiky. Kauzálnej analýze predchádzalo oboznámenie sa s problematikou APTP, vyhľadanie a spoznanie matematických nástrojov pre zisťovanie kauzálnych účinkov, vymedzenie náplne praktickej časti a premyslenie si postupu riešenia. V diplomovej práci je dokumentovaný celý proces od úvodu do problematiky cez realizáciu riešenia a zabezpečenie výstupov až po interpretáciu výsledkov. Prínosom práce je popis metodiky pre vykonanie analýzy metódou párovania v danom dátovom kontexte, porovnanie rôznych estimačných špecifikácií využívaných ako vstupné údaje do modelov a vyhodnotenie vplyvu zvoleného nástroja REPAS pre rôzne časové obdobia využitím prístupov PSM a IPW.

Kľúčové slová: kauzálna analýza, aktívna politika trhu práce, Rubinov kauzálny model, metóda párovania na základe skóre pravdepodobnosti intervencie, metóda váženia na základe inverznej pravdepodobnosti intervencie

Abstract

IVÁNOVÁ, Kristína: Impact evaluation of Slovak active labour market policy measures [Masters Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Miroslav Štefánik, PhD., Bratislava, 2025, 72 pages

My master's thesis deals with mathematical, statistical and analytical procedures in connection with labor market issues in the context of implemented active policy instruments. It focuses on evaluating the effectiveness of labor market policy instruments through counterfactual analysis. The main objective of the thesis was to conduct an analysis of the causal effects of participation in an ALMP program on employment in the labor market of Slovakia. The analysis was carried out on a data sample of generated data based on real anonymized administrative data from the Ministry of Labor, Social Affairs and Family of the Slovak Republic. The causal analysis was preceded by familiarization with the ALMP issues, researching and learning about mathematical tools for determining causal effects, defining the scope of the practical part and thinking through the solution procedure. The master's thesis documents the entire process from the introduction to the issue through the implementation of the solution and ensuring the outputs to the interpretation of the results. The contribution of the work is the description of the methodology for performing the matching analysis in a given data context, the comparison of different estimation specifications used as input data into the models, and the evaluation of the impact of the chosen REPAS instrument for different time periods using the PSM and IPW approaches.

Keywords: causal analysis, active labour market policies, Rubin causal model, propensity score matching, inverse probability weighting

Obsah

Zoznam obrázkov **9**

Zoznam tabuliek **10**

Zoznam použitých skratiek a symbolov **11**

Úvod **12**

1 Teoretická časť **14**

1.1 Úvod do problematiky aktívnej politiky na trhu práce (APTP) 14

1.1.1 Definícia pojmov a vzťahov v súvislostiach APTP 16

1.1.2 Definovanie nástrojov v súvislostiach APTP 17

A Nástroje monitoringu v súvislostiach PTP 17

B Nástroje aktívnej politiky na trhu práce 18

C Nástroje vyhodnocovania účinnosti 19

1.2 Matematické nástroje pre vyhodnocovanie účinkov APTP 19

1.2.1 Metódy vyhodnocovania účinnosti nástrojov APTP 19

1.2.2 Matematické nástroje pre zisťovanie kauzálnych účinkov APTP . . 20

1.2.3 Rubinov kauzálny model alebo teória potenciálnych výsledkov . . . 21

1.2.4 Metóda párovania 24

A Predpoklady pre využitie metódy párovania 25

B Kroky realizácie párovania 26

2 Praktická časť **38**

2.1 Situačné priblíženie oblasti praktickej časti a formulácia cieľov práce 38

2.2 Predstavenie náplne praktickej časti 40

2.2.1 Zámer praktickej časti 40

2.2.2 Predmet riešenia 40

2.2.3 Očakávané prínosy 41

2.3 Postup riešenia 41

2.3.1 Výber dátovej vzorky 41

2.3.2 Výber metodiky 42

2.3.3 Programovanie algoritmu 42

2.4 Realizácia riešenia 42

2.4.1 Voľba konkrétnej metodiky 43

2.4.2 Práca s dátovou vzorkou 43
A Tabuľka účasti v programoch 44
B Tabuľka informácie o osobách z evidencie nezamestnanosti . . . 44
C Tabuľka dátumy evidencie 44
2.4.3 Popis fungovania častí algoritmického spracovania analýzy metódou párovania 44
A Časť plánovanie a príprava algoritmu 45
B Časť implementácia metódy párovania 46
2.4.4 Voľba súboru vstupných parametrov, charakteristík a špecifikácií . . 47
A Zvolené vstupné parametre 47
B Zvolené úpravy pre súbor vstupných charakteristík 47
C Zvolené špecifikácie parametrov metódy párovania na základe skóre pravdepodobnosti intervencie 48
D Ďalšie zvolené evaluačné obdobia 49
2.5 Výsledky riešenia a ich interpretácia 50
2.5.1 Porovnávacia tabuľka 50
2.5.2 Porovnanie vyrovnanosti sád charakteristík 51
2.5.3 Numerické porovnanie modelov párovania na základe skóre pravdepodobnosti intervencie 55
2.5.4 Grafické porovnanie modelov párovania na základe skóre pravdepodobnosti intervencie 56
2.5.5 Odhad účinku ATT využitím PSM 58
2.5.6 Porovnanie odhadu účinku ATT využitím PSM a IPW 60
2.5.7 Odhad účinku ATT pre ďalšie evaluačné obdobia 61

3 Zhrnutie **63**

Záver **65**

Zoznam použitej literatúry **66**

Prílohy **70**

Zoznam obrázkov

1 Európsky pilier sociálnych práv v 20 princípoch. Zdroj: [10] 14
2 Mindmapa nástrojov politiky trhu práce spracovaná v [36]. Zdroj: [25] a [37] 18
3 Členenie oblastí kauzálnej analýzy. Zdroj: [24] 20
4 Algoritmické vyjadrenie krokov implementácie analýzy párovania v [36] . . 27
5 Grafické znázornenie distribúcie skóre pravdepodobnosti intervencie po využití párovania 1:1 metódou najbližšieho suseda. Zdroj: [32] 35
6 Príklad grafického znázornenia vyrovnanosti kovariátov využitím plného párovania. Zdroj: [19] 36
7 Grafické znázornenie vyrovnanosti kovariátov modelu m.2. 57
8 Grafické znázornenie vyrovnanosti kovariátov modelu m.3. 57
9 Porovnanie hustoty vybraných kovariátov pred a po využití metódy párovania. 58
10 Odhad ATT a smerodajnej odchýlky získaný metódou PSM pre evaluačné obdobie 2016. 59
11 Podiel jednotlivcov mimo evidencie nezamestnaných (proxy miery zamestnanosti) pre evaluačné obdobie 2016. 60
12 Porovnanie hodnôt ATT odhadnutý prístupmi PSM a IPW pre evaluačné obdobie 2016. 61
13 Porovnanie hodnôt ATT odhadnutý prístupmi PSM a IPW pre evaluačné obdobie 2017. 62
14 Porovnanie hodnôt ATT odhadnutý prístupmi PSM a IPW pre evaluačné obdobie 2018. 62

Zoznam tabuliek

1 Porovnanie účastníkov a kontrolnej skupiny pre vybrané charakteristiky . . 50
2 Sumár porovnania vyrovnanosti parametrov pri použití metódy párovania (prvý súbor charakteristík) 52
3 Sumár porovnania vyrovnanosti parametrov pri použití metódy párovania (druhý súbor charakteristík) 53
4 Sumár porovnania vyrovnanosti parametrov pri použití metódy párovania (tretí súbor charakteristík) 54
5 Prehľad hodnôt merateľných ukazovateľov pre jednotlivé modely 55

Zoznam použitých skratiek a symbolov

APTP	aktívna politika trhu práce
ATE	average treatment effect
ATT	average treatment effect on the treated group
ATU	average treatment effect on the untreated group
EU	Európska únia
IPW	inverse probability weighting
OECD	Organizácia pre hospodársku spoluprácu a rozvoj
PPTP	pasívna politika trhu práce
PSM	propensity score matching
PTP	politika trhu práce
REPAS	Rekvalifikácia, Príležitosť a Spolupráca
SDO	simple difference in means
SMD	standardized mean difference
SUTVA	stable unit treatment value assumption
TP	trh práce
UPSVaR	Úrad práce, sociálnych vecí a rodiny Slovenskej republiky

Úvod

Od lovcov a zberačov cez agrárnu spoločnosť až po industriálnu revolúciu sa nad celým historickým vývojom ľudstva vznášalo slovo práca. Rozvoj spoločnosti. Pocit užitočnosti. Finančná kompenzácia za splnenie úloh. Benefity, ktoré výkonný a fungujúci trh práce dokáže zabezpečiť. Možno aj z tohto dôvodu cyklicky medializované slovo nezamestnanosť zanecháva v ústach mnohých ľudí trpkú chuť.

Európa posledné dva veľké vrcholy zažila počas veľkej krízy v roku 2008 a počas obdobia Covid-19 v rokoch 2019 – 2021. Na základe informácií z Eurostatu [15] nezamestnanosť v eurozóne v tomto období dosahovala hodnoty okolo 12,3 % a počas vrcholu pandémie 8,5 %. Vysoká nezamestnanosť a ani hodnoty limitne sa približujúce k nule nie sú pre ekonomiku zdravé. Friedman a Phelps [7] tvrdia, že v ekonomike existuje prirodzená miera nezamestnanosti, ktorá by mala byť dlhodobo udržaná. Takáto hodnota zabezpečí efektívny chod ekonomiky a poskytne dostatočné vyjednávacie schopnosti zamestnancom aj zamestnávateľom.

Európska únia si túto hodnotu drží v hľadáčiku a využíva rôzne nástroje pre korigovanie úrovne nezamestnanosti, ktoré ako celok môžeme nazvať politika trhu práce. Politika trhu práce predstavuje komplexnú oblasť s mnohými vytýčenými cieľmi na trhu práce. Pre jej lepšie porozumenie a uchopenie je vhodné pristúpiť k určitému členeniu. Jedným z členení politiky trhu práce je na pasívnu politiku trhu práce (PPTP) a aktívnu politiku trhu práce (APTP). K obom politikám prislúchajú adekvátne súbory nástrojov. V našej práci sledujeme nástroje APTP, ktorých účelom je zabezpečiť rast pracovných príležitostí a zníženie nezamestnanosti. Primárnym zameraním je zabezpečenie efektívnej alokácie finančných prostriedkov z Európskej únie do predpísaného súboru programov a opatrení [37] na národnej úrovni. Ako ale rozhodnúť, či práve účasti v programe môžeme pripísať zlepšené zamestnanecké prognózy účastníka a nepôjde iba o náhodu alebo vplyv iných viditeľných alebo neviditeľných faktorov?

Práca sledovala viacero cieľov, pričom hlavným bolo vykonanie analýzy kauzálnych účinkov účasti v programe REPAS (séria rekvalifikačných kurzov) na zamestnanosť na Slovenskom trhu práce pre rôzne evaluačné obdobia. Za parciálne ciele sme určili voľbu a popis metodiky pre vykonanie analýzy, aplikovanie metodiky na dátovej vzorke a porovnanie rôznych vstupných špecifikácií pre využívané modely.

Prácu sme začali oboznámením sa s problematikou APTP, objasnením si matematických nástrojov pre zisťovanie kauzálnych účinkov, vymedzením náplne praktickej časti a premyslením si postupu riešenia. Pre splnenie vytýčených cieľov využijeme metódu párovania a s ňou spojenú analýzu pomocou metódy párovania. Za typ účinku pre odhad efektivity účasti v programe volíme účinok intervencie pre účastníkov (ATT). Pracujeme s dátovou vzorkou vygenerovaných dát na základe skutočných anonymizovaných administratívnych dát z Úradu práce, sociálnych vecí a rodiny Slovenskej republiky (ÚPSVaR).

Diplomová práca je štrukturovaná a členená do viacerých kapitol a podkapitol.

V prvej kapitole predstavíme teoretickú bázu diplomovej práce, ktorá bude pojednávať o úvode do problematiky APTP a matematických nástrojoch využívaných pre zisťovanie kauzálnych účinkov APTP.

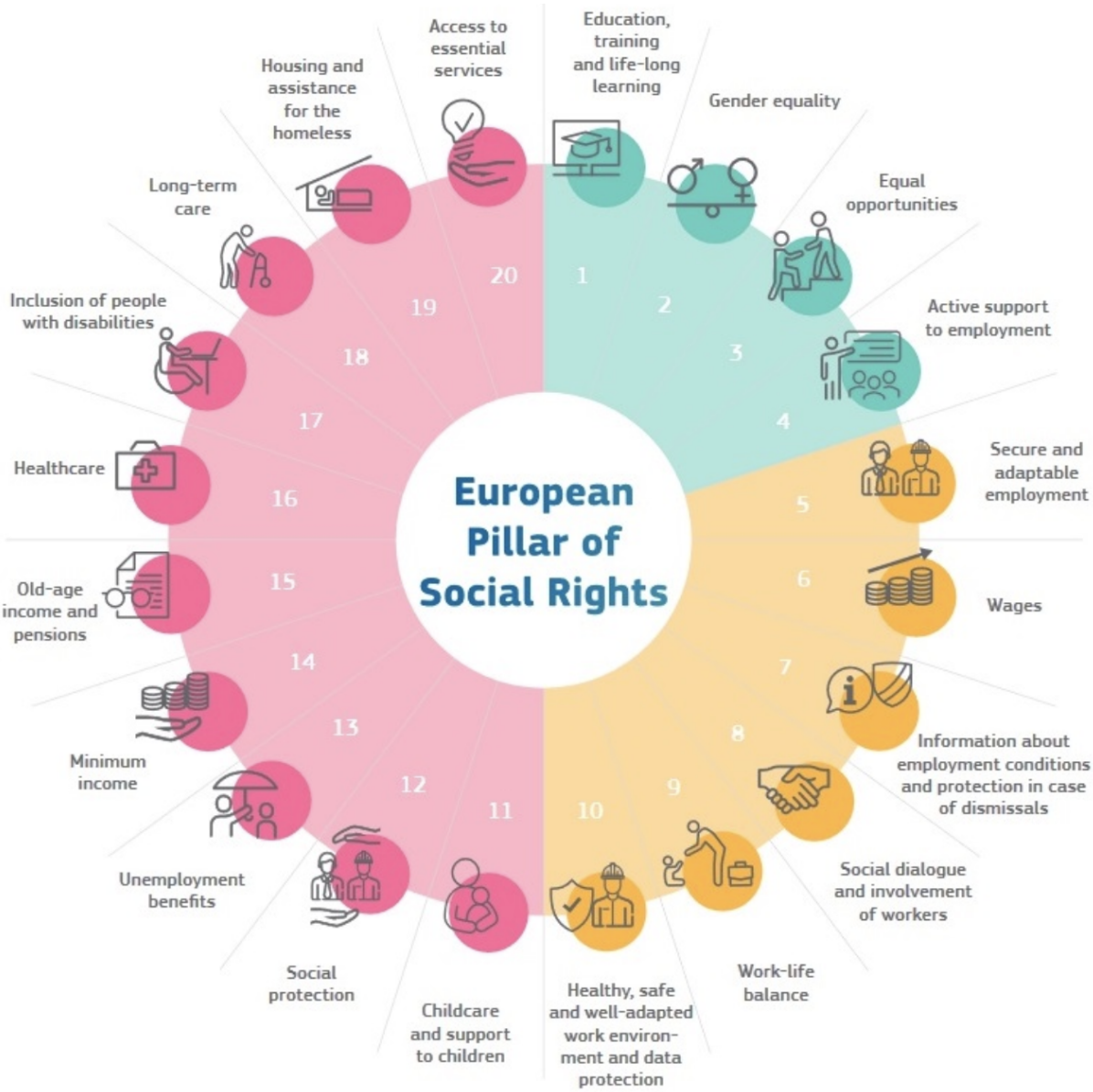
Druhá kapitola sa zaoberá praktickou časťou a situačne v nej priblížime oblasť praktickej časti, predstavíme náplň praktickej časti, vytýčime plán pre postup riešenia, uvedieme najdôležitejšie kroky realizácie riešenia a získané výsledky interpretujeme.

V tretej kapitole uvádzame zhrnutie najdôležitejších informácií a poučení získaných počas tvorby diplomovej práce.

1 Teoretická časť

1.1 Úvod do problematiky aktívnej politiky na trhu práce (APTP)

Slovensko sa vstupom do Európskej únie (EÚ) v roku 2004 stalo členom silného spoločenstva štátov. Rovnako sa zaviazalo k nasledovaniu práv, princípov a cieľov vytýčených EÚ a jej členskými štátmi. Okrem základných ľudských práv akými sú sloboda, demokracia a rovnosť členíme práva aj na sociálne práva spojené so zamestnanosťou a sociálnym začlenením obyvateľstva. Európsky pilier sociálnych práv, ktorý je zhrnutý na Obrázku 1 sumarizuje tieto práva do 20 princípov.



Obrázok 1: Európsky pilier sociálnych práv v 20 princípoch. Zdroj: [10]

Princípy sú rozdelené do troch podskupín alebo kapitol podľa zamerania. Modrý súbor princípov sa zaoberá vyrovnanými príležitosťami a prístupu na trh práce, žltá sa venuje spravodlivým pracovným podmienkam a ružová sa zameriava na sociálnu ochranu a integráciu do spoločnosti.

Z dôvodu efektívnejšieho dodržiavania princípov a zabezpečenia zvyšujúceho sa prospechu občanov v rôznych oblastiach fungovania si EÚ stanovuje 10-ročné plány pre zabezpečenie tohto efektu. Akčný plán na realizáciu Európskeho piliera sociálnych práv [11] premieňa sociálne princípy na konkrétne činnosti s merateľnými ukazovateľmi pre jednoduchšie vyhodnotenie splnenia cieľa. Aktuálne činnosti s cieľmi stanovené pre splnenie do roku 2030 sú dosiahnuť 78 %-nú zamestnanosť populácie vo veku 20 až 64 rokov, 60 %-nú každoročnú účasť dospelých na školení a znížiť počet ľudí ohrozených chudobou alebo sociálnym vylúčením aspoň o 15 miliónov. Na základe informácií z Eurostatu [14], [12], [13] v čase písania diplomovej práce dosahovala zamestnanosť populácie vo veku 20 až 64 rokov 75,9 % vo štvrtom kvartáli roku 2024, 47 % dospelých vo veku 25 - 64 sa zúčastnilo vzdelávacích a tréningových programov v roku 2022 a v čase prechodu na nový desaťročný plán hrozilo ohrozenie chudobou alebo sociálnym vylúčením každému piatemu občanovi EÚ. Celkovo sa jednalo o 96,5 milióna obyvateľov.

K dosiahnutiu cieľov do stanoveného termínu je nutné zamerať sa na celé vekové spektrum v produktívnom veku. Najmä na kritické skupiny na chvostoch tohto spektra a dlhodobo nezamestnaných obyvateľov. Pre túto skupinu obyvateľov turbulentné obdobie posledných rokov spojené s pandémiou Covid-19 a digitálnou a ekonomickou transformáciou trhu práce spôsobilo prehlbujúce sa nerovnosti medzi požiadavkami zamestnávateľa a schopnosťami a znalosťami uchádzačov o zamestnanie. Z tohto dôvodu nie je možné spoľahnúť sa iba na prirodzený dopyt a ponuku na trhu práce, ale je nutné využiť aj štátne reformy politiky trhu práce. Ich hlavným zameraním je riešenie nevyhovujúcich podmienok na trhu práce pomocou fiškálnych výdavkov do nástrojov PPTP a APTP. Ďalšia podkapitola pojednáva o najdôležitejších pojmoch v oblasti politiky trhu práce a poskytuje definície z relevantných zdrojov pre lepšie porozumenie kľúčových pojmov.

1.1.1 Definícia pojmov a vzťahov v súvislostiach APTP

Človek a jeho potreby, t. j. ľudské potreby, sú základné požiadavky, ktoré musí človek uspokojiť, aby mohol viesť plnohodnotný život. Psychológ Abraham Maslow [23] vytvoril hierarchiu potrieb, ktorá rozdeľuje ľudské potreby do piatich úrovní. K ich uspokojeniu je mnohokrát potrebné zabezpečiť relevantné tovary a služby.

Práca môže mať mnoho významov, ale v základnom slova zmysle ide o činnosť, ktorú vykonávame s cieľom dosiahnuť určitý výsledok. Môže ísť o fyzickú alebo duševnú aktivitu, za ktorú často dostávame odmenu, napríklad mzdu. Práca v ekonómii je jeden z výrobných faktorov. Je definovaná ako cieľavedomá ľudská činnosť, zameraná na tvorbu statkov a služieb, ktoré uspokojujú ľudské potreby. Jedným z najdôležitejších charakteristík práce je produktivita práce.

Trh práce je miesto, kde sa stretáva ponuka pracovných síl s dopytom po nich. Cambridgeský slovník [5] definuje trh práce ako ponuku ľudí v konkrétnej krajine alebo oblasti, ktorí sú schopní a ochotní pracovať, najmä vo vzťahu k počtu pracovných miest, ktoré sú k dispozícii. V prípade zhody ekonomicko aktívneho obyvateľstva a pracovnej pozície dochádza k nadviazaniu pracovného vzťahu a osoba sa označuje ako pracujúci. V opačnom prípade pôjde o nezamestnanú osobu.

Trh práce čelí viacerým výzvam, ktoré ovplyvňujú zamestnanosť a ekonomiku. Môžeme ich označiť pojmom **problémy na trhu práce**. Na trhu práce existuje viacero typov problémov, ktoré ovplyvňujú zamestnanosť a pracovné podmienky. Medzi najčastejšie patria nezamestnanosť, nízke mzdy, nedostatok kvalifikovanej pracovnej sily, automatizácia a digitalizácia, demografické zmeny a pracovné podmienky. Primárny problém v kontexte národnej ako aj európskej stratégie predstavuje **nezamestnanosť**.

Politika trhu práce je komplexná aktivita, ktorá zahŕňa opatrenia a stratégie, ktoré vlády a inštitúcie prijímajú na riešenie rôznych problémov trhu práce, ako sú regulácia zamestnanosti, nezamestnanosti a pracovných podmienok. Politika trhu práce podľa Wrighta [35] označuje súbor nástrojov, prostredníctvom ktorých vlády pomáhajú jednotlivcom, ktorí čelia ťažkostiam na trhu práce, najmä, ale nie výlučne, nezamestnaným ľuďom. Rozsah a povaha vládnych aktivít v tejto oblasti sa medzi rozvinutými krajinami značne líši a postupom času vyvíja. Politiku trhu práce delíme do dvoch podskupín - **pasívnu politiku** a **aktívnu politiku**.

PPTP zahŕňa súbor programov zameraných na udržanie príjmu nezamestnaných. APTP zahŕňa súbor programov zameraných na zvýšenie počtu pracovných príležitostí pre uchádzačov o zamestnanie rovnako ako aj lepšie zosúladenie voľných pracovných miest a dostupných pracovníkov. Pre lepšie priblíženie pojmu OECD na svojej webovej stránke [26] uvádza, že APTP zahŕňa poskytovanie služieb zamestnanosti s cieľom inšpirovať uchádzačov o zamestnanie, zlepšiť si zručnosti, pomôcť zamestnávateľom splniť ich kvalifikačné potreby a vytvoriť pracovné príležitosti. Na dosiahnutie všetkých týchto cieľov musia byť nástroje APTP dobre navrhnuté, cielené a nákladovo efektívne.

Na základe druhej časti zákona č. 5/2004 Z.z. o službách zamestnanosti [37] definujeme dôležité cieľové skupiny v rámci oblasti APTP: **uchádzači o zamestnanie** a **záujemcovia o zamestnanie**.

Podľa § 6 je uchádzač o zamestnanie občan, ktorý môže pracovať, chce pracovať, hľadá si zamestnanie a je vedený v evidencii uchádzačov o zamestnanie, a ktorý nie je zamestnanec, nie je v pracovnoprávnom vzťahu na základe dohody o práci vykonávanej mimo pracovného pomeru, neprevádzkuje alebo nevykonáva samostatnú zárobkovú činnosť, nevykonáva zárobkovú činnosť v členskom štáte Európskej únie alebo v cudzine. Zjednodušene, pod pojmom uchádzač rozumieme nezamestnanú osobu evidovanú na ÚPSVaR, ktorá má záujem nájsť si prácu.

Podľa § 7 je záujemca o zamestnanie občan, ktorý si hľadá iné zamestnanie, alebo ktorý má záujem o poskytovanie informačných a poradenských služieb a odborných poradenských služieb a nie je uchádzačom o zamestnanie.

1.1.2 Definovanie nástrojov v súvislostiach APTP

APTP používa niekoľko nástrojov, postupov a metodík pre realizáciu svojich úloh a cieľov. My sme sa zaoberali hlavne nástrojmi patriacimi do nasledujúcich skupín:

A Nástroje monitoringu v súvislostiach PTP

Monitoring zahŕňa udržiavanie aktuálnej bázy informácií o potenciálnych prijímateľoch nástrojov APTP uloženej v **evidencii uchádzačov o zamestnanie** a **evidencii záujemcov o zamestnanie**.

Evidenciu uchádzačov o zamestnanie definuje § 33 ods. 1 ako zoznam obsahujúci meno a priezvisko, dátum narodenia, adresu trvalého pobytu uchádzača o zamestnanie,

údaje o jeho kvalifikácii, odborných zručnostiach, o druhu doteraz vykonávanej práce, informácie o jeho záujme o určité zamestnania, údaje o jeho osobných a rodinných pomeroch súvisiacich so sprostredkovaním zamestnania, údaje o jeho zdravotnom stave oznámené uchádzačom o zamestnanie dobrovoľne.

Evidenciu záujemcov o zamestnanie definuje § 37 ods. 1 ako zoznam obsahujúci potrebné osobné údaje o záujemcovi o zamestnanie, najmä údaje o jeho kvalifikácii, odborných zručnostiach, získaných pracovných skúsenostiach v doterajších zamestnaniach, informácie o jeho záujme o určité zamestnania a ďalšie súvisiace údaje.

B Nástroje aktívnej politiky na trhu práce

APTP zahŕňa rôzne nástroje na podporu zamestnanosti a zlepšenie pracovných podmienok. O nástrojoch aktívnej politiky na trhu práce pojednáva siedma a ôsma časť zákona č. 5/2004 Z.z. o službách zamestnanosti [37]. Členenie politiky trhu práce spolu s prehľadom vybraných nástrojov je viditeľné na Obrázku 2.



Obrázok 2: Mindmapa nástrojov politiky trhu práce spracovaná v [36]. Zdroj: [25] a [37]

V zmysle spomínaného zákona sú ku dňu 15.4.2025 na webovej stránke ÚPSVaR [33] vypísané nasledujúce nástroje aktívnej politiky na trhu práce:

- Poskytovanie príspevkov na podporu rekvalifikácie uchádzačov o zamestnanie v rámci národného projektu Zručnosti pre trh práce.
- Poskytovanie príspevkov na podporu vzdelávania záujemcov o zamestnanie v rámci národného projektu Zručnosti pre trh práce.

- Príspevok na aktivačnú činnosť formou menších obecných služieb pre obec alebo menších služieb pre samosprávny kraj - § 52.
- Príspevky pre občana.
- Príspevky pre zamestnávateľa.
- Podpora zamestnávania občanov so zdravotným postihnutím.
- Vzdelávanie a príprava pre trh práce.

C Nástroje vyhodnocovania účinnosti

Existuje množstvo nástrojov na vyhodnocovanie účinnosti, ktoré pomáhajú merať výkonnosť nástrojov a procesov a umožňujú identifikovať oblasti na zlepšenie. Bližšie si ich popíšeme v ďalšej časti.

1.2 Matematické nástroje pre vyhodnocovanie účinkov APTP

Matematika je študijný odbor, ktorý nám dovoľuje ísť cez hranice našej intuície a umožňuje nám objavovať a riešiť zaujímavé problematiky využitím štruktúrovanej logiky. V tejto podkapitole bližšie predstavíme matematické pozadie a súvislosti oblasti kauzálnych účinkov, ktorá predstavuje spôsob vysporiadania sa s vyhodnocovaním účinnosti.

1.2.1 Metódy vyhodnocovania účinnosti nástrojov APTP

Po samotnej účasti v programe je kľúčovou aktivitou vyhodnotenie účinnosti nástrojov. Ideálny vyhodnocovací proces je na základe [16] založený na 3 krokoch:

1. Mikroekonomické vyhodnotenie: Vplyv účasti v programe na jednotlivca.
2. Makroekonomické vyhodnotenie: Je vplyv dostatočne veľký, aby bol dosiahnutý spoločenský zisk.
3. Analýza nákladov a výnosov: Bol dosiahnutý výsledok najlepší možný pri danom množstve alokovaných financií.

V diplomovej práci sme sa venovali mikroekonomickému vyhodnoteniu, ktoré je založené na matematickej teórii kauzálnych účinkov.

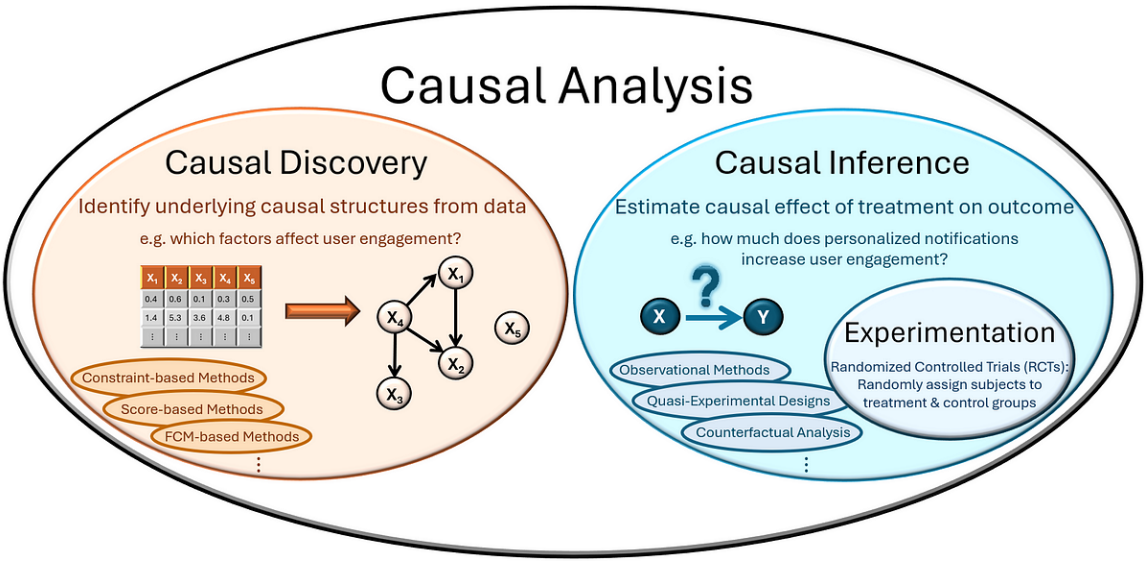
1.2.2 Matematické nástroje pre zisťovanie kauzálnych účinkov APTP

Kauzálne účinky sú podmnožinou kauzálnej analýzy, jednej z oblastí experimentálnej štatistiky, ktorej hlavným zameraním je stanovenie a dokázanie vzťahu príčiny a následku. Kauzálnu analýzu okrem kauzálnych účinkov členíme aj na kauzálne objavy. Oba členy sa zaoberajú rôznymi kauzálnymi otázkami. Huber [21] tieto podskupiny vo svojom diele definuje nasledovne:

Definícia 1.1 *Kauzálne objavy majú za cieľ naučiť sa kauzálne vzťahy medzi niekoľkými alebo dokonca mnohými premennými spôsobom založeným na údajoch. Inými slovami, kauzálny objav sa snaží pochopiť alebo odhaliť kauzálne asociácie v rámci celého systému premenných, ktoré možno znázorniť kauzálnym grafom.*

Definícia 1.2 *Kauzálne účinky, tiež známe ako liečba, účasť v programe alebo hodnotenie účinku, sa prevažne zameriavajú na hodnotenie kauzálneho účinku špecifickej premennej, ako je účasť vo vzdelávacom programe, podstúpenie zdravotnej liečby alebo marketingová intervencia a na výsledok záujmu, akým je zárobok, zdravie alebo predaj.*

Členenie kauzálnej analýzy s kľúčovými myšlienkami pre podskupiny je viditeľné na Obrázku 3.



Obrázok 3: Členenie oblastí kauzálnej analýzy. Zdroj: [24]

Na základe priblíženia problematiky APTP a vyššie uvedených definícií sme považovali za relevantné hlbšie sa ponoriť práve do druhej podskupiny, t. j. teórie kauzálnych účinkov.

V ďalších častiach práce sme sa najskôr zamerali na priblíženie Rubinovho kauzálneho modelu. Model predstavuje úvodný teoretický koncept založený na vyhodnocovaní účinkov metódou zdravého rozumu. Ďalej sme vysvetlili jeho nedostatky prichádzajúce s aplikáciou na skutočných dátových vzorkách. Nakoniec sme predstavili a vysvetlili fungovanie metódy, ktorá pri splnení setu predpokladov dokáže nedostatky Rubinovho modelu kompenzovať.

1.2.3 Rubinov kauzálny model alebo teória potenciálnych výsledkov

Teória okolo potenciálnych výsledkov je jedným zo základných konceptov využívaných v analýze kauzálnych účinkov. Predstavuje veľmi silný matematický základ pre definovanie a odhadovanie kauzálnych účinkov. Základy tohto prístupu pochádzajú z diela [31], kde sa autor zaoberal poľnohospodárskymi experimentami. Formalizáciou si prešiel v diele [30], kde bol uplatnený v sociálnych a behaviorálnych vedách. Potenciálne výsledky sú jadrom Rubinovho kauzálneho modelu.

Model je založený na základnej myšlienke, že pre akéhokoľvek jedinca existujú dva potenciálne výsledky, ktoré zodpovedajú kladnému alebo zápornému rozhodnutiu. Príkladom môže byť užitie alebo neužitie tabletky proti bolesti hlavy. Nástraha spočíva v tom, že je možné nadobudnúť práve jeden z týchto dvoch výsledkov alebo stavov. Nadobudnutý výsledok označíme ako pozorovaný výsledok.

Na základe [8] môžeme rovnicu pre pozorovaný výsledok i-tého jedinca Y_i vyjadriť v nasledujúcom tvare:

$$Y_i = D_i \cdot Y_i^1 + (1 - D_i) \cdot Y_i^0, \tag{1}$$

kde D_i je binárna premenná, ktorá nadobúda hodnoty 0, ak i-tý jedinec neprijal účasť a 1, ak i-tý jedinec prijal účasť. V prípade, že $D_i = 1$, potom $Y_i = Y_i^1$ a pozorovaný výsledok je rovný potenciálnemu výsledku v stave sveta, kde účasť i-tého jedinca nastala. V opačnom prípade sa $D_i = 0$, t. j. pozorovaný výsledok korešponduje stavu sveta, kde účasť i-tého jedinca nenastala. Rovnica (1) je v literatúre známa aj ako prepínacia rovnica.
Na základe tohto značenia môžeme definovať účinok účasti pre jednotlivca alebo kauzálny účinok nasledovne

$$\delta_i = Y_i^1 - Y_i^0, \tag{2}$$

kde Y_i^1 a Y_i^0 predstavujú potenciálne stavy jedinca. V rovniciach (1) a (2) je predstavený základný problém analýzy kauzálnych účinkov spojený s identifikáciou kontrafaktuality, teda hypotetického stavu, nad ktorým môžeme v reálnom svete iba uvažovať a nie ho jednoznačne vyjadriť. Jedným z riešení pre vysporiadanie sa s kontrafaktualitou sú náhodné experimenty, ktoré spočívajú v náhodnom priradení jedincov do účastníckej a kontrolnej skupiny. V mnohých prípadoch je ale tento prístup nekorektný. Ako príklad uvádzame priradenie študentov na štúdium vysokej školy, ktoré je nenáhodné a závisí hlavne od znalostnej a finančnej bázy jednotlivcov.

Účinok účasti pre jednotlivca z rovnice (2) sa ďalej využíva pre vyčíslenie troch základných typov účinku pre vyhodnocovanie kauzálnych účinkov. Menovite medzi ne patria priemerný účinok intervencie ATE, priemerný účinok intervencie pre účastníkov ATT a priemerný účinok intervencie pre kontrolnú skupinu ATU. Cunningham [8] ich vyjadruje nasledovnými rovnicami:

$$\begin{aligned} ATE &= E[\delta_i] \\ &= E[Y_i^1 - Y_i^0] \\ &= E[Y_i^1] - E[Y_i^0], \end{aligned} \tag{3}$$

kde $E[Y_i^1]$ vyjadruje strednú hodnotu potenciálneho výsledku pri účasti i-tého jedinca a $E[Y_i^0]$ predstavuje strednú hodnotu potenciálneho výsledku neúčasti i-tého jedinca. ATT vyjadríme ako:

$$\begin{aligned} ATT &= E[\delta_i | D_i = 1] \\ &= E[Y_i^1 - Y_i^0 | D_i = 1] \\ &= E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1], \end{aligned} \tag{4}$$

kde $E[Y_i^1 | D_i = 1]$ predstavuje podmienenú strednú hodnotu potenciálneho výsledku pri účasti i-tého jedinca za podmienky účasti a $E[Y_i^0 | D_i = 1]$ vyjadruje podmienenú strednú hodnotu potenciálneho výsledku neúčasti i-tého jedinca za podmienky účasti. Tretí účinok

ATU vyjadríme v nasledovnom tvare:

$$\begin{aligned} ATU &= E[\delta_i|D_i = 0] \\ &= E[Y_i^1 - Y_i^0|D_i = 0] \\ &= E[Y_i^1|D_i = 0] - E[Y_i^0|D_i = 0], \end{aligned} \tag{5}$$

kde $E[Y_i^1|D_i = 0]$ predstavuje podmienenú strednú hodnotu potenciálneho výsledku pri účasti i-tého jedinca za podmienky neúčasti a $E[Y_i^0|D_i = 0]$ vyjadruje podmienenú strednú hodnotu potenciálneho výsledku neúčasti i-tého jedinca za podmienky neúčasti. V praxi sú najvyužívanejšími odhadnutými typmi účinkov ATE a ATT. Z rovníc (3), (4) a (5) je viditeľná pretrvávajúca problematika ohľadom vyjadrenia oboch stavov, t. j. pozorovaného výsledku a kontrafaktuality.

Jedným z možných riešení pre odhadnutie účinku ATE je zjednodušený pohľad na problematiku cez jednoduchý rozdiel priemerov (SDO). Využitím pozorovaných hodnôt dátovej vzorky, predpisov pre vyjadrenie ATE (3), ATT (4) a ATU (5), úprav z [8] a vychádzajúc z myšlienky, že ATE môžeme vyjadriť ako váženú sumu ATT a ATU

$$ATE = \pi ATT + (1 - \pi)ATU \tag{6}$$

vieme ATE odhadnúť v nasledujúcom tvare:

$$\begin{aligned} \underbrace{\frac{1}{N_T}\sum_{i=1}^{n}(y_i|d_i = 1) - \frac{1}{N_C}\sum_{i=1}^{n}(y_i|d_i = 0)}_{SDO} &= \underbrace{E[Y^1] - E[Y^0]}_{\widehat{ATE}} \\ &+ \underbrace{E[Y^0|D = 1] - E[Y^0|D = 0]}_{Selection\ bias} \\ &+ \underbrace{(1 - \pi)(ATT - ATU)}_{Heterogeneous\ treatment\ effect\ bias}, \end{aligned} \tag{7}$$

kde na ľavej strane rovnice N_T vyjadruje počet pozorovaní v účastníckej skupine, N_C vyjadruje počet pozorovaní v kontrolnej skupine, y_i vyjadruje pozorovaný výsledok i-tého pozorovania a d_i vyjadruje rozhodnutie zúčastniť sa alebo nezúčastniť sa. Na pravej strane sa nachádza priemerný účinok účasti a dva druhy vychýlenia. Prvé, výberové vychýlenie, je vyjadrené ako rozdiel medzi strednou hodnotou potenciálneho výsledku kontrolnej sku-

piny za podmienky účasti (kontrafaktualita) a strednou hodnotou potenciálneho výsledku kontrolnej skupiny za podmienky neúčasti. Toto vychýlenie odpovedá na otázku zmeny potenciálnych výsledkov pri priradení a nepriradení účasti jedincom z kontrolnej skupiny. Druhé vychýlenie, heterogénne skreslenie účinku účasti, je vyjadrené ako rozdiel v priemernom účinku účasti medzi dvoma skupinami vážené podielom nezúčastnených. Tento typ vychýlenia reaguje na odlišnosti vplyvu účasti v rovnakom programe v závislosti od jednotlivca.

SDO na ľavej strane rovnice predstavuje vyčísliteľnú hodnotu. Problematickú časť predstavuje pravá strana, ktorá v každom člene obsahuje nevyčísliteľnú kontrafaktualitu. Z tohto dôvodu je cieľom metód kauzálnych účinkov nájsť spôsoby negácie alebo odôvodnenia vychýlenia. Metódy toto docieľujú splnením setu predpokladov, ktoré sú definované pre každú z nich a týkajú sa účastníckej a kontrolnej skupiny.

Medzi najvyužívanejšie metódy riešenia chýbajúcich kontrafaktualít zaraďujeme:

- Metódy párovania (*Matching methods*).
- Metódy regresnej diskontinuity (*Regression discontinuity*).
- Metódy inštrumentálnych premenných (*Instrumental variables*).
- Panelové dáta (*Panel data*).
- Metódy rozdielov v rozdieloch (*Differences in Differences*).

Detailnejšie sme sa zamerali na metódu párovania. Najskôr sme metódu definovali a uviedli predpoklady pre jej využitie. Následne sme predstavili 4-krokový postup pre vykonanie analýzy účinkov metódou párovania aplikovaný v praktickej časti, ktorý je využiteľný v prípade splnenia predpokladov.

1.2.4 Metóda párovania

Metóda párovania predstavuje jednu z metodík vysporiadania sa s nenáhodnými dátami a existenciou kontrafaktualít. Podľa článku [32] definujeme metódu párovania ako akúkoľvek metódu, ktorej cieľom je vyrovnať (alebo „vyvážiť“) distribúciu kovariátov X v účastníckej a kontrolnej skupine. Môže to zahŕňať párovanie 1:1, váženie alebo subklasifikáciu. Metóda

má širokú škálu využitia v štatistike, epidemiológii, sociológii, ekonómii aj politických vedách.

A Predpoklady pre využitie metódy párovania

Pre využitie spomínanej metódy musíme predpokladať, že naša dátová vzorka spĺňa tri dôležité predpoklady. Prvé dva je možné spojiť do spoločného pojmu silná ignorabilita priradenia účasti. Pri zachovaní značenia z [8] ho Rosenbaum [28] vo svojom diele definuje nasledovne:

Definícia 1.3 *Priradenie účasti D je nezávislé na potenciálnych výsledkoch Y^1 a Y^0 vzhľadom na kovariáty X:*

$$(Y^1, Y^0) \perp D|X. \tag{8}$$

Existuje kladná pravdepodobnosť priradenia účasti pre všetky kovariáty X:

$$0 \leq Pr(D = 1|X) \leq 1. \tag{9}$$

Prvú zložku predpokladu nazývame predpoklad nezavádzajúceho mechanizmu priraďovania. Za platnosti tohto predpokladu je teda medzi subjektami s rovnakými hodnotami kovariátu účasť priradená akoby náhodne. Tento predpoklad nemožno testovať, možno ho len obhájiť argumentáciou, ktorá plynie zo znalosti problematiky, v ktorej vykonávame analýzu kauzálnych účinkov. Druhú zložku nazývame predpoklad prekrytia. Úlohou predpokladu je zabezpečiť vyváženosť a dostatočné prekrytie dát na základe kovariátov. Prekrytie zabezpečí, že v skupine účastníkov aj kontrolnej skupine budú existovať záznamy s podobným profilom. Tento predpoklad je možné overiť graficky pohľadom na distribúciu záznamov. V prípade splnenia oboch predpokladov je využiteľná nasledovná definícia:

Definícia 1.4 *[22] V prípade platnosti predpokladov nezavádzajúceho mechanizmu priraďovania a prekrytia je nepozorovateľná odozva Y^0 u subjektov s aktívnou účasťou $D = 1$ „v priemere“ rovnaká ako pozorovateľná odozva subjektov z kontrolnej skupiny. Priemerný kauzálny účinok účasti teda možno na základe dát identifikovať.*

Tretím spomínaným predpokladom je predpoklad stabilnej hodnoty účasti (SUTVA) z [29]. Stuartová [32] vo svojom diele predpoklad popisuje slovami, že výsledky jednot-

livca nie sú ovplyvnené pridelením účasti iným jednotlivcom. Taktiež uvádza, že takýto predpoklad môže byť porušený napríklad v školách, kde účastníci programu a kontrolná skupina detí môžu medzi sebou interagovať, čo môže viesť k „prelievaniu" účinkov. Vierohodnosť SUTVA sa dá často zlepšiť úpravou dizajnu experimentu, napríklad znížením interakcie medzi kontrolnou a účastníckou skupinou.

B Kroky realizácie párovania

Ďalej sme sa bližšie pozreli na najdôležitejšie kroky realizácie analýzy párovania. Stuartová [32] a popis R knižnice MatchIt [19], ktorú sme využili v praktickej časti, uvádzajú nasledovné 4 kľúčové kroky pre implementáciu analýzy párovania:

1. Plánovanie.

2. Implementácia metódy párovania vzhľadom na úpravy v kroku 1.

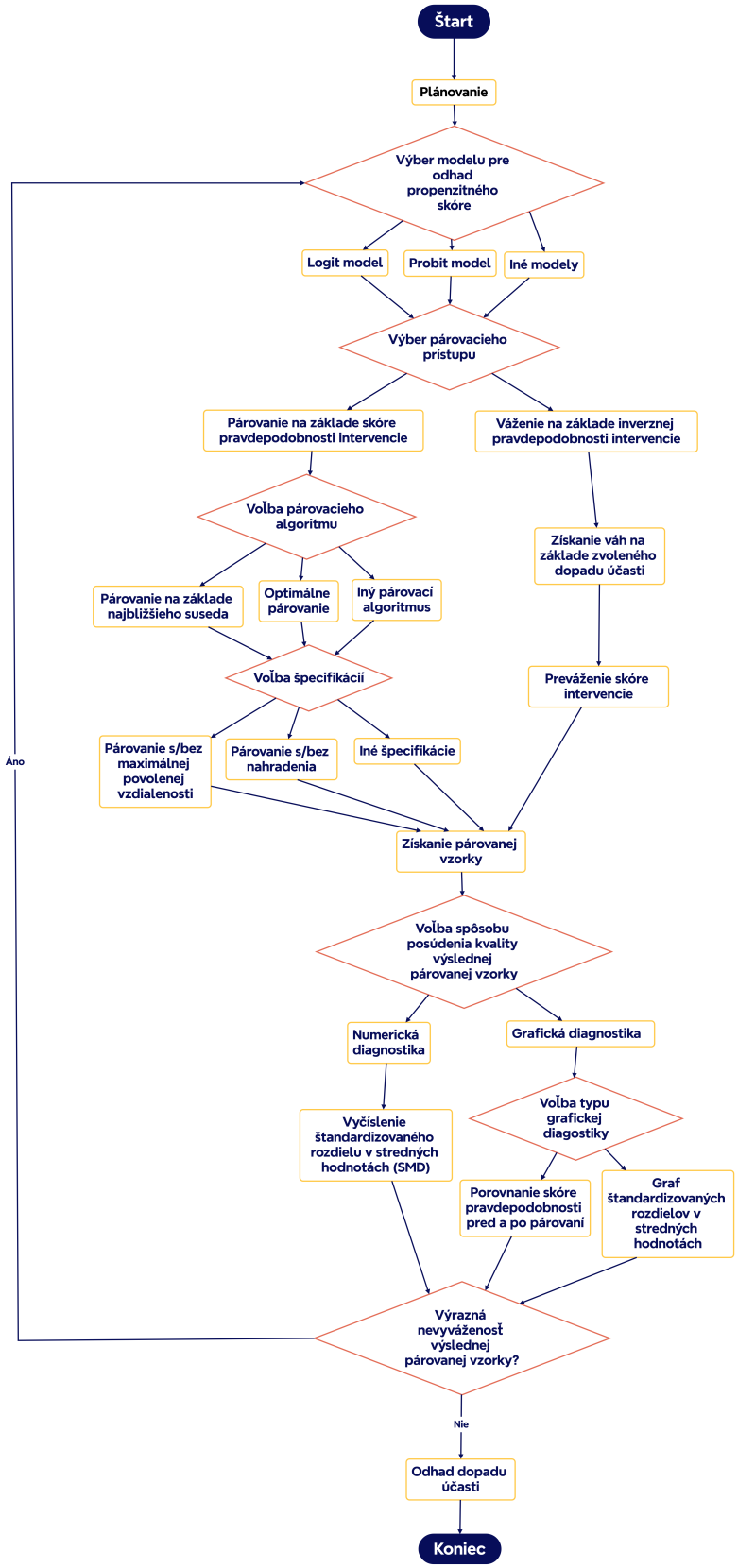
3. Posúdenie kvality výslednej párovanej vzorky a možné iterovanie medzi krokmi 1 a 2 pre dosiahnutie najlepšieho párovania vzorky.

4. Analýza výsledku a odhad vplyvu účasti vzhľadom na párovanie vykonané v kroku 3.

Pre každý z krokov sme v ďalšej podčasti uviedli kľúčové myšlienky s tým, že v kroku o implementácii metódy párovania sme detailnejšie vysvetlili fungovanie párovania na základe zvolených prístupov a parametrov párovania. Taktiež na Obrázku 4 uvádzame grafický prehľad dôležitých rozhodnutí v algoritme s kľúčovými pojmami.

1. Plánovanie

Plánovacia fáza analýzy párovania zahŕňa výber typu účinku, výber cieľovej populácie a výber kovariátov, pre ktoré je potrebná rovnováha na vyjadrenie nezaujatého odhadu účinku účasti. Každý z týchto krokov predstavuje teoretické kroky, ktoré nezahŕňajú vykonávanie analýz údajov. V ideálnom prípade by mali byť zvážené pred zberom záznamov vo fáze plánovania štúdie. Včasné premýšľanie o nich môže pomôcť pri vykonávaní úplnej a nákladovo efektívnej analýzy. Detailnejší návod ako postupovať pri výbere účinku poskytuje článok Greifera a Stuartovej [18] a pre výber kovariátov poskytuje článok VanderWeela [34]. Pred presunom do ďalšieho kroku je užitočný aj pohľad na nevyrovnanosť dát, ktoré sa metóda párovania snaží eliminovať.



Obrázok 4: Algoritmické vyjadrenie krokov implementácie analýzy párovania v [36]

2. Implementácia metódy párovania

Hlavným cieľom tohto kroku je priradenie „najpodobnejšieho" záznamu alebo „najpodobnejších" záznamov na základe kovariátov z kontrolnej skupiny k účastníkovi. Takýmto spôsobom budeme mať k dispozícii informáciu o jednotlivcovi v oboch budúcich stavoch spomínaných v Rubinovom kauzálnom modeli. V prípade malého množstva zvolených kovariátov a veľkej dátovej vzorky kontrolnej skupiny môžeme zvoliť algoritmus presného párovania. Ten je založený na párovaní podľa presnej zhody v hodnote kovariátu medzi účastníkom a jedincom v kontrolnej skupine. Následne je priemerná hodnota účinku vyčísliteľná SDO na ľavej strane rovnice (7). V prípade väčšieho množstva kovariátov a menšej dátovej vzorky narážame na problém s priraďovaním a vynárajú sa nám otázky o blízkosti alebo podobnosti medzi záznamami v skupinách. Presné párovanie v takýchto situáciách zlyháva a je výhodnejšie voliť algoritmy približného párovania. Algoritmy tohto typu sú založené na definovaní novej miery pre odhad blízkosti.

Najvyužívanejšie miery pre odhad blízkosti

Medzi štandardne využívané miery patria Euklidovská vzdialenosť, normalizovaná Euklidovská vzdialenosť a Mahalanobisova vzdialenosť. Cunningham [8] ich definuje nasledovným spôsobom. Euklidovská vzdialenosť je definovaná ako

$$\begin{aligned} \|X_i - X_j\| &= \sqrt{(X_i - X_j)'(X_i - X_j)} \\ &= \sqrt{\sum_{n=1}^{k}(X_{ni} - X_{nj})^2}, \end{aligned} \tag{10}$$

kde X_{ni} predstavuje n-tý záznam i-tého kovariátu a X_{nj} predstavuje n-tý záznam j-tého kovariátu. Normalizovaná Euklidovská vzdialenosť, ktorá efektívne reaguje na odlišnosti v mierke premenných, je definovaná nasledovne:

$$\begin{aligned} \|X_i - X_j\| &= \sqrt{(X_i - X_j)'\hat{V}^{-1}(X_i - X_j)} \\ &= \sqrt{\sum_{n=1}^{k}\frac{X_{ni} - X_{nj}}{\sigma_n^2}}, \end{aligned} \tag{11}$$

kde rozdiel kovariátov v čitateli je teraz predelený varianciou σ_n^2 n-tého záznamu v me-

novateli. Mahalanobisová vzdialenosť je definovaná nasledujúcim predpisom:

$$\|X_i - X_j\| = \sqrt{(X_i - X_j)' \widehat{\sum}_X^{-1} (X_i - X_j)}, \tag{12}$$

kde $\widehat{\sum}_X^{-1}$ predstavuje výberovú variančno-kovariančnú maticu X.

Definovanie nových mier blízkosti umožňuje párovanie záznamov aj v suboptimálnych podmienkach, ale prináša so sebou vychýlenie vo forme rozdielov v hodnotách kovariátov. Jedným z možných spôsobov vysporiadania sa s vychýlením sú techniky korekcie vychýlenia predstavené v diele [1]. Pre schopnosť robustného vysporiadania sa s vychýlením alebo nevyrovnanosťou bude pre nás relevantnejší algoritmus približného párovania založený na skóre pravdepodobnosti intervencie.

Párovanie založené na skóre pravdepodobnosti intervencie

Hlavnou myšlienkou analýzy skóre pravdepodobnosti intervencie je využitie hodnôt kovariátov záznamu na odhadnutie skalára nazývaného skóre pravdepodobnosti intervencie. Informácie z kovariátov sú stlačené do jednej hodnoty, na základe ktorej je možné vykonať všetky porovnania a párovania medzi účastníkmi a kontrolnou skupinou. Tento proces párovania vytvára rovnováhu medzi účastníckou a kontrolnou skupinou na základe skóre pravdepodobnosti intervencie.

Skóre pravdepodobnosti intervencie je podľa [28] definované ako pravdepodobnosť prijatia liečby na základe známych hodnôt kovariátov:

$$e(x) = P(D = 1|X), \tag{13}$$

kde $e(x)$ predstavuje skóre pravdepodobnosti intervencie, P je pravdepodobnosť, D je binárna premenná reprezentujúca účasť ak $D = 1$ a neúčasť ak $D = 0$, $|$ je symbol podmienenia a X sú známe kovariáty. V prípade platnosti predpokladu silnej ignorability z definície 1.3 je využiteľná veta o skóre pravdepodobnosti intervencie:

Veta 1.5 *[28] Ak priradenie účasti spĺňa predpoklad silnej ignorability pri danom X, potom spĺňa predpoklad silnej ignorability aj pri danom skóre pravdepodobnosti intervencie $e(X)$;*

t. j. $(Y^1, Y^0) \perp D|X$ *a* $0 \leq Pr(D = 1|X) \leq 1$ *pre všetky* X *implikuje*

$(Y^1, Y^0) \perp D|e(X)$ *a* $0 \leq Pr(D = 1|e(X)) \leq 1$ *pre všetky* $e(X)$.

Pri párovaní strácame potrebu kontroly celého vektora zavádzajúcich premenných X z dôvodu postačujúceho skalára skóre pravdepodobnosti intervencie. Dôkaz tejto vety, ako aj ďalšie vety o skóre pravdepodobnosti intervencie sú k dispozícii v [28].

V observačných štúdiách je nutné neznáme hodnoty skóre pravdepodobnosti intervencie odhadnúť. V našej problematike skúmaná náhodná veličina D nadobúda hodnotu 1, ak sa jednotlivec zúčastnil programu, inak má hodnotu 0. Vysvetľujúcimi premennými sú kovariáty z množiny $X = \{x_1, \ldots, x_n\}^T$. Keďže vysvetľovaná premenná je dichotomická premenná, lineárna regresia, ktorá pracuje iba so spojitými premennými, je nedostatočná. Pre tento účel sú využívané zovšeobecnené lineárne modely pre binárne dáta alebo binárna logistická regresia [3]. Očakávaná odozva je v tomto prípade modelovaná ako transformovaná pravdepodobnosť určitého faktorového výsledku. Transformácia zaručí správny rozsah pravdepodobnosti medzi 0 a 1, ako to je pre pravdepodobnosť definované. Najčastejšie používanými funkciami sú logistická a probistická funkcia, ktorých predpisy kumulatívnej distribučnej funkcie v prípade binárnej možnosti účasti alebo liečby vyzerajú nasledovne:

$$\begin{aligned} &\text{Logistická funkcia} \quad P(d_i = 1|x_i) = \frac{e^{x_i^T\beta}}{1 + e^{x_i^T\beta}} = \frac{1}{1 + e^{-x_i^T\beta}} \\ &\text{Probistická funkcia} \quad P(d_i = 1|x_i) = \Phi(x_i^T\beta) = \int_{-\infty}^{x_i^T\beta} \Phi(z)\, dz. \end{aligned} \tag{14}$$

V ďalšej časti detailnejšie popíšeme postup pre odhad skóre pravdepodobnosti intervencie na základe pravdepodobnostnej transformácie do tvaru logitov.

Nech $d_i \in \{0, 1\}$ je vysvetľovaná premenná, $X_i = \{x_i^1, x_i^2, \ldots, x_i^n\}$ sú závislé premenné a p_i je odhadnutá pravdepodobnosť, potom z tvaru logistickej regresie pre i-tý prípad platí:

$$ln\left(\frac{p_i}{1 - p_i}\right) = X_i^T\beta \rightarrow p_i = P(d_i = 1|X_i) = \Lambda(X_i^T\beta) = \frac{e^{X_i^T\beta}}{1 + e^{X_i^T\beta}} = \frac{1}{1 + e^{-X_i^T\beta}}, \tag{15}$$

kde $X_i = \{x_i^1, x_i^2, \ldots, x_i^n\}$ sú vstupné dáta, d_i je náhodná veličina detekujúca účasť a $\beta = (\beta_1, \beta_2, \ldots \beta_n)^T$ predstavuje vektor odhadnutých parametrov udávajúci váhy alebo dôležitosť jednotlivých vstupných hodnôt X_i. Skalárny súčin $X_i^T\beta$ zachytáva skóre pravdepodobnosti intervencie. Na odhadnutie parametrov sa najčastejšie využíva metóda ma-

ximálnej vierohodnosti (MMV), ktorá na základe vzorky meraní odhadne také hodnoty parametrov β, pri ktorých pravdepodobnosť nastania dosahuje najväčšie možné hodnoty. Za platnosti predpokladu nezávislosti medzi vysvetľujúcimi premennými, funkcia maximálnej vierohodnosti L pre N pozorovaní $d_1, d_2, \dots d_n$ s pravdepodobnosťami $p_1, p_2, \dots p_n$ má na základe [20] nasledujúci tvar:

$$L = \prod_{\{i|d_i=0\}} [1 - \Lambda(X_i^T\beta)] \prod_{\{i|d_i=1\}} \Lambda(X_i^T\beta) = \prod_{i=1}^{N} [1 - \Lambda(X_i^T\beta)]^{1-d_i} \Lambda(X_i^T\beta)^{d_i}.$$

Logaritmus vierohodnostnej funkcie vieme zapísať v tvare:

$$ln\ L = \sum_{i=1}^{N} [d_i ln(\Lambda(X_i^T\beta)) + (1 - d_i) ln(1 - \Lambda(X_i^T\beta))].$$

Pre nájdenie maxima vypočítame deriváciu funkcie $ln\ L$ vzhľadom na vektor β. Riešime nasledovnú rovnicu

$$\frac{\delta ln\ L}{\delta\beta} = \sum_{i=1}^{N} \left[d_i \frac{f(X_i^T\beta)}{\Lambda(X_i^T\beta)} + (1 - d_i) \frac{-f(X_i^T\beta)}{1 - \Lambda(X_i^T\beta)} \right] X_i^T = 0,$$

kde $f(z) = \frac{\delta\Lambda(z)}{\delta z}$. Zjednodušením dostaneme

$$\frac{\delta ln\ L}{\delta\beta} = \sum_{i=1}^{N} (d_i - \Lambda(X_i^T\beta)) X_i^T = 0. \tag{16}$$

Logaritmická vierohodnostná funkcia je globálne konkávna, a z tohto dôvodu je riešením rovnice (16) globálne maximum. Výraz v spomínanej rovnici je nelineárny vzhľadom na β, a z tohto dôvodu si vyžaduje špeciálne iteračné metódy (napr. Newton-Raphsonov algoritmus [17]). Odhadnuté koeficienty označíme $\hat{\beta}$. V ďalšom kroku odhadneme skóre pravdepodobnosti intervencie na základe rovnice (15) a označíme ich $\hat{p_i}$ pre $i = 1, \dots, n$.

Po odhade skóre pravdepodobnosti intervencie je možné rozhodnúť sa pre dva rôzne prístupy. Prvým z nich je párovanie na základe skóre pravdepodobnosti intervencie a druhým váženie na základe inverznej pravdepodobnosti intervencie.

Párovanie na základe skóre pravdepodobnosti intervencie je založené na párovaní hodnôt z účastníckej a kontrolnej skupiny priamo na základe odhadov $\widehat{p_i}$ vhodnou voľbou

párovacích algoritmov a ich parametrov. Tento prístup predstavuje intuitívny a jednoducho interpretovateľný spôsob párovania. Cieľ vyrovnania kovariátov dosahuje využitím podmnožiny záznamov z kontrolnej skupiny, čo ale môže viesť k strate informácií vylúčením nespárovaných záznamov. Ďalšiu nevýhodu môže predstavovať fakt, že dva záznamy s podobnou hodnotou skóre pravdepodobnosti intervencie môžu v skutočnosti disponovať odlišným súborom kovariátov.

Podľa [32] a [19] medzi najčastejšie využívané algoritmy patria párovanie na základe najbližšieho suseda (Nearest neighbour matching) a párovanie na základe optimálnych dvojíc (Optimal pair matching). Jadrom algoritmu párovania na základe najbližšieho suseda je párovanie jedného alebo k záznamov z kontrolnej skupiny na jeden záznam z účastníckej skupiny založený na skóre pravdepodobnosti intervencie. Tento algoritmus sa častokrát javí ako „chamtivý", pretože páruje skóre pravdepodobnosti intervencie na základe blízkosti, bez ohľadu na to, aké hodnoty ostanú pre vytvorenie ďalších párov. Z tohto dôvodu sa častokrát definuje, od ktorých hodnôt skóre pravdepodobnosti intervencie má párovanie začínať. V prípade párovania s nahradením, definovanie ktoré skóre pravdepodobnosti intervencie má prednosť stráca dôvod. Párovanie s nahradením má opodstatnenie v prípade malého množstva záznamov v kontrolnej skupine. Tretím parametrom, ktorý je dôležité spomenúť, je zavedenie maximálnej povolenej vzdialenosti (v anglickej terminológii caliper) pri párovaní. Tá zaistí výber párov v stanovenej medzi, ale môže spôsobiť veľké množstvo nespárovaných záznamov z účastníckej skupiny s možným dôsledkom problémov pri neskoršej interpretácii efektu ošetrenia.

Algoritmus založený na párovaní na základe optimálnych dvojíc sa snaží o minimalizáciu miery celkovej vzdialenosti úvodným pohľadom na celý súbor dát a až následným párovaním. Výhodou optimálneho párovania je teda lepšia celková kvalita párov. Nevýhody pri veľkom množstve záznamov spočívajú vo zvýšenom výpočtovom čase a náročnosti pre pamäť výpočtového zariadenia. Voľba algoritmu a parametrov je založená na porozumení problematiky a cieľoch nastavených v plánovacej fáze.

Oba spôsoby párovania s rôznymi vstupnými parametrami sme preskúmali v praktickej časti.

Druhý prístup pojednával o vážení na základe inverznej pravdepodobnosti intervencie. Prvá zmienka o tomto prístupe pochádza z diela [27] od Rosenbauma, kde bol spomenutý

ako forma priamej štandardizácie založenej na modeli. Hlavnou myšlienkou je vytvorenie syntetickej vzorky, v ktorej je rozdelenie nameraných základných kovariátov nezávislé od priradenia liečby alebo účasti. Jednotlivým záznamom sú priradené váhy na základe odhadnutého skóre pravdepodobnosti intervencie. Váhy podľa zvoleného účinku účasti priradia väčšiu dôležitosť informatívnejším údajom. A preto strácame nutnosť vylúčenia časti záznamov pre nevhodné hodnoty kovariátov, čo umožňuje prístupu IPW robustnejšie odhady efektov. V prípade priemerného účinku účasti vieme pomocou [4] a [32] váhy i-tého záznamu vyčísliť nasledovným vzťahom:

$$w_{ATE} = \frac{d_i}{\widehat{p_i}} + \frac{(1 - d_i)}{1 - \widehat{p_i}}, \tag{17}$$

kde sa inverzné pravdepodobnosti intervencie využívajú ako inverzné váhy a preváženie sa dotýka účastníckej aj kontrolnej skupiny. V prípade priemerného účinku účasti na účastníkov vieme váhy i-tého záznamu vyčísliť nasledovným vzťahom:

$$w_{ATT} = d_i + (1 - d_i)\frac{\widehat{p_i}}{1 - \widehat{p_i}}, \tag{18}$$

kde d_i reprezentuje binárnu premennú priradenia účasti a $\widehat{p_i}$ odhadnuté skóre pravdepodobnosti intervencie. Zo vzťahu je viditeľné, že vzorka zúčastnených predstavuje referenčnú vzorku a vzorka kontrolnej skupiny prechádza štandardizáciou. Váženie na základe inverznej pravdepodobnosti intervencie so sebou prináša nevýhodu veľkého rozptylu v prípade extrémnych váh. Takáto situácia nastáva, ak sú skóre pravdepodobnosti intervencie blízko k 0 alebo 1. Dve možné riešenia tohto problému zahŕňajú využitie orezávania alebo osekávania váh.

Orezávanie spočíva vo vylúčení skóre pravdepodobnosti intervencie nachádzajúcich sa na chvostoch rozdelenia rozloženia skóre pravdepodobnosti intervencie. V prípade osekávania sa váham presahujúcim určitú hraničnú hodnotu priradí práve hraničná hodnota. Hraničné hodnoty sú častokrát založené na kvantiloch rozdelenia váh. Takýmto spôsobom je dosiahnuté zníženie rozptylu odhadu účinku na úkor zvýšenia vychýlenia. Na nastavenie správnej úrovne orezávania alebo osekávania nie je v literatúre k dispozícii presný súbor odporúčaní. Váženie na základe inverznej pravdepodobnosti intervencie sme využili a výsledky popísali v praktickej časti.

3. Posúdenie kvality výslednej párovanej vzorky

Pred odhadom samotného vplyvu účasti je nutné posúdenie kvality párovania, ktoré sme získali prístupmi párovania na základe pravdepodobnosti intervencie a váženia na základe inverznej pravdepodobnosti intervencie. Posúdenie kvality je založené na vyhodnotení vyváženia kovariátov na základe empirickej distribučnej funkcie kovariátov účastníckej a kontrolnej skupiny po aplikovaní metódy párovania. Cieľom je dosiahnutie nezávislosti účasti od kovariátov, t. j. aby platil vzťah

$$F(X|D=1) = F(X|D=0), \tag{19}$$

kde F vyjadruje empirické rozdelenie. Tento cieľ je možné diagnostikovať numerickou alebo grafickou diagnostikou.

Numerická diagnostika

Numerická diagnostika je založená na číselných hodnotách získaných po implementácii metódy párovania. Stuartová [32] uvádza, že jednou z najvyužívanejších mier je štandardizovaný rozdiel v stredných hodnotách (Standardized mean difference SMD). Známa je Cohenova [6] verzia vzorca pre výpočet v tvare:

$$SMD = \frac{\widehat{X_T} - \widehat{X_C}}{\sqrt{\frac{(\sigma_T^2 + \sigma_C^2)}{2}}}, \tag{20}$$

kde sa v čitateli nachádzajú kovariáty pre účastnícku a kontrolnú skupinu a v menovateli je uvedený vážený priemer odchýlok dvoch skúmaných skupín. Priaznivé sú hodnoty najbližšie k nule.

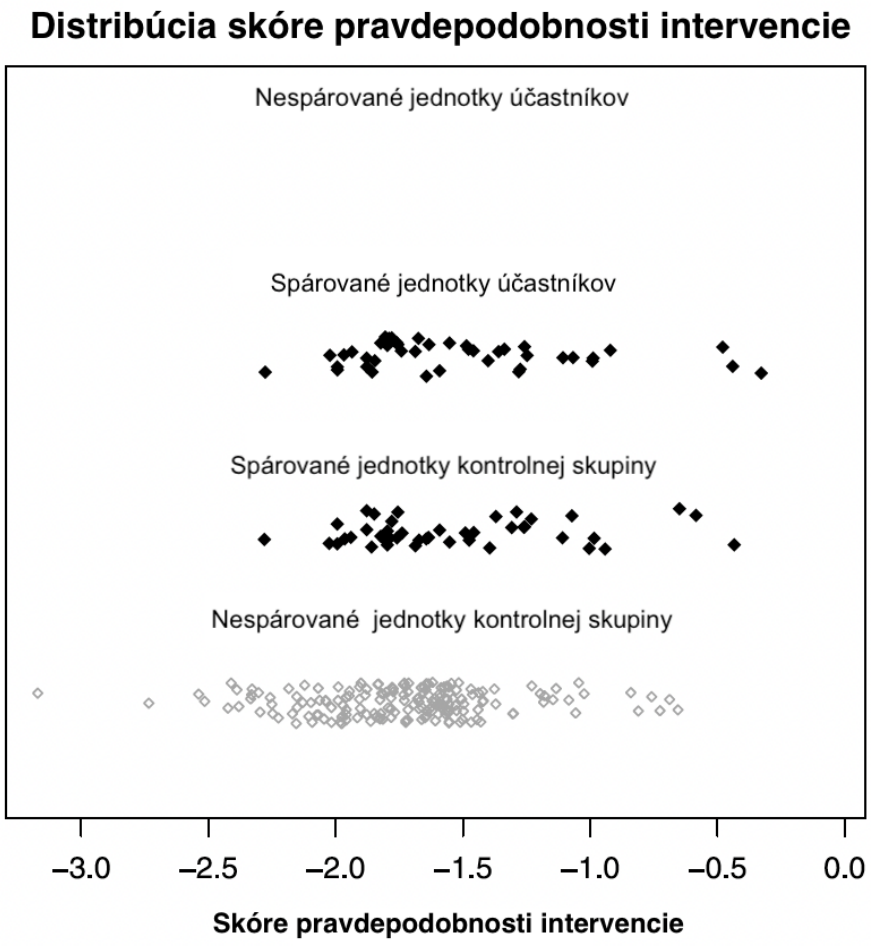
Ďalšou formou numerickej diagnostiky je využitie variačného pomeru (Variance ratio VR), ktorý pre i-tý kovariát X_i môžeme vyčísliť ako

$$VR_i = \frac{\sigma_T^2(X_i)}{\sigma_C^2(X_i)}, \tag{21}$$

kde $\sigma_T^2(X_i)$ predstavuje disperziu účastníckej skupiny pre i-tý kovariát a $\sigma_C^2(X_i)$ disperziu kontrolnej skupiny pre i-tý kovariát. Priaznivé sú hodnoty najbližšie k jednotke.

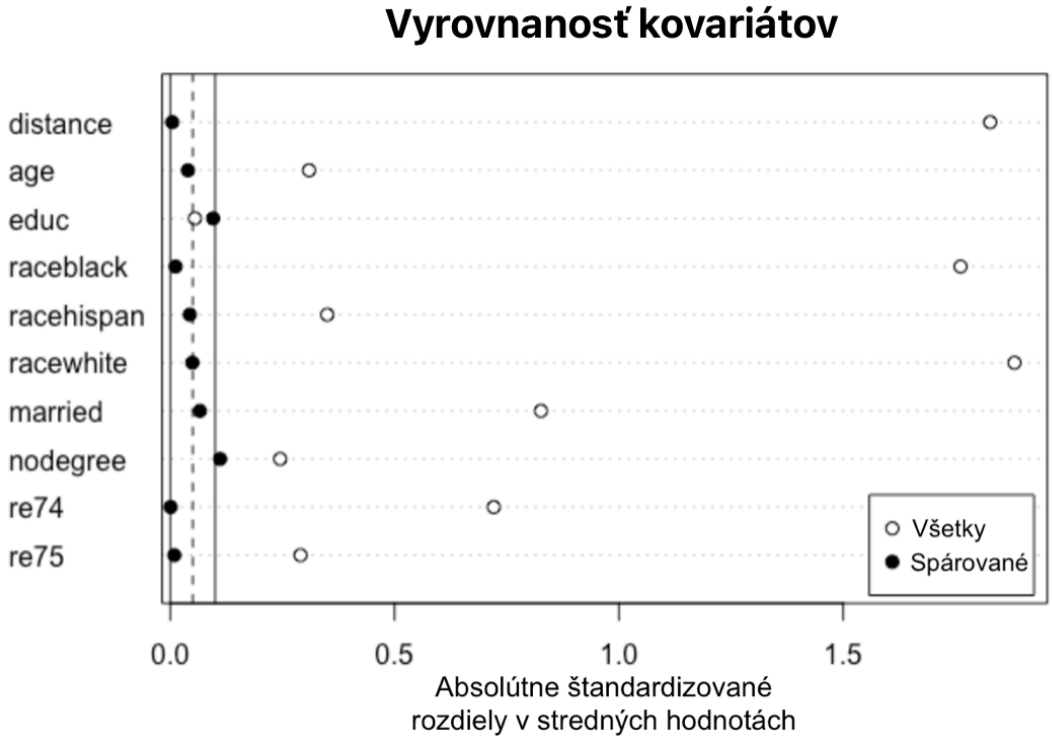
Grafická diagnostika

Grafická diagnostika je založená na vizuálnom porovnaní numerických hodnôt získaných po aplikácii metódy párovania. Tento prístup pre vyhodnotenie vyváženia je výhodnejší v prípade veľkého množstva kovariátov. Jednou z diagnostík je porovnanie skóre pravdepodobnosti intervencie v účastníckej a kontrolnej skupine pred a po využití metódy párovania. Takýto prístup predstavuje užitočný spôsob pre určenie splnenia predpokladu prekrytia. Príklad takéhoto porovnania je viditeľný na Obrázku 5, kde čierne body reprezentujú pozorovania po využití metódy párovania a šedé predstavujú pozorovania pred využitím metódy. Z obrázka je možné vyčítať, že po párovaní boli využité všetky záznamy pre účastníkov, ale iba vybrané záznamy z kontrolnej skupiny. Dôvodom bolo docielenie čo najlepšieho možného vyváženia a prekrytia skóre pravdepodobnosti intervencie medzi účastníckou a kontrolnou skupinou.



Obrázok 5: Grafické znázornenie distribúcie skóre pravdepodobnosti intervencie po využití párovania 1:1 metódou najbližšieho suseda. Zdroj: [32]

Ďalším užitočným spôsobom pre zistenie zmien vo vyvážení kovariátov je zúžitkovanie grafu štandardizovaných rozdielov v stredných hodnotách viditeľný na Obrázku 6. Tento graf poskytuje prehľad zmien pred a po využití metódy párovania. Na obrázku je viditeľné, že vo všetkých prípadoch okrem premennej educ nastalo zníženie v štandardizovaných rozdieloch v stredných hodnotách kovariátov. Zvýšenie hodnôt po párovaní môže nastať u kovariátov, ktoré nadobúdali nízke hodnoty pred využitím metódy párovania. V prípade takýchto premenných je nutné zvážiť, či je vychýlenie spôsobené týmito kovariátmi natoľko problematické, aby viedlo k úprave metódy párovania alebo zvolených kovariátov.



Obrázok 6: Príklad grafického znázornenia vyrovnanosti kovariátov využitím plného párovania. Zdroj: [19]

Vyhodnotenie vyváženia

Po aplikovaní numerickej alebo grafickej diagnostiky nasleduje posúdenie vyváženosti výslednej párovanej vzorky. V prípade výraznej nevyváženosti je nutné prehodnotiť špecifikácie využité v kroku implementácie metódy párovania. Je nutné získať novú párovanú vzorku voľbou iného modelu alebo súboru špecifikácií a parametrov. V prípade vyváženosti je možný presun do fázy odhadu a analýzy vplyvu účinkov.

4. Odhad a analýza vplyvu účasti

Odhad a analýza vplyvu účasti predstavuje posledný krok v implementácii analýzy párovania. Cunningham [8] a Abadie [2] odhadujú priemerný účinok intervencie pre účastníkov (ATT) ako

$$\begin{aligned}\widehat{ATT} &= E[Y^1 - Y^0 | X = x, D = 1] \\ &= E[Y^1 | X = x, D = 1] - E[Y^0 | X = x, D = 1] \\ &= E[Y^1 | X = x] - E[Y^0 | X = x] \\ &= E[Y^1 | X = x, D = 1] - E[Y^0 | X = x, D = 0] \\ &= E[Y | X = x, D = 1] - E[Y | X = x, D = 0].\end{aligned} \tag{22}$$

Priemerný účinok účasti pre účastníkov je teraz vyčísliteľný zohľadnením váh jednotlivých pozorovaní po využití metódy párovania. Metódy párovania sa v mnohých prípadoch nevyužívajú samostatne pre odhad kauzálnych účinkov. Najčastejšie sa výsledné páry kombinujú s regresným modelovaním. Metódy párovania dopomáhajú k zníženiu citlivosti na konkrétne špecifikácie výsledku pri odhade účinku účasti.

V ďalšej kapitole sme sa zamerali na vykonanie analýzy metódou párovania na skutočnej dátovej vzorke využitím teoretickej bázy znalostí získanej v teoretickej časti diplomovej práce.

2 Praktická časť

Druhá kapitola pojednávala o praktickej časti diplomovej práce. Tá predstavovala aplikáciu teoretických poznatkov v konkrétnom prostredí a slúžila na overenie ich využiteľnosti v praxi. Jej hlavným cieľom bola detailná analýza skúmaného problému, návrh riešenia a vyvodenie relevantných záverov na základe získaných údajov. Cieľom tejto časti bolo nielen demonštrovať schopnosť aplikácie teoretických poznatkov, ale aj prispieť k riešeniu konkrétneho problému skutočného sveta. Podkapitoly zhrnuté v tejto kapitole predstavujú prehľadný proces riešenia praktickej časti diplomovej práce.

Najskôr sme sa zamerali na situačné priblíženie oblasti praktickej časti a formuláciu cieľov práce. V tejto časti sme čitateľovi priblížili oblasť zamerania praktickej časti a nastavili ciele, ktoré v nej chceme dosiahnuť. Ďalej nasledovalo predstavenie náplne praktickej časti. V tejto časti sme čitateľovi priblížili, o čom naša praktická časť práce pojednávala a vymedzili úlohy pre ich dosiahnutie. V ďalšej podkapitole sme na základe vytýčených cieľov a úloh zhmotnili postup riešenia, ktorý slúžil ako guideline pre realizáciu riešenia. V podkapitole s totožným názvom sme sa zamerali na riešenie krokov zhrnutých v postupe. V záverečnej podkapitole sme sa venovali zhrnutiu výsledkov riešenia a ich interpretácii.

2.1 Situačné priblíženie oblasti praktickej časti a formulácia cieľov práce

V rámci našej praktickej časti sme sa detailnejšie zamerali na slovenský trh práce. Po dôkladnejšom preskúmaní domáceho trhu práce sme identifikovali viaceré relevantné informácie pre získanie lepšej predstavy o fungovaní tejto oblasti. Podarilo sa nám identifikovať štyri kľúčové skupiny, ktoré svojou činnosťou výrazne ovplyvňujú fungovanie trhu práce. Skupiny spolu s ich hlavnými funkciami uvádzame nižšie:

- **Štátne a verejné inštitúcie -** legislatívna, strategická a zastrešujúca funkcia.
- **Zamestnávatelia a podnikateľský sektor -** poskytovanie pracovných miest.
- **Zamestnanci a uchádzači o zamestnanie -** vykonávanie pracovnej činnosti alebo záujem o nájdenie pracovnej pozície.

- **Medzinárodné a európske organizácie -** financovanie nástrojov politiky trhu práce a analýza trendov pre vydávanie budúcich odporúčaní s cieľom skvalitnenia trhu práce.

Okrem toho na slovenský trh práce vplývajú aj externé faktory akými sú trhové podmienky alebo technologický pokrok. Mnohí zamestnanci a zamestnávatelia nedokážu na tieto turbulentné zmeny reagovať, čo môže viesť k zvýšeniu úrovne nezamestnanosti. Vysoká úroveň nezamestnanosti prináša so sebou ekonomické (strata príjmu, náklady na sociálnu pomoc), sociálne (vyššia kriminalita, chudoba) a psychologické dôsledky (depresia, závislosti). Taktiež má dôsledky na hospodárstvo a podniky (pokles produktivity/nižší hospodársky rast, odchod pracovných síl do zahraničia). Z tohto dôvodu má EÚ ale aj slovenská vláda motiváciu nezamestnanosť udržiavať na ideálnej úrovni okolo 3,5% až 4,5%. Jedným zo spôsobov zníženia úrovne nezamestnanosti je využitie nástrojov APTP. Tento set nástrojov sme bližšie popísali v podkapitole 1.1 teoretickej časti.

Mimoriadne dôležitou časťou procesu zúžitkovania nástrojov APTP, ktorá nastáva po fáze implementácie nástrojov, je vyhodnotenie efektivity. Opodstatnenie tohto kroku je vo vyhodnocovaní efektívnosti a účinnosti opatrení, racionalizácii verejných výdavkov, určení účinkov na zamestnanosť a ekonomiku a zabezpečení udržateľnosti opatrení.

Ďalej sme sa rozhodli pozrieť na aktuálnu situáciu v oblasti vyhodnocovania nástrojov APTP na Slovensku. Po pátraní na oficiálnych webových stránkach ÚPSVaR sme vyhľadali dva relevantné dokumenty. Prvým z nich bola záverečná hodnotiaca správa o vplyve nástrojov APTP pre programové obdobie 2014 - 2020 a druhým Správa o vyhodnotení uplatňovania aktívnych opatrení na trhu práce za rok 2021. Z týchto dokumentov je viditeľné, že chýba systematické monitorovanie a hodnotenie účinnosti jednotlivých nástrojov APTP jednoducho dostupné verejnosti, čo sťažuje identifikáciu úspešných nástrojov a oblastí vyžadujúcich zlepšenie. Zavedenie pravidelného hodnotenia založeného na kvalitnom súbore vstupných parametrov získaných overenými matematickými metódami by umožnilo lepšie prispôsobenie politík trhu práce aktuálnym potrebám a požiadavkám trhu práce a potenciálne by mohlo viesť ku skvalitneniu prístupu k práci pri nižších finančných nákladoch.

Na základe poznatkov z kapitoly 1 a situačného priblíženia sme za hlavný cieľ práce určili vyhodnotenie účinnosti vybraného opatrenia APTP pomocou numerických a gra-

fických výstupov. Prvým parciálnym cieľom pre dosiahnutie tohto hlavného cieľa bola voľba a popis metodiky vyhodnotenia účinnosti opatrenia v danom dátovom kontexte. Za druhý parciálny cieľ sme určili aplikovanie zvolenej metodiky na overenú dátovú vzorku. Tretím parciálnym cieľom bolo porovnanie estimačných špecifikácií pre rôzne modely využívané pre odhad účinku.

2.2 Predstavenie náplne praktickej časti

Predstavenie náplne zahŕňa proces prípravnej fázy pred tvorbou postupu a realizáciou riešenia praktickej časti diplomovej práce. V podkapitolách sme najskôr predstavili zámer práce a uviedli konkrétny predmet riešenia, teda časť problematiky, ktorú sme sa rozhodli riešiť. Následne sme na základe predmetu a cieľov práce uviedli očakávané prínosy, ktoré má naše riešenie potenciál priniesť.

2.2.1 Zámer praktickej časti

Zámerom praktickej časti bola aplikácia matematických znalostí a hlbokého porozumenia problematiky na vykonanie serióznej matematickej analýzy v rámci problematiky reálneho sveta. Analýza sa týkala vyhodnotenia vplyvu nástroja APTP na zamestnanosť. Pre tento účel sme využili rôzne prístupy a parametre s prihliadnutím na rôzne časové obdobia.

2.2.2 Predmet riešenia

Predmetom nášho riešenia bolo vykonanať analýzu kauzálnych účinkov rôznymi vopred zvolenými matematickými metódami s pomocou kvalitnej bázy teoretických znalostí a relevantnej dátovej vzorky. Vplyv sme sa rozhodli vyhodnotiť pohľadom na zmeny v zamestnaneckom stave účastníkov v programe REPAS v rámci trhu práce na Slovensku. Najlepší súbor vstupných charakteristík a najlepší model sme sa snažili identifikovať numerickou a grafickou diagnostikou získaných hodnôt. K získaniu konečného riešenia a zároveň splneniu hlavného ciela nám dopomohol naprogramovaný algoritmus, ktorý s minimálnymi zmenami vstupných parametrov dokázal rýchlo a efektívne vyhodnotiť účinnosť nástrojov APTP.

2.2.3 Očakávané prínosy

V tejto časti sme uviedli očakávané prínosy, ktoré by naša praktická časť mohla priniesť po dosiahnutí vytýčených cieľov:

- Zvýšenie efektívnosti APTP dlhodobým sledovaním a vyhodnocovaním uplatnenia účastníkov na trhu práce.
- Lepšia alokácia finančných prostriedkov do opatrení s najväčším účinkom.
- Zvýšenie dôvery v systém podpory zamestnanosti definovaním merateľných kritérií a zvýšením transparentnosti.
- Lepšie plánovanie a tvorba politík nastavením budúcich stratégií zamestnanosti založených na dátach a výsledkoch.
- Získanie metodiky pre odhad účinnosti nástroja APTP zvolenou metódou.

2.3 Postup riešenia

Postup riešenia predstavoval súbor krokov, ktoré slúžili ako míľniky pre dosiahnutie cieľov popísaných v podkapitole 2.1.

2.3.1 Výber dátovej vzorky

Najskôr sme sa zamerali na získanie dátovej vzorky, ktorá by súvisela s našou problematikou a predstavovala by dobrý dátový základ pre vykonanie analýzy. Ďalej bolo nutné vybrať vhodný štatistický softvér pre manipuláciu s dátovou vzorkou. Po úspešnej integrácii vzorky s vybraným softvérom bolo nutné overiť relevanciu dátovej vzorky pre zvolenú analýzu preštudovaním obsahu vzorky. Overenie prebehlo dôkladným preštudovaním popisu a obsahu rozsiahlych tabuliek. V prípade relevantných vstupných údajov sme mohli pokračovať v ďalších krokoch postupu riešenia praktickej časti. V prípade nevhodnej dátovej vzorky z dôvodu nevhodného súboru zachytených premenných alebo neprimeraného množstva záznamov sme sa pokúsili o nájdenie nového zdroja dát.

2.3.2 Výber metodiky

V ďalšom kroku sme sa zamerali na výber metodiky pre vykonanie analýzy. K dispozícii sme mali viaceré metódy, ktoré boli spomenuté v teoretickej časti práce. Pri výbere sme vzali do úvahy premenné dostupné v dátovej vzorke a časovú náročnosť implementácie jednotlivých metodík. Po výbere metodiky sme sa zamerali na vykonanie analýzy kauzálnych účinkov. Následne sme zvolili typ účinku, pomocou ktorého sme vyhodnocovali kauzálne účinky nástroja.

2.3.3 Programovanie algoritmu

Ďalej nasledovalo spísanie kódu v súlade s algoritmom zachytenom na Obrázku 4. Najskôr sme sa zamerali na úpravu dátovej vzorky pre vstup do algoritmu. V prvej fáze úprav sme vykonali filtrovanie dát na základe preddefinovaných vstupných parametrov. V druhej fáze sme túto vzorku ešte ďalej upravili na základe požiadaviek pre vstupné údaje. V ďalšom kroku sme overili úspešnosť prepojenia dátovej vzorky s vyhodnocovacím algoritmom pohľadom na numerické a grafické výstupy. V prípade výsledkov vo výraznom nesúlade s našimi teoretickými znalosťami a metódou zdravého rozumu bolo nutné pristúpiť k fáze dolaďovania algoritmu a úpravám vstupných údajov. V prípade súladu s našou bázou znalostí o problematike sme výsledky mohli extrahovať a interpretovať. Následne sme mohli cielene upravovať súbor vstupných parametrov pre získanie ďalšieho setu výsledkov alebo iného uhla pohľadu na skúmaný problém s prihliadnutím na implikácie reálneho sveta.

2.4 Realizácia riešenia

V tejto časti práce sme sa zamerali na realizáciu riešenia praktickej časti na základe krokov uvedených v predchádzajúcej podkapitole. Samotnú realizáciu sme rozdelili do štyroch kľúčových častí. Najskôr sme upresnili metodiku využívanú naprieč praktickou časťou. Ďalej sme popísali našu metodiku narábania s dátovou vzorkou. V predposlednej časti sme slovne popísali najdôležitejšie pasáže algoritmu pre odhad kauzálneho účinku využitím metódy párovania a pre splnenie ďalších parciálnych cieľov. V záverečnej časti sme upresnili zvolené parametre pre získanie výsledkov uvedených v podkapitole 2.5.

2.4.1 Voľba konkrétnej metodiky

Metódou, ktorú sme sa rozhodli aplikovať pre odhad kauzálnych účinkov, bola metóda párovania. V súčasnej dobe ide o populárnu metódu využívanú na odhad kauzálnych účinkov. Metóda taktiež predstavovala vhodný prístup pre vykonanie analýzy v oblasti našej problematiky určenia účinku účasti v programe na zamestnanosť na slovenskom trhu práce. Z predmetu riešenia, kde sme uviedli záujem o vyhodnotenie zamestnanosti účastníkov v programe, sme sa prirodzene rozhodli pre typ účinku ATT. Tento typ účinku sa zameriava na vyhodnotenie účinku účasti pre účastníkov v programe. Z dôvodu veľkého množstva rôznorodých relevantných kovariátov sme využili prístup približného párovania na základe skóre pravdepodobnosti intervencie. Pre odhad skóre pravdepodobnosti intervencie sme využili logit a probit model s účelom porovnania merateľných ukazovateľov výsledkov. Ďalej sme zvolili dva párovacie prístupy, ktorými boli párovanie na základe skóre pravdepodobnosti intervencie (PSM) a váženie na základe inverznej pravdepodobnosti intervencie (IPW). V prípade párovania na základe skóre pravdepodobnosti intervencie sme zvolili rôzne párovacie algoritmy s rôznymi špecifikáciami. Pre posúdenie kvality výslednej párovanej vzorky sme zvolili numerickú diagnostiku pomocou SMD a VR. Grafická diagnostika kandidátov pre najlepší model bola založená na porovnaní skóre pravdepodobnosti intervencie pred a po párovaní a na pohľade na grafické znázornenie štandardizovaných rozdielov v stredných hodnotách. Na základe výstupných hodnôt a grafických vizualizácií sme posúdili vyváženosť výslednej párovanej vzorky. V prípade vyhovujúcej vyváženosti sme odhadli účinok intervencie ATT.

2.4.2 Práca s dátovou vzorkou

Dátová vzorka v našej analýze predstavovala potenciálne vstupné údaje. Jednalo sa o anonymizované dáta vygenerované na základe záznamov získaných observačnou štúdiou na nezamestnaných. Táto štúdia bola realizovaná ÚPSVaR dotazníkovou formou a bola na vyžiadanie poskytnutá Ekonomickému ústavu SAV, v spolupráci s ktorým je táto práca písaná. Dátová vzorka nám bola poskytnutá ako enviroment pre softvér R. My sme sa s ňou rozhodli pracovať v programe R-Studio vo verzii 4.2.2 s príjemným užívateľským prostredím. Bližší pohľad na enviroment prezradil, že išlo o súbor o veľkosti okolo 300 MB s viac ako 100 tabuľkami, z ktorých určité tabuľky obsahovali viac ako milión záznamov.

Pre našu analýzu boli najrelevantnejšie 3 tabuľky: tabuľka účasti v programoch, tabuľka informácie o osobách z evidencie nezamestnanosti a tabuľka dátumy evidencie. Tieto tabuľky sme ľudsky interpretovali a popis najdôležitejších premenných uvádzame v prílohe.

A Tabuľka účasti v programoch

V tabuľke bolo 1 049 225 záznamov a obsahovala informácie o účasti jednotlivcov v programoch APTP.

B Tabuľka informácie o osobách z evidencie nezamestnanosti

Tabuľka obsahovala 3 024 168 záznamov získaných dotazníkovou formou obsahujúca informácie o osobách z evidencie nezamestnaných. Obsahovala 168 premenných. Pre nás boli relevantné ovplyvňujúce premenné pre výslednú zamestnanosť zachytené v 25 premenných.

C Tabuľka dátumy evidencie

Tabuľka mala 1 640 866 záznamov a zachytávala informácie o dátumoch evidencie obyvateľov na úrade práce v dňoch od 1. januára 1970. Poskytla adekvátny základ pre získanie vysvetľujúcich premenných spojených s históriou zamestnaneckých stavov jednotlivcov.

Na základe informácií zachytených a popísaných v tabuľkách účasti v programoch, informácie o osobách z evidencie nezamestnanosti a dátumy evidencie sme určili, že dátová vzorka bola dostatočne relevantná pre využitie záznamov za účelom vykonania analýzy kauzálnych účinkov metódou párovania.

2.4.3 Popis fungovania častí algoritmického spracovania analýzy metódou párovania

V tejto časti realizácie riešenia sme sa zamerali na zjednodušenie popisu najdôležitejších častí algoritmu využívaného pre vykonanie analýzy metódou párovania. Časť kódov a funkcií bola využitá z automatizovaného reportovacieho systému, ktorému sa Ekonomický ústav SAV venoval v minulosti a sú k dispozícii na ich webovej stránke [9]. Snažili sme sa využiť relevantné častí kódu, modifikovať čiastočne využiteľných častí kódu a naprogra-

movať chýbajúce funkcie alebo časti algoritmu. Pre algoritmus sme zvolili delenie na dve hlavné časti. Prvá časť kódu pojednávala o plánovaní a príprave vstupov, kým druhá časť sa zaoberala samotnou implementáciou metódy párovania.

A Časť plánovanie a príprava algoritmu

V pasáži kódu zameranej na plánovanie a prípravu sme si najskôr vymedzili súbor vstupných parametrov. Boli nimi nástroj, evaluačné obdobie a horné vekové ohraničenie záznamov. Voľba nástroja predstavovala vymedzenie programu, ktorý sme pri našej analýze vzali do úvahy. Evaluačné obdobie predstavovalo uvážený časový interval pre obdobie účasti v programe. Tretím parametrom bolo horné vekové ohraničenie záznamov alebo horná veková hranica účastníkov v programe. Zavedenie tohto parametra malo veľké opodstatnenie, pretože sme pre našu analýzu vzali do úvahy zamestnanosť po absolvovaní programu. V prípade účastníkov blízko veku odchodu do dôchodku by sa účasť v programe neodzrkadlila na úspešnosti zamestnanosti. Títo účastníci by sa po účasti ocitli v dôchodku alebo tesne pred dosiahnutím dôchodkového veku. Tento fakt by pôsobil skresľujúco na naše výsledky a neumožňoval by relevantné posúdenie dlhodobej zamestnanosti.

Ďalší krok zahŕňal vytvorenie evaluačnej vzorky alebo vzorky zloženej z relevantných záznamov účastníkov v programe a kontrolnej skupiny. Filtrácia vhodných údajov prebiehala vyradením záznamov z tabuľky účasti v programoch a tabuľky informácií o osobách z evidencie nezamestnanosti na základe zvolených vstupných parametrov a ďalších kritérií. Následne sme záznamy z oboch tabuliek spojili pre zisk evaluačnej vzorky.

V ďalšom kroku sme naprogramovali časť kódu, ktorá slúžila na predikciu dátumu účasti v programe pre kontrolnú skupinu. Táto informácia bola využívaná v implementačnej fáze algoritmu vo funkcii kontroly prítomnosti v evidencii nezamestnaných.

Ďalšia časť zahŕňala generovanie vysvetľujúcich premenných pre našu evaluačnú vzorku s cieľom získať súbor ďalších premenných s potenciálne výrazným vplyvom na zamestnanosť osoby. V rámci evaluačnej vzorky išlo o premennú obdobie nezamestnanosti, ktorá zachytávala počet dní nezamestnanosti od začiatku evaluačného obdobia. Vysoké kladné hodnoty reprezentovali dlhodobo nezamestnanú osobu, ktorá mohla stratiť pracovné návyky, čo mohlo výrazným spôsobom vplývať na budúcu zamestnanosť. Ďalej išlo o regionálnu mieru nezamestnanosti, ktorá reprezentovala nezamestnanosť v regióne osoby v roku evaluačného obdobia. Hladina nezamestnanosti v regióne predstavovala dobrý od-

had množstva voľnej pracovnej sily v danej oblasti. Spojením záznamov z evaluačnej vzorky a tabuľky dátumov evidencie sme získali dátumy príchodov a odchodov z evidencie nezamestnanosti. Vzali sme do úvahy najviac 11 období nezamestnanosti osoby, ktoré boli reprezentované číselnou hodnotou za názvami premenných. Za hlavné dôvody zohľadnenia týchto parametrov o nezamestnanosti sme považovali lepšiu kontrolu individuálnych rozdielov medzi jednotlivcami, zlepšenie kvality párovania a zníženie výberového vychýlenia.

Následne sme sa zamerali na vymedzenie vybraných charakteristík alebo ovplyvňujúcich premenných zo súboru viac ako 170 premenných dostupných pre skupinu účastníkov a kontrolnú skupinu. Snažili sme sa o zohľadnenie premenných, ktoré podľa našich uvážení najlepšie chrakterizovali rôznorodosť záznamov. Porovnanie hodnôt vybraných premenných sme reprezentovali v tabuľkovom formáte. Tento krok predstavoval poslednú časť prípravnej fázy algoritmu.

B Časť implementácia metódy párovania

Ďalej sme sa presunuli na implementačnú časť algoritmu pre metódu párovania.

Najskôr sme naprogramovali funkciu kontroly prítomnosti v evidencii nezamestnanosti v určitom kvartáli. V algoritme sme využívali dátumy evidencie z histórie nezamestnanosti osôb. V prípade, že sa osoba nachádzala v evidencii nezamestnaných pre daný kvartál, bola jej pripísaná hodnota 0, ak bola zamestnaná, získala hodnotu 1. Informácie sme uložili v premenných status prítomnosti/absencie v registry nezamestnaných s príponou čísla kvartálu od začiatku evaluačného obdobia. Jednotlivcov, pre ktorých sme pre niektorý z kvartálov nemali informácie o ich zamestnanosti, sme vyradili.

Následne sme sa zamerali na prípravu a porovnanie vstupov do metódy párovania alebo výber podmnožiny parametrov z evaluačnej vzorky zahrnutých ako ovplyvňujúce premenné. Tieto premenné sme v ďalšom kroku využili ako vstupné špecifikácie pre odhad skóre pravdepodobnosti intervencie a párovanie pomocou benchmarkovej funkcie. Porovnaním výstupných údajov tohto modelu sme zvolili najlepší súbor charakteristík pre vstup do ďalších vybraných modelov.

Ďalej sme vykonali párovanie rôznymi modelmi prístupom párovania na základe skóre pravdepodobnosti intervencie pre najlepší súbor vstupných charakteristík. Na základe vstupných parametrov funkcie matchit v programe R sme sa rozhodli uskutočniť numerické porovnanie pre rôzne špecifikácie parametrov modelov na základe merateľných

ukazovateľov (výpočtový čas, výstupné údaje modelu). Pre najlepší prístup na základe numerickej diagnostiky sme taktiež uskutočnili grafickú diagnostiku porovnaním skóre pravdepodobnosti intervencie pred a po párovaní a vykreslením grafu štandardizovaných rozdielov v stredných hodnotách.

V ďalšom kroku sme odhadli typ účinku ATT štyri kvartály (1 rok) pred nástupom do programu a 16 kvartálov (4 roky) po nástupe do programu. Následne sme grafickou vizualizáciou podielu účastníkov a kontrolnej skupiny mimo evidencie nezamestnaných porovnali mieru nezamestnanosti pre jednotlivé skupiny. Účinnosť účasti v programe sme vyhodnotili aj pomocou váženia na základe inverznej pravdepodobnosti intervencie. Tento prístup predstavoval alternatívny párovací prístup k párovaniu na základe skóre pravdepodobnosti intervencie. Získanými váhami pre účinok intervencie ATT sme uskutočnili preváženie skóre pravdepodobnosti intervencie. V ďalšom kroku sme účinok účasti získaný oboma prístupmi graficky porovnali.

V záverečnej časti sme sa rozhodli pozrieť aj na ďalšie časové obdobia blížiace sa k rokom pandémie Covid-19. Zaujímali nás zmeny vo vplyve účasti v programe na zamestnanosť približujúc sa k neštandardným rokom pre fungovanie pracovného trhu na Slovensku.

2.4.4 Voľba súboru vstupných parametrov, charakteristík a špecifikácií

A Zvolené vstupné parametre

- Nástroj: P054 REPAS
- Horné vekové ohraničenie: 55 rokov
- Evaluačné obdobie: 2016

B Zvolené úpravy pre súbor vstupných charakteristík

1. Kombinácia vybraných charakteristík z tabuľky porovnania a historických informácií o zamestnanosti jednotlivca alebo regiónu.
   Prvý súbor charakteristík: *ent*, *entry_month*, *male*, *married*, *kids*, *slovak*, *noedu*, *primary*, *lsec*, *usec*, *flang*, *drive*, *pc*, *unpast*, *min_urad*, *min_BA*, *UR_region*, *roma_share*, *population_2016*, *age*, *empl.1*, *empl.2*, *empl.3*, *empl.4*

2. Pridanie väčšieho množstva relevantných charakteristík.
   Druhý súbor charakteristík: Prvý súbor charakteristík + *relocate*, *tertiary*, *healthy*, *zaujem_vzdel*, *prax*

3. Odobratie historických informácií o zamestnanosti a ponechanie vybraných charakteristík z bodu 1.
   Tretí súbor charakteristík: Prvý súbor charakteristík - *empl.1*, *empl.2*, *empl.3*, *empl.4*

C Zvolené špecifikácie parametrov metódy párovania na základe skóre pravdepodobnosti intervencie

1. Benchmarková funkcia: Odhad skóre pravdepodobnosti intervencie logit modelom a párovanie metódou najbližšieho suseda s prednastaveným poradím párovania od najvyššieho skóre pravdepodobnosti intervencie po najnižšie bez nahradenia.

2. Odhad skóre pravdepodobnosti intervencie probit modelom a párovanie metódou najbližšieho suseda s prednastaveným poradím párovania od najvyššieho skóre pravdepodobnosti intervencie po najnižšie bez nahradenia.

3. Odhad skóre pravdepodobnosti intervencie logit modelom a párovanie metódou najbližšieho suseda s opačným poradím párovania od najnižšieho skóre pravdepodobnosti intervencie po najvyššie bez nahradenia.

4. Odhad skóre pravdepodobnosti intervencie probit modelom a párovanie metódou najbližšieho suseda s opačným poradím párovania od najnižšieho skóre pravdepodobnosti intervencie po najvyššie bez nahradenia.

5. Odhad skóre pravdepodobnosti intervencie logit modelom a párovanie metódou najbližšieho suseda s nahradením.

6. Odhad skóre pravdepodobnosti intervencie probit modelom a párovanie metódou najbližšieho suseda s nahradením.

7. Odhad skóre pravdepodobnosti intervencie logit modelom a párovanie metódou optimálneho párovania.

8. Odhad skóre pravdepodobnosti intervencie logit modelom a párovanie metódou genetického párovania.

D Ďalšie zvolené evaluačné obdobia

- 2017
- 2018

2.5 Výsledky riešenia a ich interpretácia

2.5.1 Porovnávacia tabuľka

Najskôr uvádzame porovnávaciu tabuľku charakteristík po fáze plánovania a prípravy vstupnej vzorky pri zvolených vstupných parametroch z predchádzajúcej podkapitoly.

Tabuľka 1: Porovnanie účastníkov a kontrolnej skupiny pre vybrané charakteristiky

Popis	Účastníci	Kontrolná_skupina
Základné informácie		
Počet pozorovaní	947	365 997
Vek (v rokoch)	24.9	33.8
Muži	50.9%	51.08%
Predošlá zamestnanosť	4.33%	7.37%
Deti v domácnosti	7.29%	15.63%
Úroveň vzdelania		
Bez vzdelania	0.42%	0.81%
Základné	11.72%	14.46%
Nižšie sekundárne	17.21%	27.86%
Vyššie sekundárne	50.48%	36.68%
Terciárne	20.17%	20.18%
Študijný odbor		
Lekárske a farmaceutické vedy	1.69%	1.47%
Poľnohospodárske, lesnícke a veterinárne vedy	2.96%	4.54%
Prírodné vedy	0.86%	0.72%
Spoločenské vedy a služby	34.21%	27.1%
Technické vedy	20.39%	31.27%
Vedy o kultúre a umení	1.27%	1.16%
Vojenské a bezpečnostné vedy	0.75%	0.35%
Všeobecné vedy a služby	37.91%	33.38%
Zručnosti		
Cudzí jazyk	75.82%	55.21%
Počítačové zručnosti	79.2%	56.04%
Vodičský preukaz	57.13%	50.77%
Kraj		
Banskobystrický	18.9%	13.65%
Bratislavský	1.9%	8.92%
Košický	15.52%	15.65%
Nitriansky	8.45%	12.54%
Prešovský	23.02%	18.34%
Trenčiansky	11.09%	9.83%
Trnavský	9.5%	9.07%
Žilinský	11.62%	11.98%
Národnosť		
Slovenská	88.07%	90.01%
Maďarská	11.19%	8.8%
Česká	0%	0.38%
Rómska	0.11%	0.18%
Iná	0.53%	0.48%
Prítok do programu		
1Q.2016	31.47%	27.71%
2Q.2016	31.36%	28.53%
3Q.2016	14.47%	26.6%
4Q.2016	22.7%	17.16%
Odliv z programu		
1Q.2016	6.14%	27.4%
2Q.2016	11.82%	26.5%
3Q.2016	32.27%	25.33%
4Q.2016	49.77%	20.77%

Z Tabuľky 1 vieme vyčítať, že pomer pozorovaní v účastníckej a kontrolnej skupine predstavuje vhodnú bázu pre vykonanie analýzy kauzálnych účinkov pomocou metódy párovania. Výrazné nevyrovnanosti, ktoré sme sa snažili v ďalších častiach zmierniť, sú viditeľné v premenných vek, vyššia sekundárna úroveň vzdelania, počítačové zručnosti, ovládanie cudzieho jazyka a odliv z programu v poslednom kvartáli roka. Podobnosť medzi skupinami je viditeľná v percentuálnom zastúpení pohlaví a národností. V ostatných charakteristikách sú viditeľné menšie nezrovnalosti, ktorých vyrovnanie sme taktiež uvážili v ďalších častiach.

2.5.2 Porovnanie vyrovnanosti sád charakteristík

Ďalej uvádzame výsledné porovnanie vyrovnanosti pre tri zvolené súbory charakteristík využitím benchmarkovej funkcie. Porovnanie je členené na tri časti. V hornej časti je viditeľné porovnanie pred využitím metódy párovania. V strednej časti je viditeľné porovnanie po využití metódy párovania pre párovaciu vzorku. V dolnej časti sú viditeľné zmeny vo veľkosti vzorky pre kontrolnú skupinu a účastníkov. Zelenou farbou sú zvýraznené priaznivé hodnoty, oranžovou prijateľné hodnoty a červenou nepriaznivé hodnoty v rámci porovnaní vstupných charakteristík a ideálneho rozmedzia hodnôt kľúčových mier numerickej diagnostiky.

Z porovnania pre prvý súbor charakteristík v Tabuľke 2 vieme vyčítať, že pre nevhodné vlastnosti neboli vylúčené žiadne charakteristiky. Z troch sád charakteristík sme dosiahli najnižšiu hodnotu SMD $-\,0,0000$ pre premennú vzdialenosť, v ktorej boli uložené odhadnuté skóre pravdepodobnosti intervencie pre jednotlivé záznamy. SMD v absolútnej hodnote pre ostatné premenné sa znížilo na priaznivé hodnoty menšie ako 0.1. Hodnoty variačného pomeru sa nachádzali taktiež v priaznivom rozmedzí $0,5$ – $1,5$ až na premennú vek, ktorá túto hodnotu mierne prevyšovala. Premenné, pre ktoré bola hodnota variačného pomeru nahradená bodkou, predstavovali kategorické premenné. Zvolením prvého súboru charakteristík sme dosiahli veľmi dobrú celkovú vyrovnanosť.

Tabuľka 2: Sumár porovnania vyrovnanosti parametrov pri použití metódy párovania (prvý súbor charakteristík)

Summary of Balance for All Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max
distance	0.0458	0.0025	1.1375	13.0605	0.4290	0.7055
ent	166.7558	30.2449	0.5766	0.8533	0.1259	0.3035
male	0.5085	0.5108	-0.0047	.	0.0023	0.0023
married	0.1934	0.3653	-0.4351	.	0.1719	0.1719
kids	0.0729	0.1564	-0.3208	.	0.0834	0.0834
slovak	0.8805	0.9003	-0.0609	.	0.0197	0.0197
noedu	0.0042	0.0081	-0.0604	.	0.0039	0.0039
primary	0.1173	0.1447	-0.0849	.	0.0273	0.0273
lsec	0.1723	0.2786	-0.2815	.	0.1063	0.1063
usec	0.5042	0.3668	0.2748	.	0.1374	0.1374
flang	0.7579	0.5522	0.4803	.	0.2057	0.2057
drive	0.5719	0.5077	0.1297	.	0.0642	0.0642
pc	0.7918	0.5605	0.5695	.	0.2313	0.2313
unpast	0.4197	0.8222	-0.8158	.	0.4026	0.4026
min_urad	11.9256	11.7187	0.0180	1.0545	0.0072	0.0194
min_BA	231.4132	207.6898	0.1898	0.9128	0.0542	0.1037
UR_region	10.1447	9.3776	0.2086	0.9523	0.0530	0.0994
roma_share	0.0925	0.1013	-0.0527	0.8234	0.0119	0.0266
population_2016	13920.9228	17554.8406	-0.1707	0.7606	0.0531	0.0926
age	24.9556	33.8277	-0.9944	0.7552	0.2283	0.5214
empl.1	0.2717	0.9633	-1.5548	.	0.6916	0.6916
empl.2	0.3858	0.9688	-1.1976	.	0.5830	0.5830
empl.3	0.5867	0.9783	-0.7952	.	0.3916	0.3916
empl.4	0.6882	0.9842	-0.6391	.	0.2961	0.2961

Summary of Balance for Matched Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
distance	0.0458	0.0458	-0.0000	0.9998	0.0000	0.0021	0.0002
ent	166.7558	162.1945	0.0193	0.9311	0.0154	0.0518	0.7122
male	0.5085	0.5180	-0.0190	.	0.0095	0.0095	0.9832
married	0.1934	0.1829	0.0268	.	0.0106	0.0106	0.7600
kids	0.0729	0.0761	-0.0122	.	0.0032	0.0032	0.4919
slovak	0.8805	0.8816	-0.0033	.	0.0011	0.0011	0.6421
noedu	0.0042	0.0042	0.0000	.	0.0000	0.0000	0.0085
primary	0.1173	0.1099	0.0230	.	0.0074	0.0074	0.5683
lsec	0.1723	0.1977	-0.0672	.	0.0254	0.0254	0.8174
usec	0.5042	0.4884	0.0317	.	0.0159	0.0159	0.8732
flang	0.7579	0.7632	-0.0123	.	0.0053	0.0053	0.6688
drive	0.5719	0.5909	-0.0385	.	0.0190	0.0190	0.9015
pc	0.7918	0.8044	-0.0312	.	0.0127	0.0127	0.5207
unpast	0.4197	0.4302	-0.0214	.	0.0106	0.0106	0.8697
min_urad	11.9256	11.8458	0.0069	1.1765	0.0127	0.0613	1.0328
min_BA	231.4132	226.8368	0.0366	0.8257	0.0310	0.0899	1.0945
UR_region	10.1447	9.9323	0.0578	0.9332	0.0217	0.0560	0.9745
roma_share	0.0925	0.0950	-0.0148	0.8739	0.0094	0.0507	0.8101
population_2016	13920.9228	15075.7918	-0.0542	0.8821	0.0242	0.0518	0.8254
age	24.9556	24.1776	0.0872	1.5053	0.0275	0.0497	0.7384
empl.1	0.2717	0.2717	0.0000	.	0.0000	0.0000	0.0000
empl.2	0.3858	0.3700	0.0326	.	0.0159	0.0159	0.3757
empl.3	0.5867	0.5825	0.0086	.	0.0042	0.0042	0.6869
empl.4	0.6882	0.6956	-0.0160	.	0.0074	0.0074	0.7188

Sample Sizes:

	Control	Treated
All	365551	946
Matched	946	946
Unmatched	364605	0
Discarded	0	0

Porovnaním druhého súboru charakteristík v Tabuľke 3 je viditeľné, že pre nevhodné vlastnosti neboli vylúčené žiadne charakteristiky. Pohľadom na všetky tri charakteristiky sme dosiahli druhú najvyššiu hodnotu SMD $0,0001$ pre premennú vzdialenosť. SMD v absolútnej hodnote pre ostatné premenné sa znížilo na priaznivé hodnoty menšie ako $0,1$. Hodnoty variačného pomeru boli pre všetky premenné prijateľné. Zvolením druhého súboru charakteristík sme dosiahli veľmi dobrú celkovú vyrovnanosť.

Tabuľka 3: Sumár porovnania vyrovnanosti parametrov pri použití metódy párovania (druhý súbor charakteristík)

Summary of Balance for All Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max
distance	0.0467	0.0025	1.1205	13.7939	0.4312	0.7061
ent	166.7558	30.2449	0.5766	0.8533	0.1259	0.3035
male	0.5085	0.5108	-0.0047	.	0.0023	0.0023
married	0.1934	0.3653	-0.4351	.	0.1719	0.1719
kids	0.0729	0.1564	-0.3208	.	0.0834	0.0834
slovak	0.8805	0.9003	-0.0609	.	0.0197	0.0197
noedu	0.0042	0.0081	-0.0604	.	0.0039	0.0039
primary	0.1173	0.1447	-0.0849	.	0.0273	0.0273
lsec	0.1723	0.2786	-0.2815	.	0.1063	0.1063
usec	0.5042	0.3668	0.2748	.	0.1374	0.1374
flang	0.7579	0.5522	0.4803	.	0.2057	0.2057
drive	0.5719	0.5077	0.1297	.	0.0642	0.0642
pc	0.7918	0.5605	0.5695	.	0.2313	0.2313
unpast	0.4197	0.8222	-0.8158	.	0.4026	0.4026
min_urad	11.9256	11.7187	0.0180	1.0545	0.0072	0.0194
min_BA	231.4132	207.6898	0.1898	0.9128	0.0542	0.1037
UR_region	10.1447	9.3776	0.2086	0.9523	0.0530	0.0994
roma_share	0.0925	0.1013	-0.0527	0.8234	0.0119	0.0266
population_2016	13920.9228	17554.8406	-0.1707	0.7606	0.0531	0.0926
age	24.9556	33.8277	-0.9944	0.7552	0.2283	0.5214
relocate	0.8837	0.7797	0.3245	.	0.1040	0.1040
healthy	0.9894	0.9699	0.1911	.	0.0195	0.0195
zaujem_vzdel	0.2357	0.2345	0.0028	.	0.0012	0.0012
prax	1.2569	4.0817	-0.6371	0.0000	0.0131	0.0666
empl.1	0.2717	0.9633	-1.5548	.	0.6916	0.6916
empl.2	0.3858	0.9688	-1.1976	.	0.5830	0.5830
empl.3	0.5867	0.9783	-0.7952	.	0.3916	0.3916
empl.4	0.6882	0.9842	-0.6391	.	0.2961	0.2961

Summary of Balance for Matched Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
distance	0.0467	0.0467	0.0001	1.0010	0.0000	0.0021	0.0004
ent	166.7558	152.9852	0.0582	0.8591	0.0203	0.0772	0.7970
male	0.5085	0.5148	-0.0127	.	0.0063	0.0063	0.9388
married	0.1934	0.1945	-0.0027	.	0.0011	0.0011	0.7413
kids	0.0729	0.0677	0.0203	.	0.0053	0.0053	0.4838
slovak	0.8805	0.8721	0.0261	.	0.0085	0.0085	0.6780
noedu	0.0042	0.0074	-0.0489	.	0.0032	0.0032	0.1792
primary	0.1173	0.1142	0.0099	.	0.0032	0.0032	0.5551
lsec	0.1723	0.1564	0.0420	.	0.0159	0.0159	0.6690
usec	0.5042	0.5169	-0.0254	.	0.0127	0.0127	0.8161
flang	0.7579	0.7505	0.0173	.	0.0074	0.0074	0.6639
drive	0.5719	0.5772	-0.0107	.	0.0053	0.0053	0.8909
pc	0.7918	0.7875	0.0104	.	0.0042	0.0042	0.5623
unpast	0.4197	0.3996	0.0407	.	0.0201	0.0201	0.9489
min_urad	11.9256	12.4178	-0.0429	1.0776	0.0171	0.0592	1.0211
min_BA	231.4132	222.5712	0.0707	0.9024	0.0208	0.0624	1.1353
UR_region	10.1447	9.8715	0.0743	0.9469	0.0203	0.0412	1.0104
roma_share	0.0925	0.0917	0.0050	0.9157	0.0069	0.0624	0.8159
population_2016	13920.9228	13438.8319	0.0226	1.1061	0.0116	0.0285	0.8308
age	24.9556	24.9450	0.0012	1.1339	0.0265	0.0729	0.8381
relocate	0.8837	0.8943	-0.0330	.	0.0106	0.0106	0.5144
healthy	0.9894	0.9863	0.0310	.	0.0032	0.0032	0.2171
zaujem_vzdel	0.2357	0.2230	0.0299	.	0.0127	0.0127	0.7222
prax	1.2569	1.1321	0.0281	1.2726	0.0030	0.0148	0.4897
empl.1	0.2717	0.2727	-0.0024	.	0.0011	0.0011	0.0024
empl.2	0.3858	0.4027	-0.0347	.	0.0169	0.0169	0.4386
empl.3	0.5867	0.6004	-0.0279	.	0.0137	0.0137	0.7621
empl.4	0.6882	0.7008	-0.0274	.	0.0127	0.0127	0.7987

Sample Sizes:

	Control	Treated
All	365551	946
Matched	946	946
Unmatched	364605	0
Discarded	0	0

Pri pohľade na tretí súbor charakteristík v Tabuľke 4 vieme vyčítať, že všetky charakteristiky mali vhodné vlastnosti pre to, aby mohli byť vstupné premenné. Z troch

sád charakteristík sme dosiahli najvyššiu hodnotu SMD $0,0002$ pre premennú vzdialenosť. SMD v absolútnej hodnote pre ostatné premenné sa znížilo na priaznivé hodnoty menšie ako $0,1$. Hodnoty variačného pomeru sa nachádzali taktiež v priaznivom rozmedzí $0,8 - 1,5$ až na premennú vek, ktorej hodnota sa vyšplhala až na $1,6277$. Táto hodnota predstavovala najextrémnejšiu z vyčíslených hodnôt variačného pomeru pre všetky charakteristiky. Zvolením tretieho súboru charakteristík sme z tohto dôvodu dosiahli pomerne slabú celkovú vyrovnanosť.

Tabuľka 4: Sumár porovnania vyrovnanosti parametrov pri použití metódy párovania (tretí súbor charakteristík)

Summary of Balance for All Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max
distance	0.0143	0.0026	0.5988	12.2732	0.3555	0.5516
ent	166.7558	30.2449	0.5766	0.8533	0.1259	0.3035
male	0.5085	0.5108	-0.0047	.	0.0023	0.0023
married	0.1934	0.3653	-0.4351	.	0.1719	0.1719
kids	0.0729	0.1564	-0.3208	.	0.0834	0.0834
slovak	0.8805	0.9003	-0.0609	.	0.0197	0.0197
noedu	0.0042	0.0081	-0.0604	.	0.0039	0.0039
primary	0.1173	0.1447	-0.0849	.	0.0273	0.0273
lsec	0.1723	0.2786	-0.2815	.	0.1063	0.1063
usec	0.5042	0.3668	0.2748	.	0.1374	0.1374
flang	0.7579	0.5522	0.4803	.	0.2057	0.2057
drive	0.5719	0.5077	0.1297	.	0.0642	0.0642
pc	0.7918	0.5605	0.5695	.	0.2313	0.2313
unpast	0.4197	0.8222	-0.8158	.	0.4026	0.4026
min_urad	11.9256	11.7187	0.0180	1.0545	0.0072	0.0194
min_BA	231.4132	207.6898	0.1898	0.9128	0.0542	0.1037
UR_region	10.1447	9.3776	0.2086	0.9523	0.0530	0.0994
roma_share	0.0925	0.1013	-0.0527	0.8234	0.0119	0.0266
population_2016	13920.9228	17554.8406	-0.1707	0.7606	0.0531	0.0926
age	24.9556	33.8277	-0.9944	0.7552	0.2283	0.5214

Summary of Balance for Matched Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
distance	0.0143	0.0143	0.0002	1.0035	0.0000	0.0021	0.0006
ent	166.7558	173.6850	-0.0293	0.6493	0.0477	0.0909	1.0756
male	0.5085	0.4989	0.0190	.	0.0095	0.0095	0.9452
married	0.1934	0.1977	-0.0107	.	0.0042	0.0042	0.7761
kids	0.0729	0.0719	0.0041	.	0.0011	0.0011	0.5081
slovak	0.8805	0.8795	0.0033	.	0.0011	0.0011	0.6551
noedu	0.0042	0.0085	-0.0652	.	0.0042	0.0042	0.1955
primary	0.1173	0.1047	0.0394	.	0.0127	0.0127	0.5847
lsec	0.1723	0.1903	-0.0476	.	0.0180	0.0180	0.7250
usec	0.5042	0.4926	0.0233	.	0.0116	0.0116	0.8605
flang	0.7579	0.7717	-0.0321	.	0.0137	0.0137	0.7675
drive	0.5719	0.5666	0.0107	.	0.0053	0.0053	0.9293
pc	0.7918	0.7844	0.0182	.	0.0074	0.0074	0.7576
unpast	0.4197	0.4281	-0.0171	.	0.0085	0.0085	0.6555
min_urad	11.9256	12.0365	-0.0097	0.9218	0.0086	0.0233	1.0526
min_BA	231.4132	236.0732	-0.0373	0.9135	0.0295	0.0655	1.1111
UR_region	10.1447	10.1584	-0.0037	0.9971	0.0160	0.0264	0.9839
roma_share	0.0925	0.0954	-0.0173	0.8969	0.0083	0.0423	0.8724
population_2016	13920.9228	14650.0962	-0.0343	0.9960	0.0179	0.0455	0.7805
age	24.9556	24.9281	0.0031	1.6277	0.0474	0.1268	0.6957

Sample Sizes:

	Control	Treated
All	365551	946
Matched	946	946
Unmatched	364605	0
Discarded	0	0

Na základe porovnania vyrovnanosti pre všetky tri súbory charakteristík v Tabuľke 2, Tabuľke 3 a Tabuľke 4 sme za najvhodnejšiu považovali prvý súbor charakteristík. Tento súbor napriek mierne zvýšenej hodnote VR pre premennú vek dosiahla najlepšie vyrovnanie medzi skupinami. Tieto charakteristiky predstavovali súbor vstupných premenných do modelov vytýčených v predchádzajúcej podkapitole.

2.5.3 Numerické porovnanie modelov párovania na základe skóre pravdepodobnosti intervencie

Ďalej v Tabuľke 5 uvádzame vybrané merateľné ukazovatele efektivity pre prvý súbor charakteristík využitím vybraných špecifikácií modelov. Zvolené merateľné ukazovatele boli výpočtový čas pre vykonanie párovania (Time), SMD pre premennú vzdialenosť (SMD distance), najvyššia absolútna hodnota SMD (High SMD) spolu s premennou a najvzdialenejšia hodnota variačného pomeru od jednotky (High VR) taktiež s názvom premennej. Rovnako sme uviedli aj veľkosť vzoriek pre kontrolnú skupinu (Match C) a účastníkov (Match T) po využití rôznych modelov a ich špecifikácií.

Tabuľka 5: Prehľad hodnôt merateľných ukazovateľov pre jednotlivé modely

Model	Time [s]	SMD distance	High SMD	High VR	Match C	Match T
m.1	7,65	0,0000	0,0942 unpast	1,3715 age	946	946
m.2	6,15	-0,0000	0,0872 age	1,5053 age	946	946
m.3	6,97	-0,0000	0,0873 age	1,5093 age	946	946
m.4	9,12	-0,0000	0,0921 unpast	1,3553 age	946	946
m.5	6,20	-0,0000	0,0899 age	1,4948 age	914	946
m.6	8,18	-0,0000	0,0921 unpast	1,3232 age	911	946

Z Tabuľky 5 je viditeľné, že na základe hodnôt SMD vzdialenosti dosiahli všetky modely excelentné vyrovnanie medzi skupinami vhodným výberom záznamov z kontrolnej skupiny. Najrýchlejším modelom bol model m.2. Najvyššie absolútne hodnoty SMD pre všetky modely spadali pod hraničnú hodnotu priaznivosti $0,1$. Najnižšiu hodnotu opäť dosiahol model m.2. Hornú hraničnú hodnotu $1,5$ pre VR mierne prekročili modely m.3 a m.2 a tesne sa k nej priblížil model m.5. Z najextrémnejších hodnôt VR sa najbližšie k ideálnej hodnote 1 dostal model m.6. Pri bližšom pohľade na modely m.5 a m.6, ktorým sme umožnili párovanie s nahradením, sme zistili, že pri párovaní využili iba 914 a 911

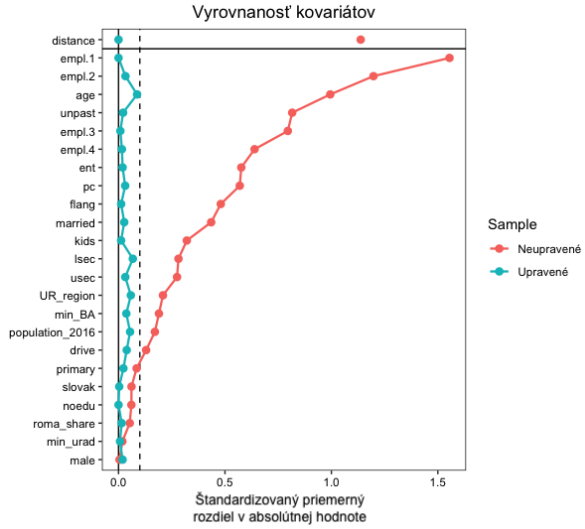
kontrolných záznamov. Spôsob párovania využitím nahradenia znižuje efektívnu veľkosť vzorky a zvyšuje rozptyl pri neskorších výpočtoch ATT. Na základe informácií z popisnej tabuľky sme taktiež vedeli, že máme k dispozícii dostatočne veľkú vzorku kontrolnej skupiny, čo docielilo priaznivé vyrovnanie aj v modeloch bez využitia nahradenia. Porovnaním všetkých merateľných ukazovateľov sme za najlepšie zvolili modely m.2 a m.3. Napriek tomu, že dosahovali mierne zvýšené hodnoty VR pre premennú vek sa im v ostatných ukazovateľoch podarilo dosiahnuť najlepšie výsledky. Modely taktiež využívali takú veľkosť vzorky pre kontrolnú skupinu, ktorá presnosť konečného odhadu získaného z efektívnej vzorky neznížila.

Vo zvolených špecifikáciách parametrov metódy párovania na základe skóre pravdepodobnosti intervencie sme uviedli aj využitie optimálneho a genetického spôsobu párovania. Tieto dva spôsoby mali predstavovať vhodnejší spôsob párovania v porovnaní s chamtivou metódou najbližšieho suseda. Výsledky týchto metód neuvádzame v tabuľke prehľadu zvolených modelov, pretože pri veľkosti vzorky okolo 366 000 záznamov a počte 20 kovariátov sa výpočtová náročnosť týchto metód vyšplhala na niekoľko hodín a vyžadovala si veľké využitie RAM počítača. Spomínané metódy poskytujú efektívnejší spôsob vysporiadania sa s párovaním menších dátových vzoriek alebo čiastkových vzoriek veľkých dátových sád. V zámere praktickej časti sme uviedli zameranie na celú dátovú vzorku, a z tohto dôvodu sme nevyužili segmentáciu vzorky pre benefit optimálneho a genetického párovania. Ďalej sme si uvedomili aj fakt, že vo všetkých modeloch predstavovali vek a predošlý výskyt v evidencii nezamestnaných problematické premenné. Uváženie ich vynechania z vybraných charakteristík mohlo so sebou priniesť lepšiu vyváženosť medzi skupinami. Rozhodli sme sa ale vytýčené tri súbory charakteristík pre zachovanie konzistencie informácií v diplomovej práci nemeniť.

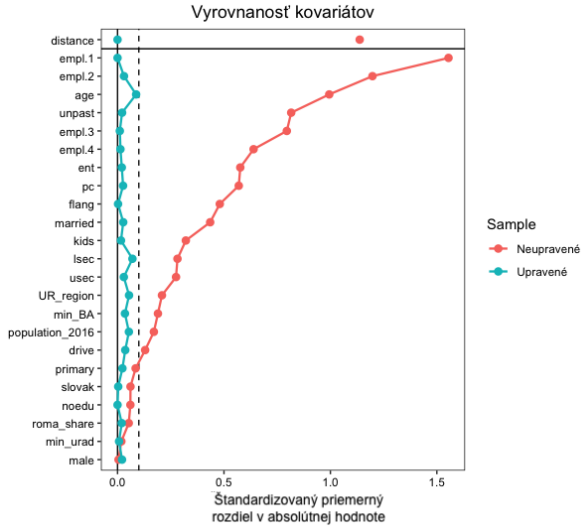
2.5.4 Grafické porovnanie modelov párovania na základe skóre pravdepodobnosti intervencie

Modely m.2 a m.3, ktoré dosahovali najlepšie výsledky, sme podrobili grafickej diagnostike. Na vyrovnanosť kovariátov pred a po využití metódy párovania sme sa pozreli pomocou štandardizovaného priemerného rozdielu v absolútnej hodnote pre jednotlivé premenné. Červená farba reprezentovala neupravené hodnoty pred využitím metódy

párovania a modrá upravené hodnoty po využití metódy párovania.



Obrázok 7: Grafické znázornenie vyrovnanosti kovariátov modelu m.2.

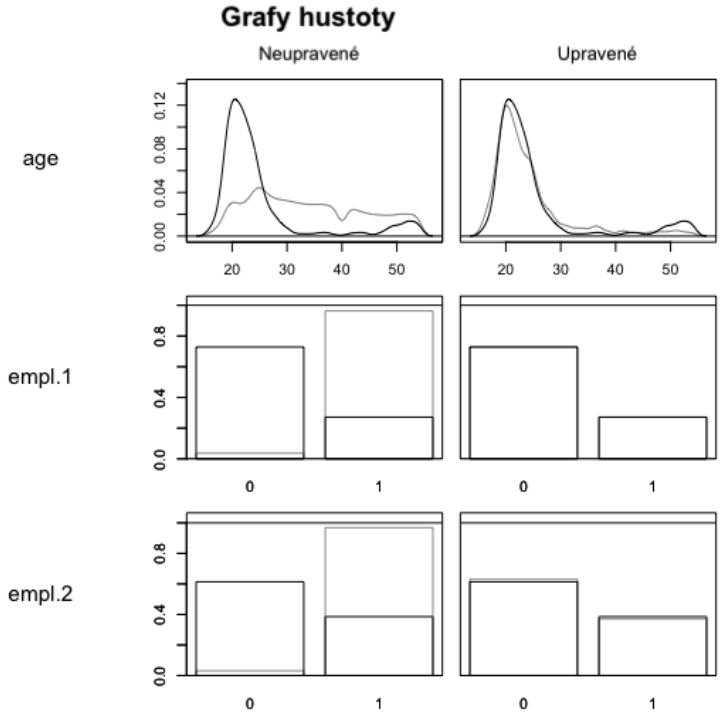


Obrázok 8: Grafické znázornenie vyrovnanosti kovariátov modelu m.3.

Pri pohľade na grafickú diagnostiku oboch modelov na Obrázku 7 a Obrázku 8 bolo viditeľné výborné zmenšenie rozdielov medzi charakteristikami skupín. Pri porovnaní modelov medzi sebou boli viditeľné veľmi malé rozdiely. Uvážením grafickej aj numerickej diagnostiky poskytol model m.2 lepšie výsledky pre výpočtový čas, štandardizovaný priemerný rozdiel a variačný pomer. Z tohto dôvodu sme sa rozhodli vybrať m.2 ako náš finálny model pre odhad ATT.

Z Obrázku 7 je možné vyčítať, že najväčšie rozdiely medzi účastníkmi a kontrolnou skupinou pred párovaním boli v premenných vek a status prítomnosti/absencie v registry nezamestnaných jeden a dva štvrťroky pred účasťou. Vizuálnym preskúmaním týchto premenných pomocou grafov hustôt pred a po využití metódy párovania je možné ešte dôkladnejšie rozlíšiť „vyrovnávaciu" silu metódy. Čiernou farbou sú reprezentované hodnoty pre účastníkov a sivou farbou hodnoty pre kontrolnú skupinu.

Z Obrázku 9 je viditeľné ako sa výrazné rozdiely medzi kontrolnou skupinou a účastníkmi zmiernili voľbou vhodných záznamov z kontrolnej skupiny na základe skóre pravdepodobnosti intervencie. V prípade premenných status prítomnosti/absencie v registry nezamestnaných sa rozdiely natoľko zmenšili, že ich graficky už nie je možné rozlíšiť.



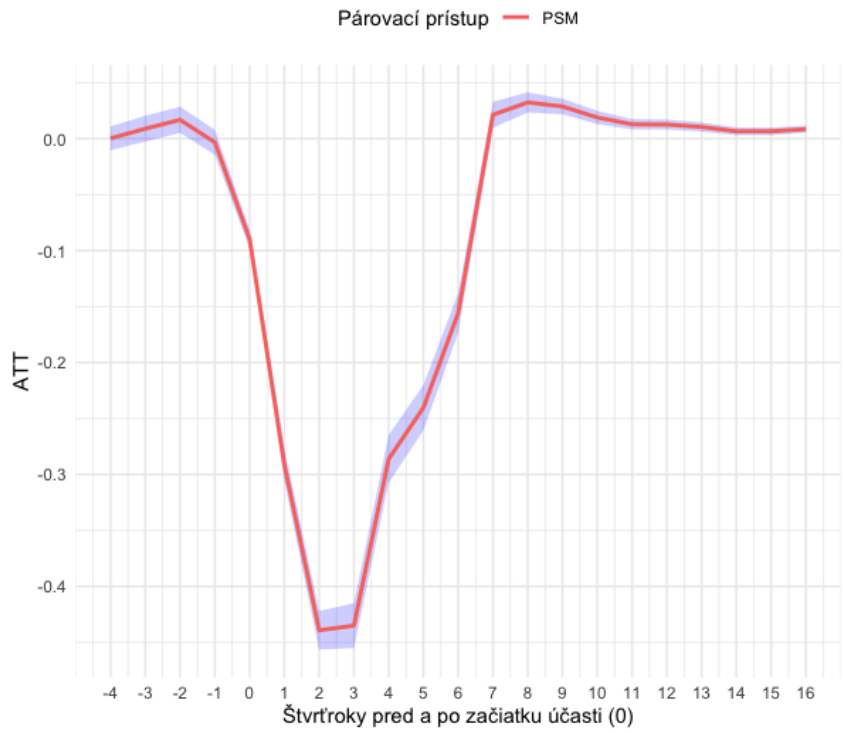
Obrázok 9: Porovnanie hustoty vybraných kovariátov pred a po využití metódy párovania.

2.5.5 Odhad účinku ATT využitím PSM

Ďalej uvádzame odhad ATT využitím prvého súboru vstupných charakteristík do modelu m.2 pre zvolené evaluačné obdobie 2016. Červenou farbou sú reprezentované odhadnuté hodnoty a modrou farbou je reprezentovaná smerodajná odchýlka.

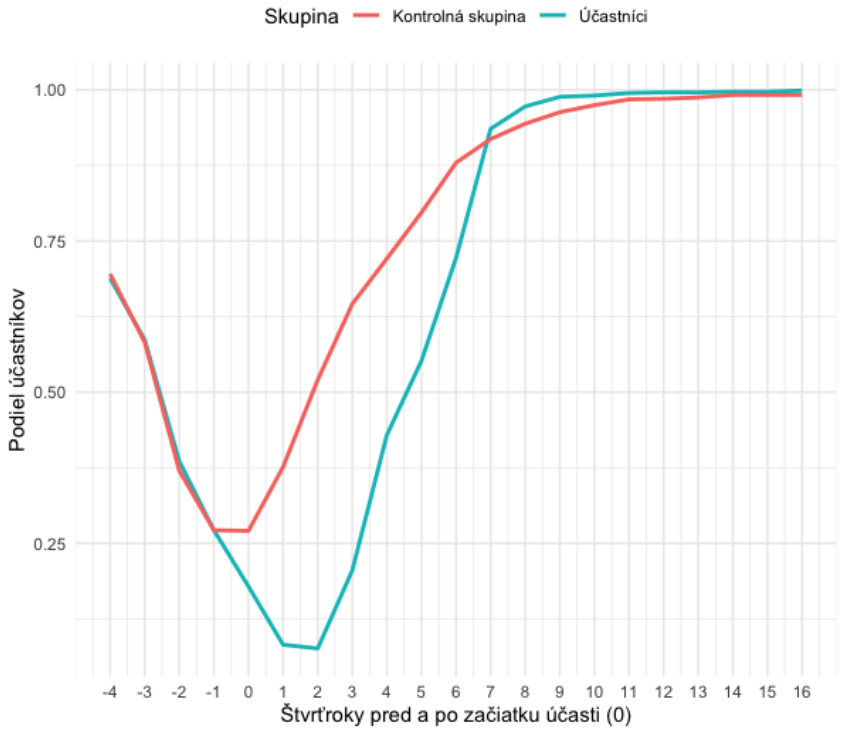
Z Obrázku 10 je v kvartáloch predchádzajúcich účasti v programe viditeľný mierny nárast efektu do druhého kvartálu pred účasťou. Následne až po kvartál 2 dochádza k obrovskému poklesu hodnôt ATT. Ten je spôsobený tzv. efektom uzavretia (lock-in effect) v opatrení, kedy v dôsledku účasti v programe alebo poklesu úsilia v hľadaní si práce účastníci vykazujú relatívne vyššiu prítomnosť v evidencií nezamestnanosti. Od druhého kvartálu je viditeľný najskôr mierny a neskôr skokový nárast účinku spôsobený postupným odlivom účastníkov z programu. Tento nárast je zmiernený v siedmom kvartáli a svoj vrchol dosiahol v ôsmom kvartáli. Od tohto bodu je viditeľný postupný pokles účinku ATT v nasledujúcich štvrťrokoch. Kladné hodnoty ATT medzi kvartálmi 7 – 16 vieme interpretovať ako pozitívny účinok alebo zvýšenú úroveň zamestnanosti pre účastníkov. Priaznivý účinok účasti v programe je ešte zrejmejší pri pohľade na podiel jednotlivcov mimo evidencie nezamestnaných, ktorý slúži ako proxy miera zamestnanosti. Červenou farbou sú

zachytené hodnoty pre kontrolnú skupinu a modrou farbou hodnoty pre účastníkov.



Obrázok 10: Odhad ATT a smerodajnej odchýlky získaný metódou PSM pre evaluačné obdobie 2016.

Z Obrázku 11 vidíme, že krivky podielu jednotlivcov mimo evidencie nezamestnaných majú totožný charakter medzi kvartálmi -4 až -1. Podobný charakter je viditeľný medzi kvartálmi 7 až 16. Výrazné rozdiely v ostatných kvartáloch sú spôsobené odlišnými kvartálmi odchodu z programu pre kontrolnú a účastnícku skupinu. Kým v kontrolnej skupine sme predpokladali odchod z programu v kvartáli 0, účastníkom bolo umožnené vystúpiť z evidencie nezamestnaných aj v neskorších kvartáloch evaluačného obdobia. Stratifikácia dátovej vzorky na základe predchádzajúceho obdobia nezamestnanosti by síce zmiernila rozdiely v dotknutých kvartáloch, ale priniesla by so sebou náročné modifikácie evaluačnej časti algoritmu. Pre nás boli pre zistenie dlhodobého účinku účasti v programe pre účastníkov najrelevantnejšie hodnoty v neskorších kvartáloch, kde je viditeľná podobnosť výsledkov pre účastníkov aj kontrolnú skupinu. Podobne, ako na Obrázku 10, v kvartáli 7 dochádza k jednoznačne viditeľnému pozitívnemu efektu účasti v programe, ktorý pretrváva aj počas nasledujúcich štvrťrokov.



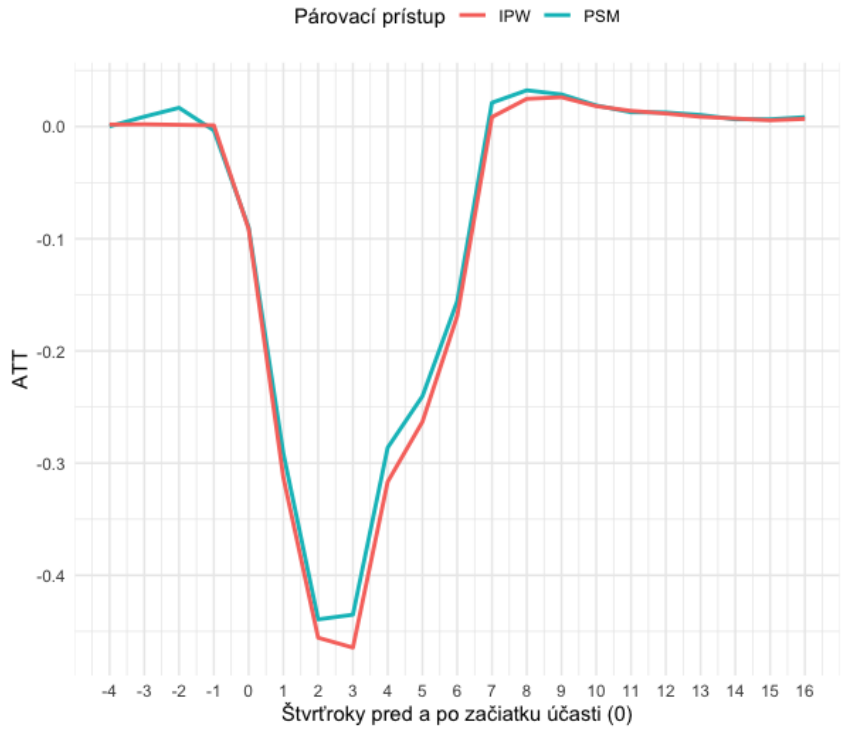
Obrázok 11: Podiel jednotlivcov mimo evidencie nezamestnaných (proxy miery zamestnanosti) pre evaluačné obdobie 2016.

2.5.6 Porovnanie odhadu účinku ATT využitím PSM a IPW

Porovnanie odhadnutého typu účinku ATT získaného metódou párovania na základe skóre pravdepodobnosti intervencie (modrá farba) a váženia na základe inverznej pravdepodobnosti intervencie (červená farba) je viditeľný na Obrázku 12.

Oba párovacie prístupy odhadli podobné hodnoty efektu pre účastníkov programu v jednotlivých kvartáloch. Výraznejšie rozdiely medzi prístupmi boli viditeľné medzi kvartálmi -4 až -1. V tomto období prístup PSM identifikoval pozitívny priemerný efekt intervencie účastníkov pred samotnou účasťou. Takýto efekt ale nemohol nastať z dôvodu, že intervencia ešte neprebehla. Prístup IPW sa týmto falošným efektom nedal oklamať a v tomto období pripísal ATT nulový efekt. Ďalšie rozdiely boli viditeľné medzi kvartálmi 2 a 3, kde prístup IPW odhadol hlbší prepad efektu ATT, čo spôsobilo odlišnosti v náraste efektu medzi prístupmi od 3 po 6 kvartál. Poslednou oblasťou s výraznejšími odlišnosťami bolo obdobie medzi 7. a 9. kvartálom. Oba prístupy vyhodnotili signifikantný pozitívny účinok priemerného efektu intervencie na účastníkov s tým rozdielom, že IPW vyhodnotilo o niečo nižšie hodnoty tohto efektu. Prístup IPW rovnako ako PSM ďalej vyhodnotili

pozitívny vplyv účasti v programe pre neskoršie kvartály bez výraznejších rozdielov.

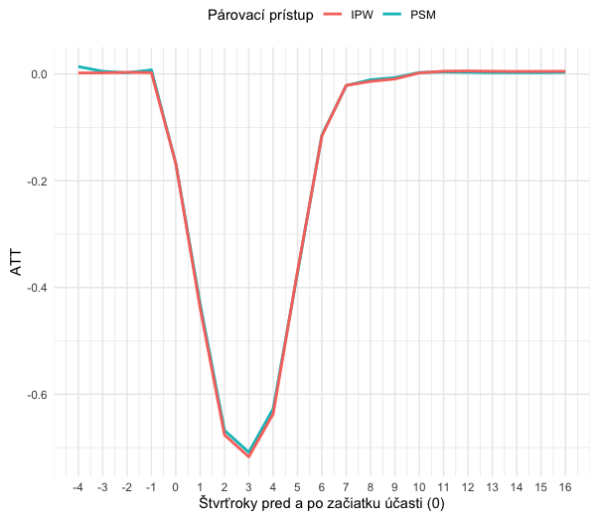


Obrázok 12: Porovnanie hodnôt ATT odhadnutý prístupmi PSM a IPW pre evaluačné obdobie 2016.

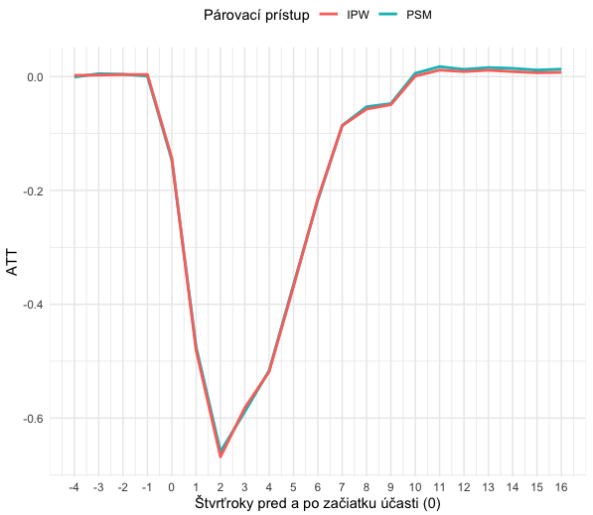
2.5.7 Odhad účinku ATT pre ďalšie evaluačné obdobia

Taktiež uvádzame výsledky pre evaluačné obdobia 2017 na Obrázku 13 a 2018 na Obrázku 14. Pre odhad efektu ATT v oboch rokoch sme pre komparatívne účely využili rovnaké špecifikácie modelu ako pre evaluačné obdobie 2016 aj na úkor mierneho zníženia vyrovnanosti medzi skupinami. Využili sme oba párovacie prístupy PSM a IPW, ktoré priniesli skoro totožné výsledky. Pre obe obdobia bol viditeľný kolísavý efekt okolo nulovej hodnoty 4 kvartály pred vstupom do programu. Ďalej nasledoval efekt uzavretia podobného charakteru ako počas evaluačného obdobia 2016. Výrazné odlišnosti nastali od 7. kvartálu, ktorý pre evaluačné obdobie 2017 spadal do obdobia okolo začiatku roka 2019 a pre evaluačné obdobie 2018 okolo začiatku roka 2020. Práve toto obdobie predstavovalo začiatky pandémie, ktorej negatívne efekty sa odzrkadľovali aj v našich výsledkoch. Účastníci, ktorí do programu vstúpili v roku 2017, pocítili prvé pozitívne výsledky účasti až v 10. kvartáli. V porovnaní s evaluačným obdobím 2016 mal efekt ATT umiernený

vplyv. Pre účastníkov, ktorí vstúpili do programu v roku 2018, bol stále viditeľný negatívny vplyv medzi kvartálmi 7 – 10. Tieto kvartály spadali do obdobia 3Q 2019 – 2Q 2020, ktoré sa prelínalo práve s príchodom ochorenia Covid-19. Trh práce bol v tomto období poznačený neistotou a prechádzal veľkými zmenami. Napriek zvýšenej kvalifikácii si absolventi programu nedokázali nájsť uplatnenie, čo spôsobilo negatívny efekt ATT počas tohto obdobia. Od 10. kvartálu, kedy sa už zamestnávatelia a trh práce prispôsobili prechodu na prácu z domu, sa začal prejavovať mierny pozitívny efekt pre účastníkov programu, ktorý sa tiahol až do konca pozorovaného obdobia.



Obrázok 13: Porovnanie hodnôt ATT odhadnutý prístupmi PSM a IPW pre evaluačné obdobie 2017.



Obrázok 14: Porovnanie hodnôt ATT odhadnutý prístupmi PSM a IPW pre evaluačné obdobie 2018.

3 Zhrnutie

Hlavným cieľom diplomovej práce bola analýza kauzálnych účinkov účasti v konkrétnom programe APTP REPAS na zamestnanosť na slovenskom trhu práce pre rôzne evaluačné obdobia. Kauzálnej analýze predchádzalo oboznámenie sa s problematikou APTP, objasnenie si matematických nástrojov pre zisťovanie kauzálnych účinkov, vymedzenie náplne praktickej časti a premyslenie si postupu riešenia. Pre splnenie vytýčeného cieľa sme využili analýzu párovania. Typ účinku pre odhad efektivity účasti v programe bol ATT. Použitou metodikou bola metóda približného párovania založená na skóre pravdepodobnosti intervencie. Pre odhad skóre pravdepodobnosti intervencie sme využili rôzne varianty logistickej regresie. Samotné párovanie záznamov bolo vykonané dvoma prístupmi, a to párovaním na základe skóre pravdepodobnosti intervencie a váženia na základe inverznej pravdepodobnosti intervencie. Kvalitatívnou analýzou numerickou a grafickou diagnostikou sme určili najlepší súbor vstupných údajov a špecifikácií modelu využitého pre oba prístupy.

Pomocou výslednej párovanej vzorky sme odhadli efekt ATT oboma prístupmi pre evaluačné obdobie 2016. Prístup PSM rovnako ako aj IPW vyhodnotili pozitívny vplyv účasti v programe pre neskoršie kvartály, čo potvrdzuje oneskorený prejav efektu účasti v programoch APTP. Z tohto dôvodu môžeme uviesť, že účasť v programe REPAS počas evaluačného obdobia 2016 priniesla účastníkom pozitívny vplyv v oblasti uplatnenia sa na pracovnom trhu. Pozitívny efekt sa prejavil takmer dva roky od začiatku evaluačného obdobia a pretrvával až do konca nášho pozorovaného obdobia, t. j. 4 roky od začiatku evaluačného obdobia.

Následne sme prístupy PSM A IPW aplikovali aj pre ďalšie evaluačné obdobia 2017 a 2018. Pre tieto evaluačné obdobia bol viditeľný oneskorený a umiernený pozitívny efekt v porovnaní s evaluačným obdobím 2016. Predpokladaným dôvodom bol práve vplyv neštandardného obdobia pandémie Covid-19, ktorý transformoval určité aspekty pracovného trhu na Slovensku.

Počas tvorby práce sme sa dopracovali aj k zaujímavým ponaučeniam a odporúčaniam. Využitie metódy párovania so sebou prinieslo užitočné výsledky, avšak táto metóda nie je schopná príliš robustne reagovať na prítomnosť neviditeľných vplyvov na zmeny v stavoch zamestnanosti. Z tohto dôvodu pohľad na dátovú vzorku cez ďalšiu metódu, napr.

inštrumentálnych premenných, mohol priniesť zaujímavý pohľad na problematiku.

Ďalej sme v praktickej časti uviedli využitie párovacích metód optimálneho a genetického párovania. Neskôr sme zistili, že tieto metódy najlepšie fungujú na menších dátových vzorkách. Z tohto dôvodu pohľad na stav zamestnanosti stratifikovaný podľa vybranej špecifikácie (geografická poloha, pohlavie, vek atď.) mohol poskytnúť alternatívne hodnoty využitím týchto dvoch ďalších metód. Taktiež mohol viesť k zaujímavým poznatkom a zisteniam na jemnejšej úrovni ako pri pohľade na celú evaluačnú vzorku.

Záver

V diplomovej práci sme sa zaoberali vyhodnocovaním účinnosti nástrojov aktívnej politiky trhu práce (APTP) na Slovensku.

V prvej fáze riešenia sme sa zamerali na oboznámenie sa s problematikou politiky trhu práce, porozumenie pojmov, faktov a súvislostí, ako aj súvisiacich problémov, požiadaviek a potrieb. Po detailnom zorientovaní sa v problematike APTP sme sa sústredili na uchopenie problematiky z matematického hľadiska. V tejto časti sme sa zamerali na teóriu a matematické nástroje pre určovanie kauzálnych účinkov, ktorými boli teoretický Rubinov kauzálny model, ako aj v praxi využívaná metóda párovania. Dôležitosť tejto fázy bola v príprave na priame využitie našich nových teoretických poznatkov pre aplikáciu v praxi. To sa nám aj následne podarilo a vytvorením bázy funkcií a príkazov implementujúcich vyššie spomínané no aplikačne prispôsobené matematické nástroje sme mohli vyhodnotiť účinnosti nástrojov aktívnej politiky trhu práce podľa skutočnej dátovej vzorky.

Diplomová práca splnila všetky stanovené ciele a očakávania. Podarilo sa nám získať potrebné teoretické poznanie, vytvoriť priestor pre praktické využitie poznatkov a následne úspešne realizovať analýzu kauzálnych účinkov pre účasť v konkrétnom programe APTP REPAS na zamestnanosť na slovenskom trhu práce.

Zoznam použitej literatúry

[1] Abadie, A., Imbens, G. W. *Bias-Corrected Matching Estimators for Average Treatment Effects*, Journal of Business & Economic Statistics, 2011, vol. 1, p. 1-11.

[2] Abadie, A., Imbens, G. W. *Large sample properties of matching estimators for average treatment effects*, Econometrica, 2006, vol. 74, no. 1, p. 235-267.

[3] Agresti, A. *Categorical Data Analysis*. New Jersey: John Wiley & Sons, 2013.

[4] Austin, P. C., Stuart, E. A. *Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies*. Statistics in medicine, 2015, vol. 34, p. 3661-3679.

[5] Cambridge Dictionary: Dostupné na internete (1.5.2025): www.dictionary.cambridge.org/dictionary/english/labour-market.

[6] Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Routledge, 1988. 567 s. ISBN 9780203771587.

[7] Cross, R. *Phelps, Hysteresis, and the Natural Rate of Unemployment*. Quarterly Journal of Business and Economics, 1986, vol. 25, no. 1, p. 56-64.

[8] Cunningham, S. *Causal Inference: The Mixtape*. Yale University Press, 2021. 584 s.

[9] Dulíková, S., Lafférs, L., Štefánik, M. *Impact evaluation of Slovak active labour market policy programmes*, Institute of Economic Research, Slovak Academy of Sciences: Dostupné na internete (1.5.2025): www.lmevidence.sav.sk/data_uploads/ALMPreps/AJ/P54R.html.

[10] European Commission: Dostupné na internete (1.5.2025): https://employment-social-affairs.ec.europa.eu/european-pillar-social-rights-20-principles_en.

[11] European Commission: Dostupné na internete (1.5.2025): https://employment-social-affairs.ec.europa.eu/policies-and-activities/european-pil

lar-social-rights-building-fairer-and-more-inclusive-european-union/european-pillar-social-rights-action-plan_en.

[12] Eurostat: Dostupné na internete (1.5.2025): www.ec.europa.eu/eurostat/statistics-explained/index.php?title=Adult_learning_-_participants.

[13] Eurostat: Dostupné na internete (1.5.2025): www.ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20211015-1.

[14] Eurostat: Dostupné na internete (1.5.2025): www.ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_market_-_quarterly_statistics.

[15] Eurostat: Dostupné na internete (1.5.2025): www.ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics.

[16] Fay, R. *Enhancing the Effectiveness of Active Labor Market Policies: Evidence from Programme Evaluations in OECD Countries*. OECD Labour Market and Social Policy Occasional Papers, Paríž, 1996.

[17] Galántai, A. *The theory of Newton's method*. Journal of Computational and Applied Mathematics, 2000, vol. 124, p. 25-44.

[18] Greifer, N., Stuart, E. A. *Choosing the Causal Estimand for Propensity Score Analysis of Observational Studies*, Dostupné na internete (1.5.2025): www.arxiv.org/abs/2106.10577.

[19] Greifer, N. *MatchIt: Getting Started*, Dostupné na internete (1.5.2025): www.cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html.

[20] Hosmer, W. D, Lemeshow, S. *Applied Logistic Regression*. New Jersey: John Wiley & Sons, 2005.

[21] Huber, M. *An Introduction to Causal Discovery*. Dostupné na internete (1.5.2025): https://www.arxiv.org/abs/2407.08602.

[22] Karlova, M. *Metody kauzální analýzy pro meření efektivity podpory prodeje*. Masarykova univerzita, 2021.

[23] Maslow, A. H. *A Theory of Human Motivation,*, Psychological Review, 1943, vol. 50, p. 370-396.

[24] Meghanath, G. *Causal analysis overview: Causal inference versus experimentation versus causal discovery*. Dostupné na internete (1.5.2025): https://www.medium.com/data-science-at-microsoft/causal-analysis-overview-causal-inference-versus-experimentation-versus-causal-discovery-d7c4ca99e3e4.

[25] Niño-Zarazúa, M., Torm, N. *Active Labour Market Policies in Asia and the Pacific: A Review of the Literature*, Nations ESCAP, Social Development Division, 1987, vol. 82, p. 387-394.

[26] OECD. Dostupné na internete (1.5.2025): www.oecd.org/en/topics/employment-services.html.

[27] Rosenbaum, P. R. *Model-Based Direct Adjustment*, Journal of the American Statistical Association, 1987, vol. 82, p. 387-394.

[28] Rosenbaum, P. R., Rubin, D. *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, Biometrika, 1983, vol. 70, no. 1., p. 41-55.

[29] Rubin, D. *Bias Reduction Using Mahalanobis-Metric Matching*, International Biometric Society, 1980, vol. 36, no. 2, p. 293-298.

[30] Rubin, D. *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*, Journal of Educational Psychology, 1974, vol. 66, p. 688–701.

[31] Splawa-Neyman, J. *On the Application of Probability Theory to Agricultural Experiments. Essay on Principles*. Roczniki nauk rolniczych, 1923.

[32] Stuart, E. A. *Matching Methods for Causal Inference: A Review and a Look Forward*, Statistical Science, vol. 25, no. 1, p. 1-21.

[33] Ústredie práce, sociálnych vecí a rodiny. Dostupné na internete (1.5.2025): www.upsvr.gov.sk/sluzby-zamestnanosti/nastroje-aktivnych-opatreni-na-trhu-prace.

[34] VanderWeele, T. J. *Principles of confounder selection*, European Journal of Epidemiology, 2019, vol. 34, p. 211-219.

[35] Wright, J. D. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, Elsevier, 2015.

[36] XMind. Dostupné na internete (1.5.2025): https://xmind.app.

[37] Zákon č. 5/2004 Z. z. Zákon o službách zamestnanosti a o zmene a doplnení niektorých zákonov §54: Dostupné na internete (1.5.2025): www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/2004/5/20040415.

Prílohy

Popis premenných tabuľky účasti v programoch

- *klient_id* - číselný id kód osoby
- *entrya* - dátum vstupu do programu
- *exita* - dátum odchodu z programu
- *nastroj* - skratka nástroja využívaného v programe
- *naklady* - náklady spojené s účasťou osoby v programe
- *projekt* - názov projektu, pod ktorým bol program implementovaný

Popis premenných tabuľky informácií o osobách v evidencii nezamestnaných

- *entry_month* - dátum vstupu do evidencie nezamestnaných
- *exit_month* - dátum odchodu z evidencie nezamestnaných
- *male* - binárne premenná reprezentujúca pohlavie
  1: muž 0: žena
- *married* - binárna premenná reprezentujúca rodinný stav
  1: ženatý/vydatá 0: neuzavreté manželstvo
- *kids* - binárna premenná určujúca prítomnosť dieťaťa pod 10 rokov v domácnosti
  1: áno 0: nie
- *slovak* - binárna premenná reprezentujúca príslušnosť osoby k slovenskej národnosti
  1: slovenská národnosť 0: iná národnosť
- *noedu* - binárna premenná reprezentujúca úroveň vzdelania
  1: žiadne vzdelanie 0: určitá úroveň vzdelania
- *primary* - binárna premenná reprezentujúca úroveň vzdelania
  1: základná úroveň vzdelania 0: nižšia úroveň vzdelania

- *lsec* - binárna premenná reprezentujúca úroveň vzdelania
  1: nižšia stredná úroveň vzdelania 0: nižšia úroveň vzdelania

- *usec* - binárna premenná reprezentujúca úroveň vzdelania
  1: vyššia stredná úroveň vzdelania 0: nižšia úroveň vzdelania

- *tertiary* - binárna premenná reprezentujúca úroveň vzdelania
  1: terciárna úroveň vzdelania 0: nižšia úroveň vzdelania

- *flang* - binárna premenná reprezentujúca znalosť cudzieho jazyka
  1: znalosť cudzieho jazyka 0: neznalosť cudzieho jazyka

- *drive* - binárna premenná reprezentujúca vlastníctvo vodičského preukazu (VP)
  1: vlastnenie VP 0: nevlastnenie VP

- *pc* - binárna premenná reprezentujúca počítačové znalosti
  1: prítomnosť počítačových znalostí 0: absencia počítačových znalostí

- *unpast* - binárna premenná reprezentujúca predošlý výskyt v evidencii nezamestnaných
  1: predošlá prítomnosť 0: predošlá neprítomnosť

- *relocate* - binárna premenná reprezentujúca ochotu presťahovať sa
  1: ochota presťahovať sa 0: neochota presťahovať sa

- *healthy* - binárna premenná reprezentujúca zdravotný stav
  1: zdravý 0: iný zdravotný stav

- *zaujem_vzdel* - binárna premenná reprezentujúca záujem o učenie sa
  1: záujem o učenie sa 0: nezáujem o učenie sa

- *min_urad* - čas cesty na najbližší úrad práce (v minútach)

- *min_BA* - čas cesty do najbližšieho krajského mesta podľa miesta bydliska (v minútach)

- *UR_region* - regionálna miera nezamestnanosti na úrovni okresu (v percentách)

- *roma_share* - segregačný podiel Rómov v mieste bydliska osoby (v percentách)

- *population_year* - populácia v mieste bydliska osoby v daný rok
- *age* - vek osoby
- *prax* - premenná reprezentujúca predošlé pracovné skúsenosti