GEOGRAFICKÝ ČASOPIS

47 1995 4

*Anton Bezák, Anna Holická**

KOMPARATÍVNA ANALÝZA EXTRAPOLAČNÝCH MODELOV PRE REGIONÁLNE POPULAČNÉ PROJEKCIE

Anton Bezák, Anna Holická: A comparative analysis of some extrapolation models for regional population projections. Geografický časopis, *47*, 4, 3 tabs., 23 refs.

The aim of this paper is to evaluate the performance of a number of extrapolation models commonly used in regional population projections. Various models are calibrated and used to project the 1994 population for each of the 63 urban regions in Slovakia. These projections are then compared with actual counts for target year in order to evaluate the model performance. The results indicate that the best forecasts are provided by either the autoregressive models or ratio models with lagged variables. Some of the elementary growth models and simple ratio models also perform well.

Key words: regional population projections, extrapolation models, ex post evaluation of forecasts, urban regions in Slovakia

ÚVOD

Populačnými projekciami nazývame všetky numerické odhady budúcej veľkosti, vnútornej štruktúry a priestorového rozloženia skúmaných populácií, odvodené na základe určitých predpokladov (Keyfitz 1972, Cohen 1986). Takéto odhady sa môžu týkať obyvateľstva celého štátu alebo rôzne veľkých regiónov vymedzených na jeho území. Metódy, ktoré sa uplatňujú v regionálnych populačných projekciách, môžeme rozdeliť do troch základných skupín (cf. Land 1986). Prvú z nich reprezentujú komponentné metódy, ktoré vychádzajú z princípu prežívania kohort a uvažujú efekty

* Geografický ústav SAV, Štefánikova 49, 814 73 Bratislava

všetkých komponentov populačného rastu vrátane migrácie. Ďalšiu skupinu tvoria rozmanité štruktúrne (resp. demo-ekonomické) metódy opierajúce sa o bohatú sieť vzájomných závislostí medzi populačnými, ekonomickými a sociálnymi charakteristikami. Do tretej skupiny môžeme napokon zaradiť komplex metód, ktoré sú založené na extrapolácii minulých a súčasných populačných trendov do budúcnosti.

Extrapolačný prístup k regionálnym populačným projekciám je pomerne často predmetom ostrej kritiky. Extrapolačné metódy sa považujú za zjednodušujúce, naivné a menej presné ako ostatné projekčné metódy. Základnou nevýhodou týchto metód je, že sa obmedzujú iba na odhad celkového počtu obyvateľov skúmaného regiónu a neposkytujú žiadnu informáciu o vekovej, pohlavnej alebo inej štruktúre danej populácie v budúcnosti. K závažným nedostatkom extrapolačných metód patrí aj takmer úplná absencia solídnej teoretickej bázy (Smith 1987). Nie je preto náhodné, že sa v mnohých regionálnych populačných projekciách dáva prednosť zložitejším metódam (Rogers 1985).

Napriek rôznym výhradám, ktoré je možné vysloviť na adresu extrapolačných metód, nemali by sa metódy tejto kategórie paušálne odmietať. Celý rad empirických štúdií potvrdzuje, že krátkodobé a strednodobé predpovede celkového počtu obyvateľov, ktoré produkujú niektoré extrapolačné metódy, sú aspoň také presné ako predpovede generované omnoho komplikovanejšími komponentnými metódami (cf. Greenberg 1972, Siegel 1972, Saboia 1974, Stoto 1983, Murdock et al. 1984 a iní). Jednoduchý spôsob aplikácie, nízke náklady a minimálne požiadavky na dáta priamo predurčujú extrapolačné metódy na odhad počtu obyvateľov menších regionálnych celkov, kde dáta požadované zložitejšími metódami jednoducho neexistujú, prípadne sú nedostupné, zastarané alebo nespoľahlivé. Pozitívnou stránkou extrapolačných metód je aj možnosť retrospektívnej aplikácie na rozsiahly súbor regiónov za identických podmienok. Tak sa môže získať veľký počet konzistentných predpovedí, ktoré sú navzájom porovnateľné v čase i v priestore (Smith a Sincich 1988).

Cieľom tejto štúdie je poskytnúť stručný prehľad najznámejších extrapolačných modelov, ktoré sa používajú v regionálnych populačných projekciách a zároveň sa pokúsiť o empirické zhodnotenie ich kvality. Prvá časť štúdie začína zhrnutím základných pojmov a termínov týkajúcich sa konštrukcie predpovedí v časových radoch a pokračuje stručným opisom 24 extrapolačných modelov. V druhej časti príspevku použijeme prediskutované modely na zostrojenie krátkodobých predpovedí počtu obyvateľov 63 funkčných mestských regiónov vymedzených na území Slovenska. Účinnosť jednotlivých modelov posúdime tak, že zostrojíme predpovede známych hodnôt počtu obyvateľov každého regiónu, ktoré potom porovnáme so skutočnými hodnotami pomocou niekoľkých kritérií. V závere štúdie načrtneme niektoré perspektívne smery výskumu extrapolačných modelov v regionálnej populačnej analýze.[1]

EXTRAPOLAČNÉ MODELY

Spomenuli sme už, že princíp extrapolácie sa v regionálnych populačných projekciách používa spravidla pri odhadoch celkového počtu obyvateľov skúmaného regiónu. Všetky extrapolačné metódy vychádzajú pritom z predpokladu, že za určité

[1] Štúdia je súčasťou široko koncipovaného výskumného projektu, ktorý si kladie za cieľ porovnať a zhodnotiť kvalitu rôznych modelov a metód používaných v populačných projekciách na regionálnej a lokálnej úrovni. Niektoré parciálne a predbežné výsledky získané v rámci projektu boli už uverejnené v osobitnom príspevku (Bezák a Holická 1995). Táto štúdia je prepracovanou verziou referátu, ktorý odznel 9.11.1994 na 11. zjazde Slovenskej geografickej spoločnosti v Častej.

historické obdobie máme k dispozícii chronologicky usporiadaný rad hodnôt P_1, P_2, ..., P_t, ktoré udávajú počet obyvateľov daného regiónu v ekvidištančných časových okamihoch, napr. na konci každého kalendárneho roka. Cieľom projekcie je získať odhad hodnoty uvažovaného časového radu v istom, presne určenom okamihu v budúcnosti. Tento odhad (t.j. konkrétnu numerickú hodnotu) budeme považovať za výsledok populačnej projekcie a nazývať *predpoveď*.

Každá predpoveď sa konštruuje v určitom časovom okamihu, ktorý sa nazýva *začiatok predpovede*. Ak za začiatok predpovede zvolíme okamih t a pokúsime sa odhadnúť počet obyvateľov daného regiónu v okamihu, ktorý je od začiatku predpovede vzdialený h časových jednotiek, potom nás zrejme zaujíma odhad hodnoty P_{t+h}. Tento odhad budeme nazývať predpoveď v čase t o h krokov dopredu a označovať symbolom $\hat{P}_{t+h}$. *Okamih* $t + h$ nazveme *cieľovým okamihom* a číslo h *časovým horizontom* predpovede (cf. Cipra 1986, pp. 22-24).

Na zostrojenie predpovede $\hat{P}_{t+h}$ môžeme použiť celý rad extrapolačných modelov. Každý z nich v zjednodušenej a formalizovanej podobe opisuje mechanizmus, na základe ktorého sa generujú hodnoty uvažovaného časového radu. Neznáme hodnoty parametrov zvoleného modelu odhadneme na základe historických dát pomocou zodpovedajúcej metódy. Pri konštrukcii predpovede potom implicitne predpokladáme, že v období, na ktoré sa predpoveď vzťahuje, sa analytický tvar a odhadnuté parametre modelu podstatne nezmenia v porovnaní s obdobím, z ktorého pochádzajú dáta použité pri kalibrácii modelu.

Rozdiel medzi predpoveďou $\hat{P}_{t+h}$ a skutočným počtom obyvateľov P_{t+h} daného regiónu v cieľovom okamihu budeme nazývať *chyba predpovede* a označovať symbolom d. Absolútnu hodnotu čísla d nazveme *absolútnou chybou* a pomer absolútnej chyby a skutočného počtu obyvateľov daného regiónu v cieľovom okamihu *relatívnou chybou* predpovede. Je zrejmé, že veľkosť chyby predpovede môžeme zistiť až vtedy, keď je známa skutočná hodnota P_{t+h}.

Z veľkého množstva extrapolačných modelov, ktoré sa uplatňujú v regionálnych populačných projekciách, sme vybrali 24 modelov zaradených do siedmich tried. V nasledujúcom prehľade podáme stručnú charakteristiku a spôsob odvodenia každého modelu, pripojíme odkaz na metódu estimácie jeho parametrov a tam, kde je to možné, uvedieme priamo predpovednú formu modelu.

Prvú triedu extrapolačných modelov tvoria *jednoduché rastové modely* odvodené na základe triviálnych predpokladov o raste regionálnych populácií. Najjednoduchší z nich je tzv. *naivný model* v tvare

$$\hat{P}_{t+h} = P_t \,, \tag{1}$$

ktorý vychádza z predpokladu, že v období, ktoré uplynie od začiatku predpovede až po cieľový okamih, nenastane žiadna zmena v počte obyvateľov skúmaného regiónu (Isserman 1977, Openshaw a van der Knaap 1983). Naivný model zohráva v regionálnych populačných projekciách úlohu určitej normy, s ktorou sa porovnáva kvalita ostatných extrapolačných modelov.

Nasledujúce dva rastové modely postihujú zmeny vo veľkosti regionálnych populácií v čase. V *modeli aritmetického rastu* s rovnicou

$$\hat{P}_{t+h} = P_t + a\,h \tag{2}$$

sa predpokladá, že absolútny prírastok danej populácie za jednotku času (parameter a) zostane v období predpovede konštantný. *Model geometrického rastu* v tvare

$$\hat{P}_{t+h} = P_t + (1 + r)^h \tag{3}$$

vychádza zasa z analogického predpokladu o konštantnom relatívnom prírastku za jednotku času (parameter r). Obidva parametre sa odhadnú z historických dát jednoduchým spôsobom. Ak za jednotku času zvolíme 1 rok, potom v prvom prípade za odhad parametra môžeme vziať aritmetický priemer ročných absolútnych prírastkov a v druhom prípade aritmetický alebo geometrický priemer ročných relatívnych prírastkov (Krueckeberg a Silvers 1974, pp. 260-264; Isserman 1977; Smith 1987 a ďalší).

Do druhej triedy sme zaradili *trendové modely*, ktoré analyticky opisujú trendovú zložku uvažovaného časového radu pomocou niektorých jednoduchých funkcií časovej premennej. Pri ich konštrukcii sa zvyčajne predpokladá, že uvažovaný časový rad je zložený z dvoch aditívnych zložiek: trendu a náhodnej zložky. Tento predpoklad nám umožňuje nielen odhadnúť parametre trendových funkcií metódami lineárnej regresnej analýzy, ale aj stotožniť predpovede budúcich hodnôt časového radu priamo s predpoveďami budúceho vývoja trendovej zložky (Cipra 1986, p. 29). V regionálnych populačných projekciách sa najčastejšie aplikujú *modely s konštantným*, *lineárnym*, *kvadratickým*, *exponenciálnym* a *mocninovým trendom* (Isard 1960, pp. 9-15; Isserman 1977). Predpovedné rovnice uvedených modelov môžeme postupne zapísať v tvare

$$\hat{P}_{t+h} = a, \tag{4}$$

$$\hat{P}_{t+h} = a + b\,(t + h), \tag{5}$$

$$\hat{P}_{t+h} = a + b\,(t + h) + c\,(t + h)^2, \tag{6}$$

$$\hat{P}_{t+h} = a\,b^{t+h} \tag{7}$$

a

$$\hat{P}_{t+h} = a\,(t + h)^b. \tag{8}$$

Jednoduchým odhadom parametra a v modeli s konštantným trendom je aritmetický priemer pozorovaných hodnôt P_i ($i = 1, 2, ..., t$). Parametre ostatných trendových modelov sa odhadnú pomocou metódy najmenších štvorcov (Cipra 1986, pp. 29-35). Upozorňujeme, že v rovniciach (7) a (8) sa najprv exponenciálny, resp. mocninový trend prevedie logaritmickou transformáciou na lineárny trend a potom sa odhadnuté parametre odlogaritmujú a dosadia do pôvodných rovníc.

Konštrukcia predpovede na základe trendového modelu sa z geometrického hľadiska dá charakterizovať ako jednoduché predĺženie regresnej krivky do budúcnosti. Ak regresnú krivku okrem predĺženia aj rovnobežne posunieme tak, aby prechádzala

bodom, ktorý zodpovedá poslednému pozorovaniu (t.j. pozorovaniu v okamihu t), dostaneme osobitnú triedu extrapolačných modelov, známu pod názvom *bodové regresné modely* (Isserman 1977). Modely tejto kategórie umožňujú prekonať jeden vážny problém, s ktorým sa často stretneme pri použití trendových modelov. V štandardných trendových modeloch môže totiž regresná krivka značne nadhodnotiť alebo podhodnotiť hodnotu posledného pozorovania. Keďže v časových radoch opisujúcich vývoj počtu obyvateľov nedochádza spravidla k náhlej zmene hodnôt susedných pozorovaní, môže sa ľahko stať, že už predpovede o jeden krok dopredu budú veľmi nepresné. Z bodových regresných modelov budeme v našej štúdii uvažovať *lineárny*, *exponenciálny* a *mocninový model* s predpovednými rovnicami

$$\hat{P}_{t+h} = P_t + b\,h, \tag{9}$$

$$\hat{P}_{t+h} = P_t\, b^h \tag{10}$$

a

$$\hat{P}_{t+h} = P_t\,[(t+h)/t]^b. \tag{11}$$

Poznamenávame, že vo všetkých troch modeloch sa hodnota parametra b zhoduje s odhadnutou hodnotou rovnomenného parametra z korešpondujúcej rovnice pre štandardný trendový model.

Štvrtú triedu extrapolačných modelov reprezentujú *autoregresné modely*. Ako naznačuje sám názov, sú to regresné modely, v ktorých hodnota predikovanej premennej vyjadrujúcej počet obyvateľov daného regiónu lineárne závisí od hodnôt tej istej premennej, ale oneskorených v čase. Najväčšie oneskorenie pritom určuje rád autoregresného modelu (Bennett 1979, pp. 40-41). V našom kontexte sa *autoregresný model k-teho rádu* dá zapísať v tvare

$$P_t = a_1 P_{t-1} + a_2 P_{t-2} + \ldots + a_k P_{t-k}. \tag{12}$$

Na zostrojenie predpovede $\hat{P}_{t+h}$ sme v našej štúdii použili autoregresné modely prvého až štvrtého rádu. Parametre zvolených modelov sa odhadli metódou najmenších štvorcov (Bennett 1979, pp. 226-238). Metóda zostrojenia predpovedí na základe autoregresných modelov je podrobne opísaná v literatúre zaoberajúcej sa analýzou časových radov (Bennett 1979, pp. 549-552; Cipra 1986, pp. 154-158).

V nasledujúcich troch triedach extrapolačných modelov sa predpokladá, že existuje určitý kvantitatívny vzťah medzi dynamikou rastu obyvateľstva skúmaného regiónu a obyvateľstva nejakého, vhodne zvoleného referenčného územia. Ak budeme predpokladať, že spomenutá závislosť pretrvá aj v budúcnosti, potom na základe predpovede počtu obyvateľov referenčného regiónu v cieľovom okamihu vieme zostrojiť aj predpoveď počtu obyvateľov skúmaného regiónu v tom istom časovom okamihu. Extrapolačné modely, ktoré vychádzajú z uvedeného princípu, budeme v zhode s N. Oppenheimom (1980, pp. 42-44) nazývať *komparatívnymi modelmi*.

Najjednoduchší komparatívny model vyjadruje predpoklad, že podiel počtu obyvateľov skúmaného a referenčného regiónu, pozorovaný na začiatku predpovede,

zostane konštantný aj v budúcnosti (Isard 1960, p. 16; Krueckeberg a Silvers 1974, p. 268; Isserman 1977). Ak symbolom C_t, resp. C_{t+h} označíme počet obyvateľov referenčného regiónu v okamihu t, resp. $t + h$, potom *model konštantného podielu (počtu obyvateľov)* môžeme zapísať v tvare

$$\hat{P}_{t+h} = (P_t / C_t)\, C_{t+h}. \tag{13}$$

Do triedy *jednoduchých komparatívnych modelov* zaradíme ešte nasledujúce dva modely. Prvý z nich sa opiera o predpoklad, že podiel absolútnych prírastkov obyvateľstva v skúmanom a referenčnom regióne za uvažované historické obdobie sa v období predpovede nezmení (Eichperger 1984, Smith a Sincich 1988). Potom platí

$$\hat{P}_{t+h} = P_t + [(P_t - P_1) / (C_t - C_1)]\, (C_{t+h} - C_t). \tag{14}$$

Druhý model získame tak, že rovnicu aritmetického rastu (2) aplikujeme na časový rad hodnôt P_i / C_i ($i = 1, 2, ..., t$). Po jednoduchej úprave dostaneme model

$$\hat{P}_{t+h} = C_{t+h}\, (P_t / C_t + a\, h)\,, \tag{15}$$

v ktorom sa predpokladá, že priemerný ročný absolútny prírastok podielu P_i / C_i pozorovaný v historickom období (parameter a) zostane v období predpovede konštantný (Smith 1987, Smith a Sincich 1988).

Aj pri odvodení ďalšej triedy komparatívnych modelov sa pripúšťa možnosť zmien podielu P_i / C_i v čase. Úplne analogicky ako v prípade extrapolačných modelov zaradených do druhej triedy môžeme pomocou niektorých jednoduchých funkcií časovej premennej vystihnúť trendovú zložku týchto zmien a po vynásobení oboch strán odhadnutej regresnej rovnice hodnotou C_{t+h} získať formulu pre predpoveď $\hat{P}_{t+h}$ (Isard 1960, p. 16; Isserman 1977). Takto sa dá odvodiť celý rad *trendových komparatívnych modelov*. Na ilustráciu sme z nich vybrali *komparatívne modely s konštantným*, *lineárnym* a *exponenciálnym trendom* v tvare

$$\hat{P}_{t+h} = c\, C_{t+h}, \tag{16}$$

$$\hat{P}_{t+h} = C_{t+h}\, [a + b\, (t + h)] \tag{17}$$

a

$$\hat{P}_{t+h} = C_{t+h}\, (a\, b^{t+h}), \tag{18}$$

kde a, b sú parametre určené metódou najmenších štvorcov a c sa odhadne ako aritmetický priemer hodnôt P_i / C_i za uvažované historické obdobie.

Poslednú triedu extrapolačných modelov tvoria tri *komparatívne modely s časovo oneskorenými premennými* v tvare

$$P_t / P_{t-1} = a + b\,(C_t / C_{t-1})\,, \tag{19}$$

$$P_t = a + b\,[(C_t / C_{t-1})\,P_{t-1}] \tag{20}$$

a

$$P_t - P_{t-1} = a + (C_t - C_{t-1})\,, \tag{21}$$

ktoré sa pôvodne navrhli na predpoveď úrovne lokálnej zamestnanosti (Hunt et al. 1980), ale neskoršie sa úspešne využili aj v regionálnych populačných projekciách (Openshaw a van der Knaap 1983). Ako vidieť, sú to lineárne regresné modely, v ktorých sa rôznym spôsobom vyjadruje vzťah medzi počtom obyvateľov skúmaného regiónu na začiatku predpovede a v okamihu bezprostredne predchádzajúcom na jednej strane a korešpondujúcimi hodnotami pre referenčný región na strane druhej. Parametre a, b sa vo všetkých troch prípadoch odhadnú metódou najmenších štvorcov. Predpovede $\hat{P}_{t+h}$ sa zostroja rekurentným spôsobom tak, že predpoveď $\hat{P}_{t+j}$ sa použije na konštrukciu predpovede $\hat{P}_{t+j+1}$ ($j = 1, 2, ..., h-1$).

Na záver tohto odseku chceme ešte pripomenúť, že všetky komparatívne modely vyžadujú nezávislú predpoveď počtu obyvateľov referenčného regiónu v cieľovom okamihu. V súvislosti s tým je potrebné upozorniť, že v každom komparatívnom modeli sa vyskytujú dva zdroje nepresnosti. Prvým z nich je chyba predpovede veličiny, ktorá vyjadruje vzťah medzi skúmaným a referenčným regiónom. Druhý, pravdepodobne menej významný zdroj chyby súvisí s predpoveďou počtu obyvateľov referenčného regiónu (cf. Smith 1987, Smith a Sincich 1988).

EMPIRICKÉ HODNOTENIE

Kvalitu populačných projekcií môžeme posúdiť dvoma principiálne odlišnými spôsobmi: teoretickou analýzou projekčného modelu alebo empirickou analýzou predpovedí generovaných daným modelom (Stoto 1983). Pri empirickom hodnotení sa zvyčajne postupuje tak, že sa najprv zostroja predpovede známych hodnôt časového radu a následne sa tieto predpovede porovnajú so skutočnými hodnotami. Túto evaluačnú stratégiu, nazývanú *hodnotenie ex post* (Keyfitz 1972), sme využili aj v našom príspevku.

Všetky extrapolačné modely, prediskutované v predchádzajúcom odseku, sa postupne použili na zostrojenie krátkodobých predpovedí počtu obyvateľov 63 funkčných mestských regiónov, vymedzených na území Slovenska v štúdii A. Bezáka (1990). Parametre testovaných modelov sa odhadli oddelene pre každý mestský región na základe časového radu, ktorý pozostával z 21 hodnôt udávajúcich počet obyvateľov príslušného regiónu na konci každého kalendárneho roka v období 1970-1990. Pri odhadoch parametrov jednoduchých rastových a jednoduchých komparatívnych modelov sa vzhľadom na pomerne veľké časové rozpätie medzi prvým a posledným pozorovaním použili iba dáta z obdobia 1980-1990.

Predpovedné formy testovaných modelov sme upravili tak, že za začiatok predpovede sme zvolili koniec roka 1990. Prostredníctvom jednotlivých modelov sme potom zostrojili predpovede počtu obyvateľov všetkých mestských regiónov na konci roka 1994. Pre každý z 24 modelov sme teda získali 63 predpovedí, ktoré sa napokon

porovnali so skutočnými hodnotami počtu obyvateľov v cieľovom okamihu. Pre úplnosť ešte uvedieme, že vo všetkých komparatívnych modeloch sme za referenčný región zvolili územie Slovenska. V snahe eliminovať druhú zložku chyby predpovede, ktorá v komparatívnych modeloch súvisí s predpoveďou počtu obyvateľov referenčného regiónu, použili sme ako hodnoty C_{t+h} údaje o skutočnom počte obyvateľov Slovenska v roku 1994.

Všetky historické dáta použité v tejto štúdii pochádzajú z priebežnej evidencie pohybu obyvateľstva podľa obcí za každý rok uvažovaného časového obdobia. Časové rady pre jednotlivé mestské regióny vznikli agregáciou hodnôt udávajúcich koncoročné stavy obyvateľstva vo všetkých obciach patriacich do príslušného regiónu. Všetky hodnoty sa pred agregáciou upravili pomocou osobitnej metódy (cf. Bezák 1994), aby sa odstránil pomerne značný nesúlad medzi výsledkami priebežnej evidencie, ktorý sa spravidla pozoruje na rozhraní dvoch intercenzových období. Z rovnakého prameňa pochádzajú aj údaje o skutočnom počte obyvateľov na konci roka 1994 za jednotlivé regióny a celé územie Slovenska.

Presnosť predpovedí generovaných jednotlivými modelmi môžeme zhodnotiť pomocou mnohých kritérií. Ako však upozorňujú viacerí autori (napr. Keyfitz 1972, Openshaw a van der Knaap 1983, Murdock et al. 1984), výber vhodných kritérií nie je jednoduchou záležitosťou, pretože presnosť z hľadiska jedného kritéria nemusí byť v súlade s presnosťou z hľadiska ostatných kritérií. Azda najvhodnejším riešením je použiť niekoľko rôznych kritérií a dúfať, že sa ich výsledky nebudú diametrálne odlišovať.

Základnú informáciu o účinnosti jednotlivých modelov poskytujú jednoduché štatistiky zhody (Openshaw a van der Knaap 1983, Smith 1987), v ktorých sa predpovede posudzujú súhrnne v rámci celého súboru mestských regiónov. Sú to:

priemerná absolútna chyba

$$S_1 = \frac{1}{n} \sum_{i=1}^{n} |d_i|, \tag{22}$$

priemerná relatívna chyba

$$S_2 = \frac{1}{n} \sum_{i=1}^{n} \frac{|d_i|}{P_i} \cdot 100\% \tag{23}$$

a *priemerná algebraická chyba*

$$S_3 = \frac{1}{n} \sum_{i=1}^{n} \frac{d_i}{P_i} \cdot 100\ \%\ , \tag{24}$$

kde d_i je chyba predpovede počtu obyvateľov i-teho regiónu, P_i je skutočný počet obyvateľov i-teho regiónu v cieľovom okamihu a n označuje počet mestských regiónov.

Štatistika S_1 vyjadruje priemernú chybu predpovede v jednotkách merania predpovedanej veličiny (t.j. v počte osôb). V bezrozmernej štatistike S_2 sa naproti tomu každá chyba predpovede uvažuje vo vzťahu k skutočnému počtu obyvateľov príslušného regiónu v cieľovom okamihu. Štatistika S_3 je mierou vychýlenia predpovedí. Kladné hodnoty S_3 naznačujú, že daný model poskytuje (v priemere) nadhodnotené predpovede počtu obyvateľov, záporné hodnoty tejto štatistiky signalizujú naopak tendenciu k podhodnoteným predpovediam.

Hodnoty všetkých troch štatistík sú pre každý testovaný model uvedené v tab. 1. Ako prezrádzajú štatistiky S_1 a S_2, najpresnejšie predpovede poskytujú autoregresné modely druhého až štvrtého rádu, ďalej všetky tri komparatívne modely s časovo oneskorenými premennými a z ostatných modelov ešte jednoduchý komparatívny model s konštantným podielom absolútneho prírastku a mocninový model bodovej regresie. Takmer rovnako dobré predpovede s malými relatívnymi chybami sa získali pomocou modelov aritmetického i geometrického rastu. Medzi úspešné modely môžeme napokon zaradiť aj ďalšie dva jednoduché komparatívne modely, ktoré produkujú predpovede s pomerne malými absolútnymi chybami.

Podstatne nižším stupňom účinnosti sa vyznačujú ostatné testované modely. Týka sa to v prvom rade oboch kategórií trendových modelov, z ktorých najmä modely s konštantným trendom priniesli veľmi nepresné predpovede s veľkými absolútnymi i relatívnymi chybami. V tejto súvislosti stojí za zmienku, že naivný model generuje omnoho presnejšie predpovede ako takmer všetky štandardné i komparatívne trendové modely. Pomerne uspokojivé predpovede poskytuje z kategórie trendových modelov iba model s kvadratickým trendom. Pozoruhodné je aj zistenie, že všetky modely bodovej regresie sú podstatne účinnejšie ako korešpondujúce trendové modely. Prekvapením je dosť nízka účinnosť autoregresného modelu prvého rádu.

Hodnoty štatistiky S_3 naznačujú, že takmer dve tretiny testovaných modelov produkujú predpovede, ktoré podhodnocujú skutočný počet obyvateľov mestských regiónov v cieľovom okamihu. Všimnime si, že mierna tendencia k podhodnoteným predpovediam je typická práve pre všetky úspešné modely. Na druhej strane treba zdôrazniť, že neexistuje žiadna konzistentná závislosť medzi priemernou presnosťou predpovede a stupňom, resp. smerom jej vychýlenia.

Priemerné hodnoty pre ten-ktorý model, uvedené v tab. 1, môžu zastierať nemalé diferencie medzi chybami individuálnych predpovedí v rámci celého súboru mestských regiónov. Z tohto dôvodu sme sústavu hodnotiacich kritérií rozšírili o ďalšie tri indikátory účinnosti testovaných modelov. Kritérium K_1 udáva pre každý model percentuálny podiel *presných* predpovedí s relatívnou chybou menšou ako 1 %, kritérium K_2 zasa percentuálny podiel *extrémnych* predpovedí s relatívnou chybou väčšou ako 5 %. Pomocou tretieho kritéria by sme mohli zistiť koľkokrát daný model poskytuje najlepšiu predpoveď (t.j. predpoveď s najmenšou relatívnou chybou) pre niektorý z mestských regiónov (cf. Bennett 1979, p. 557). Podrobná analýza individuálnych predpovedí však ukázala, že spravidla existuje niekoľko rôznych modelov, ktoré pre daný región produkujú iba nepatrne horšie predpovede ako najlepší model. V súvislosti s tým definujeme kritérium K_3 tak, že pre každý model udáva percentuálny podiel *dobrých* predpovedí, t.j. predpovedí, ktoré sa od najlepšej predpovede neodlišujú viac ako o 1 %.

Ako vyplýva z tab. 2, hodnoty troch dodatočných kritérií v podstate iba potvrdzujú závery, ku ktorým sme dospeli na základe troch štatistík zhody. Na doplnenie sa žiada uviesť azda iba poznatok, ktorý sa týka naivného modelu a dvoch modelov bodovej regresie (lineárneho a exponenciálneho). Hoci tieto modely nepatria z hľa-

Tab. 1. Porovnanie účinnosti modelov (hodnoty štatistík zhody)

Model	S_1	S_2	S_3
JEDNODUCHÉ RASTOVÉ MODELY			
naivný model	1407	1,54	-1,26
model aritmetického rastu	865	0,91	0,05
model geometrického rastu	945	0,96	0,15
TRENDOVÉ MODELY			
model s konštantným trendom	6653	6,52	-5,19
model s lineárnym trendom	2316	2,25	1,72
model s kvadratickým trendom	1059	1,42	-0,56
model s exponenciálnym trendom	2914	2,67	2,22
model s mocninovým trendom	1894	2,30	-1,21
BODOVÉ REGRESNÉ MODELY			
lineárny model	1316	1,24	0,71
exponenciálny model	1606	1,44	0,97
mocninový model	673	0,95	-0,55
AUTOREGRESNÉ MODELY			
model 1.rádu	1430	1,30	0,79
model 2.rádu	550	0,79	-0,07
model 3.rádu	595	0,80	-0,26
model 4.rádu	579	0,85	-0,39
JEDNODUCHÉ KOMPARATÍVNE MODELY			
model konštantného podielu počtu obyvateľov	777	1,27	0,29
model konštantného podielu absolútneho prírastku	671	0,79	-0,28
model aritmetického rastu podielu počtu obyvateľov	835	1,04	-0,52
TRENDOVÉ KOMPARATÍVNE MODELY			
model s konštantným trendom	3930	5,76	2,53
model s lineárnym trendom	1649	2,27	-1,13
model s exponenciálnym trendom	1628	2,05	-0,81
KOMPARATÍVNE MODELY S ČASOVO ONESKORENÝMI PREMENNÝMI			
model s rovnicou (19)	767	0,94	-0,37
model s rovnicou (20)	724	0,83	-0,21
model s rovnicou (21)	762	0,94	-0,40

diska priemernej chyby predpovede k najlepším modelom, produkujú len veľmi malý podiel extrémnych predpovedí s relatívnou chybou väčšou ako 5 %.

Posledný aspekt hodnotenia si všíma vzťah medzi účinnosťou testovaných modelov a veľkosťou mestských regiónov. Tabuľka 3 obsahuje hodnoty priemernej relatívnej chyby predpovede v troch veľkostných kategóriách regiónov, ktoré vzniknú rozčlenením pôvodného súboru podľa počtu obyvateľov regiónu na začiatku predpovede (t.j. v roku 1990). Ako vidieť, veľmi dobré predpovede vo všetkých veľkostných kategóriách poskytujú autoregresné modely druhého až štvrtého rádu, model aritmetického rastu, komparatívny model s konštantným podielom absolútneho prírastku a jeden z komparatívnych modelov s časovo oneskorenými premennými definovaný rovnicou (20). Pozoruhodné je, že posledné dva modely sú pre regióny s menej ako

Tab. 2. Porovnanie účinnosti modelov (hodnoty ostatných kritérií)

Model	K_1	K_2	K_3
JEDNODUCHÉ RASTOVÉ MODELY			
naivný model	40	2	38
model aritmetického rastu	65	0	67
model geometrického rastu	64	2	68
TRENDOVÉ MODELY			
model s konštantným trendom	10	56	8
model s lineárnym trendom	27	6	25
model s kvadratickým trendom	46	3	44
model s exponenciálnym trendom	24	10	22
model s mocninovým trendom	24	10	25
BODOVÉ REGRESNÉ MODELY			
lineárny model	48	2	48
exponenciálny model	44	2	46
mocninový model	67	0	65
AUTOREGRESNÉ MODELY			
model 1.rádu	51	2	52
model 2.rádu	78	0	75
model 3.rádu	71	0	73
model 4.rádu	67	0	70
JEDNODUCHÉ KOMPARATÍVNE MODELY			
model konštantného podielu počtu obyvateľov	44	0	46
model konštantného podielu absolútneho prírastku	73	0	75
model aritmetického rastu podielu počtu obyvateľov	57	0	57
TRENDOVÉ KOMPARATÍVNE MODELY			
model s konštantným trendom	11	51	11
model s lineárnym trendom	27	10	27
model s exponenciálnym trendom	30	10	29
KOMPARATÍVNE MODELY S ČASOVO ONESKORENÝMI PREMENNÝMI			
model s rovnicou (19)	62	0	62
model s rovnicou (20)	67	0	67
model s rovnicou (21)	60	0	65

50 tisíc obyvateľmi dokonca účinnejšie ako všetky autoregresné modely. Všimnime si, že priemerná presnosť predpovedí generovaných autoregresnými modelmi druhého až štvrtého rádu vzrastá priamo úmerne s veľkosťou mestského regiónu.

V najvyšších dvoch veľkostných kategóriách sa okrem spomenutých modelov dobre uplatňujú všetky zostávajúce komparatívne modely s výnimkou trendových modelov a z ostatných modelov ešte mocninový model bodovej regresie. K úspešným modelom pre stredne veľké regióny môžeme okrem toho zaradiť lineárny model bodovej regresie a model geometrického rastu, ktorý je vhodný aj pre regióny najnižšej veľkostnej kategórie. Najmenej presné predpovede sa vo všetkých veľkostných kategóriách získali pomocou obidvoch modelov s konštantným trendom. Vo všeobecnosti je možné konštatovať, že presnosť predpovede je v priemere najväčšia

Tab. 3. Priemerná relatívna chyba predpovede podľa veľkostných kategórií mestských regiónov

Model	Regióny s počtom obyvateľov (v tis.) 0-50	50-100	100 a viac
JEDNODUCHÉ RASTOVÉ MODELY			
naivný model	1,45	1,62	1,61
model aritmetického rastu	0,97	0,75	0,99
model geometrického rastu	0,97	0,79	1,10
TRENDOVÉ MODELY			
model s konštantným trendom	5,71	6,62	7,64
model s lineárnym trendom	2,25	1,74	2,75
model s kvadratickým trendom	1,66	1,32	1,17
model s exponenciálnym trendom	2,42	2,20	3,51
model s mocninovým trendom	2,54	2,19	2,05
BODOVÉ REGRESNÉ MODELY			
lineárny model	1,27	0,96	1,49
exponenciálny model	1,34	1,18	1,86
mocninový model	1,07	0,89	0,81
AUTOREGRESNÉ MODELY			
autoregresný model 1.rádu	1,24	1,02	1,66
autoregresný model 2.rádu	0,91	0,74	0,65
autoregresný model 3.rádu	0,90	0,65	0,77
autoregresný model 4.rádu	0,99	0,77	0,71
JEDNODUCHÉ KOMPARATÍVNE MODELY			
model konštantného podielu počtu obyvateľov	1,73	0,98	0,88
model konštantného podielu absolútneho prírastku	0,85	0,66	0,82
model aritmetického rastu podielu počtu obyvateľov	1,24	0,83	0,95
TRENDOVÉ KOMPARATÍVNE MODELY			
model s konštantným trendom	7,65	4,33	4,36
model s lineárnym trendom	3,06	1,67	1,68
model s exponenciálnym trendom	2,62	1,53	1,71
KOMPARATÍVNE MODELY S ČASOVO ONESKORENÝMI PREMENNÝMI			
model s rovnicou (19)	1,05	0,84	0,89
model s rovnicou (20)	0,76	0,88	0,88
model s rovnicou (21)	1,07	0,81	0,89

pre stredne veľké regióny a najmenšia pre regióny s menej ako 50 tisíc obyvateľmi.

ZÁVER

V tejto štúdii sme sa pokúsili zhodnotiť účinnosť niekoľkých extrapolačných modelov, ktoré sa často využívajú v regionálnych populačných projekciách. Výsledky našej analýzy potvrdzujú, že väčšina testovaných modelov sa aj v našich podmienkach môže stať spoľahlivým nástrojom regionálnej populačnej analýzy, ak sa za jediný cieľ populačnej projekcie pokladá krátkodobá predpoveď celkového počtu

obyvateľov skúmaného regiónu. Napriek tomu, že poznatky získané v tejto štúdii môžu uľahčiť výber vhodných modelov, mnohé otázky súvisiace s aplikáciou extrapolačných metód v regionálnych populačných projekciách zostávajú nepreskúmané. Jedna z nich súvisí s konštrukciou tzv. predpovedných intervalov, v ktorých sa bude s určitou pravdepodobnosťou nachádzať budúca hodnota počtu obyvateľov skúmaného regiónu. Iná skupina otázok sa týka závislosti medzi presnosťou predpovede na jednej strane a dĺžkou časového horizontu, rozsahom historického obdobia alebo rýchlosťou rastu skúmanej populácie na druhej strane. Nemenej zaujímavé sú aj rýdzo priestorové aspekty populačných projekcií, ako je napr. charakter priestorového rozloženia chýb predpovede alebo vplyv rôznych schém regionálneho členenia na účinnosť jednotlivých modelov.

LITERATÚRA

BENNETT, R.J. (1979). *Spatial time series: analysis - forecasting - control.* London (Pion).

BEZÁK, A. (1990). Funkčné mestské regióny v sídelnom systéme Slovenska. *Geografický časopis*, 42, 57-73.

BEZÁK, A. (1994). *Jednoduché metódy zabezpečujúce kompatibilitu dát z priebežnej evidencie obyvateľstva s výsledkami sčítania ľudu.* Interná štúdia. Geografický ústav SAV.

BEZÁK, A., HOLICKÁ, A. (1995). Testovanie účinnosti troch tried modelov pre populačné projekcie na regionálnej úrovni. In Hofmann, E., ed. *Postavení regionální geografie České republiky a Slovenské republiky v kontextu nových podmínek rozvoje.* Brno (Katedra geografie Pedagogické fakulty Masarykovy University), pp. 18-22.

CIPRA, T. (1986). *Analýza časových řad s aplikacemi v ekonomii.* Praha (SNTL a Alfa).

COHEN, J.E. (1986). Population forecasts and confidence intervals for Sweden: a comparison of model-based and empirical approaches. *Demography*, 23, 105-126.

EICHPERGER, C.L. (1984). Regional population forecasts: approaches and issues. In Heide, H. ter, Willekens, F.J., eds. *Demographic research and spatial policy: the Dutch experience.* London (Academic Press), pp. 235-252.

GREENBERG, M.R. (1972). A test of combination of models for projecting the population of minor civil divisions. *Economic Geography,* 48, 179-188.

HUNT, H.J., STANTON, P.G., HODGES, D. (1980). Getting an accurate forecast of local employment prospects: the Coventry experience. *Planner*, 15, 8-9.

ISARD, W. (1960). *Methods of regional analysis.* Cambridge, Mass. (MIT Press).

ISSERMAN, A.M. (1977). The accuracy of population projections for subcounty areas. *Journal of the American Institute of Planners,* 43, 247-259.

KEYFITZ, N. (1972). On future population. *Journal of the American Statistical Association*, 67, 347-363.

KRUECKEBERG, D.A., SILVERS, A.L. (1974). *Urban planning analysis: methods and models.* New York (Wiley).

LAND, K.C. (1986). Methods for national population forecasts: a review. *Journal of the American Statistical Association*, 81, 888-901.

MURDOCK, S.H., LEISTRITZ, F.L., HAMM, R.R., HWANG, S.-S., PARPIA, B. (1984). An assessment of the accuracy of a regional economic-demographic projection model. *Demography,* 21, 383-404.

OPENSHAW, S., KNAAP, G.A. VAN DER (1983). Small area population forecasting: some experience with British models. *Tijdschrift voor Economische en Sociale Geografie*, 74, 291-304.

OPPENHEIM, N. (1980). *Applied models in urban and regional analysis.* Englewood Cliffs (Prentice-Hall).

ROGERS, A. (1985). *Regional population projection models.* Beverly Hills (Sage).

SABOIA, J.L.M. (1974). Modeling and forecasting populations by time series: the Swedish case. *Demography*, 11, 483-492.

SIEGEL, J.S. (1972). Development and accuracy of projections of population and households in the United States. *Demography*, 9, 51-68.

SMITH, S.K. (1987). Tests of forecast accuracy and bias for county population projections. *Journal of the American Statistical Association*, 82, 991-1003.

SMITH, S.K., SINCICH, T. (1988). Stability over time in the distribution of population forecast errors. *Demography*, 25, 461-474.

STOTO, M.A. (1983). The accuracy of population projections. *Journal of the American Statistical Association*, 78, 13-20.

Anton Bezák, Anna Holická

A COMPARATIVE ANALYSIS OF SOME EXTRAPOLATION MODELS FOR REGIONAL POPULATION PROJECTIONS

In this paper an attempt is made to evaluate the performance of 24 extrapolation models developed for making regional population projections. The models selected for the analysis can be classified into seven broad categories: simple growth models [equations (1) to (3)], trend models [(4) to (8)], point regression models [(9) to (11)], autoregressive models of the first to fourth order (12), simple ratio models [(13) to (15)], trend ratio models [(16) to (18)], and ratio models with lagged variables [(19) to (21)]. The data base for the tests consists of a space-time population series for the 63 functional urban regions in Slovakia, covering the period 1970 to 1990.

Using these data, all extrapolation models are calibrated and used to project the 1994 population for each of the 63 urban regions. These forecasts are then compared with actual counts for target year in order to evaluate the model performance. The criteria utilized in the evaluations include mean absolute error (S_1), mean absolute percentage error (S_2), mean percentage error (S_3), percentage of errors smaller than 1% (K_1), percentage of errors greater than 5% (K_2), as well as percentage of times the model gives good forecast for the set of regions (K_3). The good forecast is defined as one within 1% of the best forecast for a given region. Tables 1 and 2 present comparisons of the accuracy of models' forecasts in terms of all criteria mentioned. In addition, the variations in mean absolute percentage error by the size of the region's population in 1990 are shown in Table 3.

The main conclusions from the results obtained can be summarized as follows. It is evident that the best forecasts are generally produced by either the autoregressive models of the second to fourth order or the ratio models with lagged variables. Both these categories of models perform well in all size groups of the urban regions. It should be noted, however, that the forecast accuracy of the autoregressive models tends to increase as population size of region increases, whereas the ratio models with lagged variables seem to be less affected by the population size of region. Sufficiently accurate forecasts are also generated by the simple growth models, the simple ratio model with equation (14), and the double logarithmic model of point regression. Although the best models tend to slightly underestimate region's population, there is no evidence of a consistent relationship between forecast accuracy and bias.

On the contrary, the poorest forecasts are provided by the standard and ratio trend models. This is especially true of both models with constant trend which produce forecasts with the greatest absolute and relative errors. It is interesting that each point regression model is more accurate than the corresponding trend model. Moreover, the no change model (1) is in fact better than all trend models (except model with quadratic trend). Finally, it is worth mentioning that forecasts for the smallest regions are generally less accurate than those for other region's size categories. The best forecasts are produced for the medium-sized regions.

English by A. Bezák