**ПРЕДВИДУВАЊЕ НА ДВИЖЕЊЕТО НА БРОЈОТ НА ЖИВОРОДЕНИ ДЕЦА ВО РЕПУБЛИКА МАКЕДОНИЈА СО ПРИМЕНА НА** $ARMA$ **МОДЕЛОТ**

**Марија Трпкова – Несторовска**

Економски факултет – Скопје, Универзитет „Св. Кирил и Методиј“ во Скопје

marijat@eccf.ukim.edu.mk

**Апстракт**

*Како составен дел на наталитетот, живородените деца се основен елемент на репродукцијата на населението и поради нивното значење тие се предмет на проучување на ова истражување. Во овој труд се настојува да се опише и анализира овој витален процес, како и да се изврши моделирање на временската серија која го прикажува движењето на бројот на живородени деца во Република Македонија со користење на авторегресивниот модел со подвижни просеци, односно* $ARMA$ *моделот. Истиот модел се користи за предвидување на идното движење на анализираната временска серија.*

*Се користат месечни податоци за периодот од јануари 2008 година до мај 2014 година. Серијата најпрво се испитува за стационарност, понатаму се врши сезонско прилагодување поради специфичниот сезонски карактер на податоците, а потоа се применува* $ARMA$ *за опишување на серијата и предвидување на нејзиниот иден развој.*

*По тестирање на повеќе модели, моделот* $ARMA(1,1)$ *се покажал како најдобар бидејќи ги исполнува потребните статистички претпоставки и има соодветни вредности за мерките на предвидување.*

Клучни зборови: број на живородени деца, $ARMA$ модел, Македонија, предвидување

**Вовед**

Речиси повеќе од една деценија Република Македонија, како и бројни земји од Балканот, се соочуваат со проблемот на опаѓачкиот наталитет. Бројни студии се обидуваат да ги лоцираат причините за овој негативен тренд, а како најчести се наведуваат лошата економска состојба, образованието и вработеноста кај жените, промената на сликата за семејството и бракот и позицијата на жената во нив, како и надворешните миграции. Намалениот наталитет со себе носи економски и социјални последици.

Статистиката на наталитетот на населението подразбира опфаќање на сите случаи на раѓање деца, без разлика дали се работи за живородено или мртвородено дете. Бидејќи само живородените деца се елемент на репродукцијата на населението, потребно е јасно да се разграничат термините живородено и мртвородено дете. Живородено дете е она кое покажало знак на живот, односно чукање на срцето и дишење (Ристески, 2009). Поради важноста на живородените деца како елемент на репродукција на населението, тие и ќе бидат предмет на истражување на овој труд.

На слика 1 е прикажано движењето на бројот на живородени деца во Република Македонија. Може да се забележи дека и покрај сезонските варијации кои се карактеристични за оваа серија, таа не бележи значителен тренд на опаѓање или раст во разгледуваниот период. Во изминатите шест години може да се каже дека серијата има стабилно движење и во рамките на интервалот од приближно 1.700 живородени деца до нешто повеќе од 2.200 деца.

Врз основа на претходно консултирана литература, целта на ова истражување е да опише и предвиди временска серија за бројот на живородени деца со користење на авторегресивниот модел со подвижни просеци, познат како $ARMA$ модел. Истражувања во кои била применета оваа методологија биле правени во земји блиски на нашата, односно во Хрватска и во Албанија. На светско ниво постојат бројни истражувања кои со $ARMA$ моделите вршат демографски истражувања кај разни витални статистики.

*Слика 1. Движење на бројот на живородени деца во Република Македонија*

Извор: Државен завод за статистика на Република Македонија

 Трудот е поделен на пет делa. Првиот дел е, секако, воведот, во кој се дефинира наталитетот и се опишува движењето на анализираната серија. Во вториот дел е елаборирана литературата која се занимава со слично истражување со примена на $ARMA$, односно $ARIMA$ при анализа на виталните статистики, со посебен акцент на трудовите кои се занимаваат со истражување на бројот на раѓањата. Во третиот дел накратко е опишана користената методологија, односно основните карактеристики на $ARMA$ моделите и пресметките според кои се вршат нивните предвидувања. Добиените резултати се презентирани и објаснети во четвртиот дел. Трудот завршува со заклучокот добиен од истражувањето на бројот на живородени деца.

**Преглед на литературата**

$ARMA$, односно $ARIMA$ моделите се често користени за анализа на наталитетот на различни земји од светот кои се карактеризираат со различни културолошки, социјални и економски услови.

Во својот труд Caleiro (2010) го користи $ARIMA$ моделот кој вклучува и регресиони променливи за да се испитаат сезонските ефекти на наталитетот во Португалија. За сезонскиот карактер на раѓањата тој претпоставува дека постојат две групи фактори: (1) социјални и културолошки, кои се однесуваат на верските обичаи, вообичаено однесување по склучување на бракот и годишните одмори и (2) климатските фактори во кои ја вклучува температурата. Конзистентни резултати од повеќе истражувања велат дека високите температури го намалуваат фекондитетот. Ова објаснува зошто во многу земји од северната хемисфера најмногу деца се родени во септември. Интересно е да се напомене дека анализата на живородените деца во Македонија ги поддржува овие резултати, бидејќи најголем број родени деца има во лето и во есен, а значително е помал бројот во пролет и во зима.

За анализа на виталните статистики во Хрватска се користени $ARIMA$ моделите (Rozga and Banovic). Истражувањето се однесува на бројот на раѓања, бројот на умирања, браковите и разводите. Сезонските и останатите варијации се отстранети од месечните временски серии и спроведени се предвидувања за пет идни периоди. Во Хрватска во последните години бројот на населението се намалува, исто како и бројот на склучени бракови, што е последица на бројни причини. Користени се два метода за сезонско прилагодување, X-12-$ARIMA$, кој е емпириски метод, и $TRAMO/SEATS$ кој е метод на основа на модел. Првиот е доста користен кај официјалните статистики, а вториот се користи кај неколку статистички агенции. Резултатите покажале дека вториот метод бележи мала предност во однос на првиот.

$ARIMA$ моделите се користени за да се опише и да се предвиди бројот на раѓања во Албанија. За еден подолг период оваа бројка бележи значаен пад, иако во последните години овој тренд бележи забавување. Трудот покажува дека со користење на $ARIMA$ моделите може да се врши истражување на демографски податоци кои се однесуваат на албанската популација (Dhamo and Puka, 2011).

Според Saboia (1974) техниките за анализа на временски серии често се користат за предвидување на популациите. Во својот труд тој ги користи авторегресивните модели и моделите на подвижни просеци со цел да ја моделира популацијата во Шведска. И двата модели даваат добри резултати. Предвидените вредности се добиени преку двата модели и се споредени со предвидувања добиени од други модели. Споредбите дават добри резултати. И покрај тоа што ова истражување користи податоци за популацијата на Шведска, постојат добри причини да се очекува дека оваа техника може успешно да се примени и на останатите параметри на популацијата.

Истиот автор (Saboia, 1977) го проширува своето истражување и ги користи $ARIMA$ моделите за анализа на временската серија за бројот на раѓањата и го испитува односот на овие модели кон класичните модели за раст на популацијата. Добиени се различни верзии на $ARIMA$ моделите во кои се содржани најзначајните информации, вкучувајќи ја и должината на генерациите на популацијата. Оваа техника е применета на временската серија за бројот на раѓања во Норвешка.

Со методот на сезонски $ARIMA$ модели извршена е анализа на бројот на умирања и бројот на раѓања во САД. Користен е Бокс-Џенкинсоновиот метод. Резултатите се финални модели кои имаат значајни авторегресивни компоненти од втор ред, како и сезонски компоненти на подвижни просеци, за двете серии. Кај серијата број на раѓања откриено е постоење на сличности во неделните движења. Овие откритија подразбираат постоење на процедура за сезонско прилагодување и таа придонесува да се подобри програмата CensusX-I1, која ја користат агенциите за витални статистики (Land and Cantor, 1983).

Во трудот на McDonald (1981) се говори за проблемот на моделирање на демографски променливи со цел нивно предвидување. Предмет на дебата е тврдењето дека теоријата ретко овозможува целосна динамичка спецификација за моделот. За да се предвиди бројот на раѓања во Австралија се користат два пристапа, пристап на временски серии и секвенцијална техника за тестирање. Добиените модели се оценуваат со споредување на нивните предвидени вредности со предвидените вредности добиени со $ARMA$ моделите.

**Методологија на истражување**

 Во трудот се користи авторегресивниот модел на подвижни средини кој уште е познат и како $ARMA$ модел. Во спроведувањето на анализата следува реализација на $ARMA$ моделот според Бокс-Џенкинсовата методологија.

 Општиот облик на $ARMA(p,q)$ процесот е:

$$y\_{t}=ϕ\_{0}+\sum\_{i=1}^{p}ϕ\_{i}y\_{t-i}+e\_{t}-\sum\_{i=1}^{q}θ\_{i}e\_{t-i}$$

каде што $ϕ\_{0},ϕ\_{i},θ\_{i}$ се параметри кои се оценуваат, додека пак $e\_{t}$ е процес на бел шум со нормален распоред. Вредностите $p$ и $q$ се ненегативни цели броеви каде што $p$ го означува редот на $AR$ процесот, додека пак $q$ го означува редот на $MA$ процесот.

 Изградбата на $ARMA$ моделот се врши според Бокс-Џенкинсоновата методологија, која е составена од три чекори: идентификување, оценување и проверка на моделот.

При првиот чекор се утврдува дали серијата е стационарна или се карактеризира со нестационарност. Кај $ARMA$ моделите е важно најпрво да се утврди дали временската серија е стационарна или нестационарна. Имено, доколку се потврди дека серијата е стационарна, се користи $ARMA$ моделот. Доколку серијата се карактеризира со нестационарност, таа ќе биде трансформирана (диференцирана од прв или втор ред) и во тој случај ќе се користи $ARIMA$ моделот (авторегресивен интегриран модел на подвижни средини). Разликата се состои во тоа што во $ARMA$ моделот се користи серијата во нејзините основни реализации, додека во $ARIMA$ моделот серијата е трансформирана. За тестирање на стационарноста во истражувањето ќе биде користен е проширениот Дики – Фулеров тест (англ. augmented Dickey Fuller test). Во оваа фаза се утврдува редот на $AR$ процесот и редот на $MA$ процесот, односно вредностите за $p$ и $q$. Ова се врши со помош на функциите на автокорелација и парцијална автокорелација или преку информациските критериуми.

Вториот чекор подразбира оценка на непознатите параметри на моделот $ϕ\_{0},ϕ\_{i},θ\_{i}$.

Во третиот чекор се проверува дали моделот е соодветен за понатамошно користење. Се врши проверка за статистичката значајност на оценетите параметри и се проверува дали корелограмот на резидуалите на моделот покажува дека меѓу нив не постои автокорелација. Доколку овие барања се исполнети, моделот може да се користи за предвидување. Доколку не се исполнети, во тој случај моделот повторно се оценува, но со нови вредности за $p$ и $q$.

 Ако со $h$ се означува почетокот на предвидувањето, а со $F\_{h}$ се означуваат информациите кои се достапни за периодот $h$, предвидувањето на вредноста на $y\_{t}$ за $l$ чекори напред, со $ARMA(p,q)$ моделот е:

$$\hat{y}\_{h}\left(l\right)=E\left(F\_{h}\right)=ϕ\_{0}+\sum\_{i=1}^{p}ϕ\_{i}\hat{y}\_{h}\left(l-i\right)-\sum\_{i=1}^{q}θ\_{i}e\_{h}(l-i)$$

каде што:

$\hat{y}\_{h}\left(l-i\right)=y\_{h+l-i}$ако $l-i\leq 0$;

$e\_{h}\left(l-i\right)=0$ ако $l-i>0$ и

$e\_{h}\left(l-i\right)=e\_{h+l-i}$ ако $l-i\leq 0$.

**Анализа на добиените резултати**

Временската серија која содржи податоци за бројот на живородените деца во Република Македонија е конструирана со месечни податоци за периодот јануари 2008 година – мај 2014 година и содржи вкупно 77 опсервации. Податоците се добиени од Месечниот статистички извештај за периодот февруари 2008 година – јули 2014 година од Државниот завод за статистика на Република Македонија.

Од графичкиот приказ на временската серија на слика 1 може да се забележи дека серијата се карактеризира со одредена сезоналност. Имено, најголем број родени деца има во лето и во есен, а значително е помал бројот во пролет и во зима. Од оваа причина потребно е да се изврши сезонско прилагодување на временската серија. За таа цел се користи постапката CensusX12. Добиените резултати се прикажани на слика 2.

*Слика 2. Серија на живородени деца изразена во изворни податоци и сезонски прилагодени податоци*

Извор: Државен завод за статистика на Република Македонија и пресметки на авторот

Може да се заклучи дека сезонските прилагодени податоци ги ублажуваат варијациите во серијата и го отстрануваат сезонското влијание. Новата серија од сезонски прилагодени податоци ќе биде тестирана за стационарност.

Тестирањето за стационарност покажа дека променливата за бројот на живородените деца се карактеризира со стационарност.

*Табела 1. Проширен Дики – Фулеров тест на единечен корен*

|  |  |
| --- | --- |
| **Променлива** | **p - вредност** |
| **Константа** | **Тренд и константа** |
| $$BIRTH$$ | 0,0000 | 0,0000 |

Извор: Пресметки на авторот

Бидејќи станува збор за стационарна серија, јасно е дека ќе се користи $ARMA(p,q)$ модел. Со цел да се утврди редот на авторегресивниот процес и редот на процесот на подвижни просеци, потребно е најпрво да се изработи корелограм на серијата од каде што ќе се анализираат функцијата на автокорелација и функцијата на парцијална автокорелација.

*Слика 3. Корелограм на серијата број на живородени деца*

**

Извор: Пресметки на авторот

Од корелограмот со 24 временски доцнења на слика 3 може да се воочи дека постојат значајни вредности на коефициентите на автокорелација и парцијална корелација. Според веќе утврденото правило, коефициентот на автокорелација е значаен ако се најде надвор од интервалот:

$$\pm 1,96∙\frac{1}{√T}$$

каде што $T$ е бројот на опсервации во серијата. Во примерот овој интервал изнесува $\left[-,223;0,223\right]$. Според ова правило, значајни коефициенти на автокорелација се јавуваат кај првото, второто и третото временско доцнење, додека значајни коефициенти за апарцијалната корелација се јавуваат кај првото временско доцнење. И 24-тото временско доцнење бележи значаен коефициент, но тоа нема да биде земено во моделот. Се оценува следниот $ARMA(3,1)$ модел:

$$y\_{t}=ϕ\_{0}+ϕ\_{1}y\_{t-1}+ϕ\_{2}y\_{t-2}+ϕ\_{3}y\_{t-3}+e\_{t}-θ\_{1}e\_{t-1}$$

*Табела 2. Оценети параметри на* $ARMA(3,1)$ *моделот*

|  |  |  |
| --- | --- | --- |
| **Променлива** | **Коефициент** | $p$ **- вредност** |
| Константа | 1960,60 | 0,0000 |
| $$AR(1)$$ | 0,5266 | 0,0519 |
| $AR(2)$  | 0,1200 | 0,4256 |
| $AR(3)$  | 0,0813 | 0,4991 |
| $MA(1)$  | -0,3398 | 0,2245 |
| Прилагоден $R^{2}$ | 0,1683 |
| Акаике информациски критериум | 11,4895 |
| Шварцов информациски критериум | 11,6451 |
| Дурбин – Вотсонова статистика | 2,0011 |

Извор: Пресметки на авторот

Од табела 2 јасно се согледува дека оценетите параметри не се статистички значајни. Од таа причина моделот не е добар и потребно е да се оцени нов модел со поинаква спецификација. По повеќе итерации во кои се вклучени само временските доцнења за статистички значајните коефициенти, добиен е моделот кој е оценет како добар. Тоа е моделот $ARMA(3,1)$:

$$y\_{t}=ϕ\_{0}+ϕ\_{1}y\_{t-1}+e\_{t}-θ\_{1}e\_{t-1}$$

*Табела 3. Оценети параметри на* $ARMA(1,1)$ *моделот*

|  |  |  |
| --- | --- | --- |
| **Променлива** | **Коефициент** | $p$ **- вредност** |
| Константа | 1959,01 | 0,0000 |
| $$AR(1)$$ | 0,8892 | 0,0000 |
| $AR(2)$  | -0,7066 | 0,0000 |
| Прилагоден $R^{2}$ | 0,1212 |
| Акаике информациски критериум | 11,5721 |
| Шварцов информациски критериум | 11,6641 |
| Дурбин – Вотсонова статистика | 1,8300 |

Извор: Пресметки на авторот

Новиот модел покажува дека оценетите параметри се статистички значајни. За да моделот биде успешен, потребно е да се испита и автокорелацијата кај резидуалите.

*Слика 4. Корелограм на резидуалите на* $ARMA(1,1)$ *моделот*

**

Извор: Пресметки на авторот

Од слика 4 се согледува дека $p$-вредноста на Јунг-Бокс $Q$ статистиката за 24 временски доцнења постојано е поголема од нивото на значајност 0,05 и од 0,1. Тоа значи дека се прифаќа нултата хипотеза која вели дека серијата на резидуалите е, всушност, процес на бел шум. Тоа значи дека моделот е добар за да ја опише серијата на живородени деца во Република Македонија и може да се користи за понатамошно предвидување на нејзините вредности.

Предвидувањето може да биде статичко и динамичко. Динамичкиот метод започнува од првата временска точка на примерокот и ги користи претходно предвидените вредности на зависните променливи за добивање на предвидувањето на тековната вредност. Статистичкиот метод ги користи вистинските вредности на зависните променливи, доколку се достапни.

*Слика 5. Изворни податоци и предвидени вредности на бројот на живородени деца во Република Македонија*

Извор: Државен завод за статистика на Република Македонија и пресметки на авторот

Од слика 5 може да се утврди дека динамичкото предвидување не ги следи идеално изворните податоци за живородени деца во Република Македонија. Помеѓу нив постојат значајни разлики, а може да се каже и дека тенденцијата на динамичкото предвидување е вредноста на предвидувањето да се приближува кон средната вредност на серијата. Моделот предвидува дека идната вредност на бројот на живородени деца ќе изнесува 1959. За разлика од него, динамичкото предвидување го следи движењето на изворната серија, но предвидените вредности не ги отсликуваат изразените варијации. Движењето на предвидените вредности е поумерено. Ова предвидување врши и предвидување за еден иден период, а тоа значи дека за јуни 2014 година се предвидува раѓање на 1977 живи новороденчиња.

За да се утврди дали вредностите добиени од статичкото и динамичкото предвидување се веродостојни, се користат седум мерки на предвидување. Добиените вредности се прикажани во табела 4.

*Табела 4. Оценети параметри на* $ARMA(1,1)$ *моделот*

|  |  |  |
| --- | --- | --- |
| **Мерка на предвидување** | **Динамичко предвидување** | **Статичко предвидување** |
| Коренувана просечна квадратна грешка | 83,92 | 75,77 |
| Просечна апсолутна грешка | 64,72 | 57,00 |
| Просечна апсолутна процентна грешка | 3,36 | 2,95 |
| Теилов коефициент на нееднаквост | 0,02 | 0,02 |
| Бијасна пропорција | 0,02 | 0,00 |
| Варијансна пропорција | 0,85 | 0,46 |
| Коваријансна пропорција | 0,11 | 0,54 |

Извор: Пресметки на авторот

Коренуваната просечна квадратна грешка и просечната апсолутна грешка се апсолутни мерки кои ги мерат отстапувањата на предвидените вредности од изворните вредности. Пониската вредност укажува на подобра прилагоденост кон изворните вредности, односно помали отстапувања. Просечната апсолутна процентна грешка и Теиловиот коефициент на нееднаквост се релативни мерки. Подобриот модел има пониски вредности на овие мерки. Теиловиот коефициент на нееднаквост се движи од 0 до 1, со тоа што колку е поблизок до 0, толку моделот подобро ги прикажува изворните податоци. Бијасната пропорција кажува колку просечната вредност на предвидувањето се разликува од просечната вредност на изворните податоци, варијансната пропорција покажува за колку се разликува варијансата на предвидувањето од варијансата на изворните податоци и коваријансната пропорција ги мери несистематските грешки на предвидувањето. Моделот е добар ако бијасната и варијансната пропорција се мали, а коваријансната пропорција е поголема (Pindyck and Rubinfeld, 1990). Според ова, статичкото предвидување има подобри вредности за мерките на предвидувањето, а со тоа и подобро ги прикажува изворните податоци.

**Заклучок**

Врз основа на спроведеното истражување може да се заклучи дека $ARMA$ моделот успешно ја опишува временската серија за движењето на бројот на живородени деца со што овозможува и алатка која во иднина може да се користи за предвидување на движењето на овој витален процес. Веќе е кажано дека живородените деца се основен елемент на репродукцијата на населението и овој тренд треба постојано да се следи, со цел да не се реализира опаѓачка тенденција. Намалувањето на бројот на живородени деца придонесува до стареење на популацијата, што од друга страна, пак, носи негативни економски, социјални и културолошки последици. Од претходните образложенија и анализи може да се каже дека во последните шест години Република Македонија бележи стабилен тек на овој витален процес, но и покрај тоа, тој треба постојано да се следи и да се стимулира неговиот раст преку низа мерки инкорпорирани во квалитетна популациона политика.

**Користена литература**

Brooks, C. (2008). “*Introductory Econometric for Finance”*, Cambridge University Press, Cambridge.

Caleiro, A. (2010) “Forecasting the number of births in Portugal”, *Conference proceedings of Joint Eurostat/UNECE Work Session on Demographic Projections*, 28-30 April, 2010 Portugal.

Dhamo, E. and Puka, L. (2011) “An $ARIMA$ birth number per month model for Albanian population”, *Conference proceedings of the 1st International Symposium on Computing in Informatics and Mathematics*, 2-4 June, 2011, Albania.

Државен завод за статистика на Република Македонија (2014) „*Месечен статистички извештај“,* за период:февруари 2008 година-јули 2014 година.

Land, K.C. and Cantor, D. (1983) “Arima models of seasonal variation in U.S. birth and death rates”, *Demography*, vol. 20 (4), pp. 541-568.

McDonald, J. (1981) “Modeling Demographic Relationships: An Analysis of Forecast Functions for Australian Births”, *Journal of American Statistical Association*, vol. 76 (376), pp. 782-792.

Pindyck, R.S., and Rubinfeld, D.L. (1990) *Econometric Models and Econometric Forecasts*, 4th edition, McGraw Hill, New York.

Ристески, С. (2009) *Демографија методи и анализа*, Економски факултет, Скопје.

Ристески, С., Тевдовски, Д. и Трпкова, М. (2012) *Вовед во анализата на временските серии*, Универзитет Св. „Кирил и Методиј“, Скопје.

Rozga, A. and Banovic, Z. (2006) “The Analysis of Seasonal Variations in Vital Statistics in Croatia: Two Approaches”, *Joint Statistical Meetings in Seattle, Washington.*

Saboia, J.L.M (1974) “Modeling and forecasting” populations by time series: The Swedish case”, *Demography*, vol. 11 (3), pp. 483-492.

Saboia, J.L.M (1977) “Autoregressive Integrated Moving Average $(ARIMA)$ Models for Birth Forecasting”, *Journal of the American Statistical Association*, vol. 72 (358), pp. 264-270.

**FORCASTING THE LIVE BIRTHS’ TREND IN THE REPUBLIC OF MACEDONIA BY USING THE** $ARMA$ **MODEL**

**Marija Trpkova – Nestorovska, PhD**

Faculty of Economics - Skopje, Ss. Cyril and Methodius University in Skopje

marijat@eccf.ukim.edu.mk

**Abstract**

*As part of natality, the live births are main element of the population reproduction, and due to their significance they are the focus of this research. This paper tends to explain and analyze this vital process, as well as to model the time series that represent the trend of the live births in the Republic of Macedonia by using the autoregressive moving averages model. The same model is used for forecasting the future trend of the analyzed time series.*

*Monthly data is used for the period January 2008 – May 2014. The series are primarily examined for stationarity, then seasonal adjustment is performed due to the specific seasonal character of the analyzed data, and then the* $ARMA$ *model is applied for explanation of the movement of the series and their forecasting.*

*After testing the several models, the* $ARMA(1,1)$ *model is proven to be the best one, meeting all the necessary statistical assumptions and having appropriate values for the forecasting measures.*

Keywords: number of live births, $ARMA$ model, Macedonia, forecasting