

# Моделирование демографических показателей с использованием SARIMA-технологии

Л.А. Шафикова  
Факультет информатики и робототехники  
Уфимский государственный авиационный  
технический университет  
Уфа, Россия  
e-mail: [shafikoval@gmail.com](mailto:shafikoval@gmail.com)

Л.С. Ямилова  
Факультет информатики и робототехники  
Уфимский государственный авиационный  
технический университет  
Уфа, Россия  
e-mail: [yles@mail.ru](mailto:yles@mail.ru)

Л.И. Яцкович  
Факультет информатики и робототехники  
Уфимский государственный  
авиационный технический университет  
Уфа, Россия  
e-mail: [ladanamaste108@gmail.com](mailto:ladanamaste108@gmail.com)

## Аннотация<sup>1</sup>

В статье рассматривается изменение такого демографического процесса, как рождаемость в Российской Федерации в 2010-2015 годы. А также осуществляется прогнозирование данного показателя на 2016 год с помощью технологии SARIMA и сравнение с фактическими данными.

## 1 Введение

Одной из наиболее актуальных проблем для национальной безопасности нашей страны является ее демографическое состояние. Известно, что рождаемость в современной России, несмотря на определенное улучшение, связанное с относительным повышением уровня жизни и некоторыми мерами государства в направлении стимулирования демографического роста, остается на достаточно низком уровне [1].

Для построения социально экономической модели развития России, необходимо знать ее демографическое состояние в будущем, одним из основных факторов которого является рождаемость. Спрогнозировав данный показатель, возможно обеспечение всех детей детскими садами и школами. Также прогнозирование рождаемости необходимо для оценки перспектив жилищного строительства.

Цель данной работы – построить прогнозную модель для краткосрочного прогнозирования динамики

показателя рождаемости РФ и обнаружить наличие или отсутствие сезонных колебаний в динамике ряда и оценить волатильность показателя.

## 2 Методология прогнозирования

Одним из современных инструментов, которые используются для прогнозирования демографических процессов, является ARIMA-моделирование. На основании работы С. Смита, Д. Таумана и Д. Суонсона [2] в исследовании в качестве прогнозной модели выбрана модель ARIMA. Эта модели относятся к классу универсальных линейных моделей для описания стационарных и нестационарных временных рядов, измеряющих, в том числе, демографические показатели [3]. Модели позволяют получать точные прогнозы, основываясь на информации, содержащейся в предыстории исследуемых рядов. Чем больше информации имеется для выполнения анализа, тем более качественным получится построенный прогноз.

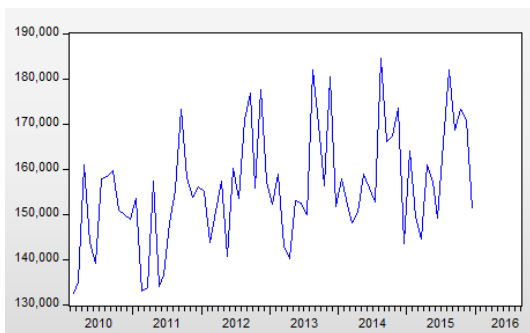
Прогнозирование с помощью моделей ARIMA успешно осуществляется уже достаточно длительное время. Если анализировать научные публикации последних десяти лет, можно заметить, что в большинстве работ по прогнозированию классическими методами используется именно эта методология, как наиболее обоснованный и достоверный алгоритм.

## 3 Результаты

В качестве исходных данных взята ежемесячная динамика показателя рождаемости в РФ с 2010 г. по 2015 г.

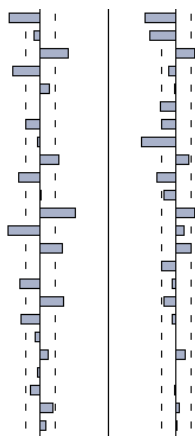
---

Труды Шестой всероссийской научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 28-31 мая, Уфа-Ставрополь, Россия, 2018



**Рисунок 1. График показателя рождаемости**

Первым этапом построения ARIMA-моделей является определение типа процесса, к которому относится временной ряд, на ретроспективе которого строится прогноз.



**Рисунок 2. Коррелограмма первых разностей**

Из анализа коррелограмм автокорреляционной функции (АКФ) и частной автокорреляционной функции (ЧАКФ) можно сделать вывод о том, что исходный ряд является стационарным на исходных условиях.

Однако пятиэтапная процедура расширенного теста Дики-Фуллера опровергла эту гипотезу при уровне значимости  $p < 0,05$ . Выяснилось, что процессы, описывающие ряды ретроспективных данных показателей рождаемости в РФ 2010-2015 г. относятся к процессам случайного блуждания первого порядка интеграции [4].

Известно, что такой показатель, как рождаемость, имеет ярко-выраженный сезонный характер. Это подтверждается и в данном случае.

Из анализа коррелограмм АКФ и ЧАКФ, представленных на рисунке 2, можем сделать вывод о наличии сезонной компоненты, т.к. коэффициенты функции для лагов 12 статистически значимы. Существуют некоторые способы, позволяющие выделить ее из ряда.

1. с помощью тригонометрических функций

В данном случае оценка сезонной составляющей происходит путем построения нелинейной регрессии

по времени, где в качестве аналитических функций выступают синус и косинус.

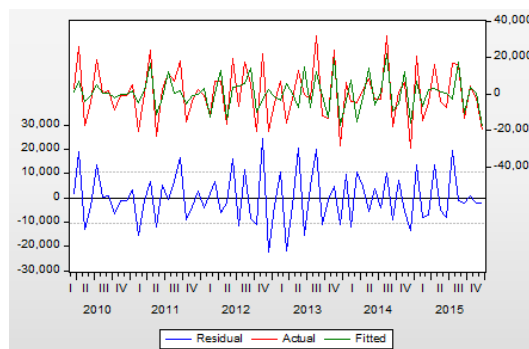
2. с помощью введения фиктивных переменных
  - 1 Для оценки сезонности с помощью фиктивных переменных, строится линейная регрессия, где в качестве зависимой переменной в зависимости от типов процессов, а в качестве независимых переменных выступают фиктивные переменные и константа.
  3. с помощью стандартного метода сезонной декомпозиции временного ряда Census.

Стандартный метод сезонной декомпозиции временного ряда Census позволяет достаточно качественно удалить сезонную компоненту из структуры временного ряда на основе комбинированного ряда.

4. с помощью модели SMA (сезонная скользящая средняя) и SAR (сезонная авторегрессия)

Этот способ предполагает выделение сезонности за счет сезонной авторегрессии с помощью модели SMA (сезонная скользящая средняя) и SAR (сезонная авторегрессия). Зависимая переменная указывается исходя из типа процесса, в качестве независимых переменных указывается сезонная скользящая средняя и сезонная авторегрессия с указанием периода сезонности.

В ходе применения всех способов, выяснилось, что корректно выделяет и устраняет сезонную составляющую из структурного ряда только четвертый способ. Результаты применения четвертого способа по выделению и устранению сезонной компоненты показали удовлетворительные статистические результаты (при уровне значимости  $p < 0,05$ ), что повысило надежность полученных прогнозных оценок.



**Рисунок 3. График фактических, расчетных значений и остатков модели**

На рисунке 3 представлены графики фактических значений с наложением моделируемых значений и остатков.

В результате была оценена модель :

$$\Delta y_t = 7.198\Delta y_{t-1} + 6.39\Delta y_{t-2} + 5.86\Delta y_{t-3} + 5.03\Delta y_{t-4} + 4.43\Delta y_{t-5} + 3.6\Delta y_{t-6} + 2.8\Delta y_{t-7} + 1.95\Delta y_{t-8} + 1.34\Delta y_{t-9} + 0.54\Delta y_{t-10} + 1738.6 \quad (1)$$

После построения модели временных рядов остатки модели были протестированы согласно тесту Уайта на наличие условной гетероскедастичности. В результате тест подтвердил наличие ARCH-эффектов (при уровне значимости  $p > 0,2$ ).

В результате проведения процедур спецификации были построены ARCH-модель, GARCH-модель и обобщенная ARCH-модель. Далее была произведена процедура селекции, основу которой составляли расчеты информационных критериев Акайке, Шварца, Хана-Квина: выбиралась модель с минимальными значениями информационных критериев [5].

**Таблица 1 – Результаты оценивания модели**

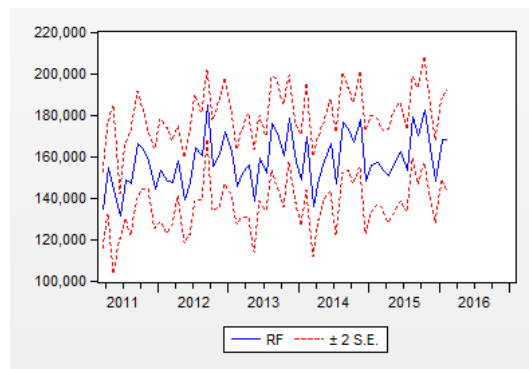
| Переменная | Коэффициент | Вероятность принятия нулевой гипотезы о равенстве нулю коэффициента |
|------------|-------------|---|
| C          | 498,16      | 0,59  |
| AR(12)     | 0,57        | 0   |
| MA(4)      | -0,77       | 0   |
| SMA(12)    | 0,31        | 0,04  |
| C          | 94278503    | 0,001   |
| GARCH(-1)  | 0,58        | 0   |
| GARCH(-2)  | -0,97       | 0   |

В таблице 1 представлены результаты оценивания модели SARMA+GARCH методом максимального правдоподобия.

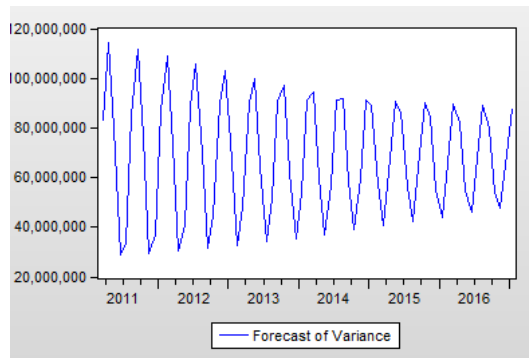
В результате селекции, была отобрана GARCH-модель. В данном случае, модель значима при уровне значимости  $p < 0,05$ .

Моделируется условная дисперсия с помощью модели GARCH второго порядка

#### 4 Обсуждение результатов



**Рисунок 4. График доверительного интервала**



**Рисунок 5. График доверительного интервала**

На рисунках 4 и 5 представлены числовые данные в виде ряда фактических и расчетных данных, упорядоченных по времени.

Доверительный интервал построен в соответствии с GARCH моделью: интервал с большей дисперсией соответствует интервалу с большим доверительным интервалом.

Для оценки прогностической ценности модели определяют следующие виды ошибок прогнозирования:

- Квадратный корень средней ошибки предсказания = 8209
- Средняя ошибка прогноза по модулю = 6780
- Средняя процентная ошибка по модулю (средняя ошибка аппроксимации) = 4,31
- Коэффициент неравенства Тейла, определяемый как:
- Доля систематической ошибки = 0,000205
- Доля вариации = 0,000085
- Доля ковариации = 0,999710

На основании полученных ошибок можно сделать заключение о высоком качестве модели, т.к. все ошибки близки к нулю, доля ковариации близка к 1, ошибка аппроксимации не превышает 7-10 % и равна 4,3%.

В результате был получен перспективный прогноз на 7 месяцев 2016 года.

**Таблица 2 – Сравнение полученных результатов с реальными данными**

| Реальные данные | Прогноз |
|-----------------|---------|
| 142052          | 162921  |
| 149012          | 162147  |
| 167141          | 162701  |
| 147711          | 164655  |
| 156589          | 164206  |
| 165872          | 164528  |
| 165340          | 165717  |

После проведения всех этапов построения модели были спрогнозированы достоверные, с определенной долей допущения, результаты.

Главной задачей данного исследования было построить прогнозную модель для краткосрочного прогнозирования показателя рождаемости в РФ. Составленный прогноз оказался близок к фактическим данным, которые предоставлены на сайте Росстата. Это свидетельствует, во-первых, о пригодности моделей SARIMA+GARCH для краткосрочного прогнозирования демографических показателей, во-вторых, о достоверности полученных краткосрочных прогнозов демографического развития в Российской Федерации.

## 5. Заключение

В ходе исследования были достигнуты следующие результаты:

- Доказана принадлежность данного процесса к процессам случайного блуждания первого порядка интеграции;
- Подтвержден сезонный характер показателя рождаемости;
- Получен прогноз рождаемости за 7 месяцев 2016 года;
- Произведено сравнение прогнозных и фактических данных;
- Сделан вывод о качестве использования технологии SARIMA;
- Сделан вывод о достоверности полученных краткосрочных прогнозов демографического развития в РФ.

## Список используемых источников

1. Ревягина Н.Г. Демографические аспекты положения современной России // Государственное и муниципальное управление. Ученые записки СКАГС. 2014. № 2. С. 214-218.

2. Бахитова Р.Х., Лакман И.А., Шамсутдинова Н.К. Применение ARIMA-моделей для краткосрочного прогнозирования показателей рождаемости в г. Уфа // Уровень жизни населения регионов России. – 2016. - № 3. – С. 214-219.
3. Stanley K. Smith. State and local population projections: Methodology a. analysis / Stanley K. Smith, Jeff Tayman, David A. Swanson. - New York: Kluwer acad. / Plenum, Cop. 2001. - XVI, 426 с.
4. Green William H. Econometric Analysis. 6th ed. Upper Saddle River, NJ, Prentice Hall, 2008.
5. Саяпова А.Р., Гусельникова Е.А., Лакман И.А., Шамуратов Н.М. Математические методы прогнозирования экономических показателей. Учебное пособие. –Уфа : БашГУ, 2000. – 128