**ПРОГНОЗИРАНЕ НАСЕЛЕНИЕТО НА БЪЛГАРИЯ ДО 2050 Г. С ИЗПОЛЗВАНЕ НА ПРОСТИ МЕТОДИ**

**Гл. ас. д-р Виктор Аврамов, НБУ**

**FORECASTING POPULATION OF BULGARIA TO 2050 WITH SIMPLE METHODS**

**Assist. Prof. Victor Avramov**

***E-mail:*** **vavramov@nbu.bg**

***Резюме:*** *Дигиталното трансформиране в бизнеса изисква развиването на компетенции за анализиране, прогнозиране на взимане на решения на база на данни. Статията цели да покаже, че сравнително прости методи за прогнозиране могат да се използват за прогнозиране на важни показатели като броя на населението, раждаемостта и коефициента на плодовитост. Смисълът от такива упражнения нараства в условия на неопределеност, когато официалните прогнози изискват повече време за актуализация, но има нужда от бързина за взимане на решение на база на дадена прогноза. Използваните прости методи са базирани на експоненциално изглаждане. Един прогнозист с минимален опит може да ги приложи и изготви прогнози за такива важни показатели в рамките на минути. Това е една необходима адаптация за организациите, чиито бизнес процеси са изцяло базирани на данни.*

***Abstract:*** *Digital transformation in business requires the development of competencies for analysis, forecasting and decision-making based on data. The paper’s goal is to show that relatively simple forecasting methods can be used to forecast important indicators such as population size, birth rate and fertility rate. The meaning of such exercises increases in conditions of uncertainty, when official forecasts require more time to update, but there is a need for speed to decision making based on a given forecast. The simple methods used here are based on exponential smoothing. A forecaster with minimal experience can apply them and produce forecasts for important indicators within minutes.* *This is a necessary adaptation for organizations whose business processes are entirely data-driven.*

**Въведение**

Анализът и извличането на знание от данни е ключова компетенция за всяка организация в дигиталния свят. Това изисква развиването на компетенции за анализиране, прогнозиране на взимане на решения на база на данни. Есенцията на това е в използването на научните методи за проверка на хипотези, аналитични умения и компетенции за обработка на данни. Друг аргумент за развиването на тези умения и компетенции е в динамиката на самата среда.

През лятото на 2022 г. се вижда, че предстои рецесия, поне в Европа, причинена от войната на Русия в Украйна, рязкото поскъпване на горивата, нарушените вериги на доставки в света и постпандемичните условия като цяло. Това е трудна ситуация за прогнозиране и в такъв контекст официалните прогнози могат бързо да загубят достоверност. Понякога обаче една навременна прогноза може да бъде от голямо значение за взимането на решения в бизнеса, политиката и другаде. Прогнозирането на важни показатели като населението, раждаемостта и коефициента на плодовитост може да бъде направено сравнително с помощта на прости методи, без от това да се загуби достоверност. Целта на тази статия е да покаже, как това може да бъде направено бързо и пестеливо.

Разглеждат се три показателя: общият брой на населението, ражданията и коефициента на плодовитост, по години, за периода 1960 – 2021 г. Източници на данните са:

* базата данни на Световната банка за общия брой на населението за целия период [1];
* базата данни на Human Fertility Database за общия брой на ражданията до 2009 г. [2], допълнени с данните на НСИ до 2021 г.;
* базата данни на Световната банка за коефициента на плодовитост [3], освен за 2021, където там липсват и са допълнение с данни от НСИ [6].

Данни от НСИ се използват само като допълнение заради липсата на времеви редове преди 2000 г. (освен при изрична заявка и срещу заплащане), както и заради неудобният за използване интерфейс на уеб страницата.

1. **Описание на данните**

*Коефициент на плодовитост*

Добре е този показател да се дефинира по-точно:

„Показателят представлява средният брой на децата, които една жена би родила живи през целия фертилен период, като се вземат под внимание повъзрастовите коефициенти на плодовитост през дадена година. Коефициентът следователно представя пълната плодовитост на едно хипотетично поколение, като се изчислява чрез събиране на повъзрастовите коефициенти на плодовитост на жените през дадена година (приема се, че броят на жените на всяка възраст е един и същ). Тоталният коефициент на плодовитост се използва също и за да се покаже плодовитостта, необходима за възпроизводството.” [4]

Неговата динамика за дългия период, който се разглежда е много интересна. Както се

**Фигура 1.** Динамика на коефициента на плодовитост.

*Източник:* The World Bank DataBank (2022). [3]

вижда от Фигура 1, стойността му е над 2 до началото на 80-те години на 20 век и се задържа около тази стойност до края на 80-те години, след което рязко спада в следващото десетилетие. Нараства и постепенно се стабилизира след 2002 г. и към края на периода достига стойност от 1.5 раждания на една жена. Тази динамика, особено след 80-те години на миналия век показва връзката му със икономическата и социална среда в страната.

*Общ брой на ражданията*

„В броя на ражданията се включват живородените и мъртвородените деца.“ [5]

**Фигура 2.** Динамика на общия брой на ражданията

*Източник:* Human Fertility Database [2] до 2009 г., допълнени NSI (2022). *Razhdaniq po mestozhiveene,statistischeski raioni i pol* за 2021 г. [5].

Тенденцията на общия брой на ражданията е силно подобна на тази на коефициента на плодовитост, както трябва да се очаква. Интересно е да се види, че за разлика от него, не нараства стабилно след 2000 г., а тенденцията е надолу след 2009 г. Може да се мисли за значима линейна връзка между двата показателя. Тя е визуализирана на Фигура 3. От диаграмата се вижда, че наистина линейната връзка между двата показателя е много силна, освен при стойностите на коефициента на плодовитостта, движещи се около 1,5 – 1,6. Те съответстват на периода, в който раждаемостта намалява. Практическият извод от тези наблюдения е, че не е добра идея единият коефициент да се използва за прогнозиране на другия. Добре е те да се прогнозират отделно.

*Общ брой на населението*

НСИ дефинира населението като:

„Съвкупностите на населението и неговите структури се дефинират като моментни и се установяват към 31.12. на съответната година. Броят и структурите на населението към края на всяка година се изчисляват на базата на данните от предходната година и данните за естественото и механично движение на населението през текущата година.“ [6]

Това е важно да се отбележи, защото годишните данни за населението представляват оценката на база наблюденията за движението на гражданите, включително миграцията в рамките на свободния пазар на труда на ЕС и глобално. Тези оценки са толкова точни, колкото данните за това движение и методологията на НСИ позволяват. Този анализ няма да се впуска допълнително в тази тема, но е достатъчно да се подчертае, че най-точните данни за населението са тези от преброяванията, които се провеждат веднъж на всеки десет години.



**Фигура 3.** Връзка между плодовитостта и броя на ражданията

*Източник:* въз основа на данните за плодовитостта и броя на ражданията.

Динамиката на населението е визуализирана на Фигура 4. Тя показва добре известния факт, че населението спира да нараства след 1988 г. и намалява рязко. От 1988 до 2021 г. общото намаление е с 2 082 321 души, или 23%.

**Фигура 4.** Динамика на населението

*Източник:* 1. The World Bank DataBank (2022). Population, total – Bulgaria, 2060 – 2021 [1] до 2020 г., NSI (2022). *Naselenie po oblasti, obshtini, mestozhiveene I pol* за 2021 г.

**II. Прогнозиране**

В рамките на един демографски анализ може да се търси потвърждение за това, че икономическите кризи оказват въздействие върху динамиката на населението, миграцията, раждаемостта, смъртността и др. Но прогнозирането с използването на предиктори, т.е. данни за въздействащите фактори изисква наличие на прогнозни данни за тях. Такива фактори могат да бъдат макроикономически като динамика на БВП, показатели за благосъстоянието, здравеопазването и качеството на живот и други. За повечето от тях, в условията на неопределеност, създаването на надеждна прогноза в момента е рисковано. Всяка грешка в прогнозите за предикторите ще се отрази и на крайната прогноза. Липсата на ориентири за развитието на икономическата ситуация обаче прави прогнозирането на макроикономически процеси много трудна задача, особено в условията на задаваща се към средата на 2022 г. рецесия.

Прогнозите тук са направени с помощта на модели, базирани на експоненциално изглаждане. Тези модели, най-общо казано, създават прогнози на база на претеглени по определен начин исторически стойности на прогнозната величина [7]. Търси се най-добър модел от група модели: просто експоненциално изглаждане, метод на Холт [8] и метод на Холт със смекчен тренд [9]. Критерият за избор на модел е минимизирането на информационния критерий на Акаике, AICC:

Тук е сумата на квадратичните грешки, е броя на наблюденията, е броят на предикторите. Той се коригира като:

Прогнозите за трите показателя са направени отделно и ще бъдат представени последователно.

Първо, за прогнозата за коефициент на плодовитост е избран прост модел на експоненциално изглаждане (SES), който има най-нисък Критерия на Акаике (Таблица 1):

**Таблица 1.** Характеристика на моделите за коефициент на плодовитост

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модел | sigma2 | log\_lik | AIC | AICc | BIC | MSE | AMSE | MAE |
| SES | 0.00523 | 35.9 | -65.8 | -65.4 | -59.5 | 0.00506 | 0.0138 | 0.0530 |
| Holt | 0.00526 | 36.8 | -63.6 | -62.5 | -53.0 | 0.00492 | 0.0131 | 0.0530 |
| Damped | 0.00516 | 37.9 | -63.8 | -62.3 | -51.1 | 0.00475 | 0.0131 | 0.0509 |

*Източник:* този анализ.

Този модел може да се формулира по следния начин:

.



 е прогнозата за период , е нивото, а е параметър на изглаждане на времевия ред. За избрания модел , . Прогнозната стойност на коефициент на плодовитост до 2015 г. е 1.58. Фигура 5 визуализира прогнозите.

**Фигура 5.** Прогнози за коефициента на плодовитост

*Източник:* този анализ.

Прогнозата за броя на ражданията използва метода на Холт със смекчен тренд, който може да се представи по следния начин:

,

,

Тук, в допълнение, е втори параметър на изглаждане, а е коефициент на изглаждане на тренда, е средната претеглена стойност на тренда в период . Такъв модел има точност, идентична на простото експоненциално изглаждане, както Таблица 2 показва. Фигура 6 обаче демонстрира, че прогнозата от

**Таблица 2.** Характеристика на моделите за ражданията

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Модел | sigma2 | log\_lik | AIC | AICc | BIC | MSE | AMSE | MAE |
| SES | 21592133. | -650. | 1307. | 1307. | 1313. | 20895613. | 58869097. | 3470. |
| Holt | 21301907. | -649. | 1308. | 1309. | 1319. | 19927590. | 55909088. | 3402. |
| Damped | 21145701. | -648. | 1306. | 1307. | 1317. | 19440403. | 56889564. | 3260. |

*Източник:* този анализ.

този модел изглежда по-реалистична. Коефициентите на избрания модел са , , ,, .

**Фигура 6.** Прогнози за броя на ражданията

*Източник:* този анализ.

Прогнозата за общият брой раждания е 58496 за 2022 г. и леко намалява до 56211 в 2050 г.

Прогнозата за населението също е направена с метода на Холт с изгладен тренд, защото прогнозният модел има най-ниска стойност на Коефициента на Акаике (Таблица 3).

**Таблица 3.** Характеристика на моделите за населението

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| .model | sigma2 | log\_lik | AIC | AICc | BIC | MSE | AMSE | MAE |
| SES | 5294824560. | -821. | 1648. | 1648. | 1654. | 5124023768. | 16725563148. | 56396. |
| Holt | 1372007242. | -778. | 1566. | 1567. | 1577. | 1283490646. | 5559902094. | 19754. |
| Damped | 958041904. | -766. | 1543. | 1544. | 1553. | 880780460. | 5153568660. | 14781. |

*Източник:* този анализ.

Най-интересно тук (и сигурно е целия текст) е да се види самата прогноза. Тя е визуализирана на Фигура 7.

В прогнозата, населението намалява, както може да се очаква. Прогнозните стойности са систематизирани в Таблица 4, като са спестени стойностите между всеки пет години.



**Фигура 7.** Прогноза за населението

*Източник:* този анализ.

**Таблица 4.** Прогноза за населението

|  |  |
| --- | --- |
| Година | Население |
| 2022 | 6865979 |
| 2025 | 6776159 |
| 2030 | 6654013 |
| 2035 | 6559499 |
| 2040 | 6486366 |
| 2045 | 6429777 |
| 2050 | 6385990 |

*Източник:* този анализ.

Общо, населението в тази прогноза намалява от 6 865 979 до 6 385 990 души в 2050 г., което е намаление със 7%.

Накрая, стойностите на моделните параметри са , , ,, .

**III. Изводи и обобщения**

Прогнозата за населението, направена тук, предвижда намаляването му до 2050 г. с почти 480 хил. души. В прогнозата на НСИ, това намаляване е два пъти по-голямо: почти 1 075 хил. души, или с 16% спрямо 2021 г. Не е лесно да се каже сега коя от двете прогнози ще бъде по-точна, но сигурно е разумно да се дава доверие на официалните прогнози. Целта на статията е не да направи дисекция на най-точните методи за прогнозиране на населението и свързани с него показатели, а да демонстрира нещо много важно: леснотата, с която може да бъдат правени подобни прогнози, когато това е необходимо. Това, което мотивира необходимостта им е несигурната среда и скоростта за взимане на решения.

***Първо,*** всяка формална организация може и трябва да прогнозира това, което е важно за нея.

***Второ,*** разчитането на външни прогнози може да игнорира конкретните нужди на организацията, защото мотивацията на изготвилите прогнозата може да бъде различна.

***Трето,*** цената на адекватното прогнозиране може да бъде нищожна на фона на ползите от него. Направената тук прогноза не отне повече от един астрономически час, заедно с оформянето на графиките.

***Четвърто,*** повече време се изразходва за оформяне на резултатите, отколкото за същинската работа по създаването им.

***Пето,*** прогнозите отчитат, че населението на България ще намалява. Организациите трябва да се подготвят за това.

***Шесто,*** в една магистърска програма като „Бизнес администрация“ има място за курс „Анализ на данни и прогнозиране за мениджъри“.

***Литература***

The World Bank DataBank (2022). *Population, total – Bulgaria, 2060 – 2021* [Data set]. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=BG

The Human Fertility Database (2022). *Total number of live births 1947-2009* [Data set]. https://www.humanfertility.org/cgi-bin/country.php?country=BGR&tab=si

The World Bank DataBank (2022). *Fertility rate, total (births per woman) – Bulgaria, 2060 – 2020* [Data set]. https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=BG

NSI (2022). *Totalen koeficitsient na plodovitost* [Total fertility coefficient]. https://www.nsi.bg/sites/default/files/files/metadata/SDI%204.1.MD\_bg.pdf

NSI (2022). *Razhdaniq po mestozhiveene,statistischeski raioni i pol* [Births by place of residence, statistical regions, districts and sex]. https://nsi.bg/bg/content/2956/раждания-по-местоживеене-статистически-райони-области-и-пол

NSI (2022). *Naselenie po oblasti, obshtini, mestozhiveene I pol* [Population by regions, municipalities, place of residence and gender]. https://nsi.bg/bg/content/2975/население-по-области-общини-местоживеене-и-пол

R. J. Hyndman and G. Athanasopoulos (2021). *Forecasting: principles and practice*, 3 ed., OTexts: Melbourne, Australia, 2021.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages (ONR Memorandum No. 52). *Carnegie Institute of Technology, Pittsburgh USA. Reprinted in the International Journal of Forecasting, 2004.*

1. Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31(10), 1237–1246.*