



УДК 519.246.8; 314.1; 311.17

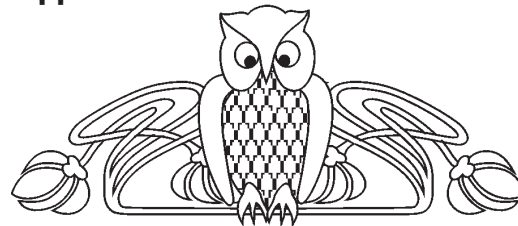
СТАТИСТИЧЕСКИЕ ПОДХОДЫ К АНАЛИЗУ И ПРОГНОЗИРОВАНИЮ ДЕМОГРАФИЧЕСКИХ ДАННЫХ

Е. Д. Копнова

кандидат технических наук, доцент департамента статистики и анализа данных, Национальный исследовательский университет «Высшая школа экономики», Москва
E-mail: ekopnova@hse.ru

Л. А. Родионова

кандидат экономических наук, доцент департамента статистики и анализа данных, Национальный исследовательский университет «Высшая школа экономики», Москва
E-mail: lrodionova@hse.ru



Введение. В статье рассмотрены возможности применения ARIMA-моделирования к анализу и прогнозированию временных рядов демографических показателей. Зарубежные исследования показали, что наряду с традиционными демографическими методами ARIMA-модели дают хорошие результаты при прогнозировании таких показателей, как численность населения, уровень рождаемости и смертности, продолжительность жизни населения. **Методика исследования.** В работе использовалась методология Бокса – Дженкинса анализа и прогнозирования временных рядов, в частности, применительно к демографическим данным: коэффициенту суммарной рождаемости в России (1990–2014 гг.), количеству заключенных браков по месяцам в России (2005–2015 гг.), общему коэффициенту рождаемости во Франции (1740–2014 гг.) и уровню безработицы в России (1996–2016 гг.). Были проанализированы ARIMA-, SARIMA- и ARFIMA-модели в зависимости от характера динамики исследуемого показателя. **Результаты.** Проведенный анализ показал, что оцененные ARIMA-модели для коэффициента суммарной рождаемости и количества заключенных браков являются адекватными, имеют хорошие статистические и прогностические свойства. На основании полученных моделей построены прогнозы. В случае длинных рядов наличия свойств процессов с длинной памятью выявлено не было.

Ключевые слова: анализ временных рядов, ARIMA-модели, SARIMA-модели, ARFIMA-модели, статистические методы.

DOI: 10.18500/1994-2540-2016-16-3-306-315

Введение

Демографическое прогнозирование является неотъемлемой частью анализа социально-экономических процессов в целом. Необходимость построения прогноза движения населения и предвидение будущей демографической ситуации: как изменится численность населения, возрастно-половая структура, рождаемость, продолжительность жизни, смертность – все это важные составляющие при развитии и планировании различных сфер экономики и общества (социальной инфраструктуры, здравоохранения, образования, пенсионной системы) при принятии важных политических решений.

В демографическом прогнозировании используют различные подходы: математические методы, метод передвижки возрастов (когортно-компонентный метод), статистические методы. Одним из статистических методов прогнозирования в демографии является одномерное ARIMA-моделирование. В основе этого метода лежит подход «чистого временного ряда»: текущие значения демографического показателя связаны с его прошлыми значениями, а для прогнозирования будущих значений используется информация о прошлых значениях изучаемого показателя. Подход ARIMA-моделирования был предложен Боксом и Дженкинсом в 1970 г. [1] к анализу экономических процессов, кроме того, авторы ввели реализацию данного подхода в компьютерные пакеты статистических программ по всему миру. Вскоре ARIMA-модели (Autoregressive Integrated Moving Average Model) стали использовать во многих областях науки, в том числе и демографии. В работах 1970-х гг. демографы моделировали численность населения, уровень рождаемости и смертности (Hiorns, 1972; Lee, 1974; Pollard, 1970; Saboia, 1974) [2–5]. Кашьяп и Рао в 1976 г. сравнили различные методы анализа и прогнозирования временного ряда численности населения в США за 1900–1971 гг., для которого была получена удовлетворительная модель ARIMA(1,1,0) в логарифмах [6]. Сабой применил ARIMA-модели к анализу уровня рождаемости по данным Норвегии за 1919–1974 гг., наиболее адекватными моделями оказались ARIMA(4,1,1) и ARIMA(3,1,2), на основании которых был построен прогноз на период 1976–2000 гг. с 50%-ным доверительным интервалом [7]. Пфлаумер в 1992 г. проанализировал численность населения США за 1900–1988 гг., на основании модели ARIMA(2,2,0) и модели ARIMA(1,1,0) в логарифмах был построен прогноз до 2080 г. с 95%-ным доверительным интервалом [8].

За последние два десятилетия во многих странах заметно увеличилась средняя продолжительность жизни мужчин и женщин. Ряд авторов применили ARIMA-модели к анализу



средней продолжительности жизни мужчин и женщин в Италии и США за 1900–2006 гг. Наиболее удовлетворительными моделями оказались ARIMA(2,1,1) для женщин и ARIMA(1,1,1) для мужчин, на основании которых были построены прогнозы до 2050 г. с 80%-ным и 95%-ным доверительными интервалами [9–11]. Интересным является результат, что средняя продолжительность жизни к 2050 г. для женщин составит 97 лет, для мужчин 88 лет.

Отметим, что отечественные демографы при построении прогнозов по демографическим данным, как правило, отдают предпочтение математическим методам и когортно-компонентному анализу. Основной целью данной работы является развитие статистической методологии анализа

и прогнозирования демографических данных. В своем исследовании подробнее остановимся на применении ARIMA-моделей к демографическим данным по России и продемонстрируем возможности прогнозирования с использованием данного класса моделей.

Методика исследования

ARIMA-модели: теоретические предпосылки. Рассмотрим временной ряд наблюдений y_1, y_2, \dots, y_n некоторого демографического показателя, в частности, суммарного коэффициента рождаемости в России за 1990–2014 гг. (рис. 1). Эти наблюдения будут рассматриваться как реализация случайных величин, которые описываются некоторым стохастическим процессом.

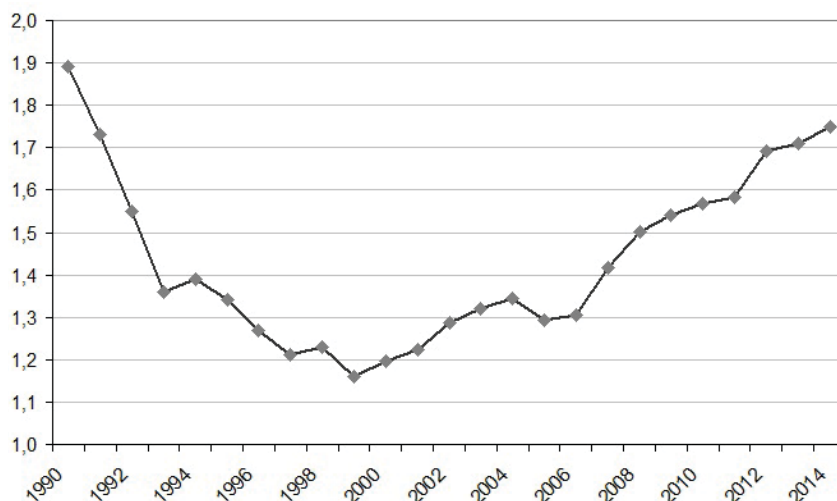


Рис. 1. Суммарный коэффициент рождаемости в России [12]

Ключевым понятием при ARIMA-моделировании является стационарность случайного процесса. Случайный процесс принято называть *слабо стационарным*, если среднее, дисперсия и ковариация y_t не зависят от времени t : $E(y_t) = \mu < \infty$, $V(y_t) = \gamma_0$, $\text{cov}(y_t, y_{t-k}) = \gamma_k$ [13].

Заметим, что многие демографические временные ряды не являются стационарными, имеют сложную структуру, и их элементы формируются под воздействием различных факторов, как и многие другие явления нашего общества. В исследованиях, как правило, выделяют четыре типа факторов [14]. Долговременные факторы формируют общую тенденцию изменения анализируемого показателя y_t , описываемую неслучайной функцией – функцией тренда. Например, ряды коэффициента суммарной рождаемости (см. рис. 1), ожидаемой продолжительности жизни для мужчин и женщин (рис. 2) в России и г. Москве за период 1990–2014 гг. содержат трен-

ды. Однако, как показывает визуальный анализ графиков, функции, описывающие трендовые составляющие для рассматриваемых процессов, могут различаться.

Сезонные факторы формируют периодически повторяющиеся в определенное время года колебания анализируемого показателя. В качестве примера можно рассмотреть количество браков, заключенных по месяцам, в России в период 2005–2015 гг. (рис. 3). Легко заметить, что данный показатель носит сезонный характер.

Как отмечают демографы, в современной России в мае регистрируется в три-четыре раза меньше браков, чем в августе. Сравнительно мало браков заключается также в январе–марте, а больше всего – в августе и сентябре, снижение частоты заключения браков в декабре и апреле связывают с распространением церковной регистрации и обряда венчания (на эти месяцы обычно приходятся посты). В 2010 г. больше всего браков

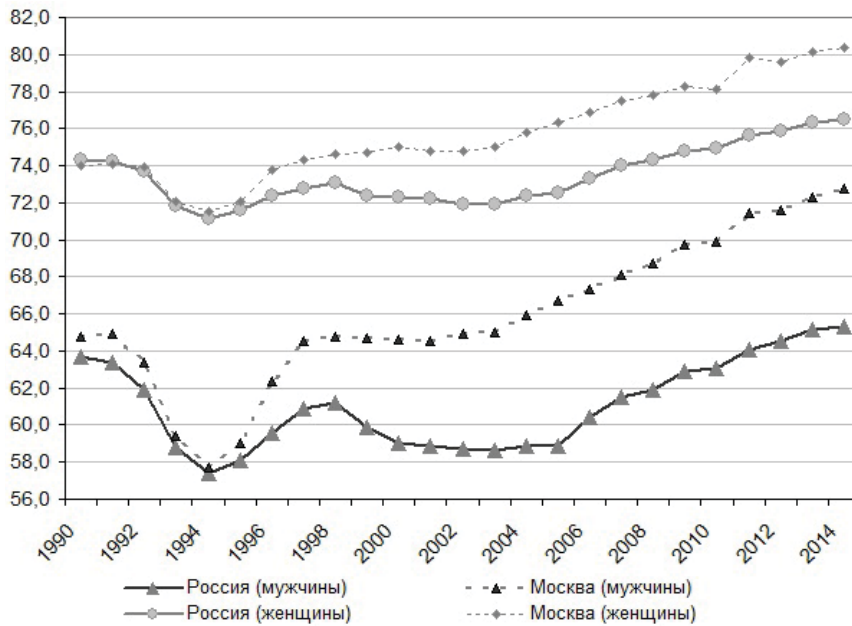


Рис. 2. Ожидаемая продолжительность жизни мужчин и женщин при рождении в России и г. Москве (число лет) [12]

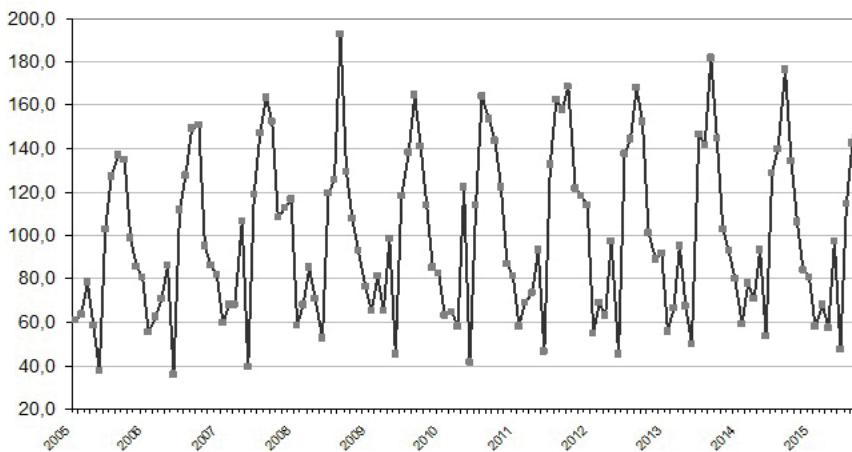


Рис. 3. Число зарегистрированных браков (тыс.) по месяцам регистрации [12]

было зарегистрировано в июле (164,3 тыс., что 4 раза больше, чем в мае), а в 2012 г. – в августе (168,0 тыс. – в 3,7 раза больше, чем в мае) [15]. Похожая динамика сохранилась и в 2013–2015 гг. Отметим, что сезонность присутствует также в ряде других демографических показателей: рождаемости, смертности, заболеваемости, миграции. Третий тип – циклические факторы – формируют изменения анализируемого признака, обусловленные действием долговременных циклов (в частности демографических). Отметим, что для изучения демографических циклов требуются длинные временные ряды. И последний тип факторов – случайные факторы – не поддаются учету, их воздействие на формирование значений временного ряда обуславливает стохастическую природу элементов ряда.

Существует несколько способов определения, является ли ряд стационарным. Во-первых, графический анализ: наличие тренда или периодической составляющей, возрастание или убывание разброса наблюдений со временем является индикатором нестационарности ряда. Так, на рис. 1–2 изучаемые временные ряды – коэффициент суммарной рождаемости, ожидаемая продолжительность жизни мужчин и женщин – являются нестационарными временными рядами из-за наличия тренда. Во-вторых, анализ графика выборочной автокорреляционной функции (ACF) (коррелограммы) и частной автокорреляционной функции (PACF), характеризующей «чистую корреляцию» между y_t и y_{t-k} при исключении влияния промежуточных значений $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$. Для стационарных временных рядов графики автокорреляционной



функции и частной автокорреляционной функции «быстро убывают» после нескольких первых значений лагов. В-третьих, использование статистических тестов на стационарность, например, расширенного ADF-теста Дики – Фуллера (Dickey, Fuller), PP-теста Филлипса – Перрона (Phillips, Perron), KPSS-теста Квятковского – Филлипса – Шмидта – Шина (Kwiatkowski, Philips, Schmidt, Shin) [13].

Модель авторегрессии и скользящего среднего ARMA(p, q) (Autoregressive Moving Average) в общем виде выглядит следующим образом:

$$y_t - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p} = \beta + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad (1)$$

где e_t – белый шум, обладающий свойствами: математическое ожидание $\mu=0$, дисперсия $\gamma_0 = \sigma^2$, ковариации $\gamma_k = 0, k > 0$. На практике используют, как правило, ARMA-модели не выше 2-го порядка. При моделировании демографических явлений исследователи сталкиваются с нестационарными процессами, которые могут быть приведены к стационарному виду при помощи оператора последовательной разности. По определению, первая разность – это $\Delta y_t = y_t - y_{t-1}$. Пусть случайный процесс y_t после того, как к нему применили d раз оператор последовательной разности, стал стационарным процессом $\Delta^d y_t$, удовлетворяющим одной из ARMA(p, q)-моделей. В этом случае процесс y_t принято называть *интегрированным* процессом авторегрессии и скользящего среднего ARIMA(p, d, q). Такой процесс в общем случае имеет вид:

$$\Delta^d y_t = \beta + \sum_{i=1}^p \alpha_i \Delta^d y_{t-i} + e_t + \sum_{j=1}^q \theta_j e_{t-j}. \quad (2)$$

Также в литературе распространена запись ARIMA(p, d, q)-модели через лаговый оператор:

$$\alpha_p(L) \Delta^d y_t = \beta + \theta_q(L) e_t, \quad (3)$$

где $Ly_t = y_{t-1}, L^k y_t = y_{t-k}, \alpha_p(L) = 1 - \alpha_1 L - \alpha_2 L^2 \dots - \alpha_p L^p, \theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 \dots + \theta_q L^q$.

Отметим, что на практике встречаются процессы с порядком интегрирования d не более двух. В основе прогнозирования с использованием ARIMA-моделей лежит методология Бокса – Дженкинса, которая содержит три этапа: идентификация модели, оценивание и проверка адекватности модели, прогнозирование. На этапе идентификации модели необходимо протестировать ряды на стационарность и подобрать ARIMA-модель для дальнейшего оценивания. На втором этапе параметры ARIMA-моделей оцениваются методом максимального правдоподобия. Для проверки адекватности полученных ARIMA-моделей и их сравнения используют

несколько критериев: оценки коэффициентов модели должны быть статистически значимыми, остатки модели должны обладать свойствами белого шума (анализ автокорреляционной функции $\rho_k = 0$, критерий Льюинга – Бокса, тест Бройша – Годфри). В случае, когда несколько ARIMA-моделей оказываются адекватными, необходимо выбрать модель с наименьшим количеством параметров и наилучшими статистическими характеристиками качества подгонки модели, для чего обычно используют информационный критерий Акаике (AIC) и критерий Шварца (BIC) [13]. Отметим, что данная методика реализована в большинстве статистических пакетов, например Stata, R, Gretl, Eviews и др.

SARIMA: моделирование сезонных эффектов в демографических данных. Одним из расширений ARIMA-моделей является учет сезонности, оценивание так называемых SARIMA-моделей, сочетающих в себе сезонную составляющую и обычную ARIMA-модель. В сезонных SARIMA-моделях необходимо также оценить дополнительные параметры сезонных компонент: SARIMA(p, d, q)(P_s, D_s, Q_s), P_s – сезонный порядок авторегрессии, D_s – порядок сезонной разности, Q_s – сезонный порядок скользящего среднего. В общем виде мультипликативная сезонная модель SARIMA(p, d, q)(P_s, D_s, Q_s) в записи через лаговый оператор имеет вид

$$\alpha_p(L) \alpha_{ps}(L) \Delta^d \Delta_s^{D_s} S d y_t = \theta_q(L) \theta_{qs}(L) e_t, \quad (4)$$

где $\alpha_{ps}(L) = 1 - \alpha_{s1} L^s - \alpha_{s2} L^{2s} \dots - \alpha_{sp} L^{ps}$,

$\theta_{qs}(L) = 1 + \theta_{s1} L^s + \theta_{s2} L^{2s} \dots + \theta_{sq} L^{qs}$.

Построение SARIMA-моделей аналогично анализу ARIMA-моделей.

ARFIMA: моделирование процессов с долгой памятью в демографии. Обобщением модели ARIMA(p, d, q), когда параметр d может принимать любое действительное число, являются модели ARFIMA(p, d, q) (Fractionally Integrated Autoregression and Moving Average), представимые в виде

$$\alpha_p(L) (1-L)^d y_t = \theta_q(L) e_t, \quad (5)$$

где $Ly_t = y_{t-1}, L^k y_t = y_{t-k}, \alpha_p(L) = 1 - \alpha_1 L - \alpha_2 L^2 \dots - \alpha_p L^p,$

$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 \dots + \theta_q L^q, -1 < d < 1$. Показано,

что в случае $-0.5 < d < 0.5$ ARFIMA-процесс является стационарным и обратимым, а автокорреляции ρ_k убывают с гиперболической скоростью. Формально стационарный процесс определяется как процесс с длинной памятью, если существуют такие числа α, c ($0 < \alpha < 1, c > 0$), что для автокорреляционной функции исследуемого процесса выполняется соотношение



$$\lim_{k \rightarrow \infty} \frac{\rho_k}{ck^{-\alpha}} = 1. \quad (6)$$

Для идентификации процессов с длинной памятью традиционно используется R/S-анализ, предложенный Херстом [16] и Мандельбротом [17]. В предложенном подходе анализируется R/S-статистика, вычисляемая как отношение размаха частичных сумм (R) центрированных значений ряда к его стандартному отклонению (S). Идея R/S-анализа состоит в том, что в предположении полного отсутствия памяти размах ряда увеличивается пропорционально корню из величины временного промежутка, т.е. $R/S \approx ct^H$, $H = 0,5$ (показатель степени H называют показателем Херста). В противном случае показатель Херста H характеризует наличие памяти во временном ряду. При $0,5 < H < 1$ процесс является персистентным, т.е. монотонно следует вдоль детерминированного тренда, а при $0 < H < 0,5$ процесс называют антиперсистентным – обладающим свойством высокочастотной изменчивости значений относительно тренда. Для проверки гипотезы об отсутствии длинной памяти в ряду используется модифицированная статистика, предложенная Ло [18] и рассчитываемая на основе R/S-статистики.

Результаты

Анализ динамики коэффициента суммарной рождаемости (КСР) с использованием ARIMA-моделей. Рассмотрим более подробно содержание методологии Бокса – Дженкинса на примере российских демографических данных – коэффициента суммарной рождаемости за период 1990–2014 гг. Длина временного ряда составила

24 наблюдения. ADF-тест Дики – Фуллера показал, что исследуемый процесс является нестационарным для всех спецификаций (с константой, с константой и трендом, без них) и имеет порядок интегрирования $d = 1$ (ряд был приведен к стационарному виду взятием первой разности). Анализ графика и его автокорреляционной и частной автокорреляционной функций свидетельствует о нестационарности изучаемых процессов, что также подтвердили ADF-тесты Дики – Фуллера, в частности, статистика $DF = -2,387$ (для случая с константой и трендом). В результате оценивания и сравнения нескольких ARIMA-моделей (разных порядков авторегрессии и скользящего среднего) наиболее адекватной оказалась модель ARIMA(1,1,0) для КСР, имеющей вид (7). В круглых скобках указаны стандартные ошибки.

$$\Delta y_t = -0,011 + 0,629\Delta y_{t-1} + e_t, \sigma = 0,068. \quad (7)$$

(0,04) (0,182)

Для рассматриваемой модели КСР Q-статистика = 7,04, AIC = -49,261 BIC = -45,988. Значения автокорреляционных функций остатков равны нулю, Q-статистика Льюинга – Бокса и статистика теста Бройша – Годфри свидетельствуют об отсутствии автокорреляции в процессе остатков, это позволяет сделать вывод о том, что процесс остатков обладает свойствами белого шума.

Наряду с оценкой адекватности модели, для оценки качества и точности прогноза по полученной модели использовались характеристики, представленные в табл. 1. Отметим, что полученные характеристики точности прогноза свидетельствуют о достаточно высоких прогностических свойствах полученной ARIMA(1,1,0)-модели для КСР.

Таблица 1

Характеристики точности прогноза в модели КСР

Характеристика точности прогноза	Формула вычисления	Значение характеристики
Средняя квадратичная ошибка	$MSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$	0,004
Средняя абсолютная процентная ошибка	$MAPE = \frac{1}{n} \sum_{t=1}^n \left \frac{\hat{y}_t - y_t}{y_t} \right \cdot 100\%$	3,82
Средняя процентная ошибка	$MPE = \frac{1}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t} \cdot 100\%$	0,31

На рис. 4 представлены наблюдаемые значения КСР и предсказанные по ARIMA-модели для России, а также построен 80%-ный и 95%-ный доверительные интервалы прогноза до 2020 г. В табл. 2 приведены прогнозные значения КСР в России на основе полученной ARIMA-модели с 80%-ным и 95%-ным

доверительными интервалами. Результаты моделирования показывают, что на протяжении 2015–2022 гг. в России будет наблюдаться замедление роста коэффициента суммарной рождаемости до значения 1,73 на одну женщину в 2022 г. и наметится тенденция к снижению данного показателя.

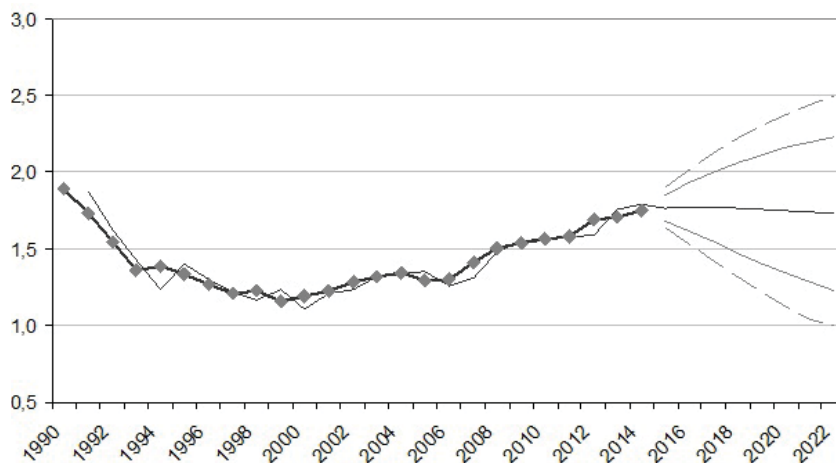


Рис. 4. КСР в России: наблюдаемые и предсказанные значения на основе ARIMA(1,1,0)-модели с 80%-ным и 95%-ным доверительными интервалами

Таблица 2

Прогнозные значения КСР в России на основе ARIMA(1,1,0)-модели с 80%-ным и 95%-ным доверительными интервалами

Год	Прогноз	Статистическая ошибка	95%-ный доверительный интервал	80%-ный доверительный интервал
2015	1,77	0,07	(1,63, 1,89)	(1,68, 1,85)
2016	1,78	0,13	(1,53, 2,02)	(1,61, 1,93)
2017	1,78	0,18	(1,42, 2,12)	(1,54, 2,01)
2018	1,77	0,23	(1,31, 2,22)	(1,47, 2,06)
2019	1,76	0,28	(1,21, 2,30)	(1,40, 2,11)
2020	1,75	0,32	(1,12, 2,37)	(1,34, 2,16)
2021	1,74	0,36	(1,04, 2,44)	(1,28, 2,19)
2022	1,73	0,39	(0,96, 2,50)	(1,22, 2,23)

Анализ числа зарегистрированных браков в России с помощью SARIMA-моделей. Как было отмечено выше, показатель числа зарегистриро-

ванных браков в России носит ярко выраженный сезонный характер с пиками в июле–сентябре и резкими спадами в мае (рис. 5).

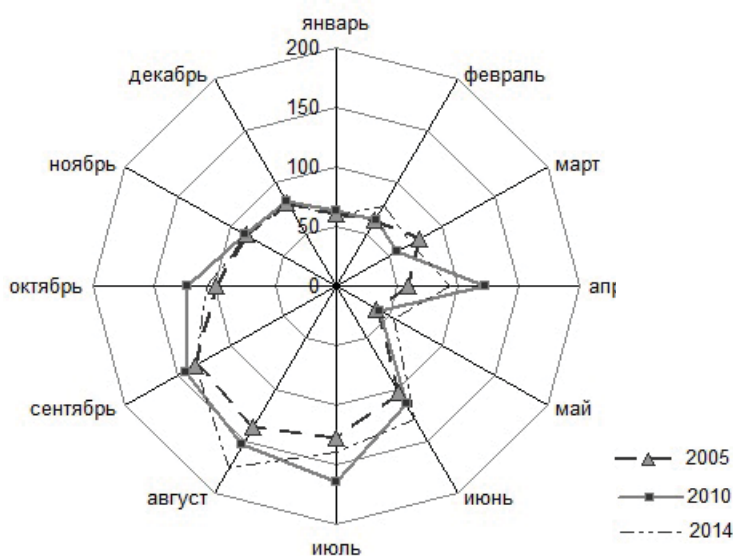


Рис. 5. Число зарегистрированных браков (тыс.) по месяцам регистрации в 2005, 2010, 2014 гг. [12]



Анализ автокорреляционной и частной автокорреляционной функций исходного ряда позволил сделать предположение о периоде сезонности $S = 12$. Для анализа был рассмотрен временной интервал с января 2003 г. по август

2015 г. Длина временного ряда для данного показателя составила 115 наблюдений. При моделировании была взята одна сезонная разность. Удовлетворительной оказалась модель $SARIMA(0,0,1)(3,1,0)_{12}$ в виде

$$(1 + 0,57L^{12} + 0,48L^{24} + 0,46L^{36})\Delta_{12}y_t = 0,78 + (1 + 0,34L^3)e_t, \quad \sigma = 11,47. \quad (8)$$

(0,08) (0,09) (0,10) (0,64) (0,09)

На основании теста Льюинга – Бокса для 24 лагов (Q-статистика = 26,66) был сделан вывод, что процесс остатков для оцененной модели обладает свойствами белого шума. В табл. 3 представлены прогнозные значения числа зарегистрированных браков на 2016 г. Результаты показывают, что в целом сезонный характер показателя на будущий год останется примерно таким же, как и в предыдущие годы.

Таблица 3

Прогнозные значения числа зарегистрированных браков в России в 2016 г. по месяцам регистрации (тыс.) на основе SARIMA-модели

Месяц	Прогноз	95%-ный доверительный интервал
Январь	58,66	(34,86, 82,45)
Февраль	71,34	(47,54, 95,13)
Март	63,98	(40,18, 87,77)
Апрель	98,29	(74,49, 122,08)
Май	49,02	(25,23, 72,81)
Июнь	129,21	(105,41, 153,01)
Июль	145,11	(121,31, 168,91)
Август	167,86	(144,06, 191,65)
Сентябрь	151,17	(125,49, 176,85)
Октябрь	107,73	(82,04, 133,40)
Ноябрь	95,27	(69,58, 120,95)
Декабрь	95,34	(69,43, 121,23)

Моделирование процессов с долгой памятью в демографии. Эмпирические демографические исследования показывают, что достаточно часто встречаются процессы с длинной памятью (long-memory processes), когда существует сильная зависимость между отдаленными наблюдениями [19]. Гранджер в 1980 г. одним из первых определил, что процессы с длинной памятью могут появляться в результате макроагрегирования традиционных стационарных временных рядов [20]. Можно предположить, что при анализе длинных временных рядов демографических данных также встречаются процессы с длинной памятью. На сайте журнала «Демоскоп» [21] представлены длинные ряды демографических данных (за 250 лет) по некоторым показателям, таким как общий коэффициент рождаемости, общий коэффициент смертности, коэффициент суммарной рождаемости, коэффициент младенческой смертности по некоторым европейским странам. В качестве примера рассмотрим ряд общего коэффициента рождаемости во Франции за период 1740–2014 гг. (годовые данные) (рис. 6). Длина временного ряда составила 275 наблюдений. Динамика ряда характеризуется убывающим трендом и с резким провалом в годы Первой мировой войны. Вторым показателем для исследования процессов с длинной памятью был взят уровень безработицы в России за период с января 1996 г. по январь 2016 г. (рис. 7). Длина временного ряда составила 265 наблюдений.

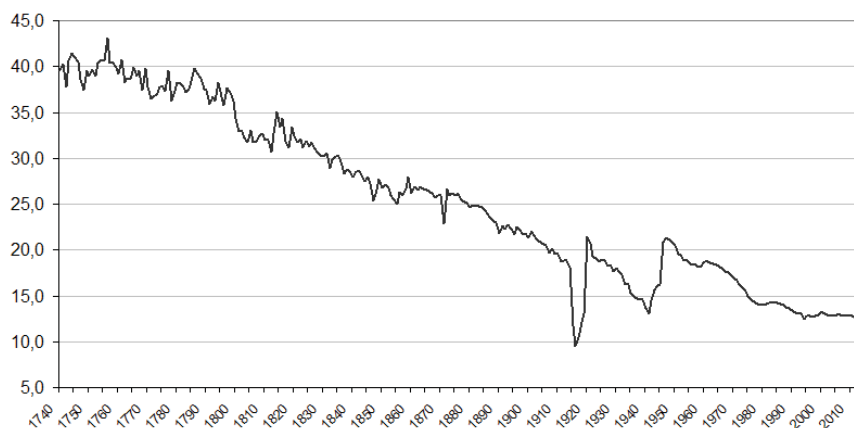


Рис. 6. Общий коэффициент рождаемости во Франции (1740–2014 гг.) [21, 22]

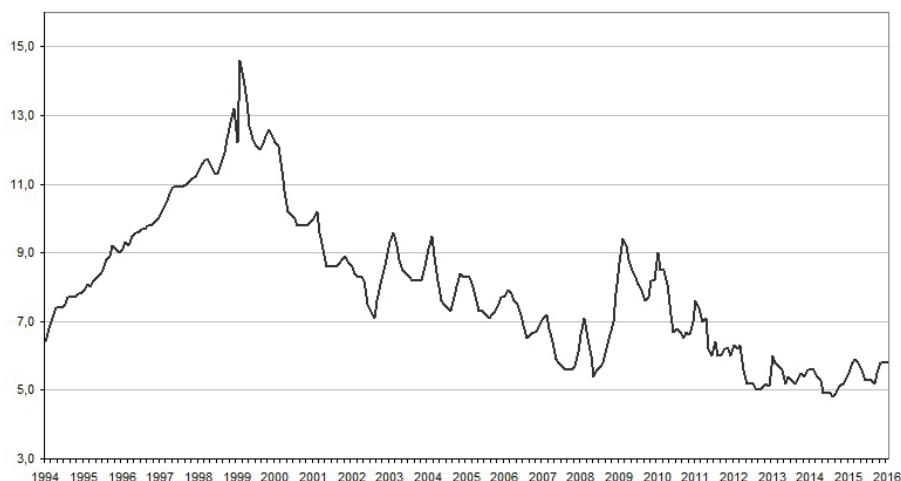


Рис. 7. Уровень безработицы в России (январь 1996 – январь 2016 гг.) [12]

Отметим, что для исследуемых процессов значения автокорреляционных функций убывают достаточно медленно (рис. 8), что может свиде-

тельствовать в пользу моделей с долгой памятью. В случае уровня безработицы наблюдается также слабовыраженная сезонность.

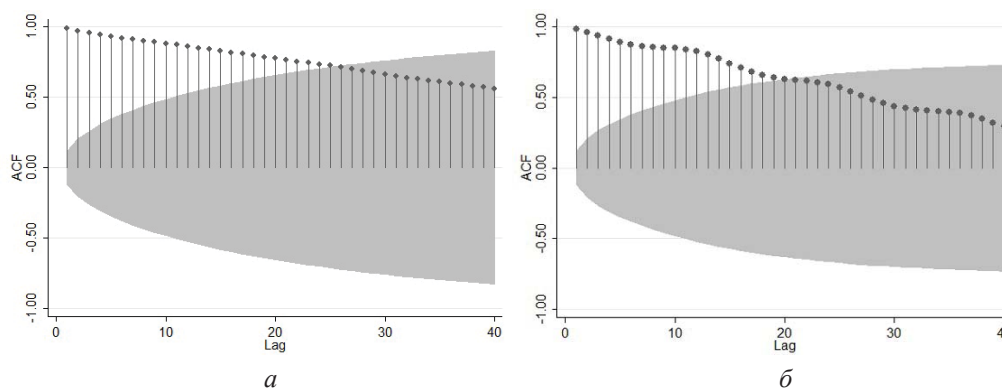


Рис. 8. ACF для общего коэффициента рождаемости во Франции (а) и уровня безработицы в России (б)

В нашем исследовании R/S-анализ показал, что ряд общего коэффициента рождаемости во Франции и уровня безработицы в России не обладают свойствами процессов с длинной памятью. Рассчитанная тестовая L_0 -статистика показала, что на уровне значимости 0,01 гипотеза об отсутствии длинной памяти не отвергается: $L_{01} = 1,59$; $L_{02} = 1,46$ соответственно; значения тестовой статистики обоих показателей лежат вне интервала (0,721, 2,098). Для анализируемых показателей были подобраны обычные ARIMA-модели. Анализ качества оцененных моделей показал, что наиболее удовлетворительной моделью для общего коэффициента рождаемости во Франции является ARIMA(1,0,0) с детерминированным линейным трендом, а для уровня безработицы в России – SARIMA(2,1,0)(1,1,1)₁₂. Остатки полученных моделей обладают свойствами белого шума.

В работе были рассмотрены несколько примеров моделирования демографических процессов с помощью ARIMA-моделей и их обобщений: SARIMA и ARFIMA-моделей. Было показано, что исследуемый в работе демографический процесс – коэффициент суммарной рождаемости удовлетворительно описывается процессом авторегрессии 1-го порядка, несмотря на небольшой объем выборки. В случае сезонности для показателя числа браков, заключенных по месяцам, в России удовлетворительной моделью оказалась SARIMA(0,0,1)(3,1,0)₁₂. Все полученные модели оказались адекватными, а характеристики точности прогноза показали достаточно высокие прогностические свойства полученных моделей. При анализе длинных временных рядов: общего коэффициента рождаемости во Франции ($T = 275$) и уровня безработицы в России ($T = 265$) – при помощи R/S-теста свойств процессов с длинной



памятью выявлено не было, данные процессы удовлетворительно были описаны моделями ARIMA(1,0,0) с детерминированным линейным трендом и SARIMA(2,1,0)(1,1,1)₁₂ соответственно. Развитая в работе статистическая методология анализа и прогнозирования демографических процессов может быть использована как один из способов демографического прогнозирования, которое на сегодняшний день приобретает все большую значимость как во внутренней политике при социально-экономическом планировании развития страны в целом или отдельных субъектов, так и на международной арене.

Список литературы

1. Box G. P., Jenkins G. M. Time Series Analysis Forecasting and Control. San Francisco : Holden-Day, 1970. 554 p.
2. Hiorns R. W. Mathematical Models in Demography // The Structure of Human Populations. Ed. by G. A. Harrison, A. Z. Boycl. Oxford : Clarendon Press. 1972. P. 110–127.
3. Lee R. Forecasting Births in Post-Transition Populations // Journal of the American Statistical Association. 1974. Vol. 69. P. 607–617.
4. Pollard J. H. On Simple Approximate Calculations Appropriate to Populations with Random Growth Rates // Theoretical Population Biology. 1970. Vol. 1. P. 208–218.
5. Saboia J. L. M. Modeling and Forecasting Populations by Time Series – The Swedish Case // Demography. 1974. Vol. 11. P. 483–492.
6. Kashyap R. L., Rao A. R. Dynamic Stochastic Models from Empirical Data. N.Y. ; L. : Academic Press, 1976. 352 p.
7. Saboia J. L. M. Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting // Journal of the American Statistical Association. 1977. Vol. 72 (358). P. 264–270.
8. Pflaumer P. Forecasting US population totals with the Box – Jenkins approach // International Journal of Forecasting. 1992. Vol. 8. P. 329–338.
9. Tiziana T., Vaupel J. W. Forecasting life expectancy in an international context // International Journal of Forecasting. 2012. Vol. 28. P. 519–531.
10. Alho J. M., Spencer B. D. Statistical demography and forecasting. Springer, 2005. 410 p.
11. Booth H. Demographic forecasting : 1980 to 2005 in review // International Journal of Forecasting. 2006. Vol. 22. P. 547–581.
12. Росстат. URL: <http://www.gks.ru> (дата обращения: 22.03.2016).
13. Мхитарян В. С., Архипова М. Ю., Балаш В. А., Балаш О. С., Дуброва Т. А., Сиротин В. П. Эконометрика / под общ. ред. В. С. Мхитаряна. М. : Проспект, 2014. 384 с.
14. Айвазян С. А. Методы эконометрики. М. : ИНФРА-М, 2010. 510 с.
15. Щербакова Е. Число зарегистрированных браков и разводов в 2012 году снизилось // Демоскоп weekly. 2013. № 541–542. URL: <http://demoscope.ru/weekly/2013/0541/barom04.php> (дата обращения: 20.12.2015).
16. Hurst H. E. Long term Storage Capacity of Reservoirs // Transactions of the American Society of Civil Engineers. 1951. Vol. 116. P. 770–799.
17. Mandelbrot B. Statistical methodology for non-periodic cycles : From the covariance to R/S analysis // Annals of Economic and Social Measurement. 1972. Vol. 1. P. 259–290.
18. Lo A. W. Long-term memory in stock market prices // Econometrica. 1991. Vol. 59. P. 1279–1313.
19. Hyndman R. J., Booth H., Ysmeen F. Coherent Mortality Forecasting : The Product-Ratio Method With Functional Time Series Models // Demography. 2013. Vol. 50, iss. 1. P. 261–283.
20. Granger C. W. Long memory relationships and the aggregation of dynamic models // Journal of Econometrics. 1980. Vol. 14, iss. 2. P. 227–238.
21. Демоскоп : [сайт]. URL: http://www.demoscope.ru/weekly/app/long_cbr.php (дата обращения: 22.03.2016).
22. World Bank : [сайт]. URL: <http://www.worldbank.org> (дата обращения: 22.03.2016).

The Statistical Approach to the Analysis and Forecasting of Demographic Data

E. D. Kopnova

National Research University Higher School of Economics,
31, Shabolovka str., Moscow, 115162, Russia
E-mail: ekopnova@hse.ru

L. A. Rodionova

National Research University Higher School of Economics,
31, Shabolovka str., Moscow, 115162, Russia
E-mail: lrodionova@hse.ru

Introduction. Possibilities of application ARIMA-models to analysis and forecasting of demographic time series were considered in the article. Foreign studies had shown that the ARIMA-models give good results for forecasting indicators such as population, birth rates and death rates, life expectancy, along with the traditional demographic methods (cohort-component approach). **Research technique.** Box – Jenkins methodology of the analysis and forecasting of time series, particularly with regard to demographic data: total fertility rate in Russia (1990–2014), the number of marriages by months in Russia (2005–2015), total fertility rate in France (1740–2014) and the unemployment rate in Russia (1996–2016) was used in the work. ARIMA-, ARIMA- and ARIMA-models, depending on the nature of the dynamics of the studied indicators were analyzed. **Results.** The analysis had shown that the estimated ARIMA-models for the total fertility rate and number of marriages were adequate and had good statistical and prognostic properties. Forecasts were built on basis of the obtained models. In the case of long series availability of properties with a long memory processes have not been identified.

Key words: time series analysis, ARIMA-model, SARIMA-model, ARFIMA-model, statistical methods.



References

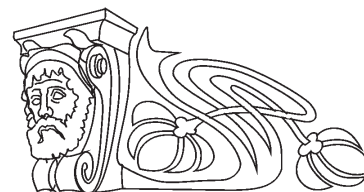
1. Box G. P., Jenkins G. M. *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day, 1970. 554 p.
2. Hiorns R. W. *Mathematical Models in Demography. The Structure of Human Populations*. Ed. by G. A. Harrison, A. Z. Boycl. Oxford, Clarendon Press, 1972, pp. 110–127.
3. Lee R. Forecasting Births in Post-Transition Populations. *Journal of the American Statistical Association*, 1974, vol. 69, pp. 607–617.
4. Pollard J. H. On Simple Approximate Calculations Appropriate to Populations with Random Growth Rates. *Theoretical Population Biology*, 1970, vol. 1, pp. 208–218.
5. Saboia J. L. M. Modeling and Forecasting Populations by Time Series – The Swedish Case. *Demography*, 1974, vol. 11, pp. 483–492.
6. Kashyap R. L., Rao A. R. *Dynamic Stochastic Models from Empirical Data*. New York, London, Academic Press, 1976. 352 p.
7. Saboia J. L. M. Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting. *Journal of the American Statistical Association*, 1977, vol. 72 (358), pp. 264–270.
8. Pflaumer P. Forecasting US population totals with the Box – Jenkins approach. *International Journal of Forecasting*, 1992, vol. 8, pp. 329–338.
9. Tiziana T., Vaupel J.W. Forecasting life expectancy in an international context. *International Journal of Forecasting*, 2012, vol. 28, pp. 519–531.
10. Alho J. M., Spencer B. D. *Statistical demography and forecasting*. Springer, 2005. 410 p.
11. Booth H. Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 2006, vol. 22, pp. 547–581.
12. Rosstat. Available at: <http://www.gks.ru> (accessed 22 Mart 2016).
13. Mhitarjan V. S., Arhipova M. Ju., Balash V. A., Balash O. S., Dubrova T. A., Sirotin V. P. *Ekonometrika [Econometrics]*. Ed. by V. S. Mhitarjan]. Moscow, Prospekt Publ., 2014. 384 p.
14. Ajvazjan S. A. *Metody jekonometriki [Econometrics Methods]*. Moscow, INFRA-M Publ., 2010. 510 p.
15. Scherbakova E. Chislo zaregistrirovannykh brakov i razvodov v 2012 godu snizilos' (The number of registered marriages and divorces decreased in 2012). *Demoskop weekly (Demoscope weekly)*, 2013, № 541–542. Available at: <http://demoscope.ru/weekly/2013/0541/barom04.php> (accessed 20 December 2015).
16. Hurst H. E. Long term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, 1951, vol. 116, pp. 770–799.
17. Mandelbrot B. Statistical methodology for non-periodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement*, 1972, vol. 1, pp. 259–290.
18. Lo A. W. Long-term memory in stock market prices. *Econometrica*, 1991, vol. 59, pp. 1279–1313.
19. Hyndman R. J., Booth H., Yasmee F. Coherent Mortality Forecasting: The Product-Ratio Method With Functional Time Series Models. *Demography*, 2013, vol. 50, iss. 1, pp. 261–283.
20. Granger C. W. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 1980, vol. 14, iss. 2, pp. 227–238.
21. *Demoskop*. Site. Available at: http://www.demoscope.ru/weekly/app/long_cbr.php (accessed 22 Mart 2016).
22. *World Bank*. Site. Available at: <http://www.worldbank.org> (accessed 22 Mart 2016).

УДК 336.225.66 (470.44)

К ВОПРОСУ О ЛЬГОТНОМ РЕЖИМЕ НАЛОГООБЛОЖЕНИЯ ДЛЯ СУБЪЕКТОВ МАЛОГО ПРЕДПРИНИМАТЕЛЬСТВА В САРАТОВСКОЙ ОБЛАСТИ

М. В. Голубниченко

кандидат экономических наук, доцент кафедры финансов и кредита,
Саратовский национальный исследовательский государственный
университет имени Н. Г. Чернышевского
E-mail: mvl04@mail.ru



Введение. В условиях экономического кризиса в России остро ощущается необходимость государственной поддержки субъектов малого предпринимательства, особенно в виде снижения налоговой нагрузки на предприятия на уровне региона. Необходимо проанализировать особенности льготного режима налогообложения для субъектов малого предпринимательства в Саратовской области и его влияние на развитие малого бизнеса в последние годы. Этим определяется актуальность исследования. **Теоретический анализ.** В Саратовской области

льготный налоговый режим для субъектов малого предпринимательства представлен в ряде региональных законов. В частности, установлены дифференцированные налоговые ставки при применении упрощенной системы налогообложения для отдельных категорий налогоплательщиков, введены двухлетние «налоговые каникулы» по 11 видам деятельности при патентной системе налогообложения и 72 видам деятельности при упрощенной системе налогообложения. С 1 января 2016 г. снижена стоимость патентов по отдельным видам деятельности, а также