UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



MODELOVANIE PONUKY PRÁCE NA SLOVENSKU V KONTEXTE STARNUTIA POPULÁCIE, EXISTUJÚCICH A OČAKÁVANÝCH MIGRAČNÝCH TOKOV

DIPLOMOVÁ PRÁCA

2025 Bc. Yelyzaveta HORBACHOVA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MODELOVANIE PONUKY PRÁCE NA SLOVENSKU V KONTEXTE STARNUTIA POPULÁCIE, EXISTUJÚCICH A OČAKÁVANÝCH MIGRAČNÝCH TOKOV

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	1114 Aplikovaná matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	Mgr. Miroslav Štefánik, PhD.

Bratislava 2025 **Bc. Yelyzaveta HORBACHOVA**





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Yelyzaveta Horbachova
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: Modelovanie ponuky práce na Slovensku v kontexte starnutia populácie, existujúcich a očakávaných migračných tokov
Modelling the Slovak labour supply in the context of population ageing, existing and expected labour migration

Anotácia: Dynamické mikrosimulačné modely ponúkajú vhodnú platformu pre rozšírenie demografických modelov o dodatočné funkcie. Diplomantka bude mať k dispozicí kód existujúceho model slovenskej ponuky práce SLAMM-microsim s jeho vstupnou databázou. Úlohou bude návrh a kalibrácia funkcií imigrácie a emigrácie pracovnej sily na a zo Slovenského trhu práce. Výsledná verzia modelu by mala byť schopná modelovať demografické procesy, ako aj individuálne rozhodnutia týkajúce sa ekonomickej aktivity, či migrácie. Jeho tvorba vyžaduje zvládnutie odbornej literatúry z viacerých vedných oblastí, počnúc tvorbou mikrosimulačných modelov, cez koncepty demografie, či spoločenskovedného výskumu v oblasti trhu práce.

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

Touto cestou sa chcem poďakovať svojmu vedúcemu diplomovej práce, Mgr. Miroslavovi Štefánikovi, PhD., za jeho odborné vedenie, cenné rady, ochotu a podporu počas celého procesu vypracovania tejto práce.

Abstrakt v štátnom jazyku

HORBACHOVA, Yelyzaveta: Modelovanie ponuky práce na Slovensku v kontexte starnutia populácie, existujúcich a očakávaných migračných tokov [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Miroslav Štefánik, PhD., Bratislava, 2025, 68s.

Modelovanie ponuky práce je dôležitý analytický nástroj, ktorý umožňuje simulovať vývoj trhu práce na základe rôznych scenárov a hodnotiť vplyv politík. Táto diplomová práca sa zameriava na kalibráciu funkcií migrácie pracovnej sily v dynamickom mikrosimulačnom modeli SLAMM, určenom na analýzu trhu práce na Slovensku. Predstavujeme koncept mikrosimulačných prístupov so zameraním na dynamické mikrosimulačné modely. Popisujeme štruktúru modelu SLAMM, jeho vstupné a výstupné údaje, ako aj základné mechanizmy. Hlavnou časťou práce je implementácia vylepšení pravdepodobnostného rozdelenia migrantov podľa dĺžky ich pobytu, ktoré počítame pomocou parametrického a neparametrického modelovania. Porovnávame tieto prístupy z hľadiska presnosti na základe historických dát.

Kľúčové slová: Modelovanie ponuky práce, Dynamické mikrosimulačné modely, SLAMM model, Pravdepodobnostné rozdelenie migrantov, Parametrické a neparametrické modelovanie

Abstract

HORBACHOVA, Yelyzaveta: Modelling the Slovak labour supply in the context of population ageing, existing and expected labour migration [Diploma Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Miroslav Štefánik, PhD., Bratislava, 2025, 68p.

Modeling labor supply is an important analytical tool that allows simulating the development of the labor market based on various scenarios and assessing the impact of policies. This thesis focuses on calibrating migration functions of the labor force in the dynamic microsimulation model SLAMM, which is designed for labor market analysis in Slovakia. We present the concept of microsimulation approaches, focusing on dynamic microsimulation models. We describe the structure of the SLAMM model, its input and output data, as well as its basic mechanisms. The main part of the thesis is the implementation of improvements to the probability distribution of migrants based on the length of their stay, which we calculate using parametric and non-parametric modeling. We compare these approaches in terms of accuracy based on historical data.

Keywords: Modeling labor supply, Dynamic microsimulation models, SLAMM model, Probability distribution of migrants, Parametric and non-parametric modeling

Obsah

Úvod **8**

1 Mikrosimulačné prístupy k modelovaniu ponuky práce **10**

1.1 Koncept mikrosimulačných modelov 10

1.2 Výhody a obmedzenia mikrosimulačných modelov 12

2 SLAMM model **15**

2.1 Architektúra modelu 15

2.2 Predpoklady modelu 19

2.3 Modifikácie modelu 20

3 Modelovanie migrácie v SLAMM modeli **22**

3.1 Opis dátového problému a výzvy pri modelovaní 22

3.2 Prístupy k modelovaniu pravdepodobnostného rozdelenia 23

3.2.1 Parametrické modelovanie 23

3.2.2 Neparametrické modelovanie 36

3.3 Transformácia spojitého rozdelenia na diskrétnu formu 40

3.4 Finalizácia pravdepodobností pre funkciu *time_to_leave* 43

3.5 Odhad pravdepodobností odchodu migrantov 46

4 Predikcie a validácia modelu **50**

4.1 Predikcie modelu na základe rôznych pravdepodobnostných rozdelení . . . 50

4.2 Validácia výsledkov s historickými dátami 56

Záver **63**

Zoznam použitej literatúry **65**

Príloha A **69**

Úvod

Slovensko, podobne ako mnohé krajiny Európy, čelí nepriaznivým demografickým trendom, ktoré majú výrazný vplyv na ekonomiku a stabilitu krajiny. Jedným z hlavných problémov je dlhodobo nízka pôrodnosť. Počet obyvateľov dosiahol vrchol v roku 2021 a podľa predpokladov Eurostatu už nebude ďalej rásť. Hoci pôrodnosť v posledných rokoch mierne vzrástla, projekcia naznačuje, že do konca storočia neprekročí úroveň 1,7. [23]

Ďalšou významnou výzvou je starnutie populácie, čo sa prejavuje zhoršovaním miery ekonomickej závislosti – pomeru medzi počtom ľudí v produktívnom veku (15 – 64 rokov) a počtom osôb v neproduktívnom veku. V nasledujúcich 15 rokoch sa predpokladá pokles počtu ľudí v produktívnom veku o približne 280-tisíc, čo v dôsledku zníži pomer osôb v produktívnom veku na jedného seniora z 3,7 na úroveň 2,8. [23] Tieto demografické trendy predstavujú vážny problém pre trh práce a udržateľnosť ekonomiky.

Vzhľadom na tieto výzvy je dôležité venovať sa modelovaniu ponuky práce. Takéto modely poskytujú cenné nástroje na lepšie pochopenie toho, ako sa tieto zmeny prejavia na trhu práce, a umožňujú vypracovanie politík, ktoré môžu zmierniť ich negatívne dopady.

V tejto diplomovej práci bude použitý dynamický mikrosimulačný model SLAMM, vyvinutý na modelovanie a analýzu slovenskej ponuky práce. Hlavným cieľom práce je kalibrácia funkcií migrácie pracovnej sily v rámci modelu SLAMM.

V prvej kapitole sa podrobne oboznámime s konceptom mikrosimulačných modelov, pričom hlavný dôraz bude kladený na dynamické mikrosimulačné modely. Následne sa zameriame na výhody a nevýhody týchto prístupov, aby sme poskytli pohľad na ich využiteľnosť a limity v praxi. Pri spracovaní tejto kapitoly budeme vychádzať najmä z prehľadovej literatúry o dostupných mikrosimulačných modeloch, konkrétne zo zdrojov [1], [15], [24].

V druhej kapitole predstavíme model SLAMM so zameraním na jeho architektúru, vstupné a výstupné údaje a hlavné predpoklady, na ktorých je založený. Opíšeme jeho mechanizmy simulácie trhu práce a spôsob, akým zohľadňuje migráciu pracovnej sily. Zároveň načrtneme modifikácie modelu, ktoré sú predmetom tejto práce a ktoré majú za cieľ zlepšiť jeho predikčnú presnosť. Obsah kapitoly vychádza najmä z dokumentácie a materiálov týkajúcich sa modelu SLAMM, ako sú [22] a [29].

V tretej kapitole sa zameriavame na implementáciu navrhnutých modifikácií modelu.

Hlavne popíšeme a aplikujeme prístupy k modelovaniu pravdepodobnostných rozdelení, ktoré nám umožnia zlepšiť presnosť simulácií v rámci daného modelu.

V štvrtej kapitole budeme riešiť implementáciu výsledkov pravdepodobnostného rozdelenia, ktoré získame v predchádzajúcej kapitole. Následne porovnáme rôzne prístupy z hľadiska ich presnosti a posúdime ich efektivitu a spoľahlivosť na základe historických dát.

1 Mikrosimulačné prístupy k modelovaniu ponuky práce

Modelovanie ponuky práce predstavuje efektívny nástroj na pochopenie dynamiky trhu práce a identifikáciu faktorov, ktoré ovplyvňujú jeho fungovanie. V tejto kapitole sa zameriavame na mikrosimulačné prístupy k modelovaniu, s dôrazom na dynamické mikrosimulačné modely. Takéto modely umožňujú detailnú analýzu správania jednotlivcov na trhu práce. Sú obzvlášť užitočné pri predikcii dôsledkov politických zmien a určovaní dlhodobých trendov v pracovnej sile.

Na úvod predstavíme mikrosimulačné prístupy k modelovaniu trhu práce, následne sa podrobne zameriame na koncept dynamických mikrosimulačných modelov. V druhej časti sa budeme zaoberať ich výhodami a nevýhodami.

1.1 Koncept mikrosimulačných modelov

Na účel modelovania ponuky práce sa využívajú rôzne metódy, od jednoduchých ekonometrických modelov až po komplexné simulačné prístupy. Jedným z nich sú mikrosimulačné modely, ktoré simulujú správanie na mikroúrovni, ako sú jednotlivci alebo firmy. Tieto modely vychádzajú z individuálnych charakteristík, ako sú napríklad demografické údaje, úroveň vzdelania, typ pracovného povolania či socio-ekonomické podmienky.

Mikrosimulačné modelovanie sa začalo rozvíjať v polovici 20. storočia, keď americký ekonóm a štatistik Guy H. Orcutt v roku 1957 publikoval významný článok v Review of Economics and Statistics. [20] V ňom predstavil mikrosimulačný prístup ako nový nástroj na analýzu ekonomických a sociálnych procesov na mikroúrovni. Orcuttova práca kládla dôraz na potrebu predbežného hodnotenia politických opatrení pred ich implementáciou, čo umožňuje predvídať ich potenciálne dopady a optimalizovať politické rozhodovanie. Napriek tomu, že tento prístup priniesol inovatívne možnosti v oblasti analýz, pokrok v rozvoji modelov bol v nasledujúcich desaťročiach obmedzený. Hlavnými faktormi boli technologické prekážky a nedostatok dostupných mikrodát, ktoré boli nevyhnutné pre realizáciu komplexných simulácií. [1]

Výraznejší rozvoj týchto analytických nástrojov nastal až v 80. rokoch 20. storočia, keď sa rozšírilo používanie osobných počítačov a zlepšila dostupnosť mikrodát. [1] Tieto

technologické pokroky umožnili realizovať komplexné analýzy a pracovať s väčšími vzorkami dát, čím sa otvorili nové možnosti pre politické simulácie. Mikrosimulačné modely sa začali intenzívne využívať v rôznych krajinách, predovšetkým v Spojených štátoch, Spojenom kráľovstve a Nemecku, kde sa uplatnili pri analýze daňových systémov, sociálnej politiky a verejných financií. [1]

Od 90. rokov sa simulácie rozšírili aj do ďalších oblastí, ako sú trh práce, zdravotná politika či environmentálne otázky. [24] Ich schopnosť modelovať dlhodobé ekonomické a sociálne dopady z nich robí užitočný nástroj pre analýzu politických rozhodnutí. Súčasný vývoj v oblasti týchto modelovacích techník je čoraz viac multidisciplinárny a zahŕňa aplikácie v oblastiach, ako sú poľnohospodárstvo, zdravotníctvo a pracovná ekonomika, pričom na ich rozvoji sa podieľajú odborníci z rôznych vedných disciplín.

V odbornej literatúre sa mikrosimulačné modely najčastejšie delia na statické a dynamické. [24] Pri popise a charakteristike týchto modelov sme vychádzali z [24].

Statické mikrosimulačné modely sa zameriavajú na analýzu trhových podmienok v konkrétnom čase a neposkytujú možnosť predikovať, ako sa správanie jednotlivcov vyvíja v priebehu času. Tieto modely sú relatívne jednoduché na implementáciu a poskytujú rýchle výsledky pri skúmaní krátkodobých dopadov zmien v politických opatreniach, ako sú úpravy v daniach alebo dávkach. Avšak, neberú do úvahy dlhodobé zmeny v správaní jednotlivcov ani ich adaptáciu na nové podmienky trhu práce. Príkladom je model EUROMOD, ktorý sa využíva na hodnotenie krátkodobých daňových a dávkových zmien. [24] V tejto práci sa budeme venovať iba dynamickým mikrosimulačným modelom.

Dynamické mikrosimulačné modely predstavujú pokročilejší prístup, ktorý umožňuje integráciu časovej dimenzie do analýz. Simulujú správanie jednotlivcov v priebehu ich života, pričom zohľadňujú faktory, ako sú starnutie populácie, zmeny povolania či ekonomické cykly. Tieto nástroje poskytujú komplexný pohľad na dlhodobé dopady ekonomických zmien. Sú užitočné pri analýze trendov v pracovnej sile a modelovaní dôsledkov politických opatrení, ako sú reformy dôchodkového systému alebo úpravy v daňových politikách. Príkladom takéhoto prístupu je DYNASIM, ktorý predpovedá vývoj reprezentatívnej vzorky populácie USA, pričom simuluje rôzne demografické faktory, ako sú narodenia, úmrtia, sobáše, rozvody a zdravotný stav. Okrem toho zohľadňuje aj ekonomické aspekty, ako je zapojenie do pracovného trhu, výška príjmov, pracovná doba a

proces odchodu do dôchodku. [2] Ďalším významným príkladom je MOSART, model používaný v Nórsku na simuláciu životných udalostí a ekonomických procesov. Slúži ako nástroj na analýzu politík v oblasti dôchodkového systému. [2]

Dynamické mikrosimulačné modely môžu byť vytvorené na základe rôznych typov dát. Medzi najčastejšie využívané patria administratívne údaje, ktoré poskytujú rozsiahly počet pozorovaní a vysokú presnosť v daňových a ekonomických premenných, ako sú mzdy a odvody. Ich využitie na výskumné účely však môže byť obmedzené legislatívnymi a etickými reguláciami, ako aj administratívnymi prekážkami, čo sťažuje prístup k týmto dátam a ich efektívne spracovanie. Ďalším zdrojom sú prieskumy, ktoré poskytujú detailné informácie, ale majú obmedzené vzorky, pričom váhy môžu ovplyvniť ich reprezentatívnosť. Tretím typom sú syntetické dáta, ktoré sa generujú na základe preddefinovaných scenárov a modelových predpokladov. Môžu sa však objaviť riziká spojené s nadmerným zjednodušením reality a obmedzenou presnosťou výsledkov.

V priebehu posledných rokov sa dynamické mikrosimulačné modely vylepšujú novými prístupmi, ktoré zvyšujú ich schopnosti a efektivitu. Tieto inovácie zahŕňajú integráciu s agentovými modelmi (angl. agent-based models), ktoré umožňujú simulovať interakcie medzi jednotlivcami a analyzovať ich vplyv na širšie socioekonomické procesy. [24] Niektoré nové modely sú tiež spojené s technológiami ako strojové učenie, čo prispieva k presnejšej predikcii ekonomických výsledkov, najmä v oblastiach ovplyvnených vonkajšími šokmi, ako sú cenové výkyvy. [15]

1.2 Výhody a obmedzenia mikrosimulačných modelov

Ako bolo uvedené v predchádzajúcej časti, mikrosimulačné modely umožňujú podrobnú analýzu správania jednotlivcov alebo domácností na mikroúrovni. Dynamické mikrosimulačné modely tento prístup rozširujú pridaním časovej dimenzie, čím dokážu simulovať zmeny správania v čase. Okrem ich výhod je však dôležité analyzovať aj potenciálne nevýhody a obmedzenia. Táto kapitola sa zameriava na vyhodnotenie týchto modelov. Cieľom je podrobnejšie opísať ich silné a slabé stránky a prispieť k efektívnejšiemu rozhodovaniu pri ich využívaní vo výskume a praxi. Pri jej spracovaní sme vychádzali z [24].

Jednou z hlavných výhod mikrosimulačných modelov je ich flexibilita, ktorá umožňuje ich prispôsobenie na rôzne účely – od hodnotenia redistribučných politík až po analýzu de-

mografických zmien. Vďaka tejto univerzálnosti nachádzajú uplatnenie v širokom spektre výskumných oblastí, ako je analýza trhu práce, verejné zdravie či hodnotenie ekonomických politík. Dynamické verzie týchto modelov navyše poskytujú možnosť tvorby dlhodobých prognóz, napríklad v oblasti dôchodkových systémov, kde sa sledujú ich dopady na rôzne vekové skupiny v čase.

Mikrosimulačné modely tiež umožňujú simulovať rôzne scenáre, čím predstavujú mechanizmus na ex-ante hodnotenie politík ešte pred ich implementáciou a pomáhajú predvídať možné nežiaduce účinky. Tieto schopnosti ich robia silným nástrojom na podporu rozhodovacích procesov, v ktorých je potrebné zohľadniť množstvo faktorov a rizík.

Ďalšou z významných výhod týchto modelov je ich schopnosť integrovať sa s makroekonomickými modelmi, čo umožňuje zosúladiť mikroekonomické simulácie s makroekonomickými predpoveďami pomocou kalibračných algoritmov. Tento prístup vedie k lepšej konzistencii medzi individuálnymi rozhodnutiami a širšími ekonomickými trendmi.

Napriek svojim výhodám sa mikrosimulačné modely stretávajú aj s určitými obmedzeniami. Jednou z hlavných nevýhod je vysoká závislosť na kvalitných vstupných údajoch, pričom často chýbajú dostatočne podrobné a dlhodobé dáta potrebné na simuláciu. Nedostatok spoľahlivých údajov o rôznych populačných skupinách alebo ekonomických faktoroch môže viesť k nepresnostiam modelov a obmedziť ich prediktívnu schopnosť.

Ďalšou nevýhodou je, že vývoj a implementácia takýchto modelov sú časovo a finančne náročné, pričom si vyžadujú tím odborníkov z rôznych oblastí, čo robí ich vývoj zdrojovo náročným. Napríklad, vývoj modelov ako MINT, CBOLT a POLISIM v USA si vyžiadal viac ako 6 miliónov USD každý. [14] Okrem vysokých nákladov na vývoj môže byť problémom aj dlhý čas potrebný na vytvorenie plne funkčného modelu, pričom v priebehu tohto procesu môžu nastať legislatívne alebo ekonomické zmeny, ktoré si vyžiadajú prepracovanie modelu ešte pred jeho dokončením.

Tieto modely sú tiež náchylné na citlivosť voči predpokladom a parametrom, čo môže ovplyvniť ich prediktívnu schopnosť pri nečakaných zmenách v správaní populácie alebo ekonomických podmienkach. Zmeny v kľúčových predpokladoch alebo parametroch, ako sú zmeny v politických preferenciách alebo makroekonomické šoky, môžu mať významný vplyv na výstupy modelu, čo sťažuje presné predpovede a ich schopnosť verne odrážať realitu v rôznych obdobiach.

Taktiež modely môžu vyžadovať úpravy na zohľadnenie nových foriem správania alebo vonkajších šokov, ako sú ekonomické krízy alebo demografické zmeny. Rovnako môže byť potrebné prispôsobiť modely neočakávaným faktorom, ako sú rýchle technologické pokroky alebo globálne pandémie, čo si môže vyžadovať častú rekalibráciu na zabezpečenie ich aktuálnosti a presnosti. Validácia dynamických mikrosimulačných modelov môže byť problematická, najmä ak nie sú k dispozícii historické údaje na porovnanie predikcií. Bez vhodných historických dát na overenie predikcií môže byť spoľahlivosť týchto modelov spochybnená, čo ešte viac komplikuje schopnosť modelov poskytovať presné predpovede v rôznych podmienkach.

Napriek týmto obmedzeniam zostávajú mikrosimulačné modely jedným z hlavných nástrojov v socioekonomickom výskume. Ich schopnosť simulovať komplexné interakcie na mikroúrovni ich robí neoceniteľnými pri analýze politík s dlhodobými dopadmi. Tieto modely sú užitočné pre riešenie problémov, ktoré ovplyvňujú ekonomiku, spoločnosť a jednotlivcov, najmä v súčasnom svete rýchlo sa vyvíjajúcich výziev.

2 SLAMM model

V tejto kapitole sa zameriame na dynamický mikrosimulačný model SLAMM, ktorý slúži na modelovanie a analýzu slovenskej ponuky práce. V úvode vysvetlíme jeho architektúru, hlavné predpoklady, vstupy a výstupy, ako aj mechanizmy, ktorými simuluje dynamiku trhu práce. Následne navrhneme modifikácie jeho štruktúry, ktoré sú súčasťou tejto diplomovej práce. Úpravy sa sústredia na kalibráciu funkcií migrácie pracovnej sily.

2.1 Architektúra modelu

Model SLAMM (Slovenský model trhu práce, angl. Slovak Labour Market Model) je mikrosimulačný model určený na predikciu dynamiky pracovnej sily na Slovensku. Bol vyvinutý Ekonomickým ústavom Slovenskej akadémie vied. Kombinuje mikrosimulačné a makroekonomické prístupy, čím poskytuje komplexný rámec na analýzu dopytu po pracovnej sile. Jeho hlavným cieľom je kvantifikácia potreby nahradzovania zamestnancov, ktorí opúšťajú trh práce, a vyhodnotenie potreby zahraničnej pracovnej sily. [29]

Tento model je charakterizovaný ako dynamický mikrosimulačný model fungujúci v diskrétnom čase s ročnou periodicitou. Jednotkami modelu sú jednotlivci, ktorých rozhodnutia o ekonomickej aktivite sú ovplyvnené procesom starnutia, ako aj ďalšími faktormi, ktoré sa vyvíjajú v čase. Tieto vlastnosti umožňujú detailné modelovanie individuálnych a demografických zmien, ktoré priamo ovplyvňujú trh práce. Informácie uvedené v tejto kapitole sú čerpané z [22] a [29].

Model SLAMM je možné rozdeliť na dve autonómne časti – ponukovú a dopytovú stranu trhu práce. Dopytová strana je určovaná makroekonomickými faktormi, pričom sa opiera o klasický vstupno-výstupný makroekonomický rámec. Predikcia zahŕňa potrebný počet pracovných síl pre zabezpečenie rovnovážnej úrovne produkcie v jednotlivých segmentoch trhu práce.

Na druhej strane, ponuková časť modelu je založená na simulácii demografických procesov a rozhodnutí účastníkov trhu práce o ekonomickej aktivite. Tieto rozhodnutia sú simulované pomocou deterministických aj stochastických funkcií, ktoré zohľadňujú vek jednotlivcov a údaje z predchádzajúcich období. V tejto diplomovej práci sa zaoberáme výlučne ponukovou časťou.

Model SLAMM využíva údaje z prieskumu pracovnej sily (angl. Labour Force Survey, LFS), ktorý je pravidelným štatistickým prieskumom realizovaným na úrovni Európskej únie. Na Slovensku je za jeho vykonávanie zodpovedný Štatistický úrad Slovenskej republiky. LFS poskytuje cenné informácie o štruktúre pracovného trhu a demografických charakteristikách populácie. Tieto dáta slúžia ako vstup do modelu SLAMM a zahŕňajú osem základných atribútov na individuálnej úrovni, medzi ktoré patrí vek, pohlavie, najvyššie dosiahnuté vzdelanie, odvetvie zamestnania, ekonomický status a ďalšie relevantné faktory, ktoré sú podrobne uvedené v Tabuľke 1. Na zabezpečenie reprezentatívnosti simulovaných výsledkov sú údaje vážené podľa štruktúry celkovej populácie Slovenska.

Model sa skladá zo štyroch hlavných modulov:

- Demografický modul
- Modul vzdelávania
- Modul ekonomickej aktivity
- Modul zamestnania

Demografický modul slúži na simuláciu procesov, ako sú pôrodnosť, úmrtnosť a starnutie populácie. Migrácia je v rámci tohto modulu nastavená na nulovú hodnotu, čím sa umožňuje modelovanie populácie výlučne na základe vnútroštátnych demografických zmien. Pre novonarodené osoby sa automaticky generujú základné atribúty, ako vek a pohlavie. Modul zároveň zohľadňuje dĺžku pobytu zahraničných pracovníkov a poskytuje možnosť kalibrácie na základe administratívnych údajov o povoleniach na prácu.

Modul vzdelávania simuluje proces získavania vzdelania na základe proporcií pozorovaných v najnovšej dostupnej kohorte vo veku 30 až 35 rokov, ktorá už ukončila svoj vzdelávací proces. Tento modul umožňuje modelovanie alternatívnych predpokladov vývoja vzdelanostnej štruktúry populácie.

Modul ekonomickej aktivity simuluje rozhodovanie o zapojení do ekonomickej činnosti na základe predchádzajúceho stavu, vekovej štruktúry a proporcií zaznamenaných v poslednom cenzuse. Osoby, ktoré nie sú ekonomicky aktívne, patria medzi dôchodcov, predčasných dôchodcov, zdravotne postihnutých alebo študentov. Prechody medzi niektorými statusmi sú možné, avšak dôchodok predstavuje konečný stav, z ktorého už nie je návrat

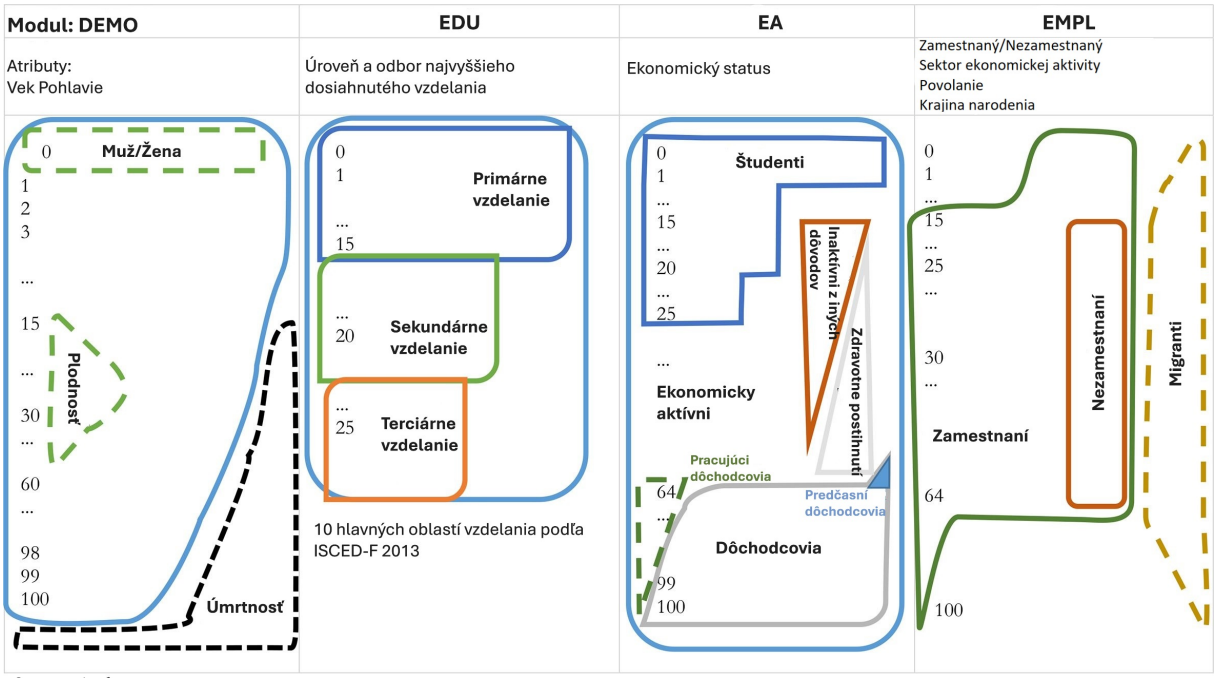
Premenná	**Hodnoty**
Pohlavie	Muž, Žena
Vek	V rokoch
Najvyšší dosiahnutý stupeň vzdelania	Primárne Nižšie sekundárne Vyššie sekundárne Postsekundárne Terciárne
Odbor najvyššieho dosiahnutého vzdelania	10 odborov (podľa ISCED-F 2013)
Ekonomický status v súčasnom období	Zamestnaný Nezamestnaný Študent Dôchodca Zdravotne postihnutý Neaktívny
Ekonomický status pred rokom	*Rovnaké kategórie ako vyššie*
Ekonomický sektor zamestnania v súčasnom období	18 sektorov (podľa NACE 2.0)
Ekonomický sektor zamestnania pred rokom	*Rovnaké kategórie ako vyššie*
Skupina povolania	3 skupiny (podľa úrovní zručností ISCO-08) (ISCO 1-3; ISCO 4-8; ISCO 9)
Skupina povolania pred rokom	*Rovnaké kategórie ako vyššie*
Krajina narodenia	Slovensko EÚ Mimo EÚ

Tabuľka 1: Premenné a ich hodnoty vo vstupných dátach modelu SLAMM, zdroj: [29]

medzi aktívnych. Študenti vstupujú do ekonomickej aktivity po dosiahnutí stanoveného veku ukončenia vzdelania. Každý, kto nespadá do kategórií neaktivity, je považovaný za ekonomicky aktívneho.

Modul zamestnania rozdeľuje ekonomicky aktívnych ľudí na zamestnaných a nezamestnaných. Okrem toho určuje ich odvetvie a povolanie na základe exogénnej makroekonomickej predikcie. V prípade nedostatku vhodných domácich pracovníkov sú do modelu pridávaní zahraniční pracovníci, čím sa rieši nesúlad medzi dopytom a ponukou práce. Alokácia sa vykonáva na základe výpočtu skóre jednotlivca podľa individuálnych pracovných charakteristík.

Na Obr. 1 je znázornená štruktúra modelu SLAMM.



Obr. 1: Štruktúra modelu SLAMM, zdroj: [29]

Vzhľadom na komplexnosť trhu práce model SLAMM kombinuje mikrosimulačné a makroekonomické prístupy, čím umožňuje prepojenie predikcií ekonomicky aktívnej populácie s dopytom po pracovnej sile. Tento prístup poskytuje rámec pre simuláciu politických opatrení, ako sú legislatívne zmeny týkajúce sa dôchodkového veku alebo stratégie na podporu zahraničnej pracovnej sily. SLAMM zároveň analyzuje vplyv ekonomických faktorov, ako sú inflácia a hospodársky rast, na trh práce, čo zvyšuje presnosť predikcií a umožňuje prispôsobenie politík potrebám rôznych segmentov populácie.

Medzi hlavné prínosy modelu SLAMM patrí jeho schopnosť simulovať scenáre, ktoré odrážajú aktuálne politické a demografické trendy. Model testuje vplyv zmien, ako je skrátenie pobytu zahraničných pracovníkov alebo predĺženie veku odchodu do dôchodku, a analyzuje ich dopad na pracovný trh. Tieto scenáre sú kľúčové pre tvorbu efektívnych politík reagujúcich na starnutie populácie a meniace sa potreby trhu práce. SLAMM tiež poskytuje nástroje na sledovanie distribučných efektov politických a ekonomických zmien, čo umožňuje lepšie pochopiť ich vplyv na rôzne skupiny populácie.

2.2 Predpoklady modelu

SLAMM model je postavený na viacerých kľúčových predpokladoch, ktoré určujú jeho fungovanie a schopnosť simulovať dynamiku trhu práce. Tieto predpoklady vychádzajú z demografických a ekonomických trendov a zohľadňujú interakciu medzi ponukou a dopytom na trhu práce.

Demografické predpoklady modelu SLAMM sú založené na projekciách EUROPOP 2018 [10], ktoré poskytujú údaje o úmrtnosti, plodnosti a vekovej štruktúre populácie. Tieto údaje vytvárajú základ pre realistické modelovanie demografickej dynamiky. Pripomíname, že v demografickom module sa predpokladá nulová čistá migrácia, pričom zahraniční pracovníci sú pridávaní do simulácií na základe požiadaviek trhu práce v module zamestnanosti.

Vzdelanostné predpoklady modelu sú založené na údajoch z LFS. Vzdelanostná štruktúra populácie sa simuluje podľa proporcií pozorovaných v konkrétnych vekových kohortách. Na simuláciu odborného zamerania populácie sa využíva klasifikácia ISCED-F 2013 [9], čo umožňuje lepšie pochopenie vzťahu medzi vzdelaním a požiadavkami pracovného trhu.

Z **ekonomických predpokladov** aplikovaný scenár predpokladá stabilný nárast celkovej zamestnanosti medzi 1-1.5 % ročne do roku 2025, po ktorom sa očakáva nulový rast. Po tomto období by sa mala zamestnanosť ustáliť na stabilnej úrovni, pričom sa nepredpokladajú výrazné zmeny v jednotlivých sektoroch trhu práce.

Vzhľadom na rastúcu potrebu zahraničnej pracovnej sily SLAMM model zahrňuje **predpoklady týkajúce sa migrácie**. Počet zahraničných pracovníkov nie je stanovený exogénne, ale je odvodený od deficitu pracovnej sily v jednotlivých segmentoch trhu práce.

Dĺžka pobytu migrantov je určená na základe údajov o povoleniach na prácu, pričom ich odchod z krajiny ovplyvňuje náhradný dopyt po pracovnej sile. Model predpokladá, že dĺžka zotrvania migrantov je kalibrovaná iba na údaje z roku 2019, pričom v nasledujúcich rokoch sa dynamika migračných tokov považuje za stabilnú.

Tieto predpoklady tvoria základ pre simuláciu trhu práce na Slovensku a poskytujú dôležité údaje pre tvorbu politík, ktoré zabezpečia efektívnu alokáciu pracovných síl. Model SLAMM tak ponúka komplexný pohľad na vývoj trhu práce, zohľadňujúci demografické, vzdelanostné a ekonomické faktory.

2.3 Modifikácie modelu

Hlavným cieľom tejto diplomovej práce je kalibrácia funkcií migrácie pracovnej sily v rámci modelu SLAMM. Kľúčovým parametrom, na ktorý sa zameriame, je dĺžka pobytu imigranta na Slovensku. Tento parameter má zásadný vplyv na potrebu zabezpečiť dostatočný prítok pracovnej sily. Za tieto procesy sú v modeli zodpovedné dve funkcie – *time_to_leave* a *leave*, ktoré sú súčasťou demografického modulu.

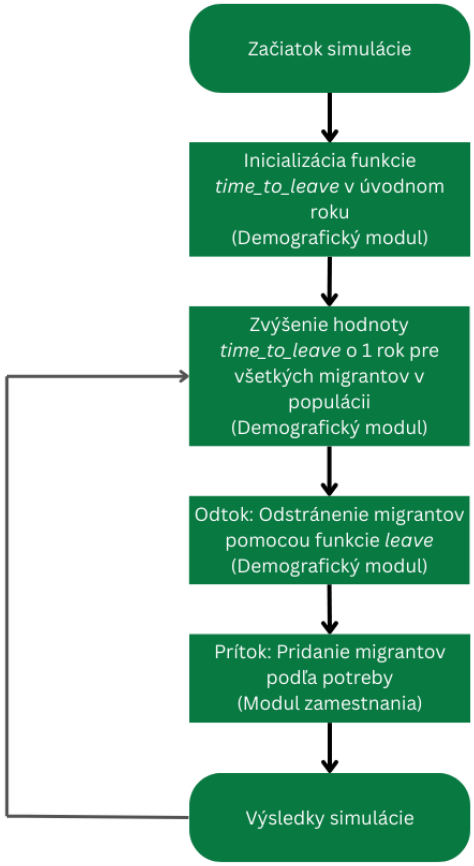
Vstupné údaje z LFS neobsahujú informácie o dĺžke pobytu migranta na Slovensku. Tento proces bol preto implementovaný do modelu prostredníctvom funkcie *time_to_leave*. V počiatočnej fáze modelovania je každému migrantovi priradená náhodná hodnota v rozmedzí od 0 do 9 rokov na základe stanovenej pravdepodobnosti. V nasledujúcich obdobiach, pokiaľ migrant nie je odstránený zo systému, sa jeho hodnota *time_to_leave* každoročne zvyšuje o jeden rok. Tento proces sa následne opakuje v každom ďalšom období simulácie.

Funkcia *leave* potom rozhoduje, kedy migrant opustí Slovensko. Pravdepodobnosť jeho odchodu závisí od hodnoty *time_to_leave*. Pre každú hodnotu tejto premennej je priradená pravdepodobnosť, s ktorou migrant buď zostane v krajine, alebo ju opustí. Ak funkcia *leave* vráti hodnotu `true`, migrant zostane na Slovensku. Ak vráti hodnotu `false`, migrant bude odstránený z modelu a prestáva byť súčasťou pracovného trhu. Týmto spôsobom funkcia *leave* simuluje emigráciu, čím priamo ovplyvňuje dynamiku pracovného trhu.

Tieto dve funkcie spolu tvoria systém, ktorý simuluje dĺžku pobytu migrantov a ich rozhodnutie o odchode. Interakcia týchto funkcií je kľúčová pre správne nastavenie migračnej funkcie v rámci modelu SLAMM. Pre lepšie pochopenie fungovania systému je nižšie

uvedený diagram simulačného procesu (viď Obr. 2).

Simulácia sa najprv inicializuje. V úvodnej fáze funkcia *time_to_leave* priradí každému migrantovi dĺžku zotrvania na základe definovaného rozdelenia. Počas simulačného obdobia sa hodnota *time_to_leave* každého migranta v populácii postupne zvýši o 1 rok, čím sa simuluje plynutie času. Následne sa uplatní funkcia *leave*, ktorá upraví populáciu modelu odstránením tých cudzincov, ktorí spĺňajú definované podmienky odchodu. Potom sa podľa potreby do modelu pridajú noví migranti, čo je riadené modulom zamestnania. Na záver sa vyhodnotia výsledky simulácie.



Obr. 2: Postup simulácie migrácie v modeli SLAMM, zdroj: vlastná analýza

V ďalších kapitolách sa budeme zaoberať kalibráciou týchto funkcií a ich dopadom na výsledky modelu.

3 Modelovanie migrácie v SLAMM modeli

V tejto kapitole sa zameriavame na zlepšenie presnosti modelu SLAMM prostredníctvom kalibrácie migračných funkcií. Ako bolo uvedené v predchádzajúcej kapitole, migrácia je v modeli ovplyvnená predovšetkým dvoma funkciami: *time_to_leave* a *leave*.

Na odhad pravdepodobnostného rozdelenia pre funkciu *time_to_leave* využívame tri prístupy. Prvým je základný prístup, v ktorom sú pravdepodobnosti odvodené z podielu pracujúcich migrantov v jednotlivých intervaloch dĺžky pobytu. Druhým je parametrický prístup, ktorý predpokladá, že dĺžka pobytu migrantov sleduje konkrétne teoretické rozdelenie. Tretím je neparametrický prístup, ktorý využíva metódy jadrového odhadu hustoty bez nutnosti predpokladať konkrétny tvar rozdelenia. Po odhade pravdepodobnostného rozdelenia zotrvania migrantov vypočítame pravdepodobnosti odchodu z krajiny pre funkciu *leave*.

3.1 Opis dátového problému a výzvy pri modelovaní

Predtým, než sa pustíme do samotného modelovania pravdepodobností, je nevyhnutné pripraviť vstupné dáta. Problém však nastáva pri získavaní potrebných informácií. Ideálnym riešením by bolo mať k dispozícii detailné záznamy na individuálnej úrovni, ktoré by obsahovali údaje o príchodoch a odchodoch osôb na a zo Slovenska. Z týchto dát by bolo možné presne vypočítať dĺžku pobytu jednotlivcov, čo by umožnilo presnejšiu kalibráciu modelu.

Dostupnosť týchto informácií je však výrazne limitovaná. Problémy spôsobuje nielen ich časový rozsah, ale aj citlivá povaha údajov, ktorá bráni ich širšiemu využitiu. Získavanie takýchto detailných údajov je navyše administratívne náročné.

Avšak bežne sú dostupné agregované tabuľky, ktoré bývajú zverejňované online s relatívne krátkym časovým oneskorením. Použitie agregovaných dát preto často predstavuje jedinú reálnu možnosť na realizáciu takýchto štúdií. Hoci tieto údaje neposkytujú rovnakú úroveň detailnosti ako individuálne záznamy, umožňujú získať dostatočný prehľad o hlavných trendoch a vzorcoch migrácie.

Preto v rámci tejto diplomovej práce na výpočet rozdelenia pravdepodobnosti budeme využívať agregované údaje poskytované Ústredím práce, sociálnych vecí a rodiny (ÚPS-

VaR) [30]. Tieto dáta zahŕňajú mesačné počty cudzincov na Slovensku, rozdelené podľa dĺžky ich pracovného pomeru.

Zároveň sformulujeme výskumné hypotézy, ktoré následne overíme v rámci tejto diplomovej práce. Celkovo sa zameriame na dve hypotézy:

- Hypotéza 1 (H1): Je možné spresniť simulácie a predikcie presnejšou implementáciou pravdepodobnostného rozdelenia premennej dĺžky zotrvania v krajine na základe dostupných údajov?
- Hypotéza 2 (H2): Ktorý zo spôsobov implementácie pravdepodobnosti zotrvania je preferovaný z hľadiska zvýšenia presnosti predikcií?

Testovanie hypotéz poskytne cenné poznatky o tom, ako rôzne prístupy k pravdepodobnostnému rozdeleniu ovplyvňujú presnosť modelu a prispievajú k lepším výsledkom pri predikcii migrácie pracovnej sily.

3.2 Prístupy k modelovaniu pravdepodobnostného rozdelenia

V tejto podkapitole sa venujeme modelovaniu pravdepodobnostného rozdelenia dĺžky pobytu migrantov na Slovensku. Predstavujeme dva hlavné metodické prístupy: parametrické a neparametrické modelovanie. Tieto prístupy následne aplikujeme na reálne dáta. Proces modelovania prebieha v softvéri R, pričom sa využívajú rôzne štatistické metódy na optimalizáciu rozdelenia a validáciu jeho zhody s empirickými údajmi.

3.2.1 Parametrické modelovanie

Parametrické modelovanie predstavuje jeden z hlavných prístupov v štatistickej analýze dát. Jeho podstatou je predpoklad, že analyzované dáta pochádzajú z teoretického pravdepodobnostného rozdelenia, ktorého vlastnosti sú určené obmedzeným počtom parametrov. [17] Tento prístup umožňuje efektívne opisovať a interpretovať štruktúru dát pomocou presne definovaných matematických modelov.

Kľúčové kroky parametrického modelovania zahŕňajú:

- Získanie a príprava dát.
- Výber teoretického pravdepodobnostného rozdelenia.

- Odhad parametrov rozdelenia z dát.
- Validácia modelu.

Ešte predtým, než pristúpime k výberu vhodného teoretického rozdelenia, je potrebné zadefinovať náhodnú premennú, ktorú budeme modelovať. Pre účely tejto diplomovej práce bola definovaná náhodná premenná X, ktorá nadobúda nezáporné hodnoty. Dôležitým aspektom je určenie, či ide o diskrétnu alebo spojitú náhodnú premennú. Na jednej strane môže dĺžka pobytu v krajine vykazovať logické spojité správanie. Avšak, z druhej strany, dáta, ktoré použijeme na modelovanie, sú agregované podľa intervalov, ako sú obdobia 0-6 mesiacov, 7-12 mesiacov, 12-24 mesiacov, 25-36 mesiacov, 37-48 mesiacov, 49-72 mesiacov a nad 72 mesiacov. Struktúra týchto dát obsahuje diskrétne kategórie, ktoré zachytávajú dĺžku pracovného pomeru v uvedených intervaloch.

Okrem toho, v súčasnej verzii modelu je dĺžka pobytu cudzincov reprezentovaná ako diskrétna náhodná premenná, ktorá nadobúda hodnoty od 0 do 9 rokov. Z tohto hľadiska by bolo vhodnejšie použiť diskrétnu náhodnú premennú. Napriek tomu bolo rozhodnuté testovať obe možnosti — diskrétne aj spojité teoretické rozdelenie. V tomto prípade však preferujeme spojité rozdelenie, keďže lepšie odráža reálne správanie migrantov a plynulý charakter dĺžky ich zotrvania v krajine.

V tomto bode čelíme problému, že dostupné údaje sú prezentované v intervaloch rôznej dĺžky, čo znemožňuje ich priame použitie na odhad pravdepodobnostného rozdelenia v diskrétnej alebo spojitej podobe. Porovnanie požadovanej agregácie a vstupných údajov je zobrazené v Tabuľke 2.

Na riešenie tohto problému intervaly pretransformujeme na diskrétne hodnoty, pričom každá z nich bude zodpovedať jednému roku. Tento prístup je v súlade s konštrukciou modelu SLAMM, keďže tento model vykonáva simulácie na ročnej báze.

Prvé dva intervaly spojíme a priradíme hodnotu 1, tretí a štvrtý necháme nezmenené a priradíme im hodnoty 2, resp. 3. Piaty interval, ktorý pokrýva dve po sebe idúce obdobia, rozdelíme na dva roky (5 a 6), pričom počet imigrantov rovnomerne alokujeme medzi tieto roky. Posledný interval, ktorý zahŕňa trvanie pracovného pomeru nad 72 mesiacov, označíme hodnotou 7.

Ďalšou výzvou pri modelovaní je skutočnosť, že údaje nepredstavujú migračné toky, ale stavy, ktoré zachytávajú aktuálny počet imigrantov v jednotlivých kategóriách za daný

ÚPSVaR agregácia	**Modelová agregácia**
0-6 mesiacov 7-12 mesiacov	0 rokov
12-24 mesiacov	1 rok
25-36 mesiacov	2 roky
37-48 mesiacov	3 roky
49-72 mesiacov	4 roky 5 rokov
nad 72 mesiacov	6 rokov 7 rokov 8 rokov 9 rokov

Tabuľka 2: Porovnanie štruktúry dát ÚPSVaR a požadovanej agregácie v modeli SLAMM

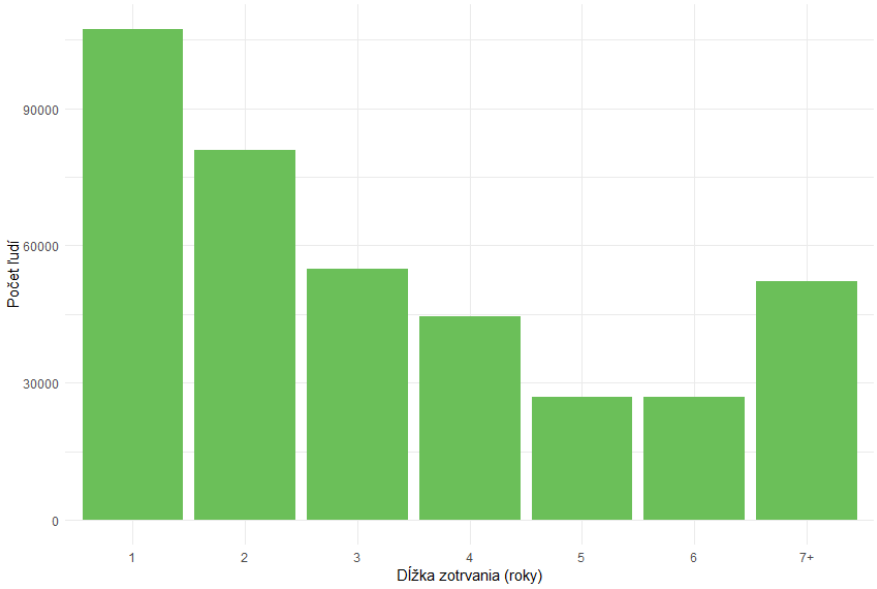
mesiac. To znamená, že pre každý rok máme k dispozícii 12 súborov dát, ktoré sú medzi sebou silne závislé. Táto závislosť vyplýva z toho, že napríklad osoba, ktorá patrí do intervalu 0-6 mesiacov, sa po niekoľkých mesiacoch môže presunúť do intervalu 7-12 mesiacov. Napriek týmto obmedzeniam sme sa rozhodli ponechať údaje v ich pôvodnej podobe. Tento prístup nám umožňuje využiť dostupné stavové údaje na odhad pravdepodobnostného správania imigrantov.

Keďže model SLAMM funguje na ročnej báze, je nevyhnutné, aby dáta vložené do modelu boli agregované na ročnej úrovni. Preto mesačné údaje za každý rok zlučíme do jedného celkového súboru pre daný rok. Tento postup zabezpečuje, že vstupné dáta spĺňajú požiadavky modelu a sú v súlade s jeho časovým rozsahom.

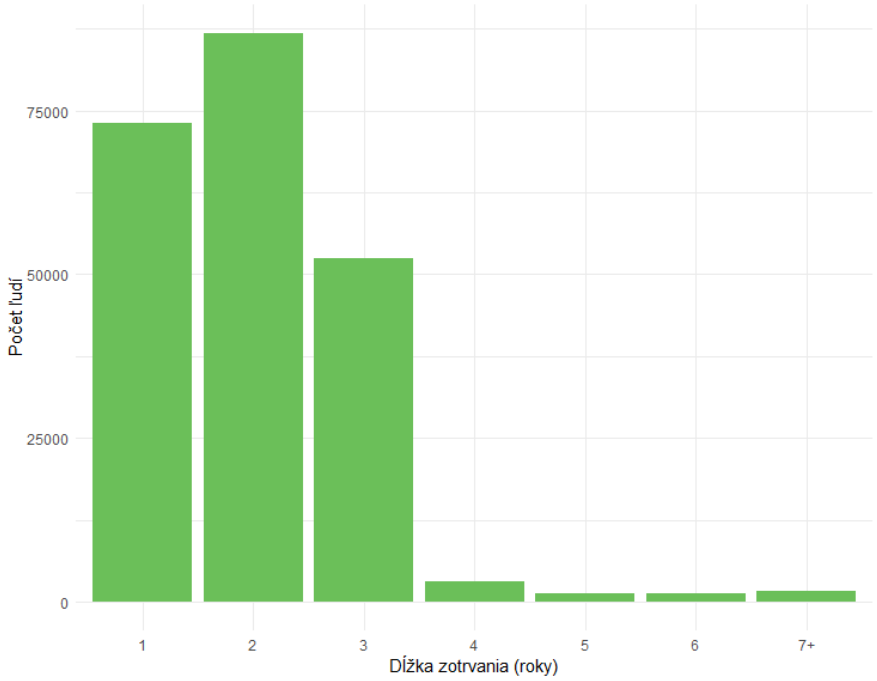
Taktiež, v rámci tejto diplomovej práce, rozdelíme krajiny pôvodu cudzincov do jednotlivých skupín, pre ktoré následne vypočítame príslušné pravdepodobnostné rozdelenia. Tieto skupiny sú definované nasledovne:

- Európska únia (EÚ),
- Ukrajina,
- non-EÚ: krajiny mimo EÚ, s výnimkou Ukrajiny.

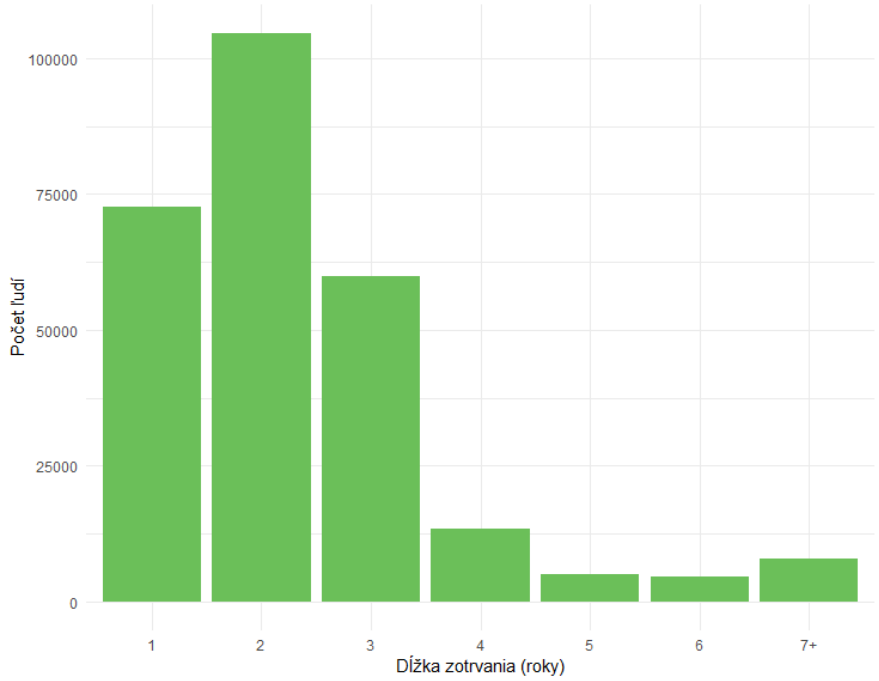
Na Obr. 3, 4 a 5 sú zobrazené histogramy, ktoré ilustrujú rozdelenie dát pre jednotlivé skupiny.



Obr. 3: Rozdelenie dĺžky zotrvania cudzincov z EÚ na Slovensku v roku 2019, zdroj: vlastná analýza



Obr. 4: Rozdelenie dĺžky zotrvania cudzincov z Ukrajiny na Slovensku v roku 2019, zdroj: vlastná analýza



Obr. 5: Rozdelenie dĺžky zotrvania cudzincov z non-EÚ na Slovensku v roku 2019, zdroj: vlastná analýza

Týmto sme dokončili prvý krok parametrického modelovania.

Po príprave vstupných dát môžeme prejsť k ďalšiemu kroku modelovania, ktorým je výber vhodného teoretického pravdepodobnostného rozdelenia. Niektoré rozdelenia boli vybrané na základe [6], resp. [21]. Pre účely tejto diplomovej práce sme zvolili nasledujúce teoretické rozdelenia, ktoré sú uvedené v Tabuľke 3.

Všetky tieto rozdelenia sú definované na nezáporných, respektíve kladných hodnotách a môžu byť vhodné na modelovanie vstupných údajov z hľadiska ich štruktúry. Detailné informácie o týchto rozdeleniach boli čerpané z [11].

Okrem toho jedno z rozdelení je definované na intervale $(1, \infty)$, konkrétne ide o log-gama rozdelenie. Všeobecne je toto rozdelenie definované na intervale $(-\infty, +\infty)$, avšak pre túto diplomovú prácu používame jeho špeciálnu verziu. Táto verzia bola implementovaná v R balíku `actuar` a je určená pre akturárske modelovanie. [27].

Je dôležité zdôrazniť, že pri práci s údajmi, ktoré predstavujú časové trvanie, by mohlo byť z hľadiska konzistencie vhodné uvažovať rozdelenia definované na intervale od 0. To by zabezpečilo, že všetky vstupné hodnoty by boli v súlade s celou časovou osou. Avšak vzhľadom na to, že analyzujeme dĺžky zotrvania, ktoré sú pre naše dáta vždy väčšie

Spojité rozdelenia
Exponenciálne rozdelenie
Gama rozdelenie
Loggama rozdelenie
Loglogistické rozdelenie
Lognormálne rozdelenie
Paretovo rozdelenie
Weibullovo rozdelenie
Diskrétne rozdelenia
Geometrické rozdelenie
Poissonovo rozdelenie

Tabuľka 3: Vybrané teoretické pravdepodobnostné rozdelenia

alebo rovné 1, môže byť použitie loggama rozdelenia definovaného na intervale $(1, \infty)$ vhodnejšie. Toto rozdelenie má potenciál poskytovať presnejšie výsledky v porovnaní s ostatnými rozdeleniami, keďže lepšie zodpovedá charakteru našich vstupných údajov.

Avšak, aby sa zachovali údaje o dĺžke zotrvania 1 rok, ktoré by inak pre dané rozdelenie spadali na hranicu otvoreného intervalu, tieto hodnoty boli upravené pridaním malej odchýlky ϵ. Táto úprava zabezpečuje, že všetky relevantné dáta sú zahrnuté do výpočtov rozdelenia, čím sa dosahuje vyššia presnosť a úplnosť modelovania.

Následujúci krok modelovania spočíva v odhade parametrov rozdelenia z dostupných dát. Na tento účel existuje viacero štandardných metód. Informácie použité pri spracovaní tejto problematiky pochádzajú z [11]. Medzi najčastejšie používané metódy patrí:

- metóda momentov,
- metóda maximálnej vierohodnosti (angl. maximum likelihood estimation, *MLE*),
- kvantilová metóda.

Pre odhad parametrov teoretického rozdelenia zvolíme metódu maximálnej vierohodnosti. Táto metóda je preferovaná pre svoje vlastnosti, keďže poskytuje efektívne odhady, ktoré sú pri dostatočne veľkej vzorke optimálne. Je flexibilná, univerzálne aplikovateľná na širokú škálu rozdelení a jednoducho implementovateľná pomocou moderných softvérových

nástrojov. V porovnaní s alternatívnymi metódami, ako je napríklad metóda momentov, umožňuje presnejšie odhady.

Metóda maximálnej vierohodnosti je založená na princípe maximalizácie pravdepodobnosti, že pozorované dáta pochádzajú z modelu s konkrétnymi parametrami. Cieľom je nájsť také hodnoty parametrov, ktoré maximalizujú funkciu vierohodnosti.

Predstavme si náhodný vektor $Z = (Z_1, Z_2, \ldots, Z_N)^T$, pre ktorý poznáme jeho hustotu pravdepodobnosti v prípade spojitého rozdelenia, alebo pravdepodobnostnú funkciu v prípade diskrétneho rozdelenia, vyjadrenú vzťahom $f(z, \boldsymbol{\theta})$. V tomto prípade $z = (z_1, z_2, \ldots, z_N)$ predstavuje vektor pozorovaní, zatiaľ čo $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)^T$ je vektor neznámych parametrov, ktoré chceme odhadnúť. Tento vektor $\boldsymbol{\theta}$ je však obmedzený na konkrétnu známu množinu $\Theta \subset \mathbb{R}^m$.

Vierohodnostná funkcia má tvar:

$$L(\boldsymbol{\theta}, z) = f(z, \boldsymbol{\theta})$$

Na zjednodušenie výpočtov sa v praxi často využíva logaritmická verzia vierohodnostnej funkcie (angl. log-likelihood function):

$$\ell(\boldsymbol{\theta}, z) = \ln L(\boldsymbol{\theta}, z) = \ln f(z, \boldsymbol{\theta})$$

Optimalizáciou funkcie $\ell(\boldsymbol{\theta}, z)$ získame odhady parametrov $\hat{\boldsymbol{\theta}}$, ktoré najlepšie zodpovedajú pozorovaným dátam.

Dôležitou výhodou *MLE* je možnosť efektívneho porovnávania rôznych teoretických rozdelení medzi sebou na základe kritérií, ako je napríklad logaritmus funkcie vierohodnosti. Čím vyššia je jeho hodnota, tým presnejšie dané rozdelenie odráža skutočné údaje.

Okrem log-likelihood funkcie sa na posudzovanie modelov využívajú aj iné ukazovatele, ako sú Akaikeho informačné kritérium (*AIC*) a Bayesovské informačné kritérium (*BIC*). Definície a vlastnosti týchto kritérií sme prevzali z [7].

Definícia *AIC* je nasledovná:

$$AIC = 2k - 2\ln(L),$$

kde k je počet parametrov modelu a L je maximálna hodnota vierohodnostnej funkcie pre daný model.

Definícia BIC je vyjadrená nasledujúcim vzorcom:

$$BIC = \ln(n)k - 2\ln(L),$$

kde n je počet pozorovaní, k je počet parametrov modelu a L je maximálna hodnota vierohodnostnej funkcie pre daný model.

Akaikeho informačné kritérium hodnotí modely na základe rovnováhy medzi presnosťou a zložitosťou, pričom nižšia hodnota AIC naznačuje lepší model. Podobne, Bayesovské informačné kritérium znevýhodňuje modely s väčším počtom parametrov, ale uplatňuje silnejšiu penalizáciu, čím uprednostňuje jednoduchosť. Nižšia hodnota BIC tiež naznačuje lepší model.

Štvrtým krokom parametrického modelovania je posúdenie zhody medzi teoretickými rozdeleniami a pozorovanými údajmi. Na tento účel existuje viacero prístupov, vrátane štatistických testov dobrej zhody, grafických metód, simulačných techník a informačných kritérií. [5], [7] V rámci tejto diplomovej práce sa zameriame na testy zhody, ktoré patria medzi najpoužívanejšie metódy na overenie, ako dobre sa teoretické rozdelenie hodí na opísanie empirických údajov. Zvažujeme aplikáciu Anderson-Darlingovho testu a Cramér-von Misesovho testu pre spojité rozdelenia a Pearsonovho chí-kvadrát testu pre diskrétne rozdelenia. Informácie o týchto testoch sme čerpali z [5]. Pri aplikácii týchto testov sa overujú nasledujúce hypotézy:

- Nulová hypotéza (H_0): Rozdelenie dát je rovnaké ako predpokladané teoretické rozdelenie, t. j. dáta pochádzajú z tohto teoretického rozdelenia.

- Alternatívna hypotéza (H_1): Rozdelenie dát sa líši od predpokladaného teoretického rozdelenia, t. j. dáta nepochádzajú z tohto rozdelenia.

Pri týchto testoch zvolíme hladinu významnosti $\alpha = 0.05$. To znamená, že hypotézu H_0 zamietame, ak p-hodnota (angl. p-value), ktorá vyjadruje pravdepodobnosť, že za predpokladu pravdivosti nulovej hypotézy by sme získali pozorované výsledky alebo ešte extrémnejšie hodnoty, je menšia ako 5 %. [19]

Okrem toho použijeme aj informačné kritériá AIC, BIC a hodnoty log-likelihood funkcie na výber najlepšieho rozdelenia zo zoznamu uvedeného v Tabulke 3.

V tejto fáze práce sme dokončili teoretický rámec, ktorý sa zameriava na popis modelovania pravdepodobnostného rozdelenia. Následne prechádzame k implementácii. Pripomíname, že modelovanie sa vykonáva pre tri skupiny: EÚ, Ukrajina a non-EÚ (bez Ukrajiny). V Tabuľkách 4, 5, 6 sú uvedené výsledky testov zhody (konkrétne hodnoty p-value), informačných kritérií a hodnôt logaritmu vierohodnostnej funkcie pre jednotlivé skupiny.

Na základe výsledkov testov dobrej zhody zamietame nulovú hypotézu pre všetky testované teoretické rozdelenia. Predpokladáme, že jedným z možných vysvetlení takto nízkej hodnoty p-value je vysoký počet pozorovaní, ktorý v každej skupine presahuje 200 000. Vzhľadom na veľký objem dát testy môžu mať vyššiu tendenciu zamietať nulovú hypotézu, aj keď rozdiely medzi teoretickým a empirickým rozdelením sú minimálne. [3]

S ohľadom na tento jav na posúdenie vhodnosti rozdelenia použijeme iba informačné kritériá a log-likelihood funkciu. Tento prístup zvolíme, pretože predikcie a simulácie sa zameriavajú na schopnosť modelu verne reprezentovať realitu a generovať realistické predpovede. Preto je postačujúce, ak pravdepodobnostné rozdelenie zachytí základné charakteristiky dát, čím sa stáva vhodným na použitie ako vstup do SLAMM modelu.

Z Tabuliek 4, 5, 6 vyplýva, že najvhodnejšie rozdelenie pre každú zo skupín je loggama. V Tabuľke 7 sú uvedené parametre pre jednotlivé rozdelenia. Na Obr. 6, 7 a 8 sú vykreslené hustoty rozdelení.

Tabuľka 4: Porovnanie vhodnosti teoretických rozdelení EÚ

Spojité rozdelenia						
Rozdelenie	AD p-val	CvM p-val	ChiSq p-val	AIC	BIC	LogLik
Exponenciálne	1.5e-09	0	-	1709744	1709754	-854871
Gama	1.5e-09	0	-	1582885	1582907	-794140
Loggama	1.5e-09	0	-	1515722	1515743	-757859
Loglogistické	1.5e-09	0	-	1619096	1619118	-809364
Lognormálne	1.5e-09	0	-	1580069	1580091	-790032
Pareto	1.5e-09	0	-	1709756	1709778	-854876
Weibullovo	1.5e-09	0	-	1590052	1590074	-795021
Diskrétne rozdelenia						
Rozdelenie	AD p-val	CvM p-val	ChiSq p-val	AIC	BIC	LogLik
Geometrické	-	-	0	1820548	1820559	-910273
Poissonovo	-	-	0	1649248	1649259	-824623

Tabuľka 5: Porovnanie vhodnosti teoretických rozdelení pre Ukrajinu

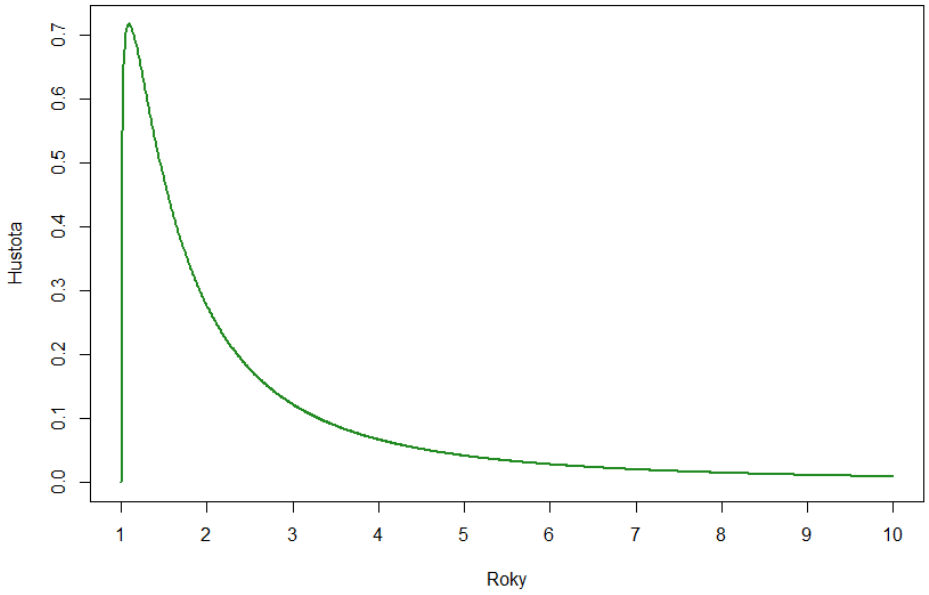
Spojité rozdelenia						
Rozdelenie	AD p-val	CvM p-val	ChiSq p-val	AIC	BIC	LogLik
Exponenciálne	0	2.73432e-09	-	744670	744680	-372334
Gama	0	2.73432e-09	-	552757	552778	-276377
Loggama	0	2.73432e-09	-	490392	490413	-245194
Loglogistické	0	2.73432e-09	-	562408	562429	-281202
Lognormálne	0	2.73432e-09	-	546178	546199	-273087
Paretovo	0	2.73432e-09	-	744672	744692	-372334
Weibullovo	0	2.73432e-09	-	576766	576786	-288381
Diskrétne rozdelenia						
Rozdelenie	AD p-val	CvM p-val	ChiSq p-val	AIC	BIC	LogLik
Geometrické	-	-	0	839336	839346	-419667
Poissonovo	-	-	0	648564	648574	-324281

Tabuľka 6: Porovnanie vhodnosti teoretických rozdelení non-EÚ

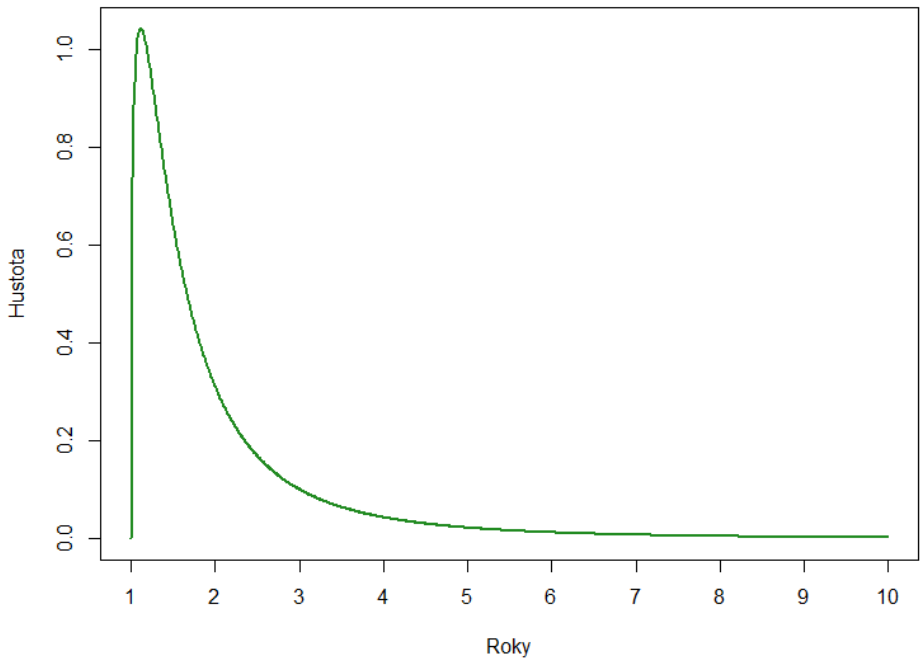
Spojité rozdelenia						
Rozdelenie	AD p-val	CvM p-val	ChiSq p-val	AIC	BIC	LogLik
Exponenciálne	2e-09	0	-	987021	987031	-493509
Gama	2e-09	0	-	807161	807182	-403578
Loggama	2e-09	0	-	737296	737317	-368646
Loglogistické	2e-09	0	-	799225	799246	-399611
Lognormálne	2e-09	0	-	788659	788680	-394328
Paretovo	2e-09	0	-	987023	987044	-493509
Weibullovo	2e-09	0	-	839986	840007	-419991
Diskrétne rozdelenia						
Rozdelenie	AD p-val	CvM p-val	ChiSq p-val	AIC	BIC	LogLik
Geometrické	-	-	0	1088685	1088695	-544341
Poissonovo	-	-	0	881535	881546	-440767

Tabuľka 7: Rozdelenia a parametre pre rôzne skupiny

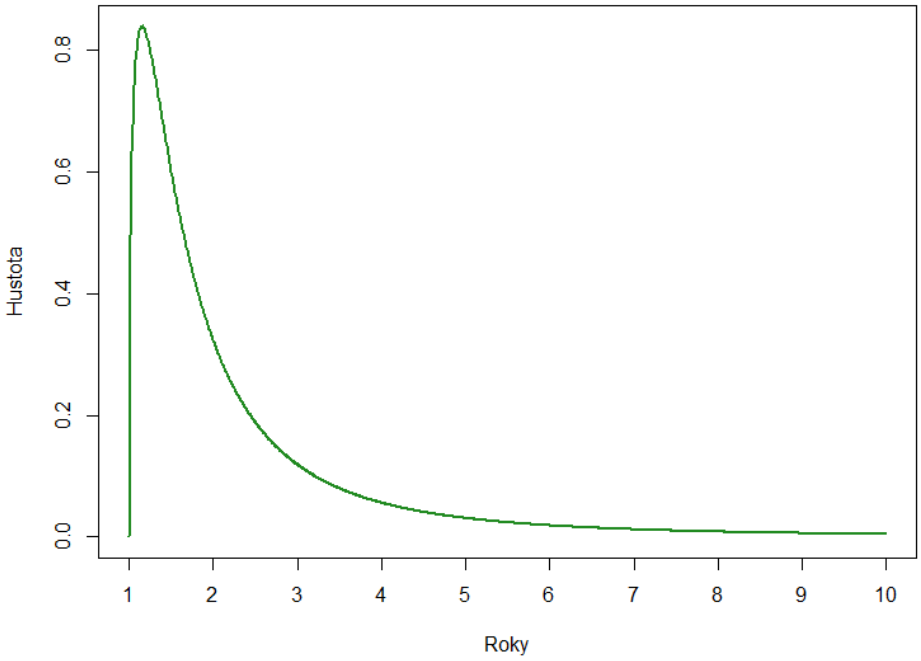
Skupiny	**Rozdelenie**	**Parametre rozdelenia**
EÚ	Loggama	Shape: 1.19; rate: 1.23
Ukrajina	Loggama	Shape: 1.35; rate: 2.18
Non-EÚ	Loggama	Shape: 1.43; rate: 1.96



Obr. 6: Odhad hustoty dĺžky pobytu cudzincov z EÚ pomocou parametrického prístupu, zdroj: vlastná analýza



Obr. 7: Odhad hustoty dĺžky pobytu cudzincov z Ukrajiny pomocou parametrického prístupu, zdroj: vlastná analýza



Obr. 8: Odhad hustoty dĺžky pobytu cudzincov z non-EÚ pomocou parametrického prístupu, zdroj: vlastná analýza

3.2.2 Neparametrické modelovanie

Ďalším prístupom k modelovaniu pravdepodobnostného rozdelenia, ktorý použijeme v našej práci, je neparametrické modelovanie. Tento prístup nevyžaduje predpoklad konkrétneho tvaru rozdelenia ani definovanie parametrov. Je vhodný na analýzu neznámych alebo zložitejších dátových štruktúr, najmä v prípadoch, kde chýbajú predchádzajúce štúdie alebo poznatky o danej problematike. Nevýhodou niektorých neparametrických metód je, že výsledky získané ich použitím môžu byť zložitejšie na interpretáciu v porovnaní s výsledkami parametrických metód.

Postup modelovania pre účely našej diplomovej práce zvolíme podobný ako pri parametrickom prístupe, avšak s niekoľkými úpravami. Kľúčové kroky sú nasledovné:

1. Získanie a príprava dát.
2. Výber vhodnej metódy pre neparametrické modelovanie.
3. Odhad hustoty rozdelenia.

Prvý krok sme už vykonali v predchádzajúcej časti. Následne zvolíme metódu, pomocou ktorej budeme odhadovať tvar hustoty rozdelenia vstupných dát. Existujú rôzne prístupy, ako napríklad jadrové odhady hustoty (angl. Kernel Density Estimation, *KDE*), metóda k-najbližších susedov, simulácie pomocou bootstrapu alebo metódy Monte Carlo. [13], [28] Pre účely tejto diplomovej práce zvolíme jadrové odhady, keďže táto metóda umožňuje získať hladkú hustotu pravdepodobnosti na základe diskrétnych pozorovaní. Táto vlastnosť robí z *KDE* praktický a efektívny nástroj pre analýzu dát, kde je potrebná vizualizácia alebo aproximácia neznámeho rozdelenia pravdepodobnosti. Pri formulovaní teoretických základov sme sa opierali o [12].

Matematicky je jadrový odhad hustoty definovaný ako:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$$

kde:

- n je počet pozorovaní,
- h je šírka pásma (angl. bandwidth),

- K je jadrová funkcia, ktorá je symetrická a jej integrál je rovný 1,
- x_i sú jednotlivé pozorovania z dátovej vzorky.

Hlavným prvkom metódy je jadrová funkcia, ktorá slúži na vyhladenie údajov a vytvorenie spojitého odhadu hustoty. Existuje viacero možností pre voľbu vhodnej funkcie K, medzi ktoré patria napríklad Gaussova, Epanechnikovova alebo uniformná funkcia. V tejto práci zvolíme najbežnejšie používanú Gaussovu funkciu, ktorá je definovaná nasledovne:

$$K(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

Je však dôležité zdôrazniť, že výber konkrétnej jadrovej funkcie má relatívne menší vplyv na presnosť odhadu v porovnaní s voľbou šírky pásma h. [12] Táto hodnota zásadne ovplyvňuje mieru vyhladenia a kvalitu výsledného modelu. Ak je h príliš malé, odhad hustoty bude príliš detailný a citlivý na šum v dátach. Naopak, ak je h príliš veľké, hustota bude príliš vyhladená, čo vedie k strate detailov. [4]

Na určenie optimálnej hodnoty šírky pásma sa používajú rôzne metódy, medzi ktoré patrí Silvermanovo pravidlo, Scottovo pravidlo, ako aj rôzne optimalizačné techniky, ako napríklad minimalizácia integrovaného štvorcového chyby alebo krížová validácia. [4] V rámci tejto diplomovej práce sa zameriame na Silvermanovo pravidlo, ktoré je osvedčenou a efektívnou metódou. Jeho výhodou je nielen jednoduchosť implementácie, ale aj spoľahlivosť pri poskytovaní kvalitných výsledkov, aj v prípade rozsiahlych datasetov.

Odhad šírky pásma h podľa Silvermanovho pravidla je definovaný nasledovne [12]:

$$h = 0.9 \min\left(\hat{\sigma}, \frac{IQR}{1.34}\right) n^{-1/5},$$

kde:

- $\hat{\sigma}$ je odhad štandardnej odchýlky vzorky,
- IQR je medzikvartilový rozsah (rozdiel medzi 75. a 25. percentilom),
- n je počet pozorovaní.

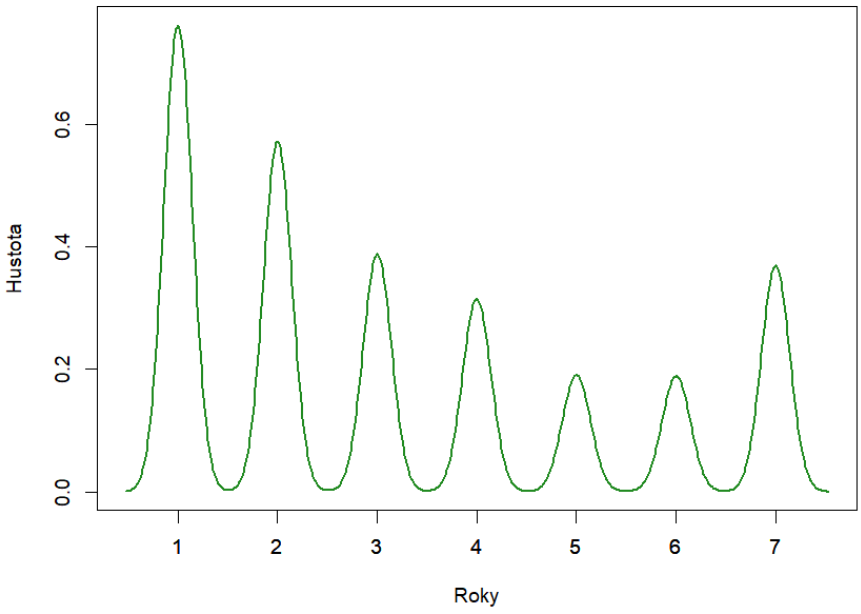
Podobne ako pri parametrickom modelovaní, aj v tomto prípade rozdelíme krajiny do troch skupín. Odhady šírky pásma pre jednotlivé skupiny sú uvedené v Tabuľke 8. Na

Obr. 9, 10 a 11 sú zobrazené grafické znázornenia odhadov hustoty rozdelenia získané pomocou *KDE*.

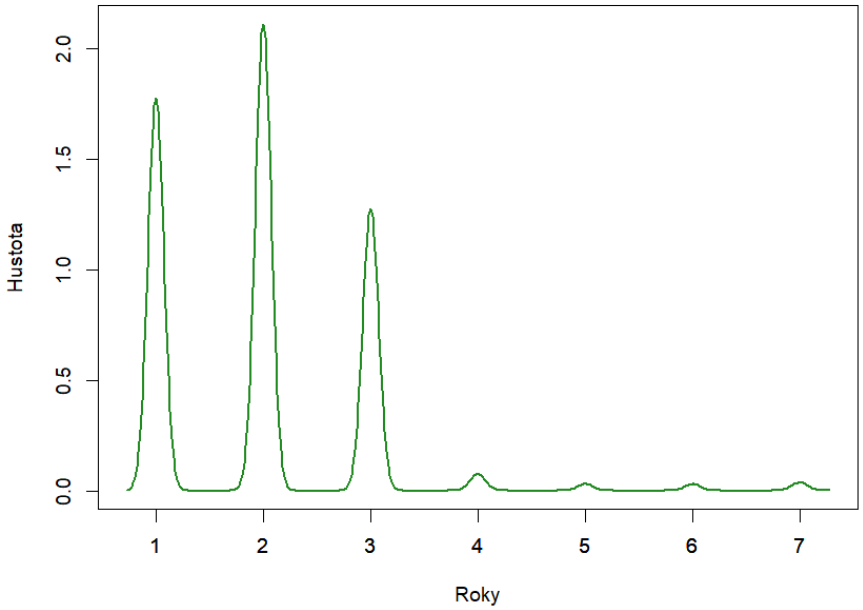
Na základe prezentovaných výsledkov možno pozorovať, že neparametrický prístup zachytáva diskrétny charakter dát až príliš presne. Táto vlastnosť spôsobuje, že výsledné hustoty obsahujú viaceré ostré vrcholy, pričom medzi nimi sa objavujú intervaly s veľmi nízkymi hodnotami hustoty. Tento jav naznačuje, že zvolená metóda má problém s dostatočným vyhladením rozdelenia pre naše vstupné dáta.

Skupiny	**Šírka pásma**
EÚ	0.14
Ukrajina	0.07
non-EÚ	0.1

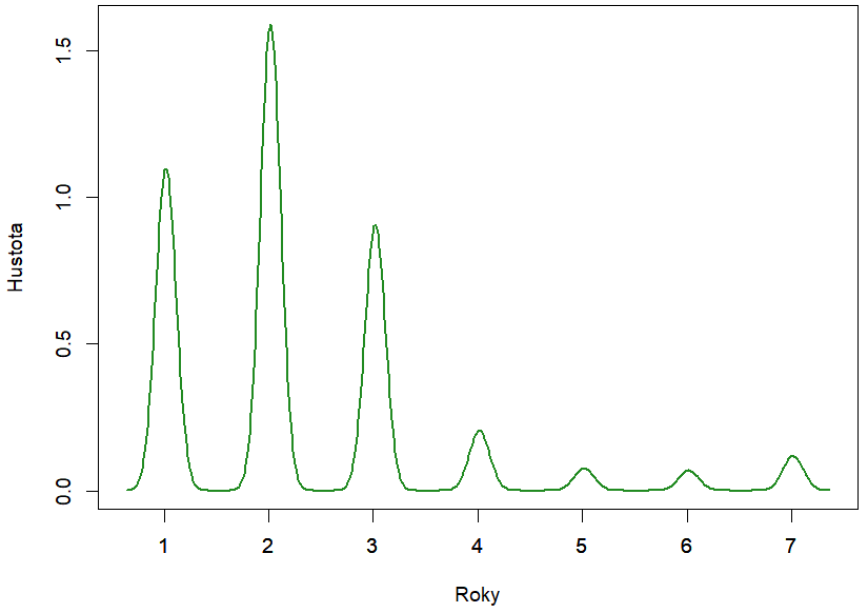
Tabuľka 8: Odhad šírky pásma podľa Silvermanovho pravidla



Obr. 9: Odhad hustoty dĺžky pobytu cudzincov z EÚ pomocou neparametrického prístupu, zdroj: vlastná analýza



Obr. 10: Odhad hustoty dĺžky pobytu cudzincov z Ukrajiny pomocou neparametrického prístupu, zdroj: vlastná analýza



Obr. 11: Odhad hustoty dĺžky pobytu cudzincov z non-EÚ pomocou neparametrického prístupu, zdroj: vlastná analýza

3.3 Transformácia spojitého rozdelenia na diskrétnu formu

V predchádzajúcich častiach sme odhadli tvar hustoty dĺžky pobytu cudzincov. Model SLAMM však nepracuje priamo so spojitou pravdepodobnostnou hustotou, ale vyžaduje diskrétnu pravdepodobnostnú funkciu ako vstup do migračnej simulácie. Preto je potrebné transformovať spojité pravdepodobnostné rozdelenie na diskrétnu formu, ktorú môže model efektívne využiť.

Jednou z možností, ako tento problém vyriešiť, je diskretizácia hustoty. Táto transformácia nám umožní získať pravdepodobnosti pre jednotlivé ročné intervaly, čo je v súlade s časovým krokom simulácie v modeli SLAMM.

Na diskretizáciu môžeme využiť základnú vlastnosť hustoty pravdepodobnosti. Táto vlastnosť hovorí, že plocha pod krivkou hustoty v intervale $[a, b]$ zodpovedá pravdepodobnosti, že hodnota náhodnej premennej X sa nachádza v tomto intervale. [16] Matematicky je tento vzťah vyjadrený nasledovne [16]:

$$P(a \leq X < b) = \int_a^b f(x)\,dx,$$

kde $f(x)$ je hustota pravdepodobnosti náhodnej premennej X.

Pre potreby modelu SLAMM je vhodné pracovať s ročnými intervalmi, pričom každému roku bude priradená diskrétna pravdepodobnosť. Pravdepodobnosť, že hodnota náhodnej premennej X spadne do intervalu $[i, i+1]$, je vyjadrená nasledovnými vzorcami:

$$P(i \leq X < i+1) = \int_i^{i+1} f(x)\,dx, \quad \text{kde } i = 0, 1, 2, \ldots, 8.$$

$$P(9 \leq X < \infty) = \int_9^{\infty} f(x)\,dx.$$

V kontexte modelu bude každý ročný interval reprezentovaný svojou ľavou hranicou i, čo zabezpečí konzistentnosť so vstupmi modelu.

Výsledky diskretizácií pre parametrický a neparametrický prístup sú prezentované v Tabuľkách 9, 10 a 11. Pre zvýšenie prehľadnosti boli výsledky zaokrúhlené, čo môže spôsobiť, že suma pravdepodobností bude mierne menšia alebo väčšia ako 1.

Dĺžka zotrvania	Parametrický prístup	Neparametrický prístup
0	-	0.1362
1	0.4880	0.2394
2	0.1845	0.1726
3	0.0908	0.1264
4	0.0532	0.0908
5	0.0347	0.0683
6	0.0243	0.1003
7	0.0179	0.0660
8	0.0136	0.0000
9	0.0929	0.0000

Tabuľka 9: Výsledky diskretizácií pre EÚ

Dĺžka zotrvania	Parametrický prístup	Neparametrický prístup
0	-	0.1660
1	0.6623	0.3647
2	0.1815	0.3177
3	0.0674	0.1265
4	0.0319	0.0097
5	0.0175	0.0055
6	0.0106	0.0062
7	0.0069	0.0035
8	0.0047	0.0000
9	0.0173	0.0000

Tabuľka 10: Výsledky diskretizácií pre Ukrajinu

Dĺžka zotrvania	Parametrický prístup	Neparametrický prístup
0	-	0.1353
1	0.5886	0.3311
2	0.1996	0.3071
3	0.0823	0.1368
4	0.0418	0.0344
5	0.0242	0.0177
6	0.0153	0.0230
7	0.0103	0.0146
8	0.0073	0.0000
9	0.0308	0.0000

Tabuľka 11: Výsledky diskretizácií pre non-EÚ

3.4 Finalizácia pravdepodobností pre funkciu *time_to_leave*

V tejto časti finalizujeme výsledky pravdepodobnostných rozdelení, ktoré budú použité ako vstupné hodnoty pre funkciu *time_to_leave*. V predchádzajúcich častiach sme vypočítali pravdepodobnostné rozdelenia pre parametrický aj neparametrický prístup. V základnom prístupe odhadujeme pravdepodobnosti zotrvania migrantov na základe podielu pracujúcich migrantov v jednotlivých intervaloch dĺžky pobytu.

Výsledky tohto prístupu sú uvedené v Tabuľke 12. Táto metóda poskytuje jednoduchý empirický odhad, ktorý slúži ako referenčný bod pre porovnanie s parametrickým a neparametrickým modelovaním.

Dĺžka zotrvania	**EÚ**	**Ukrajina**	**non-EÚ**
0	0.2731	0.3331	0.2711
1	0.2055	0.3960	0.3908
2	0.1395	0.2390	0.2232
3	0.1131	0.0139	0.0502
4	0.0682	0.0055	0.0177
5	0.0682	0.0055	0.0177
6+	0.1324	0.007	0.0292

Tabuľka 12: Pravdepodobnosti pre dĺžky zotrvania podľa skupín pre základný prístup

Model SLAMM vyžaduje pravdepodobnosti pre jednotlivé roky zotrvania $0, 1, \ldots, 9$. Avšak v základnom prístupe sú pravdepodobnosti pre roky 6 až 9 agregované do jednej skupiny označenej ako 6+, čo spôsobuje nekonzistentnosť údajov. Na odstránenie tejto nekonzistencie a zabezpečenie presnosti údajov použijeme loggama rozdelenia, ktoré boli predtým vypočítané v rámci parametrického prístupu. Tento postup nám umožňuje priradiť pravdepodobnosti jednotlivým rokom (6, 7, 8, 9) pomerovo, a to na základe tvaru loggama rozdelenia.

Podobná nekonzistencia sa vyskytuje aj v parametrickom prístupe, kde pravdepodobnostné rozdelenia nezačínajú od dĺžky zotrvania 0, ale od 1. Tento problém vyriešime tak, že pravdepodobnosti pre dĺžku 1 budú proporčne rozdelené medzi 0 a 1 na základe pomerov v základnom prístupe. Tieto úpravy zabezpečujú plnú konzistenciu dát.

Výsledky finalizovaných pravdepodobnostných rozdelení podľa jednotlivých skupín sú uvedené v Tabuľkách 13, 14 a 15.

Pri porovnaní prístupov je vidieť, že základný a parametrický prístup vykazujú podobné hodnoty vo všetkých skupinách, zatiaľ čo neparametrický prístup vykazuje odlišnosti, najmä pre hodnoty 0 a 2. Okrem toho predpokladáme, že neparametrický prístup môže byť najmenej vhodný na odhad rozdelenia dĺžky zotrvania, pretože pre roky 8 a 9 sú pravdepodobnosti pre všetky skupiny nulové, čo nezodpovedá reálnemu správaniu migrantov. Podrobnejšie porovnanie jednotlivých prístupov bude vykonané v nasledujúcej kapitole, kde sa zhodnotí, ktorý prístup poskytuje najpresnejšie výsledky.

Týmto krokom bol dokončený proces výpočtu a úprav vstupných hodnôt pre funkciu *time_to_leave*.

Dĺžka zotrvania	**Základný**	**Parametrický**	**Neparametrický**
0	0.2731	0.2785	0.1362
1	0.2055	0.2095	0.2394
2	0.1395	0.1845	0.1726
3	0.1131	0.0908	0.1264
4	0.0682	0.0532	0.0908
5	0.0682	0.0347	0.0683
6	0.0429	0.0243	0.1003
7	0.0278	0.0179	0.0660
8	0.0184	0.0136	0.0000
9	0.0433	0.0929	0.0000

Tabuľka 13: Výsledky pravdepodobnostného rozdelenia pre funkciu *time_to_leave* v rámci EÚ

Dĺžka zotrvania	Základný	Parametrický	Neparametrický
0	0.3331	0.3026	0.1660
1	0.3960	0.3597	0.3647
2	0.2390	0.1815	0.3177
3	0.0139	0.0674	0.1265
4	0.0055	0.0319	0.0097
5	0.0055	0.0175	0.0055
6	0.0044	0.0106	0.0062
7	0.0016	0.0069	0.0035
8	0.0006	0.0047	0.0000
9	0.0004	0.0173	0.0000

Tabuľka 14: Výsledky pravdepodobnostného rozdelenia pre funkciu *time_to_leave* v rámci Ukrajiny

Dĺžka zotrvania	Základný	Parametrický	Neparametrický
0	0.2711	0.2411	0.1353
1	0.3908	0.3475	0.3311
2	0.2232	0.1996	0.3071
3	0.0502	0.0823	0.1368
4	0.0177	0.0418	0.0344
5	0.0177	0.0242	0.0177
6	0.0156	0.0153	0.0230
7	0.0070	0.0103	0.0146
8	0.0033	0.0073	0.0000
9	0.0033	0.0308	0.0000

Tabuľka 15: Výsledky pravdepodobnostného rozdelenia pre funkciu *time_to_leave* v rámci non-EÚ

3.5 Odhad pravdepodobností odchodu migrantov

Nasledujúcim krokom na zlepšenie rozdelenia migrantov v modeli SLAMM je úprava funkcie *leave*, ktorá určuje rozhodovanie migrantov o odchode z krajiny počas jednotlivých rokov simulácie. Pripomíname, že rozhodnutia sú založené na hodnote *time_to_leave*, priradenej každému jednotlivcovi. Pri kalibrácii funkcie *leave* sme sa rozhodli využiť funkciu prežitia (angl. survival function, $S(t)$).

V rámci tejto diplomovej práce bude survival funkcia $S(t)$ vyjadrovať pravdepodobnosť, že čas zotrvania migranta T v krajine bude dlhší ako určitý čas t. Matematicky je táto funkcia definovaná ako [8]

$$S(t) = P(T > t),$$

kde T je náhodná premenná reprezentujúca čas zotrvania migranta. Táto funkcia má nasledovné vlastnosti [8]:

- čas t je definovaný na intervale $[0, \infty)$,
- $S(0) = 1$,
- funkcia $S(t)$ je klesajúca, t.j. $S(t_1) \geq S(t_2)$ pre $t_1 \leq t_2$.

Táto survival funkcia však poskytuje kumulatívnu pravdepodobnosť zotrvania. V rámci modelu SLAMM je však dôležité mať pravdepodobnosti viazané na jednotlivé roky zotrvania t, a nie na kumulatívne nerovnosti $P(T > t)$.

Taktiež, na základe vstupných dát je vidieť, že pre Ukrajinu a krajiny mimo EÚ sa najväčší počet migrantov nenachádza v kategórii 0-12 mesiacov, ale v tej nasledujúcej, 12-24 mesiacov. Tento jav môže byť dôsledkom obmedzení v kvalite dostupných dát a spôsobom ich spracovania do agregovanej podoby. Tento faktor znemožňuje aplikáciu tradičných modelov prežitia, keďže je porušený jeden z ich kľúčových predpokladov: že najväčší počet osôb sa nachádza na začiatku sledovaného obdobia, pričom neskôr klesá v dôsledku odchodov.

Rovnako, vzhľadom na povahu týchto dát, zlyháva aj výpočet pravdepodobnosti, ktorý by bol založený na podiele počtu migrantov medzi dvoma časovými bodmi. Ako príklad možno uviesť výpočet počtu migrantov v kategórii 12–24 mesiacov k decembru 2022, ktorý by mal zodpovedať počtu osôb v kategórii 0–12 mesiacov k decembru 2021, po odpočítaní

tých, ktorí krajinu opustili. Tento prístup by však viedol k negatívnym podielom, ktoré nevyhovujú definícii pravdepodobnosti.

Preto sme sa rozhodli využiť alternativny prístup. Na rozdiel od survival funkcie tento prístup nepočíta pravdepodobnosť prežitia, ale priamo pravdepodobnosť odchodu.

Pre každú dĺžku zotrvania t vypočítame pravdepodobnosť, že jednotlivec opustí Slovensko, ako pomer medzi počtom cudzincov s dĺžkou zotrvania t a súčtom počtu cudzincov s dĺžkou zotrvania od t do poslednej možnej dĺžky zotrvania T. Tento vzorec vyzerá nasledovne:

$$P(\text{odchod pri dĺžke zotrvania } t) = \frac{\text{Počet cudzincov s dĺžkou zotrvania } t}{\sum_{i=t}^{T} \text{Počet cudzincov s dĺžkou zotrvania } i},$$
$$t = 1, \ldots, T, \quad T = 10.$$

Pre lepšie pochopenie výpočtu uvedieme konkrétne príklady pre $t = 1$ a $t = 2$. Pre výpočet pravdepodobnosti odchodu pri dĺžke zotrvania $t = 1$ použijeme nasledujúci vzorec:

$$P(\text{odchod pri dĺžke zotrvania 1}) = \frac{\text{Počet cudzincov s dĺžkou zotrvania 1}}{\sum_{i=1}^{10} \text{Počet cudzincov s dĺžkou zotrvania } i}.$$

Podobne, pre výpočet pravdepodobnosti odchodu pri dĺžke zotrvania $t = 2$, vzorec bude:

$$P(\text{odchod pri dĺžke zotrvania 2}) = \frac{\text{Počet cudzincov s dĺžkou zotrvania 2}}{\sum_{i=2}^{10} \text{Počet cudzincov s dĺžkou zotrvania } i}.$$

Výsledky výpočtu pravdepodobnosti odchodu pre funkciu *leave* pomocou tohto prístupu, rozdelené podľa jednotlivých skupín, sú zobrazené v Tabuľkách 16, 17 a 18. Pri dĺžke zotrvania 10 rokov dosahuje pravdepodobnosť odchodu hodnotu 100 %. Tento výsledok odráža skutočnosť, že po uplynutí tohto obdobia sa cudzinec stáva občanom Slovenska a už nie je považovaný za imigranta. Týmto boli ukončené všetky potrebné kalibrácie pre migračnú funkciu SLAMM modelu.

Dĺžka zotrvania	Pravdepodobnosť odchodu
1	0.2731
2	0.2826
3	0.2675
4	0.2962
5	0.2538
6	0.3401
7	0.3242
8	0.3108
9	0.2982
10	1.0000

Tabuľka 16: Výsledky pravdepodobnosti odchodu pre EÚ

Dĺžka zotrvania	Pravdepodobnosť odchodu
1	0.3331
2	0.5937
3	0.8821
4	0.4366
5	0.3058
6	0.4406
7	0.6350
8	0.6151
9	0.5958
10	1.0000

Tabuľka 17: Výsledky pravdepodobnosti odchodu pre Ukrajinu

Dĺžka zotrvania	Pravdepodobnosť odchodu
1	0.2711
2	0.5362
3	0.6603
4	0.4369
5	0.2739
6	0.3773
7	0.5915
8	0.6530
9	0.8816
10	1.0000

Tabuľka 18: Výsledky pravdepodobnosti odchodu pre non-EÚ

4 Predikcie a validácia modelu

V tejto kapitole sa zameriame na analýzu výsledkov mikrosimulačného modelu. Implementujeme pravdepodobnostné rozdelenia pre premennú dĺžky zotrvania na Slovensku, ktoré boli vypočítané v predchádzajúcej časti. Následne overíme hypotézy H1 a H2, formulované v sekcii 3.1. Cieľom je zistiť, či vylepšená implementácia pravdepodobnostných rozdelení vedie k zlepšeniu výsledkov modelu v porovnaní s historickými dátami a určiť najefektívnejší prístup z hľadiska presnosti.

4.1 Predikcie modelu na základe rôznych pravdepodobnostných rozdelení

V tejto časti aplikujeme pravdepodobnosti odvodené v predchádzajúcej kapitole. Výpočty zahŕňajú dosadenie výsledkov pre funkciu *leave*, ako aj implementáciu troch prístupov pre funkciu *time_to_leave*. Na overenie hypotéz H1 a H2, definovaných v kapitole 3.1, porovnáme výsledky jednotlivých simulácií s historickými dátami z ÚPSVaR.

Keďže začiatočný rok analýzy je 2019, simulácie začínajú od roku 2020 a pokračujú až do posledného roku, pre ktorý máme k dispozícii kompletné reálne údaje, konkrétne do roku 2024.

Model SLAMM využíva stochastické funkcie pri výpočtoch, čo znamená, že jedna simulácia neposkytuje dostatočne presný obraz o možných výsledkoch. Preto pre každý z troch prístupov spustíme model 100-krát, pričom z výsledkov simulácií vypočítame strednú hodnotu. Tento postup nám umožňuje získať reprezentatívne a robustné výstupy, ktoré slúžia ako základ na analýzu a porovnanie jednotlivých prístupov.

Výsledky skúmame z dvoch hľadísk:

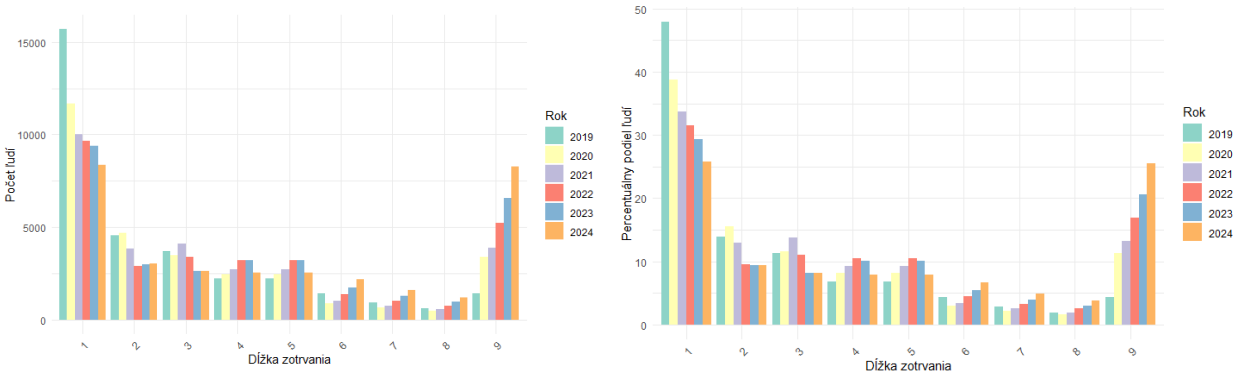
- Absolútny počet cudzincov podľa dĺžky zotrvania na Slovensku.
- Rozdelenie cudzincov podľa dĺžky zotrvania.

Tieto dva postupy poskytujú užitočný pohľad na presnosť modelu. Posúdenie počtov cudzincov podľa dĺžky zotrvania umožňuje overiť, či model správne zachytáva rozloženie cudzincov v čase a predpovedá dynamiku zmien v počtoch ľudí zotrvávajúcich na Slovensku.

Na druhej strane, rozdelenie cudzincov podľa dĺžky zotrvania poskytuje detailnejší pohľad na správanie jednotlivých podskupín, ako sú krátkodobí alebo dlhodobí cudzinci. Tento prístup je vhodný pri hodnotení schopnosti modelu predpovedať, ako sa tieto podskupiny vyvíjajú v jednotlivých rokoch simulácie.

Na základe predpokladov a architektúry modelu SLAMM sa očakáva, že počet cudzincov bude najvyšší pri najkratšej dĺžke zotrvania. Táto hypotéza však nie vždy koreší s reálnymi dátami. Tento nesúlad môže negatívne ovplyvniť presnosť výsledkov modelu. Na zmiernenie tohto vplyvu pri porovnávaní výsledkov s historickými dátami zlúčime počty cudzincov so dĺžkou zotrvania 0 a 1. Týmto spôsobom zabezpečíme, že najvyšší počet cudzincov sa bude nachádzať pri najkratšej dĺžke zotrvania.

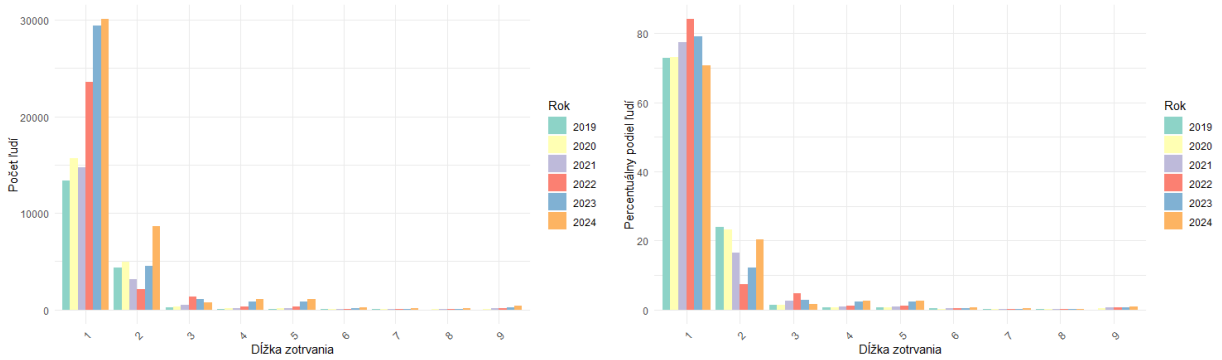
Pre každú zo skúmaných skupín vykreslíme počty cudzincov a ich rozdelenia podľa dĺžky zotrvania na základe historických údajov za obdobie rokov 2019–2024. Výsledky sú prezentované na Obr. 12 až Obr. 14. Do grafov sme zahrnuli aj rok 2019, aby sme mohli porovnať rozdelenie cudzincov podľa dĺžky zotrvania pre jednotlivé roky. Tento krok nám umožňuje overiť, či model spĺňa svoj predpoklad o konzistentnosti rozdelenia simulovaných rokov s historickými údajmi z roku 2019.



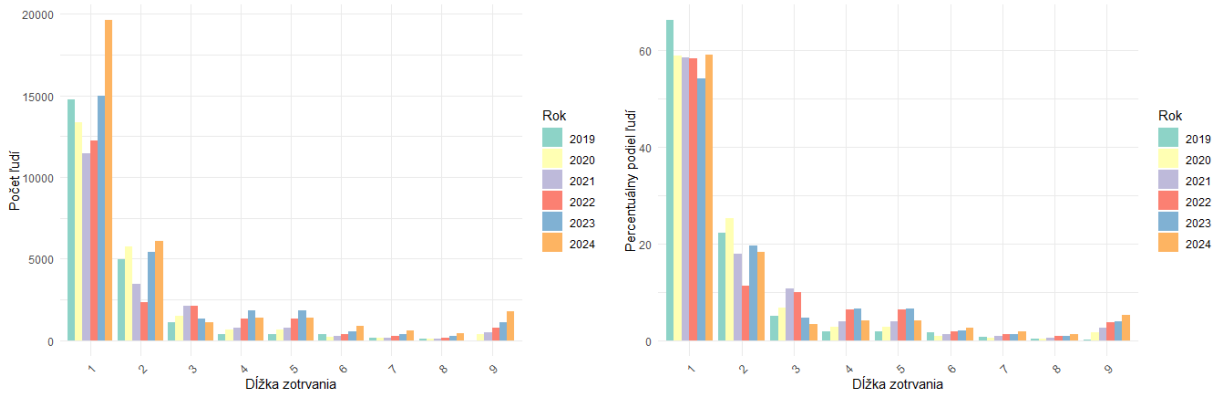
Obr. 12: Rozdelenie cudzincov z EÚ podľa dĺžky zotrvania na základe historických dát: Absolútne hodnoty a Percentuálne pomery, zdroj: vlastná analýza

Na základe grafov môžeme pozorovať, že rozdelenia pre roky 2020–2024 nie vždy presne zodpovedajú rozdeleniu za rok 2019. Najvýraznejšie odchýlky sú viditeľné na Obr. 12 pri cudzincoch z EÚ. Táto skutočnosť môže negatívne ovplyvniť presnosť modelu pri porovnávaní jeho výsledkov so skutočnými historickými údajmi.

Ďalej sú prezentované výsledky simulácií pre tri prístupy funkcie *time_to_leave* v jednotlivých rokoch (viď Obr. 15 až Obr. 23).

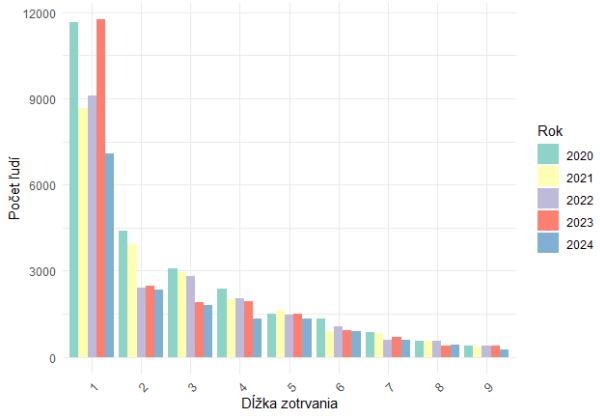


Obr. 13: Rozdelenie cudzincov z Ukrajiny podľa dĺžky zotrvania na základe historických dát: Absolútne hodnoty a Percentuálne pomery, zdroj: vlastná analýza

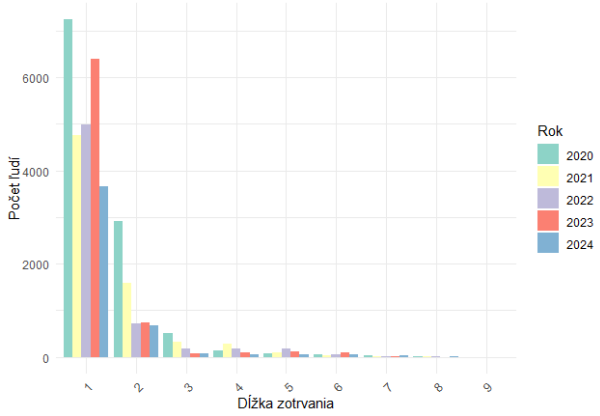


Obr. 14: Rozdelenie cudzincov z non-EÚ podľa dĺžky zotrvania na základe historických dát: Absolútne hodnoty a Percentuálne pomery, zdroj: vlastná analýza

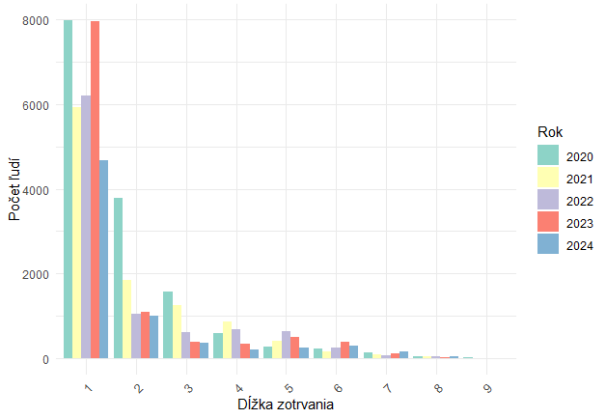
Avšak, samotné grafické porovnanie výsledkov nie je postačujúce na komplexnú analýzu. Preto je dôležité následne aplikovať vhodné metriky na validáciu týchto výsledkov v porovnaní s historickými údajmi, čím zabezpečíme presnosť a spoľahlivosť celej validácie.



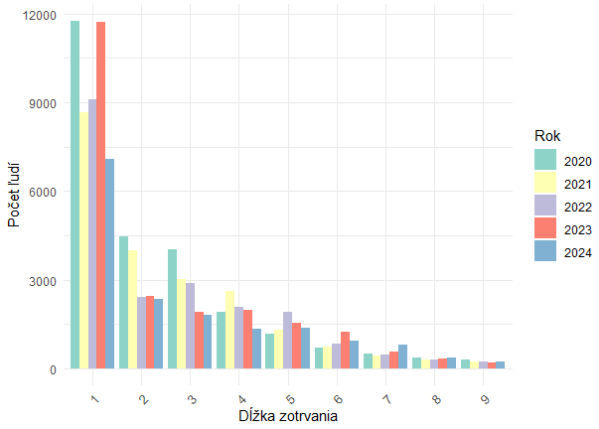
Obr. 15: Simulácie rozdelenia cudzincov z EÚ podľa dĺžky zotrvania pomocou základného prístupu: Absolútne hodnoty, zdroj: vlastná analýza



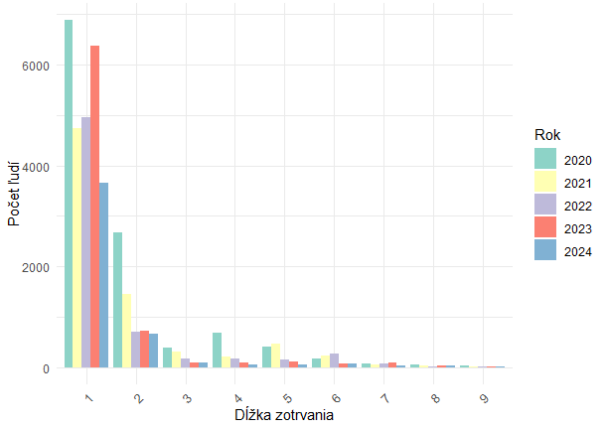
Obr. 16: Simulácie rozdelenia cudzincov z Ukrajiny podľa dĺžky zotrvania pomocou základného prístupu: Absolútne hodnoty, zdroj: vlastná analýza



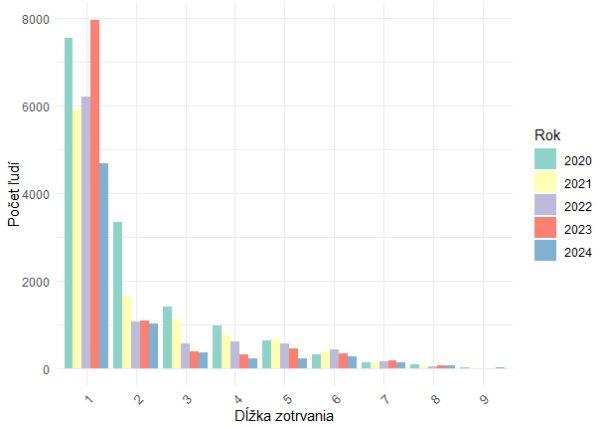
Obr. 17: Simulácie rozdelenia cudzincov z non-EÚ podľa dĺžky zotrvania pomocou základného prístupu: Absolútne hodnoty, zdroj: vlastná analýza



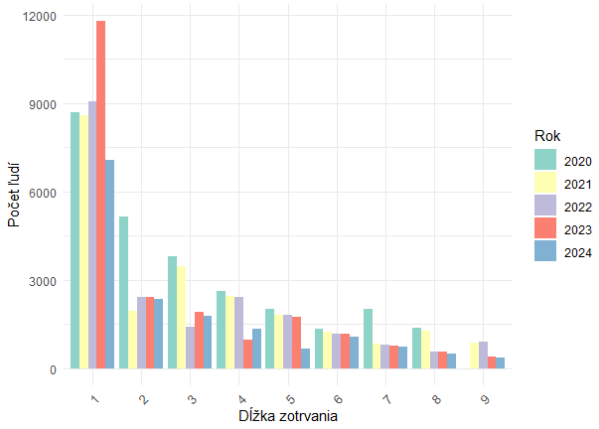
Obr. 18: Simulácie rozdelenia cudzincov z EÚ podľa dĺžky zotrvania pomocou parametrického prístupu: Absolútne hodnoty, zdroj: vlastná analýza



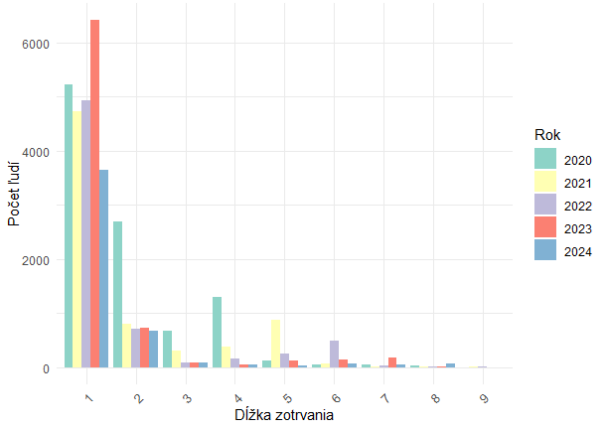
Obr. 19: Simulácie rozdelenia cudzincov z Ukrajiny podľa dĺžky zotrvania pomocou parametrického prístupu: Absolútne hodnoty, zdroj: vlastná analýza



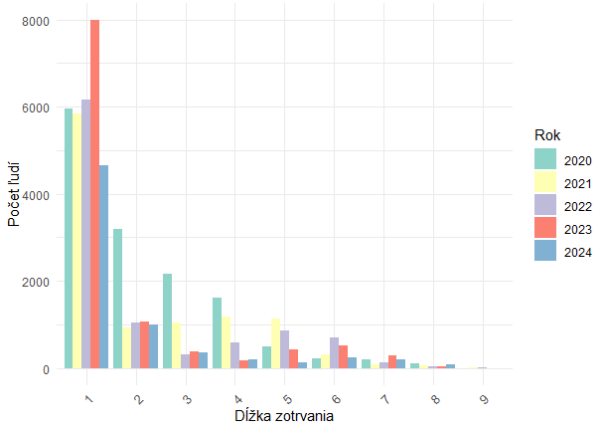
Obr. 20: Simulácie rozdelenia cudzincov z non-EÚ podľa dĺžky zotrvania pomocou parametrického prístupu: Absolútne hodnoty, zdroj: vlastná analýza



Obr. 21: Simulácie rozdelenia cudzincov z EÚ podľa dĺžky zotrvania pomocou neparametrického prístupu: Absolútne hodnoty, zdroj: vlastná analýza



Obr. 22: Simulácie rozdelenia cudzincov z Ukrajiny podľa dĺžky zotrvania pomocou neparametrického prístupu: Absolútne hodnoty, zdroj: vlastná analýza



Obr. 23: Simulácie rozdelenia cudzincov z non-EÚ podľa dĺžky zotrvania pomocou neparametrického prístupu: Absolútne hodnoty, zdroj: vlastná analýza

4.2 Validácia výsledkov s historickými dátami

Na validáciu výsledkov je možné využiť rôzne metriky. Na účel tejto diplomovej práce volíme tri metriky: *MAE* (stredná absolútna chyba, angl. Mean Absolute Error), *MAPE* (stredná absolútna percentuálna chyba, angl. Mean Absolute Percentage Error) a *KLD* (Kullbackova-Leiblerova divergencia, angl. Kullback-Leibler divergence). Každá z týchto metrík poskytuje odlišný pohľad na presnosť odhadov získaných zo simulácií v porovnaní s reálnymi údajmi.

MAE analyzuje absolútne odchýlky medzi predikovanými a skutočnými hodnotami, zatiaľ čo *MAPE* hodnotí relatívne odchýlky, čím poskytuje lepší pohľad na chyby v kontexte veľkosti dát. Kullback-Leiblerova divergencia sa zameriava na odchýlky v rozdelení pravdepodobnosti medzi simulovanými a reálnymi dátami, čím umožňuje lepšie pochopenie štruktúrnych rozdielov medzi týmito súbormi údajov.

V prípade *MAE* je vzorec nasledovný [26]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|,$$

kde y_i predstavuje skutočné hodnoty, $\hat{y}_i$ odhadované hodnoty a n je počet pozorovaní.

Pre výpočet *MAPE* platí tento vzorec [18]:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Na výpočet Kullback-Leiblerovej divergencie sa používa nasledujúci vzorec [25]:

$$D_{KL}(P||Q) = \sum_{i}^{n} P(i)\log\frac{P(i)}{Q(i)},$$

kde $P(i)$ a $Q(i)$ predstavujú pravdepodobnostné rozdelenia reálnych a simulovaných dát. Vyššie hodnoty *KLD* naznačujú väčšiu nerovnosť medzi distribúciami. Hodnota Kullback-Leiblerovej divergencie je vždy nezáporná. To znamená, že $D_{KL}(P \parallel Q) \geq 0$ pre všetky P a Q, pričom hodnota 0 nastáva iba v prípade, že obe distribúcie sú identické. Treba však poznamenať, že pri výpočte môžu nastať problémy v prípadoch, keď $Q(i) = 0$. Na riešenie tohto problému budú nulové hodnoty $Q(i)$ nahradené malým kladným číslom $\varepsilon = e^{-10}$, aby sa predišlo numerickým chybám a singularitám pri výpočte.

Tieto tri metriky použijeme na validáciu simulovaných dát pre každý z prístupov *time_to_leave*, pričom index 1 označuje základný prístup, index 2 parametrický prístup a index 3 neparametrický prístup. Navyše, tieto metriky budú vypočítané samostatne pre absolútne počty a pre rozdelenie cudzincov (viď Tabuľku 19, resp. 20).

Tabuľka 19: Validácia výsledkov: Absolútne hodnoty

EÚ									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	610	28%	0.149	698	29%	0.184	1160	67%	2.289
2021	909	25%	0.172	969	35%	0.234	1021	39%	0.105
2022	1139	35%	0.256	1165	40%	0.346	1129	32%	0.143
2023	1636	45%	0.397	1636	45%	0.526	1671	44%	0.415
2024	1808	50%	0.529	1792	49%	0.582	1831	50%	0.466
Ukrajina									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	1212	42%	0.030	1365	156%	0.093	1603	145%	0.238
2021	1340	57%	0.046	1385	76%	0.087	1507	95%	0.145
2022	2412	68%	0.040	2429	73%	0.065	2458	104%	0.121
2023	3287	80%	0.034	3282	69%	0.032	3291	78%	0.060
2024	4202	89%	0.054	4201	87%	0.038	4202	83%	0.061
Non-EÚ									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	916	34%	0.050	1010	35%	0.062	1355	57%	0.418
2021	1016	48%	0.059	1020	45%	0.094	1183	54%	0.159
2022	1267	63%	0.087	1268	57%	0.148	1303	66%	0.187
2023	1867	72%	0.133	1872	70%	0.225	1856	64%	0.295
2024	2916	81%	0.244	2922	81%	0.187	2924	80%	0.208

Tabuľka 20: Validácia výsledkov: Percentuálne pomery

EÚ									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	2.7	35%	0.149	3.3	27%	0.184	4.2	79%	2.289
2021	3.0	30%	0.172	3.6	30%	0.234	3.3	52%	0.105
2022	4.2	28%	0.256	4.2	31%	0.346	4.1	30%	0.143
2023	5.9	36%	0.397	5.9	38%	0.526	6.1	34%	0.415
2024	6.0	37%	0.529	6.0	36%	0.582	6.2	37%	0.466
Ukrajina									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	1.7	77%	0.030	2.8	323%	0.093	4.9	342%	0.238
2021	2.5	69%	0.046	3.3	220%	0.087	3.9	222%	0.145
2022	1.8	76%	0.040	2.4	221%	0.065	3.2	291%	0.121
2023	1.4	63%	0.034	1.4	104%	0.032	1.6	164%	0.060
2024	2.1	60%	0.054	2.1	69%	0.038	2.2	113%	0.061
Non-EÚ									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	1.5	45%	0.050	2.3	66%	0.062	4.5	105%	0.418
2021	1.2	31%	0.059	1.9	60%	0.094	3.5	74%	0.159
2022	1.9	35%	0.087	2.0	42%	0.148	2.6	63%	0.187
2023	4.7	54%	0.133	4.7	47%	0.225	5.2	69%	0.295
2024	2.6	42%	0.244	2.6	36%	0.187	2.8	42%	0.208

V prípade občanov EÚ v rokoch 2020 až 2022 základný prístup dosahuje najnižšie hodnoty metrík *MAE* a *MAPE*, čo naznačuje, že najvernejšie odráža skutočný vývoj migrácie. V týchto rokoch sa tiež vyznačuje nižšími hodnotami Kullback-Leiblerovej divergencie, čo potvrdzuje jeho schopnosť nielen reprodukovať celkové počty cudzincov, ale aj ich rozdelenie podľa dĺžky zotrvania. Parametrický prístup vykazuje v priebehu väčšiny sledovaných rokov mierne vyššie odchýlky, no udržiava stabilné a konzistentné výsledky. V prípade percentuálnych pomerov dokonca dosahuje lepšie hodnoty *MAPE* než základný prístup. Naopak, neparametrický prístup vykazuje nižšiu presnosť podľa metrík *MAE* a *MAPE* v úvodných rokoch simulovaného obdobia, pričom najväčšie odchýlky sú zaznamenané v roku 2020. V poslednom roku simulácie sa však výsledky všetkých troch prístupov začínajú stabilizovať.

Pri migrantoch z Ukrajiny sú vo všetkých troch prístupoch zaznamenané vyššie chyby podľa metrík *MAE* a *MAPE*. Hodnota *MAPE* dosahuje pri absolútnych počtoch cudzincov aj pri percentuálnych pomeroch veľmi vysoké hodnoty. Vysoké hodnoty pri percentuálnych pomeroch sú spôsobené citlivosťou *MAPE* na odchýlky pri malých číslach, kde aj minimálny nesúlad môže viesť k výrazným percentuálnym rozdielom. Taktiež tvar rozdelenia zohráva kľúčovú rolu – pre Ukrajincov je vysoká pravdepodobnosť krátkodobého zotrvania na Slovensku, pričom pri dĺžke zotrvania nad 3 roky sa pozoruje len minimálne percento migrantov.

Napriek vysokým hodnotám *MAE* a *MAPE* hodnoty *KLD* zostávajú relatívne nízke v porovnaní s EÚ, čo naznačuje, že model efektívnejšie predikuje tvar rozdelenia dĺžky zotrvania. Najnižšie odchýlky sú dosahované pri základnom prístupe, zatiaľ čo neparametrický vykazuje najväčšie odchýlky.

V prípade migrantov z krajín mimo EÚ sú rozdiely medzi prístupmi menej výrazné. V skorších rokoch dosahuje základný prístup najnižšie hodnoty chýb. Od roku 2022 však parametrický prístup v niektorých prípadoch dosahuje lepšie alebo porovnateľné výsledky. Neparametrický prístup sa v tejto skupine ukazuje ako najmenej presný v počiatočných rokoch, pričom v roku 2024 sa výsledky pre všetky prístupy stávajú dostatočne podobné.

V rámci tejto diplomovej práce dodatočne analyzujeme výsledky validácie z Tabuliek 19 a 20 agregovaním podľa skupín migrantov, resp. rokov, pričom agregovanie realizujeme prostredníctvom priemerovania hodnôt. Táto analýza poskytuje hlbší pohľad na efektív-

nosť jednotlivých prístupov a umožňuje identifikovať trendy, ktoré naznačujú, v ktorých lokalitách a časových obdobiach model dosahuje lepšie alebo horšie výsledky.

Podľa Tabuliek 21 a 22 sa ukazujú výrazné rozdiely medzi jednotlivými skupinami migrantov. Najväčšie odchýlky v hodnotách *MAE* a *MAPE* sa vyskytujú pri migrantoch z Ukrajiny, pričom hodnoty *KLD* pre túto skupinu sú najlepšie zo všetkých. Najnižšie chyby v *MAE* a *MAPE* sa zaznamenávajú pri migrantoch z EÚ. Pre všetky tri skupiny dosiahol najlepšie výsledky základný prístup podľa všetkých metrík, pričom parametrický prístup v niektorých prípadoch dosiahol porovnateľné výsledky.

Tabuľka 21: Validácia výsledkov, agregovaných podľa skupín: Absolútne hodnoty

EÚ								
MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
1220	37%	0.301	1252	40%	0.374	1362	46%	0.684
Ukrajina								
MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2491	67%	0.041	2532	92%	0.063	2612	101%	0.121
Non-EÚ								
MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
1596	60%	0.115	1618	58%	0.143	1724	64%	0.253

Tabuľka 22: Validácia výsledkov, agregovaných podľa skupín: Percentuálne pomery

EÚ								
MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
4.4	33%	0.301	4.6	32%	0.374	4.8	46%	0.684
Ukrajina								
MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
1.9	69%	0.041	2.4	187%	0.063	3.2	226%	0.125
Non-EÚ								
MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2.4	41%	0.115	2.7	50%	0.143	3.7	71%	0.253

Z časového hľadiska sa pozoruje postupný pokles presnosti predikcií, čo je viditeľné v Tabuľkách 23 a 24. Tento trend je najvýraznejší v hodnotách *MAE* a *KLD*. V rokoch 2023 a 2024 sa však rozdiely medzi jednotlivými prístupmi zmenšujú, pričom najlepšie výsledky aj naďalej dosahuje základný prístup, rovnako ako pri agregácii podľa skupín.

Tabuľka 23: Validácia výsledkov, agregovaných podľa rokov: Absolútne hodnoty

Priemery podľa rokov									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	913	35%	0.076	1024	73%	0.113	1373	90%	0.982
2021	1088	43%	0.092	1125	52%	0.138	1237	63%	0.136
2022	1606	55%	0.128	1621	57%	0.186	1630	67%	0.150
2023	2263	66%	0.188	2263	61%	0.229	2273	62%	0.257
2024	2975	73%	0.276	2972	72%	0.269	2986	71%	0.245

Tabuľka 24: Validácia výsledkov, agregovaných podľa rokov: Percentuálne pomery

Priemery podľa rokov									
Rok	MAE_1	$MAPE_1$	KLD_1	MAE_2	$MAPE_2$	KLD_2	MAE_3	$MAPE_3$	KLD_3
2020	2.0	52%	0.076	2.8	139%	0.113	4.5	175%	0.982
2021	2.2	43%	0.092	2.9	104%	0.138	3.6	116%	0.136
2022	2.6	46%	0.128	2.9	98%	0.186	3.3	128%	0.150
2023	4.0	51%	0.188	4.0	63%	0.261	4.3	89%	0.257
2024	3.6	46%	0.276	3.6	47%	0.269	3.7	64%	0.245

Na základe dosiahnutých výsledkov možno konštatovať, že základný prístup k modelovaniu dĺžky zotrvania preukázal najvyššiu mieru konzistentnosti a presnosti vo väčšine analyzovaných prípadov. Napriek svojej flexibilite, neparametrický prístup vykazoval výrazne vyššiu mieru odchýlok a ukázal sa ako najmenej vhodný nástroj. Na základe týchto zistení sú hypotézy H1 a H2 zamietnuté, keďže najlepšie výsledky dosiahol základný modelovací prístup.

Pri hodnotení výsledkov simulácií je dôležité zohľadniť širší kontext a možné obmedzenia, ktoré mohli ovplyvniť presnosť modelovania. Jedným z týchto faktorov bola vyššia

miera variability v simulovaných rokoch. Tieto roky boli menej stabilné v porovnaní so základným rokom 2019, ktorý slúžil ako východisko pre vytvorenie migračnej funkcie. Táto nestabilita mohla prispieť k zvyšujúcej sa predikčnej chybe, keďže sa postupne prenášali odchýlky z predchádzajúcich rokov.

Z toho dôvodu by bolo vhodné rozšíriť model SLAMM o funkcie, ktoré umožňujú zohľadniť rôzne externé šoky – napríklad pandémiu, ktorá migráciu obmedzila, alebo vojnový konflikt, ktorý naopak spôsobil jej nárast. Tieto rozšírenia by mohli zvýšiť robustnosť modelu voči neočakávaným udalostiam a lepšie zachytiť meniace sa migračné trendy.

Zároveň sa javí ako vhodné zaviesť korekčné mechanizmy pri dlhodobých predikciách, ktoré by dokázali upravovať model v prípade rastúcich odchýlok v čase. Ďalším možným zlepšením je zahrnutie emigrácie, čo by mohlo zvýšiť presnosť výsledkov.

Záver

Modelovanie ponuky práce predstavuje dôležitý nástroj na analýzu trhu práce, pretože umožňuje simuláciami skúmať budúci vývoj na základe rôznych scenárov, identifikovať potenciálne problémy a hodnotiť dopady rôznych politík. Hlavným cieľom tejto diplomovej práce je kalibrácia funkcie migrácie v modeli SLAMM.

V prvej kapitole sme sa zamerali na mikrosimulačné prístupy k modelovaniu ponuky práce, najmä na dynamické mikrosimulačné modely. Opísali sme ich vývoj, hlavné využitia, ako aj výhody, medzi ktoré patrí flexibilita, schopnosť simulovať rôzne scenáre a prepojenie s makroekonomickými modelmi. Zároveň sme poukázali na obmedzenia, ako sú závislosť od kvalitných a dostupných dát, vysoké náklady na vývoj a citlivosť na predpoklady.

V druhej kapitole sme sa sústredili na mikrosimulačný model SLAMM, jeho štruktúru a fungovanie, vrátane vstupov a výstupov modelu. Popísali sme štyri hlavné moduly a predpoklady, ktoré model využíva pri simulácii trhu práce. Vysvetlili sme, ako fungujú funkcie *time_to_leave* a *leave*, ktoré modelujú dĺžku pobytu migrantov na Slovensku a ich rozhodovanie o odchode.

V tretej kapitole sme sa venovali výpočtu vstupných hodnôt pre migračné funkcie v modeli SLAMM. Pri funkcii *time_to_leave* sme okrem základného prístupu, založeného na empirických podieloch pracujúcich migrantov v jednotlivých intervaloch dĺžky pobytu, aplikovali aj alternatívne metódy – parametrické a neparametrické modelovanie. Oba prístupy sme podrobne popísali a následne aplikovali na dostupné dáta. V rámci parametrického modelovania sme pre každú skupinu migrantov (EÚ, Ukrajina a krajiny mimo EÚ) vybrali najvhodnejšie rozdelenie na základe informačných kritérií *AIC*, *BIC* a hodnoty log-likelihood funkcie. Ako najvhodnejšie sa vo všetkých prípadoch ukázalo rozdelenie loggamma, implementované v R knižnici `actuar`. Pre neparametrický prístup sme zvolili jadrový odhad hustoty, kde šírky pásma sme odhadovali pomocou Silvermanovho pravidla. Výsledné rozdelenia z oboch prístupov sme následne diskretizovali, aby boli plne kompatibilné so štruktúrou modelu.

V štvrtej kapitole sme sa sústredili na predikciu a validáciu výstupov modelu prostredníctvom porovnania s historickými údajmi za roky 2020 až 2024. V tejto časti sme aplikovali pravdepodobnostné rozdelenia pre premennú dĺžky zotrvania na Slovensku, ktoré

boli vypočítané v predchádzajúcej kapitole. Na validáciu sme využili tri metriky: *MAE*, *MAPE* a Kullback-Leiblerovu divergenciu. Na základe týchto metrík sme dospeli k záveru, že základný prístup vykazuje najvyššiu presnosť, pričom parametrický prístup v niektorých prípadoch dosiahol porovnateľné výsledky. Naopak, neparametrický prístup vykázal najväčšie odchýlky, najmä pri migrantoch z Ukrajiny.

Za hlavný prínos danej práce môžeme považovať kalibráciu migračných funkcií v modeli SLAMM. Porovnaním rôznych metód odhadu dĺžky pobytu migrantov a validáciou na historických dátach sa zvyšuje presnosť modelu, ktorý môže byť využitý pri tvorbe politík v oblasti migrácie a trhu práce.

Prirodzeným pokračovaním tejto práce by bolo rozšírenie modelu o emigráciu a implementácia funkcií zohľadňujúcich vplyv externých šokov. Takéto vylepšenia by mohli prispieť k presnejším predikciám migračných tokov a lepšiemu pochopeniu ich dopadov na trh práce.

Zoznam použitej literatúry

[1] Bourguignon, F., Spadaro, A.: *Microsimulation as a tool for evaluating redistribution policies*, The Journal of Economic Inequality 4 (2006), 77-106, dostupné na internete (25.4.2025):

https://link.springer.com/article/10.1007/s10888-005-9012-6

[2] Bronka, P., et al.: *SimPaths: An open-source microsimulation model for life course analysis*, CeMPA Working Paper Series, No. CeMPA WP 6/23, University of Essex, Centre for Microsimulation and Policy Analysis (CeMPA), Colcheste, 2023, dostupné na internete (25.4.2025):

https://www.econstor.eu/handle/10419/306901

[3] Chávez-Campos, G. M., et al.: *A sample size statistical analysis and its impact on decarburization measurements metrics*, JOM 73 (2021), 2031–2038, dostupné na internete (25.4.2025):

https://link.springer.com/article/10.1007/s11837-021-04697-9

[4] Chen, Y.C.: *A tutorial on kernel density estimation and recent advances*, Biostatistics & Epidemiology 1, no. 1 (2017), 161–187, dostupné na internete (25.4.2025):

https://www.tandfonline.com/doi/abs/10.1080/24709360.2017.1396742

[5] D'Agostino, R. B., Michael A. S.: *Goodness-of-fit-techniques*, Marcel Dekker Inc., New York, 1986

[6] De Menezes, A.G., Moniz, A.: *Determinants of length of stay: A parametric survival analysis*, Tourism Analysis 16 no. 5 (2011), 509–524, dostupné na internete (25.4.2025):

https://repositorio.uac.pt/bitstream/10400.3/2589/1/TA16-5_519.pdf

[7] Emiliano, P.C., Vivanco, M.J., De Menezes, F.S.: *Information criteria: How do they behave in different models?*, Computational Statistics & Data Analysis 69 (2014), 141–153, dostupné na internete (25.4.2025):

https://www.sciencedirect.com/science/article/pii/S0167947313002776?casa_token=d0fm4w34OhUAAAAA:hmO-55kv18Z7F-sEZDrzL83X0OKo02hGC6AUr0XEMQ-g-SXL2YdktM7CbMOzUAhUZy_gEjEgXB76

[8] Emmert-Streib, F., Dehmer, M.: *Introduction to Survival Analysis in Practice*, Machine Learning and Knowledge Extraction 1 (2019), 1013–1038, dostupné na internete (25.4.2025):

https://doi.org/10.3390/make1030058

[9] Eurostat: Medzinárodná štandardná klasifikácia vzdelania (ISCED), dostupné na internete (25.4.2025):

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED)

[10] Eurostat: Predpokladaná populácia EÚ do roku 2100, dostupné na internete (25.4.2025):

https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190710-1

[11] Evans, M., Hastings, N., Peacock, B.: *Statistical Distributions 4-th Edition*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2011

[12] Harpole, J.K., et al.: *How bandwidth selection algorithms impact exploratory data analysis using kernel density estimation*, Psychological methods 19 no. 3 (2014), 428–443, dostupné na internete (25.4.2025):

https://psycnet.apa.org/record/2014-21924-001

[13] Izenman, A.J.: *Review papers: Recent developments in nonparametric density estimation*, Journal of the american statistical association 86, no. 413 (2020), 205–224, dostupné na internete (25.4.2025): https://www.tandfonline.com/doi/abs/10.1080/01621459.1991.10475021

[14] Harding, A.: *Challenges and Opportunities of Dynamic Microsimulation Modelling*, International Journal of Microsimulation 16 (2023), 1-13, dostupné na internete (25.4.2025):

https://www.microsimulation.pub/articles/00280?

[15] Hughes, N., et al.: *Improving the performance of micro-simulation models with machine learning: The case of Australian farms*, Economic Modelling 115 (2022), dostupné na internete (25.4.2025):

https://www.sciencedirect.com/science/article/pii/S0264999322002036

[16] Janková, K., Pázman, A.: *Pravdepodobnosť a štatistika*, Univerzita Komenského v Bratislave, Bratislava, 2011

[17] Katina, S., Králík M., Hupková A.: *Aplikovaná štatistická inferencia I. Biologická antropológia očami matematickej štatistiky*, Masarykova univerzita, Brno, 2015

[18] Kim, S., Kim, H.: *A new metric of absolute percentage error for intermittent demand forecasts*, International Journal of Forecasting 32 (2016), 669–679, dostupné na internete (25.4.2025):

https://www.sciencedirect.com/science/article/pii/S0169207016000121

[19] Lowitz, B. B., Casciato D. A.: *Principles, definitions, and statistics*, Manual of clinical oncology (2009), 3–18

[20] Lutz, W.: *FAMSIM-Austria: Feasibility Study for a Dynamic Microsimulation Model for Projections and the Evaluation of Family Policies Based on the European Family and Fertility Survey*, Austrian Institute for Family Studies, Vienna, 1997, dostupné na internete (25.4.2025):

https://core.ac.uk/download/pdf/33896591.pdf

[21] Marazzi, A., et al.: *Fitting the distributions of length of stay by parametric models*, Medical care 36 (1998), 915–927, dostupné na internete (25.4.2025):

https://journals.lww.com/lww-medicalcare/abstract/1998/06000/fitting_the__distributions_of_length_of_stay_by.14.aspx

[22] Martiška, P.: *Modelling the Sustainability and Adequacy of the Pension System in Slovakia*, Project of dissertation thesis, University of Economics in Bratislava, Bratislava, 2024

[23] Národná banka Slovenska: *Štrukturálne výzvy*, NBS, 2024, dostupné na internete: (25.4.2025):

https://nbs.sk/publikacie/strukturalne-vyzvy/

[24] O'Donoghue, C.: *Handbook of microsimulation modelling*, Emerald Group Publishing, Bingley, 2014

[25] Ponti, M., et al: *A decision cognizant Kullback–Leibler divergence*,Pattern Recognition 61 (2017), 470-478, dostupné na internete (25.4.2025): https://www.sciencedirect.com/science/article/abs/pii/S003132031630231X?via%3Dihub

[26] Robeson, S.M., Willmott, C.J.: *Decomposition of the mean absolute error (MAE) into systematic and unsystematic components*, PloS one 18 (2023), dostupné na internete (25.4.2025):

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0279774

[27] R Project Search: Stránka poskytuje vyhľadávací nástroj pre dokumentáciu, balíky a funkcie v prostredí R, dostupné na internete: (25.4.2025):

https://search.r-project.org/CRAN/refmans/actuar/html/Loggamma.html

[28] Silverman, B. W.: *Density estimation for statistics and data analysis*, Routledge, New York, 1998

[29] Štefánik M., Miklosovic T.: *Modelling foreign labour inflows using a dynamic microsimulation model of an ageing country-Slovakia*, International Journal of Microsimulation 13, no. 2 (2020), 102-113, dostupné na internete (25.4.2025):

https://microsimulation.pub/articles/00220

[30] Ústredie práce, sociálnych vecí a rodiny: Údaje o zamestnávaní cudzincov na území Slovenskej republiky, dostupné na internete: (25.4.2025)

https://www.upsvr.gov.sk/statistiky/zamestnavanie-cudzincov-statistiky/zamestnavanie-cudzincov-na-uzemi-slovenskej-republiky-za-rok-2024.html?page_id=1335245

Príloha A

Rok	Skupina	1	2	3	4	5	6	7	8	9
2020	EÚ	11636	4393	3068	2375	1517	1344	869	573	383
2021	EÚ	8676	3921	2956	1992	1633	913	837	554	373
2022	EÚ	9099	2396	2824	2035	1463	1056	608	565	384
2023	EÚ	11742	2468	1902	1941	1488	949	696	405	390
2024	EÚ	7084	2338	1808	1329	1338	906	595	445	261
2020	Ukrajina	7233	2921	514	145	69	54	30	12	4
2021	Ukrajina	4749	1595	332	279	99	38	20	11	5
2022	Ukrajina	4980	712	182	187	189	55	14	8	4
2023	Ukrajina	6400	738	88	101	128	106	19	5	3
2024	Ukrajina	3663	682	84	50	67	67	38	7	2
2020	non-EÚ	7981	3778	1579	587	270	231	132	51	8
2021	non-EÚ	5921	1855	1240	860	416	165	91	44	6
2022	non-EÚ	6210	1041	619	691	625	256	69	32	6
2023	non-EÚ	7971	1088	373	340	486	386	105	23	4
2024	non-EÚ	4669	1010	371	206	239	292	156	35	3

Tabuľka 25: Výsledky simulácií počtov cudzincov podľa dĺžky zotrvania: základný prístup

Rok	Skupina	1	2	3	4	5	6	7	8	9
2020	EÚ	11751	4478	4037	1906	1176	686	484	367	286
2021	EÚ	8669	3986	3010	2608	1305	722	431	307	234
2022	EÚ	9088	2409	2871	2089	1917	843	478	289	214
2023	EÚ	11717	2463	1909	1982	1531	1246	558	324	202
2024	EÚ	7103	2347	1808	1325	1373	937	785	358	211
2020	Ukrajina	6882	2667	390	689	409	182	72	51	32
2021	Ukrajina	4743	1453	302	217	468	225	63	28	20
2022	Ukrajina	4952	714	171	173	149	262	81	23	11
2023	Ukrajina	6385	730	92	93	117	83	94	31	9
2024	Ukrajina	3663	673	86	50	62	65	28	36	11
2020	non-EÚ	7549	3343	1413	968	636	310	125	78	18
2021	non-EÚ	5907	1658	1097	776	683	383	125	39	8
2022	non-EÚ	6189	1057	559	615	566	420	156	44	5
2023	non-EÚ	7942	1084	381	310	442	346	165	54	6
2024	non-EÚ	4671	1010	364	211	218	262	134	54	6

Tabuľka 26: Výsledky simulácií počtov cudzincov podľa dĺžky zotrvania: parametrický prístup

Rok	Skupina	1	2	3	4	5	6	7	8	9
2020	EÚ	8684	5138	3786	2634	2009	1338	2000	1364	0
2021	EÚ	8593	1944	3470	2465	1815	1228	834	1276	882
2022	EÚ	9051	2407	1401	2404	1803	1185	816	562	888
2023	EÚ	11782	2433	1915	974	1757	1167	783	550	388
2024	EÚ	7065	2358	1791	1338	675	1079	733	505	355
2020	Ukrajina	5221	2691	677	1296	122	58	44	25	0
2021	Ukrajina	4718	809	306	373	881	70	21	16	9
2022	Ukrajina	4936	715	95	168	258	492	25	8	7
2023	Ukrajina	6408	721	92	54	117	142	176	9	3
2024	Ukrajina	3639	670	85	51	36	63	47	68	4
2020	non-EÚ	5948	3192	2172	1616	508	232	200	105	0
2021	non-EÚ	5856	934	1050	1188	1139	307	94	66	12
2022	non-EÚ	6168	1054	314	592	861	700	125	32	8
2023	non-EÚ	7986	1066	380	174	420	524	282	42	3
2024	non-EÚ	4667	1002	354	210	122	254	204	93	5

Tabuľka 27: Výsledky simulácií počtov cudzincov podľa dĺžky zotrvania: neparametrický prístup