

СТАТИСТИЧЕСКИЕ ПОДХОДЫ К АНАЛИЗУ И МОДЕЛИРОВАНИЮ СЕЗОННОСТИ В ДЕМОГРАФИЧЕСКИХ ДАННЫХ

Лилия Родионова, Елена Копнова

Согласно майскому указу Президента (2018 г.), в число национальных целей и стратегических задач развития РФ на период до 2024 г. входят «обеспечение устойчивого естественного роста численности населения РФ и повышение ожидаемой продолжительности жизни до 78 лет». Возросшая необходимость мониторинга текущей демографической ситуации, изучение структуры демографических показателей, пристальное внимание научного сообщества к реализации национальных целей обусловили выбор темы настоящего исследования.

В работе исследовались проблемы моделирования сезонности демографических показателей РФ (числа рождений, числа умерших, младенческой смертности, числа заключенных браков) по ежемесячным данным Росстата за период 2007-2018 гг. Зарубежные исследования показали, что, наряду с традиционными демографическими методами, ARIMA-модели дают хорошие результаты при прогнозировании демографических показателей (численности населения, уровня рождаемости и смертности, продолжительности жизни населения). Использование статистического подхода на основе SARIMA-моделей в данной работе позволило получить адекватные модели с хорошими статистическими и прогностическими свойствами. Стационарность процессов с учетом сезонности анализировали на основе HEGY-теста. Исследуемые в работе показатели имели ряд особенностей, которые были учтены при моделировании. Ряды числа рождений и числа умерших имели второй и первый порядки интегрируемости соответственно и содержали детерминированную сезонность, ряд числа заключенных браков имел первый порядок обычной и сезонной интегрируемости, ряд младенческой смертности не содержал сезонность, что было подтверждено на основе анализа автокорреляционной функции и периодограммы. Для анализируемых показателей в работе были построены точечные и интервальные оценки прогноза на 2019 г. Для сравнения качества прогнозирования SARIMA-моделей в работе были оценены также сезонные модели Хольта-Уинтерса.

Ключевые слова: демографические данные, прогнозирование, анализ временных рядов, ARIMA-модели, SARIMA-модели, HEGY-тест, статистические методы.

ВВЕДЕНИЕ

Прогнозирование в демографии является неотъемлемой частью анализа и прогнозирования социально-экономических процессов в обществе в целом.

Лилия Анатольевна Родионова (lrوديонova@hse.ru), Национальный исследовательский университет «Высшая школа экономики», Россия.

Елена Дмитриевна Копнова (ekopnova@hse.ru), Национальный исследовательский университет «Высшая школа экономики», Россия.

Статья поступила в редакцию в марте 2019 г.

Необходимость построения прогноза изменения численности населения, предвидение 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, пенсионной системы, при принятии важных политических решений, что особенно актуально ввиду пристального внимания сообщества к реализации майских указов президента (2018 г.) относительно роста численности населения и повышения ожидаемой продолжительности жизни населения в Российской Федерации¹.

Основной целью данной работы является знакомство широкой аудитории читателей с существующей статистической методологией прогнозирования, а также развитие статистической методологии анализа и прогнозирования сезонности в демографических данных. В своем исследовании подробнее остановимся на применении SARIMA-моделей к демографическим данным России и продемонстрируем возможности статистического инструментария прогнозирования с использованием данного класса моделей.

ОБЗОР ЛИТЕРАТУРЫ

Исторический аспект проблемы. Вопрос сезонности демографических показателей отражен в исследованиях с середины XIX века. Большая часть ранних исследований была посвящена изучению сезонности смертности (включая младенческую смертность), числа браков, количества рождений. Ф. Гиляровский заметил, что пики максимума и минимума числа рождений в Новгородской губернии приходились на определенные месяцы года, наибольшее и наименьшее число детей умирало в определенные «стадии младенчества», вне зависимости от месяца года (Гиляровский 1866). С. Ершов исследовал сезонные колебания браков, рождений и смертей и выявил значительную связь между этими показателями. Также были отмечены межнациональные различия в детской смертности: в России в XIX веке детская смертность в 2 раза превышала уровень этого показателя в Татарстане (Ершов 1888). В более поздних работах по историческим данным отмечается, что цикличность рождаемости вызвана цикличностью брачности. Сезонность браков объяснялась религиозными и институциональными запретами. Пики максимального и минимального числа зарегистрированных браков совпадали с пиками рождения первых детей в семьях и приходились на осенние месяцы, которые были наиболее благоприятными для выживаемости младенцев. Цикличность рождения последующих детей была менее выражена и не зависела от брачности. Сезонность младенческой смертности авторы объясняли природно-климатическими факторами, вызывающими всплеск желудочно-кишечных инфекций в летний период (Авдеев, Блум, Троицкая 2002; Винник 2012).

Сезонность демографических данных в современной России. В современной России также наблюдается сезонность демографических показателей: рождаемости, смертности, заболеваемости, миграции, числа браков и разводов. Е. Кваша отмечает, что в XX веке

¹ Указ Президента Российской Федерации от 7 мая 2018 года № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года».

уровень *младенческой смертности* в России снизился в 20 раз и произошло ослабление ее зависимости от времени года. В XIX – начале XX века заметное повышение младенческой смертности приходилось на летние месяцы из-за желудочных заболеваний и на зимние месяцы из-за простудных болезней. С развитием медицины в экономических развитых странах с 60-х годов наблюдалась зимняя сезонность младенческой смертности или ее отсутствие вообще (Кваша 2003). Максимальные темпы снижения младенческой смертности в России наблюдались в последние десятилетия. За 1990–2000 гг. ее показатель снизился на 12%, за 2000–2012 гг. – на 43,8%. Авторы отмечают, что, несмотря на серьезные социально-политические потрясения и рост смертности населения в целом, в 1990-е годы показатель младенческой смертности продолжал снижаться (Баранов и др. 2014). В современной России летняя и зимняя сезонность в младенческой смертности обусловлена колебаниями в постнеонатальный период, на который приходится большая часть умерших от экзогенных причин (например, от пневмоний) (Кваша 2003). По данным за январь-декабрь 2017 г. наибольшее число умерших в возрасте до года приходилось на май (863 человека), наименьшее – на февраль (748 человек), наблюдалось снижение помесечных значений в 2017 г. по сравнению с предшествующими годами (Щербакова 2018b).

При анализе *смертности* демографы отмечают, что тенденция снижения числа умерших преобладала в 2004–2016 гг., ранее наблюдался рост числа умерших и общего коэффициента смертности (в 1992–1994 гг., затем спад 1995–1998 гг., а затем возобновление роста). Наибольшее число смертей приходится на зимне-весенние месяцы, чаще на январь, наименьшее – на летне-осенние месяцы (Щербакова 2016). Смертность от основных классов причин смерти также подвержена сезонному фактору. Показатель смертности от болезней органов дыхания и болезней системы кровообращения повышается на фоне перенесенных острых респираторных заболеваний, и для него характерна сезонная зависимость – больше всего умерших регистрируется в зимние и весенние месяцы, меньше всего – в июле-сентябре. По данным 2001–2017 гг. наибольшее число умерших в результате самоубийств регистрируется в мае (минимум – в июне или июле), от случайных отравлений алкоголем – чаще всего в январе (минимум – июль-август), на август-октябрь приходится пик смертности от транспортных несчастных случаев (Щербакова 2019).

Отмечаются также и сезонные характеристики *рождаемости*, хотя число рождений мало подвержено ярко выраженной сезонности, но в течение года наблюдаются пики и спады числа рождений. В 1990-е годы наибольшее число рождений отмечалось в январе и марте, в 2000-е годы – в летние месяцы, а меньше всего в 1990-е и 2000-е годы – в октябре-декабре. В 2017 г. наибольшее число рождений приходилось на август (159,6 тыс.), спад – в апреле (123,7 тыс.) (Щербакова 2018a).

Е. Щербакова исследует сезонность *числа зарегистрированных браков* в период 1995–2017 гг. В 2017 г. наибольшее число браков приходилось на июль (148 тыс.), август (137 тыс.) и сентябрь (123 тыс.), тогда как наименьшее число заключенных браков было в мае (40 тыс.) и январе (50 тыс.). Динамика регистрации браков по месяцам сохраняет общие тенденции, оставшиеся с XIX века. Автор объясняет данную сезонность традициями, сложившимися в связи с религиозными и культурными особенностями, а также, возможно, особенностью климата нашей страны. Прослеживается сезонность и в числе разводов:

ежегодно наименьшее число разводов приходится на такие месяцы, как январь, февраль и май, а наибольшее – на март, октябрь или декабрь, однако количество разводов гораздо меньше зависит от сезонного фактора, чем регистрация браков. В 2017 г. наименьшее число разводов было зарегистрировано в феврале (43 тыс.), наибольшее – в марте (55 тыс.) (Щербакова 2018а).

Сезонные эффекты наблюдаются в *миграции* населения. Так, например, при изучении маятниковой миграции в Московском регионе по данным РЖД было выявлено, что сезонные колебания объема пассажиропотока между февралем (наименьшее число трудовых мигрантов) и июлем, в который наблюдается наибольшее их число, составляет приблизительно 2,3 млн человек (Махрова, Бочкарев 2017).

Статистическое моделирование сезонности в демографии. В демографическом прогнозировании современные исследователи используют различные подходы: математические методы, метод передвижки возрастов (когортно-компонентный метод), статистические методы. Статистический подход ARIMA-моделирования (*англ. autoregressive integrated moving average model*) был предложен Боксом и Дженкинсом в 1970 г. для анализа экономических процессов (Box, Jenkins 1970: 1-231). Однако вскоре ARIMA-модели стали использовать во многих областях науки, в том числе и демографии. В работах 70-х годов демографы моделировали численность населения, уровень рождаемости и смертности (Hiorns 1972; Lee 1974; Pollard 1970; Saboia 1974). Для прогнозирования численности населения США за 1900-1971 гг. использовали ARIMA(1,1,0) в логарифмах (Kashyap, Rao 1976), для анализа уровня рождаемости в Норвегии за период 1919-1974 гг. – ARIMA(4,1,1) и ARIMA(3,1,2) (Saboia 1977). Ряд авторов использовали ARIMA-модели для прогнозирования средней продолжительности жизни мужчин и женщин в Италии и США по данным за период 1900-2006 гг. (ARIMA(2,1,1) для женщин и ARIMA(1,1,1) для мужчин) (Torri, Vaupel 2012; Alho, Spencer 2005; Booth 2006).

Дальнейшим развитием методологии было обобщение моделей на случай сезонных данных - так называемые SARIMA-модели. На сегодняшний день данный класс моделей активно используется исследователями для прогнозирования и лежит в основе алгоритмов сезонной корректировки X-12-ARIMA (Findley et al. 1998), созданных в Бюро переписей США, и метода TRAMO/SEATS (Gomez, Maravall 1996), разработанного Евростатом и Национальным банком Бельгии. Приведем несколько примеров исследований, где используются статистическое моделирование сезонности, в частности SARIMA-модели и их модификации.

Сезонность смертей по возрасту и причинам в США анализировали по данным 1994-1998 гг. Было выявлено, что степень сезонности смертей снизилась с течением времени для молодых когорт и увеличилась для старших когорт (Feinstein 2002). Влияние сезонных изменений на болезни и причины смерти в Японии по ежемесячным данным с 1970 по 1999 г. оценивали на основе модели с гармоническими сезонными составляющими (Nakaji et al. 2004). Аналогичные модели использовали при анализе связи между летней температурой и смертностью среди пожилых людей (в возрасте 65 лет и старше) с предыдущим уровнем зимней смертности по данным Рима в 1987–2005 гг. (Stafoggia et al.

2009). По данным Аландских островов (Финляндия) за 1650-1950 гг. наблюдалась заметно уменьшающаяся тенденция в сезонных колебаниях рождаемости, но все же было выделено два пика (в марте-апреле и в сентябре-октябре), а за период 1901-1950 гг. сезонные колебания практически исчезли (Eriksson et al. 2008). По данным Испании за период 1941-2000 гг. методами спектрального анализа Фурье было выявлено, что сезонность числа рождений изменилась после 1960-х годов с уменьшением амплитуды и последующей потерей сезонности в 1990-х, сезонность рождений отсутствовала в 1991-2000 гг. (Sancho-Candela et al. 2007). Большое распространение SARIMA-модели получили при анализе заболеваемости: исследование сезонных эффектов смертности от диареи у детей в возрасте до 5 лет в Бразилии (штате Рио-де-Жанейро) по ежемесячным данным с 1980 по 1998 г. (Kale, Andreozzi, Nobre 2004), взаимосвязи между климатическими колебаниями и инфекцией сальмонеллы в Аделаиде (Австралия) (Zhang, Bi, Hiller 2008), взаимосвязи применения антибиотиков на устойчивость к антибиотикам с учетом сезонных эффектов по данным США 1999 -2007 (Sun, Klein, Laxminarayan 2012) и др.

МЕТОДИКА ИССЛЕДОВАНИЯ И ДАННЫЕ

Теоретические предпосылки моделей. В данном разделе введем ряд определений и обозначений. Временной ряд наблюдений y_1, y_2, \dots, y_n некоторого демографического показателя, в частности, числа родившихся в России, по месяцам (тыс. человек) за 2007-2018 гг. (рисунок 1), рассматривается как реализация случайных величин, которые описываются некоторым стохастическим процессом. Согласно методологии Бокса-Дженкинса моделирование и прогнозирование на основе ARIMA/SARIMA-модели включает в себя несколько этапов.

Этап 1. Определение стационарности² временных рядов. Многие демографические процессы не являются стационарными, их элементы формируются под воздействием различных факторов четырех типов (Айвазян 2010: 293-409). Долговременные факторы формируют общую тенденцию изменения анализируемого показателя y_t . Например, ряд числа родившихся по месяцам (рисунок 1) за период 2007-2018 гг. визуально содержит параболический тренд, а ряд числа умерших по месяцам в России (рисунок 2) – линейный, убывающий тренд. Отметим, что в дальнейшем анализе для правильной спецификации модели важно различать два типа тренда: детерминированный и стохастический (Канторович 2002).

Сезонные факторы формируют периодические повторяющиеся в определенное время года колебания анализируемого показателя. На рисунках 1-2 помимо трендовых составляющих наблюдаются сезонные компоненты. Как уже было отмечено, рождаемость и смертность в России подвержены влиянию сезонного фактора. Заметим, что демографические процессы не всегда содержат ярко выраженные трендовые компоненты, так, например, число зарегистрированных браков в России (тыс.) по месяцам регистрации

² Случайный процесс называют *слабо стационарным*, если среднее, дисперсия и ковариация y_t не зависят от времени t : $E(y_t) = \mu < \infty$, $V(y_t) = \gamma_0$, $\text{cov}(y_t, y_{t-k}) = \gamma_k$.

(рисунок 3) содержит сезонную компоненту, но визуально сложно сделать вывод о существовании устойчивой тенденции, однако заметно снижение показателя после 2015 г. Анализ периода сезонности возможен на основе графика сезонной волны, значений автокорреляционной функции (Кендалл, Стюарт 1976). Периодическая зависимость может быть определена как корреляционная зависимость порядка k между i -м элементом ряда и $(i+k)$ -м элементом ряда, которая описывается корреляционной функцией $\rho(k)$ в зависимости от лага k^3 . Наряду с «обычной» $\rho(k)$ (ACF) рассматривается частная автокорреляционная функция (PACF)⁴, характеризующая «чистую корреляцию» между y_t и y_{t-k} при исключении влияния промежуточных значений $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ (по аналогии с частной корреляцией в статистике). *Коррелограммы* (графики ACF и PACF) используют для выявления сезонности. На графиках ACF и PACF для сезонных ежемесячных данных, например, наблюдаются пики на «сезонных» лагах: 12, 24, 36 и т.д. Содержательно это означает корреляцию на лагах через 12, 24, 36 месяцев. Анализ коррелограмм будет приведен при анализе стационарности рядов.

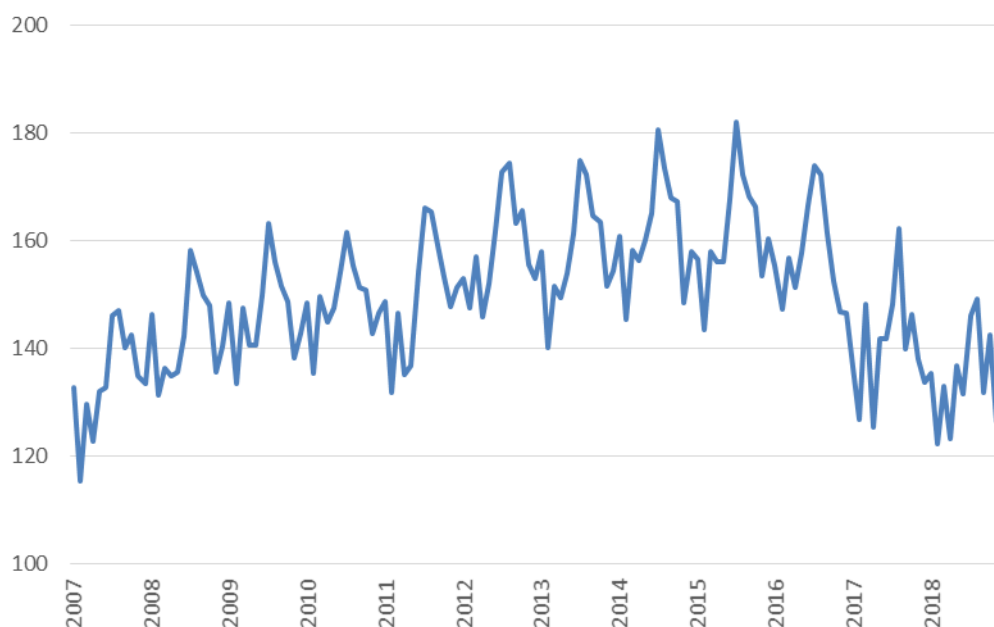


Рисунок 1. Число родившихся в России по месяцам рождений, тыс.

Источник: ЕМИСС – URL: <https://fedstat.ru/indicator/33555> (дата обращения: 01.02.2019).

³ Формула вычисления $\rho(k)$ приведена в Приложении 1.

⁴ ACF - autocorrelation function, PACF - partial autocorrelation function.

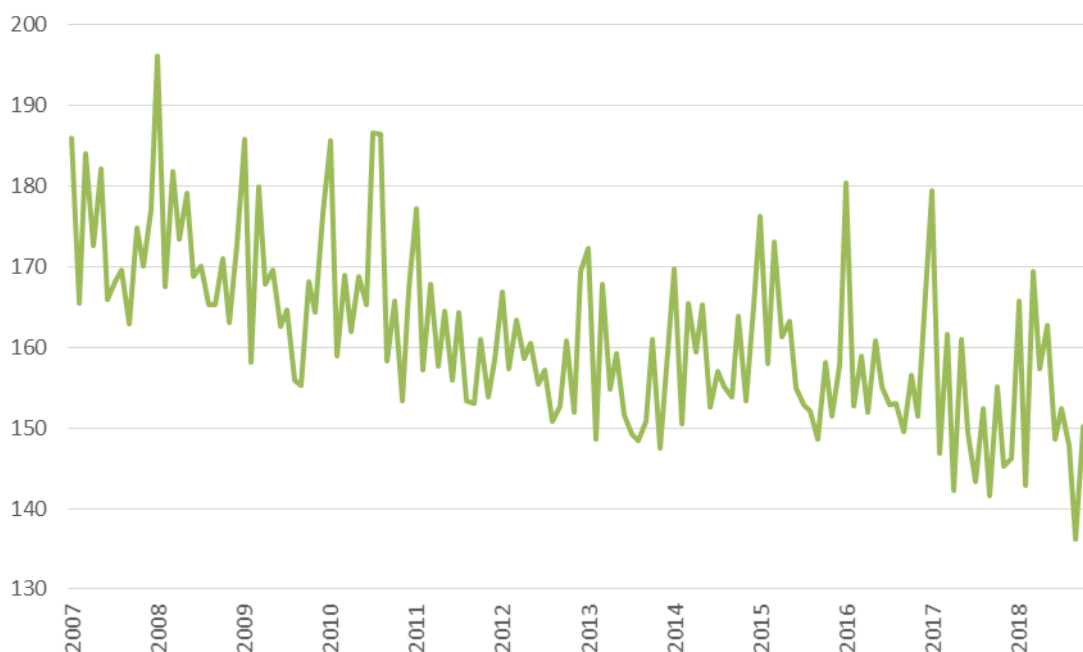


Рисунок 2. Число умерших в России по месяцам, тыс.

Источник: ЕМИСС – URL: <https://fedstat.ru/indicator/33556> (дата обращения: 01.02.2019).

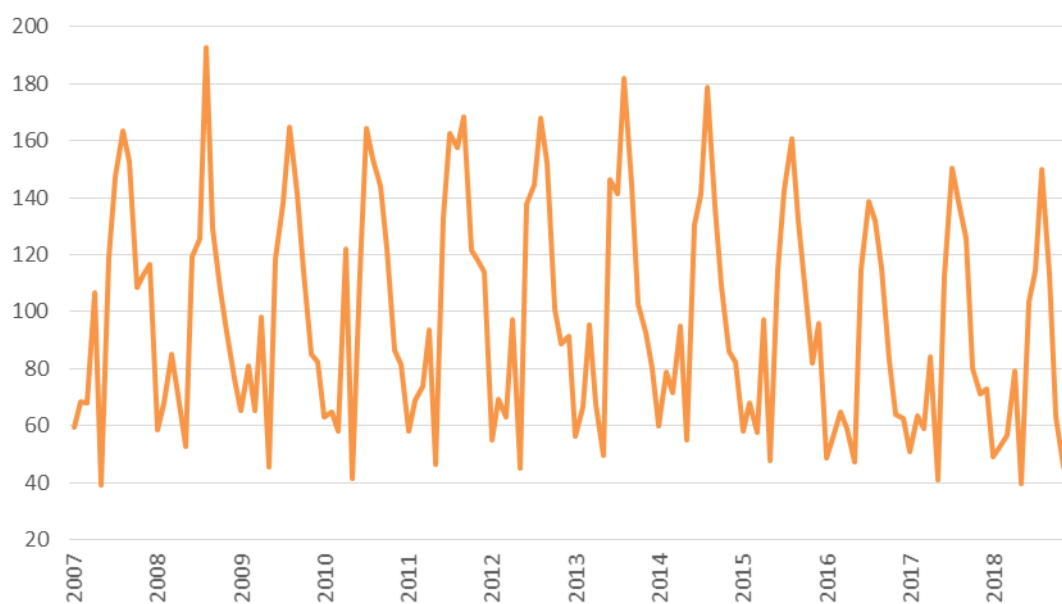


Рисунок 3. Число зарегистрированных браков в России по месяцам регистрации, тыс.

Источник: ЕМИСС – URL: <https://fedstat.ru/indicator/33553> (дата обращения: 01.02.2019).

Циклические факторы формируют изменения анализируемого признака, обусловленные действием долговременных циклов (в частности, демографических). 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 своеобразным индикатором нестационарности ряда. Во-вторых, анализ графиков ACF и PACF: для стационарных временных рядов коррелограммы «быстро убывают» после нескольких первых значений лагов (Айвазян 2010: 293–409). В-третьих, использование статистических тестов, так называемых тестов *единичного корня*⁵, например, расширенного ADF-теста Дики-Фуллера (Dickey, Fuller 1979)⁶, PP-теста Филлипса-Перрона (Phillips, Perron 1988), KPSS-теста Квятковского-Филлипса-Шмидта-Шина (Kwiatkowski et al. 1992). Содержательно в случае ADF-теста и PP-теста нулевая гипотеза состоит в предположении нестационарности ряда. Альтернативной гипотезой является предположение, что ряд стационарен (с константой и/или трендом). Для KPSS-теста – наоборот: нулевой гипотезой является стационарность ряда (отсутствие единичного корня в характеристическом уравнении).

Этап 2. Идентификация и оценивание модели. Модель авторегрессии и скользящего среднего ARMA(p, q) (*Autoregressive Moving Average*) в общем виде имеет вид (1).

$$\alpha_p(L)y_t = \theta_q(L)\varepsilon_t, \quad (1)$$

где ε_t – белый шум – процесс, обладающий свойствами: математическое ожидание и ковариации равны нулю, дисперсия постоянна: $\gamma_0 = \sigma^2$;

$Ly_t = y_{t-1}$, $L^k y_t = y_{t-k}$, – лаговый оператор,

$$\alpha_p(L) = 1 - \alpha_1 L - \alpha_2 L^2 \dots - \alpha_p L^p, \quad \theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 \dots + \theta_q L^q.$$

При моделировании демографических явлений исследователи сталкиваются с нестационарными рядами, которые, могут быть приведены к стационарному виду при помощи взятия последовательной разности. По определению первая разность это $\Delta y_t = y_t - y_{t-1}$. Содержательно первая разность – это переход от исходного показателя к его приростам. 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 назвать *интегрированным* процессом авторегрессии и скользящего среднего ARIMA(p, d, q). Такой процесс в общем случае имеет вид (2).

$$\alpha_p(L)\Delta^d y_t = \theta_q(L)\varepsilon_t. \quad (2)$$

Отметим, что на практике встречаются временные ряды с порядком интегрируемости d не более двух, т. е. ряды становятся стационарными при переходе к первым и вторым разностям (приростам). Параметры ARIMA-моделей оцениваются методом максимального правдоподобия.

⁵ Определение представлено в Приложении 1.

⁶ Формальное описание теста представлено в Приложении 1.

Этап 3. Оценка адекватности модели. Понятие «адекватности» модели подразумевает выполнение предпосылок ARIMA/SARIMA-модели (анализ корней соответствующих характеристических уравнений⁷), анализ остатков модели на наличие автокорреляции и соответствие нормальному закону распределения. Для проверки адекватности полученных ARIMA-моделей и их сравнения используют несколько критериев: оценки коэффициентов модели должны быть статистически значимыми, ряд остатков модели должен быть реализацией процесса белого шума (что позволяет сделать анализ ACF: для белого шума $\rho_k=0$ для любого k , критерий Льюинга-Бокса). 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).

Этап 4. Прогнозирование. Прогнозные значения показателей определялись как условные математические ожидания модельных процессов с максимальным горизонтом прогноза, равным 12 месяцам. Оптимальную модель для прогнозирования выбирали на основе анализа ее прогностических возможностей с применением расчета показателей ошибки, характеризующих точность прогноза, представленных в таблице 1 (Турунцева 2011). MAPE и MPE являются абсолютными мерами качества прогноза: прогноз считается хорошим, если значения меньше 10%. RMSE и MAE являются относительными мерами, зависят от единиц измерения показателя. RMSE наиболее часто используется для сравнения прогнозов по разным моделям.

Таблица 1. Характеристики точности прогноза

Характеристика точности прогноза	Формула вычисления
Средняя квадратичная ошибка, RMSE (Root Mean Squared Error)	$RMSE = \sqrt{\frac{\sum_{i=1}^h (\hat{y}_{T,i} - y_{T+i})^2}{h}}$
Средняя абсолютная процентная ошибка, MAPE (Mean Absolute Percent Error)	$MAPE = \frac{1}{h} \sum_{i=1}^h \left \frac{\hat{y}_{T,i} - y_{T+i}}{y_{T+i}} \right \cdot 100\%$
Средняя процентная ошибка, MPE (Mean Percent Error)	$MPE = \frac{1}{h} \sum_{i=1}^h \frac{y_{T+i} - \hat{y}_{T,i}}{y_{T+i}} \cdot 100\%$
Средняя абсолютная ошибка, MAE (Mean Absolute Error)	$MAE = \frac{1}{h} \sum_{i=1}^h \hat{y}_{T,i} - y_{T+i} $

*Примечание: $\hat{y}_{T,i}$ - предсказанные значения ряда в момент времени T на i шагов вперед,
 y_t - наблюдаемые значения ряда в момент времени $T+i$, h -горизонт прогнозирования.*

Отметим, что для вычисления характеристик точности прогноза исходная выборка ($T = 155$ наблюдений) была разделена на две части: обучающую (первые 125 наблюдений) и тестовую (последние 30 наблюдений). Для тестовой выборки обычно берут последние наблюдения, составляющие 20% исходной выборки (Hyndman, Koehler 2006). Характеристики качества прогноза рассчитывали по тестовой выборке.

⁷ Определение дано в Приложении 1.

Для сравнения качества прогнозов использовали сезонные модели Хольта-Винтерса, основанные на экспоненциальном сглаживании (Holt-Winters Model - HW) (Holt 2004). Вид модели HW и оцененные параметры сглаживания для каждого из показателей приведены в Приложении 1.

Статистическую значимость различия прогнозов для разных моделей также определяли с использованием теста Диболда-Мариано (Diebold, Mariano 1995).

SARIMA. 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), где 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 (3).

$$\alpha_p(L)\alpha_{P_s}(L)\Delta^d\Delta_s^D y_t = \theta_q(L)\theta_{Q_s}(L)\varepsilon_t, \quad (3)$$

где $\alpha_{P_s}(L) = 1 - \alpha_{s1}L^s - \alpha_{s2}L^{2s} \dots - \alpha_{sP}L^{P_s}$, $\theta_{Q_s}(L) = 1 + \theta_{s1}L^s + \theta_{s2}L^{2s} \dots + \theta_{sQ}L^{Q_s}$.

Этапы анализа SARIMA-моделей аналогичны анализу ARIMA-моделей, однако есть ряд особенностей, связанных с определением порядка сезонной разности D_s. Для определения порядка сезонной интегрируемости используют специальные тесты сезонных единичных корней. Отметим, что обычные тесты единичного корня (ADF-тест, PP-тест, KPSS-тест) в случае сезонности неприменимы, как будет показано, однако существуют модификации данных тестов на случай сезонной интегрированности, например тест DHF (Dickey, Hansa, Fuller 1984). Одним из современных подходов тестирования наличия сезонных и несезонных единичных корней является тест HEGY (Hylleberg et al. 1990), который был применен в нашем исследовании.

Исходные данные и анализ стационарности временных рядов. Введем обозначения для дальнейшего представления результатов:

- birth - ряд числа родившихся в России, по месяцам (тыс. человек);
- mort - ряд числа умерших по месяцам в России (тыс. человек);
- mar – ряд числа зарегистрированных браков в России (тыс.) по месяцам регистрации;
- inf – ряд числа зарегистрированных умерших детей на 1-ом году жизни (человек).

Выбор временного интервала для исследования (январь 2007 г. – ноябрь 2018 г., длина временного ряда T = 155 наблюдений) обусловлен ограничением публикуемых сезонных данных по выбранным показателям на сайте ЕМИСС на момент исследования.

Стационарность рядов первоначально анализировали на основании трех тестов: расширенного ADF-теста, PP-теста, KPSS-теста для исходных данных в различных спецификациях (с константой; константой и трендом; константой, трендом и сезонными фиктивными переменными). Результаты формальных тестов показали, что нельзя сделать однозначного вывода об интегрируемости исследуемых рядов, так как наличие сезонности в данных значительно влияет на свойства рядов и искажает результаты тестирования

стационарности рядов с применением стандартных тестов (ADF, PP, KPSS), по своему содержанию не учитывающих возможность существования случайной сезонности, которую нельзя описать с помощью детерминированных составляющих (например, с помощью сезонных дамми-переменных).

Ряд birth. График временного ряда *birth* представлен на рисунке 1. Наблюдается ярко выраженный параболический тренд и сезонность, что явно свидетельствуют о нестационарности ряда.

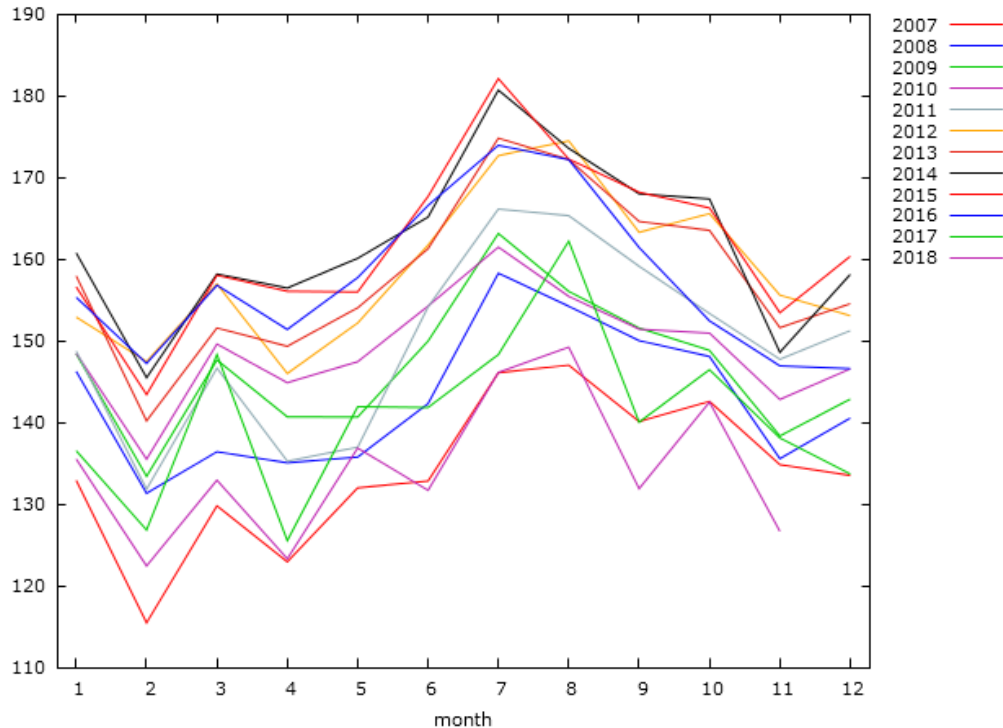


Рисунок 4. Число зарегистрированных родившихся в России по месяцам рождений: сравнение по годам, тыс.

Источник: Расчеты авторов.

Ярко выраженные пики числа родившихся с 2007 г. приходятся, в среднем, на июль, август, октябрь, спады наблюдаются в феврале, апреле (рисунок 4). Анализ коррелограммы ряда *birth* (рисунок П2.1 Приложения 2) показывает наличие тренда (убывание ACF на первых лагах) и сезонности (пики в ACF на лагах, кратные периоду сезонности ($k = 12, 24, 36$), с затуханием).

Результаты расширенного теста Дикки-Фуллера (ADF), теста Филлипса-Перрона (PP), теста Квятковского-Филлипса-Шмидта-Шина (KPSS), проведенных для показателя *birth* в уровнях, для первой и второй разности, представлены в Приложении 3 и свидетельствуют о противоречивости выводов. На 5%-ном уровне значимости можно утверждать, согласно ADF-тесту, что нулевая гипотеза единичного корня ряда в уровнях не отвергается для всех спецификаций. Это свидетельствует о нестационарности ряда и отсутствии линейного или параболического детерминированного тренда. Аналогичный результат получен и в случае KPSS-теста, где нулевой гипотезой является стационарность исследуемого процесса и нулевая гипотеза для трех спецификаций отвергается. При этом

результаты PP-теста говорят в пользу стационарности ряда (нулевая гипотеза единичного корня отклоняется). При анализе результатов ADF-теста для первой разности ряда birth можно заметить наличие детерминированного линейного тренда (отклонение нулевой гипотезы на 5%-ном уровне значимости для спецификации 3). Таким образом, можно предположить, что порядок интегрируемости ряда birth $d = 2$ либо $d = 1$ и в дальнейшие модели необходимо включать линейный детерминированный тренд. Однако ввиду противоречивости полученных результатов также необходимо исследовать сезонную интегрируемость ряда.

Ряд mort. График временного ряда mort представлен на рисунке 2. Как было отмечено выше, в данных наблюдаются убывающий линейный тренд и сезонность, что также свидетельствует о нестационарности ряда.

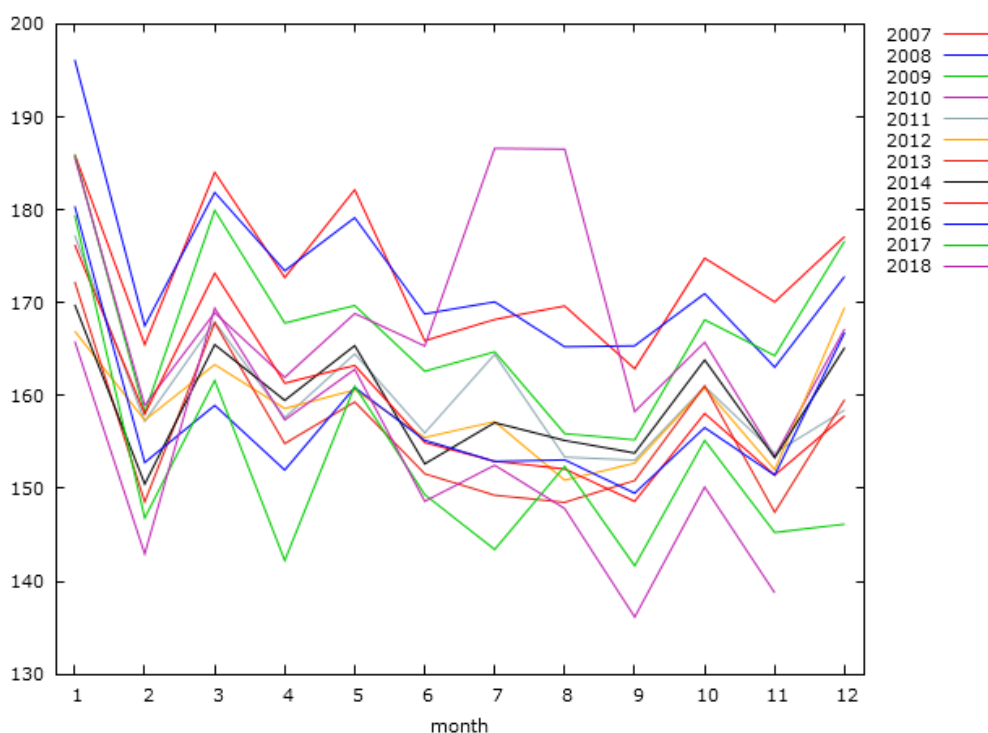


Рисунок 5. Число умерших в России по месяцам: сравнение по годам, тыс.

Источник: Расчеты авторов.

На рисунке 5 показано, что наибольшее число умерших приходится на январь, март, спад приходится на летне-осенний период (июль – октябрь), что соответствует многолетней тенденции. Исключением стал 2010 г., когда наблюдалось значительное превышение показателя в июле-августе по сравнению со среднегодовым значением, что было связано с экстремальными природно-климатическими условиями и большим количеством пожаров на территории России (Щербакова 2017).

Анализ коррелограммы ряда mort (рисунок П2.2 Приложения 2) показывает наличие тренда (убывание ACF на первых лагах) и сезонности (пики в ACF на лагах, кратные периоду сезонности ($k = 12, 24, 36$) с затуханием).

Результаты тестирования ряда mort на стационарность приведены в Приложении 3. Из-за сезонности в данных опять были получены противоречивые результаты: ADF и KPSS

тесты указывают на нестационарность процесса, PP-тест свидетельствует в пользу его стационарности (нулевая гипотеза единичного корня отклоняется при $\alpha = 0,01$).

Ряд mar. График временного ряда *mar* представлен на рисунке 3. Визуальный анализ графика ряда показывает отсутствие тренда, однако наблюдается ярко выраженная сезонность с пиками в июле-сентябре и резкими спадами в мае на протяжении всего изучаемого периода (рисунок 6).

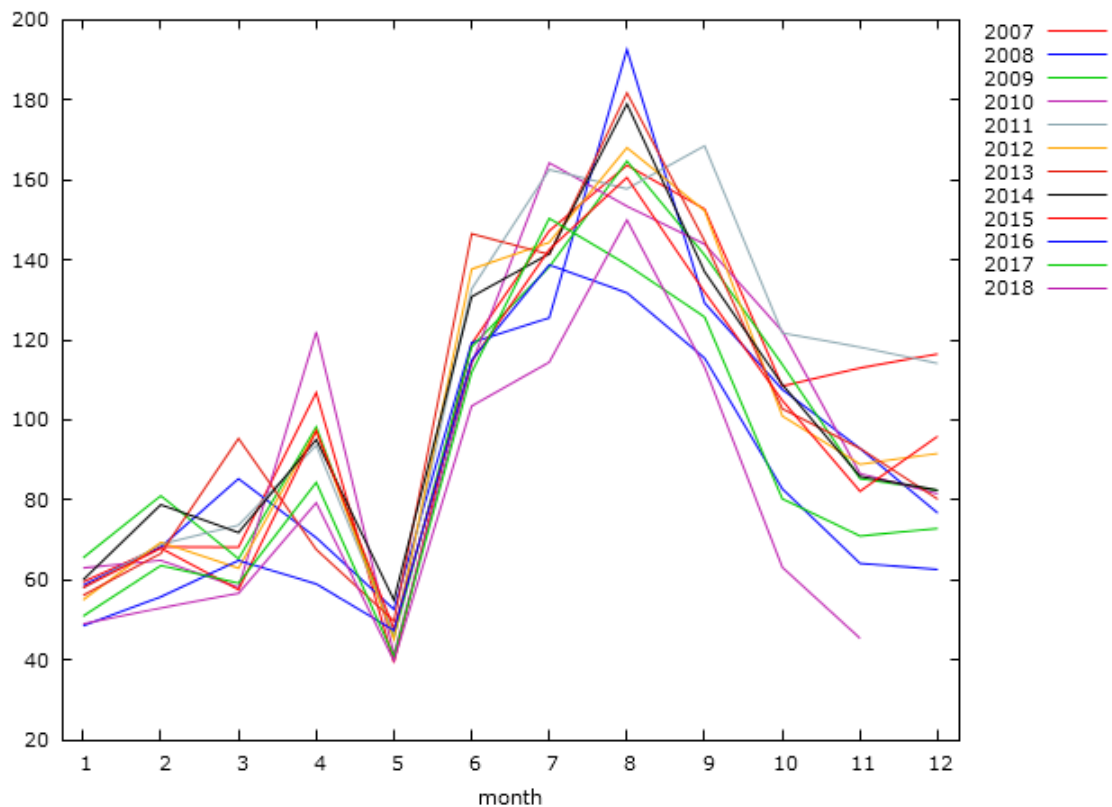


Рисунок 6. Число зарегистрированных в России браков по месяцам регистрации: сравнение по годам, тыс.

Источник: Расчеты авторов.

Анализ коррелограммы ряда *mar* (рисунок П2.3 Приложения 2) также позволил сделать предположение о периоде сезонности $S = 12$ – пики в ACF на лагах, кратные периоду сезонности ($k = 12, 24, 36$), и значения автокорреляции на данных лагах со временем не затухают в отличие от графиков ACF для рядов *birth* и *mort*.

Результаты тестирования ряда *mar* на стационарность приведены в Приложении 3. Из-за сезонности в данных нельзя сделать однозначный вывод о стационарности процесса.

Ряд inf. График временного ряда *inf* представлен на рисунке 7. Начиная с 2013 г., наблюдается устойчивый резко убывающий тренд.

Несмотря на то, что в данных можно выделить месяцы с высокой и низкой младенческой смертностью (рисунок 8), все же надо отметить, что данный показатель не подвержен сезонному фактору.

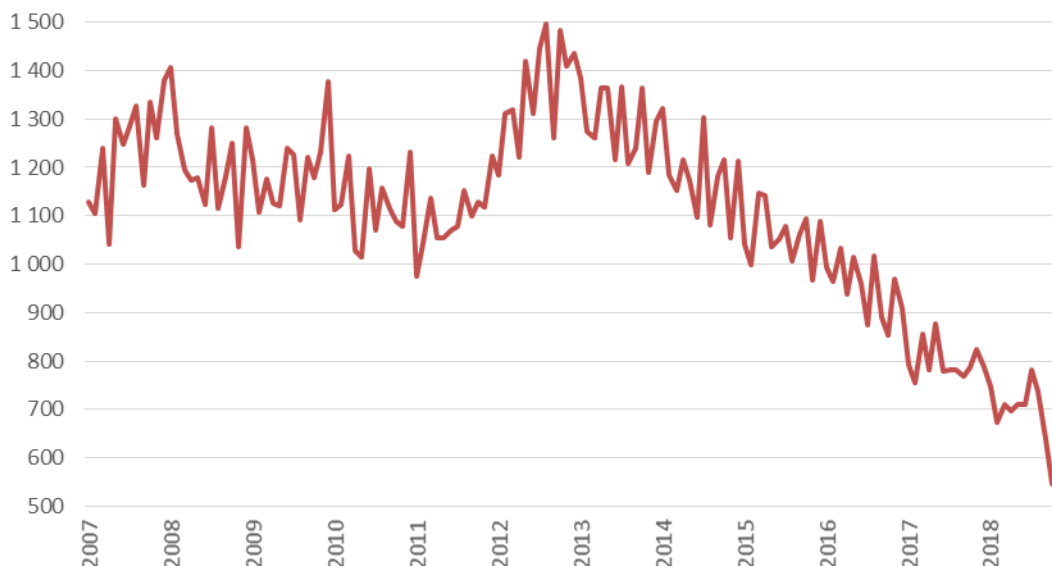


Рисунок 7. Число зарегистрированных в России детей, умерших на первом году жизни, чел.

Источник: ЕМИСС – URL: <https://fedstat.ru/indicator/33557> (дата обращения: 01.02.2019).

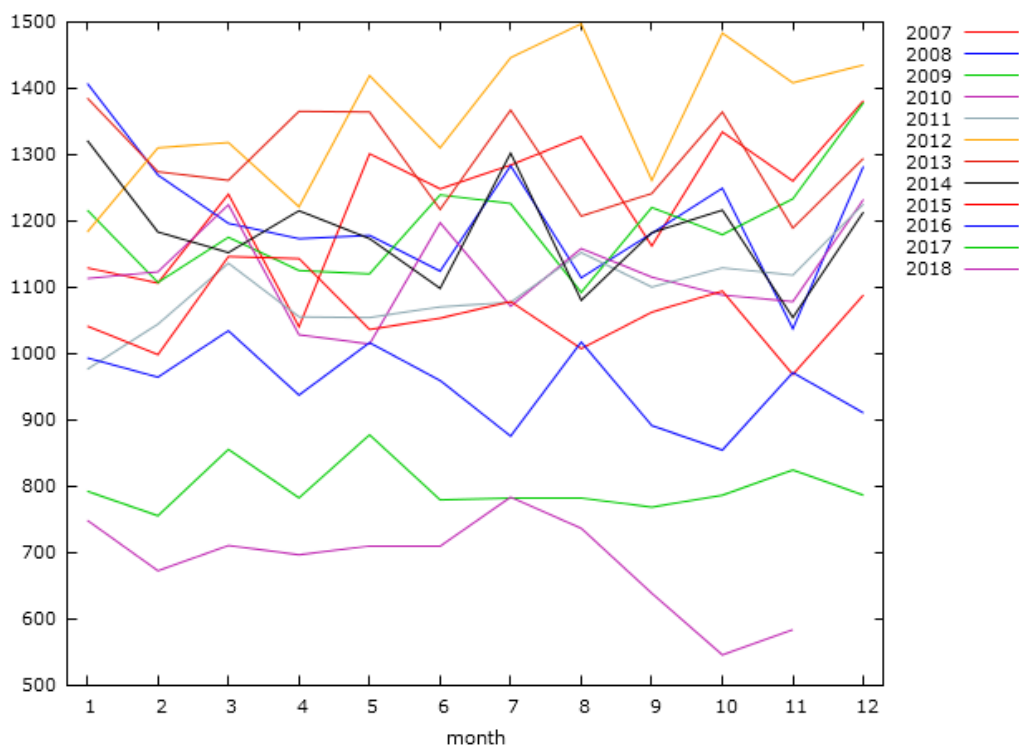


Рисунок 8. Число зарегистрированных в России детей, умерших на первом году жизни: сравнение по годам, чел.

Источник: Расчеты авторов.

Анализ коррелограммы ряда $\ln f$ (рисунок П2.4а Приложения 2) и его первой разности (рисунок П2.4б Приложения 2; для устранения влияния на график трендовой компоненты) также подтверждает выводы демографов относительно отсутствия сезонности. Характерные пики, связанные с сезонностью, на лагах, кратных периоду

сезонности ($S = 12$), отсутствуют (рисунки П2.4а, б Приложения 2). Более того можно заметить некую периодичность с лагом 3 (рисунок П2.4б Приложения 2), на что также указывают пики в периодограмме (рисунок П2.5 Приложения 2), однако для моделирования такой периодичности необходимы дополнительные обоснования.

Таким образом, ряд $\ln f$ был исключен из дальнейшего анализа из-за отсутствия сезонной компоненты, а целью настоящей работы являлось как раз моделирование сезонности.

Анализ демографических временных рядов на наличие сезонных единичных корней. Результаты тестирования с помощью классических критериев единичного корня в случае сезонных данных приводят к противоречивым результатам и не позволяют сделать однозначный вывод относительно порядка интегрируемости исследуемых процессов. Ввиду этого в дальнейшем анализе были проведены тесты HEGY, позволяющие одновременно исследовать наличие сезонных и несезонных единичных корней. В случае месячных данных тест HEGY был модифицирован в работе Больё и Мирон (Beaulieu, Miron 1993). Для проверки устойчивости результатов теста к смене модификации были рассмотрены различные вариации теста. Сводные результаты теста HEGY для исследуемых нами демографических рядов представлены в таблице 2. Отметим, что единичные корни были указаны в случае неотвержения нулевой гипотезы на 5%-ном уровне значимости.

Таблица 2. Сводные результаты тестирования рядов на сезонные единичные корни

Временной ряд	С константой	С константой и сезонными дамми-переменными	С константой и трендом	С константой, трендом, сезонными дамми-переменными
birth	единичный полугодовой месячный	единичный	единичный полугодовой месячный	единичный
mort	единичный полугодовой месячный	единичный	единичный полугодовой	единичный
mar	единичный полугодовой месячный	единичный месячный	единичный полугодовой месячный	единичный месячный

Примечание: Тестовое уравнение сезонной разности критерия HEGY для месячных данных предполагает наличие 12 единичных корней: корень $+1$ соответствует несезонному единичному корню, -1 – полугодовому корню, комплексные корни $\pm i$ – квартальным корням, $(0,5(1,44 \pm i))$ – месячным корням и т.д.

Источник: Расчеты авторов.

В таблице 2 жирным шрифтом выделены выводы по наличию единичных корней для каждого ряда, а сами статистики тестов для выбранной модификации представлены в таблице 3.

Результаты тестирования на наличие несезонных/сезонных единичных корней позволяют выделить некоторые особенности исследуемых временных рядов. Так, все ряды являются нестационарными с порядком интегрируемости $d = 1$ для *mort* и *mar*, $d = 2$ для

birth (результаты теста HEGY⁸ для первой и второй разности представлены в Приложении 3). Ряды birth и mort содержат детерминированную сезонность, что возможно учесть в SARIMA-модели включением сезонных дамми-переменных для соответствующих месяцев. Ряд mar обладает ярко выраженной случайной сезонностью с порядком сезонной интегрируемости $D_s = 1$.

Таблица 3. Результаты HEGY-теста на сезонные единичные корни с учетом модификации

Временной ряд	Модификация теста	Значение статистик	P-значение	Вывод
birth	С константой и трендом	t1 = 0,09	0,9959	Единичный, полугодовой единичный корень, детерминированная сезонность
		t2 = -1,26	0,1679	
		F1 = 0,16	0,8399	
mort	С константой	t1 = -1,42	0,5539	Единичный, полугодовой единичный корень, детерминированная сезонность
		t2 = -1,04	0,2328	
		F1 = 2,86	0,0590	
mar	С константой	t1 = 0,63	0,9919	Единичный, полугодовой и месячный единичные корни
		t2 = -0,91	0,2754	
		F1 = 0,46	0,6314	

Примечание: Статистика t1 соответствует гипотезе о наличии несезонного единичного корня, t2 – гипотезе о наличии полугодового сезонного единичного корня, F1 – гипотезе о наличии месячного сезонного единичного корня.

Источник: Расчеты авторов.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Прогнозирование числа родившихся. В результате оценивания и сравнения нескольких SARIMA-моделей (разных порядков авторегрессии и скользящего среднего) с учетом того, что результаты HEGY-теста указывают на детерминированную сезонность, наиболее адекватной оказалась модель SARIMA(2,2,1/3)(1,0,0) с добавлением сезонных дамми-переменных dm_i в виде уравнения (4). Дамми-переменная месяца dm_i принимает два значения: 1 для месяца i , 0 – для всех остальных месяцев. В таблице 4 представлены оценки данной модели.

$$(1 - \alpha_1 L - \alpha_2 L^2)(1 - \alpha_{12} L^{12}) \Delta^2 y_t = \theta_0 + (1 + \theta_1 L + \theta_3 L^3) \varepsilon_t + \sum_{i=2}^{12} \beta_i dm_i. \quad (4)$$

Для рассматриваемой модели все корни характеристического уравнения по модулю больше 1, что свидетельствует об обратимости и стационарности процесса. Значения автокорреляционных функций остатков равны нулю, статистика Льюинга–Бокса $Q = 20,51$ (p-значение = 0,15) (для первых 20 лагов) также свидетельствует об отсутствии автокорреляции в остатках модели, это позволяет сделать вывод о том, что ряд остатков является реализацией процесса белого шума. В Приложении 4 также приведена коррелограмма остатков модели, демонстрирующая отсутствие автокорреляции до 36 лага. Нормальность распределения остатков проверялась на основе критерия χ^2 : $\chi^2(2) = 4,11$,

⁸ Ряд birth становится стационарным при взятии второй разности.

что говорит в пользу неотклонения нулевой гипотезы о том, что ряд остатков является реализацией процесса, подчиняющегося нормальному закону распределения, на уровне значимости 5%. Наряду с оценкой адекватности модели для оценки качества прогноза по полученной модели использовали характеристики точности прогноза: RMSE = 3,49; MPE = 0,06; MAPE = 1,84. Отметим, что полученные характеристики точности прогноза свидетельствуют о достаточно высоких прогностических свойствах полученной SARIMA-модели.

Таблица 4. Оценки SARIMA-модели (4) для birth

	Коэффициент	Ст. ошибка	z	P-значение
Зависимая переменная: $\Delta^2 \text{birth}$				
θ_0	3,948**	1,672	2,361	0,018
α_1	-1,050***	0,067	-15,704	0,000
α_2	-0,800***	0,083	-9,664	0,000
α_{12}	0,166*	0,096	1,742	0,082
θ_1	-0,476***	0,103	-4,632	0,000
θ_3	-0,455***	0,087	-5,220	0,000
dm2	-8,487***	2,513	-3,378	0,0007
dm3	4,105**	1,755	2,339	0,019
dm4	-8,773***	2,181	-4,022	0,000
dm5	6,636**	2,914	2,278	0,023
dm6	-4,528**	2,225	-2,035	0,042
dm7	10,454***	2,629	3,977	0,000
dm8	-9,849***	2,360	-4,174	0,000
dm9	-18,659***	2,948	-6,330	0,000
dm10	-12,110***	3,049	-3,972	0,000
dm11	-10,068***	2,337	-4,308	0,000
dm12	3,895	3,917	0,994	0,320
$\sigma = 3,48$; AIC = 713,45; BIC = 764,65; Q(20) = 20,51 (p-значение = 0,15); $\chi^2(2) = 4,11$ (p-значение = 0,12)				

Примечание: *** – статистическая значимость коэффициента на 1%-ном уровне, ** – на 5%-ном уровне, * – на 10%-ном.

Источник: Расчеты авторов.

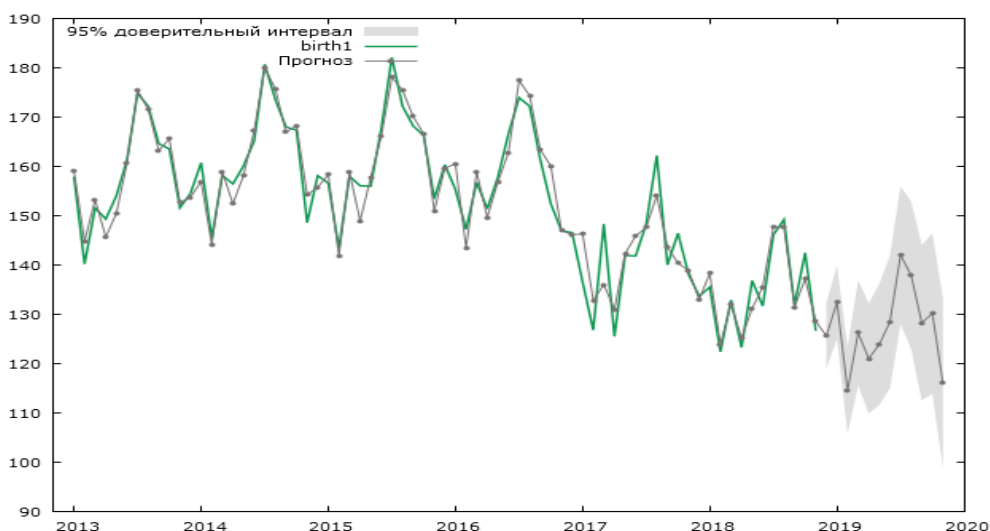


Рисунок 9. Число родившихся в России: наблюдаемые и предсказанные значения на основе SARIMA-модели с 95%-ным доверительным интервалом, тыс.

Источник: Расчеты авторов.

На рисунке 9 представлены наблюдаемые значения числа родившихся и предсказанные по SARIMA-модели для России, а также построен 95%-ный доверительный интервал прогноза до декабря 2019 г. В таблице 5 приведены прогнозные значения числа родившихся в России на основе полученной SARIMA-модели со значениями 95%-ных доверительных интервалов.

Таблица 5. Прогнозные значения числа родившихся в России на основе SARIMA-модели

Дата	Прогноз, тыс.	Ст. ошибка	95%-ный доверительный интервал прогноза	Сезонные приросты (к соответствующему месяцу предыдущего года)	
				тыс.	%
Декабрь 2018	125,648	3,416	(118,953, 132,343)	-8,038	-6,01
Январь 2019	132,466	3,759	(125,098, 139,834)	-3,062	-2,26
Февраль 2019	114,565	4,490	(105,764, 123,365)	-7,860	-6,42
Март 2019	126,267	5,429	(115,626, 136,908)	-6,670	-5,02
Апрель 2019	121,042	5,690	(109,889, 132,195)	-2,255	-1,83
Май 2019	123,941	6,274	(111,643, 136,238)	-12,926	-9,44
Июнь 2019	128,487	6,860	(115,042, 141,933)	-3,210	-2,44
Июль 2019	142,01	7,111	(128,072, 155,949)	-4,132	-2,83
Август 2019	137,965	7,616	(123,039, 152,891)	-11,268	-7,55
Сентябрь 2019	128,283	8,041	(112,524, 144,042)	-3,614	-2,74
Октябрь 2019	130,193	8,296	(113,933, 146,453)	-12,299	-8,63
Ноябрь 2019	116,09	8,736	(98,9681, 133,213)	-10,578	-8,35

Источник: Расчеты авторов.

Результаты моделирования показывают, что на протяжении 2019 г. на фоне общего снижения числа родившихся в России (о чем пишут демографы) будет продолжаться и снижение числа родившихся по месяцам, о чем свидетельствуют рассчитанные сезонные приросты. Наибольшее число родившихся ожидается в июле 2019 г. (142,01 тыс.), наименьшее – в феврале 2019 г. (114,565 тыс.).

Прогнозирование числа умерших. Для *mort* хорошие статистические свойства показала модель ARIMA(1,1,1) с добавлением сезонных дамми-переменных dm_i в виде уравнения (5). Дамми-переменная месяца dm_i принимает два значения: 1 – для месяца i , 0 – для всех остальных месяцев. В таблице 6 представлены оценки модели (5).

$$(1 - \alpha_1 L)\Delta y_t = \theta_0 + (1 + \theta_1 L)\varepsilon_t + \sum_{i=2}^{12} \beta_i dm_i. \quad (5)$$

Для рассматриваемой модели все корни характеристического уравнения по модулю больше 1, что свидетельствует об обратимости и стационарности процесса. Значения автокорреляционных функций остатков равны нулю, статистика Льюинга–Бокса $Q = 17,23$ (p -значение = 0,51) (для первых 20 лагов) также свидетельствует об отсутствии автокорреляции в остатках, коррелограмма остатков модели до 36 лага приведена в Приложении 4. Характеристики точности прогноза ($RMSE = 4,54$; $MPE = -0,04$; $MAPE = 2,11$) свидетельствуют о высоких прогностических свойствах полученной ARIMA-модели.

Таблица 6. Оценки ARIMA-модели (5) для mort

	Коэффициент	Ст. ошибка	z
Зависимая переменная: $\Delta mort$			
θ_0	-0,164	0,038	-4,291
α_1	0,467	0,108	4,310
θ_1	-0,956	0,061	-15,685
dm2	-22,924	1,494	-15,346
dm3	-7,886	1,810	-4,358
dm4	-17,971	1,940	-9,264
dm5	-11,291	1,998	-5,652
dm6	-20,389	2,023	-10,080
dm7	-17,480	2,031	-8,608
dm8	-19,707	2,026	-9,727
dm9	-24,767	2,005	-12,351
dm10	-14,770	1,954	-7,559
dm11	-23,142	1,835	-12,612
dm12	-12,687	1,544	-8,219
$\sigma = 4,53$; AIC = 863,58; BIC = 907,92; Q(20) = 17,23 (p-значение = 0,51); $\chi^2(2) = 1,92$ (p-значение = 0,38)			

Примечание: Все коэффициенты статистически значимы на 1%-ном уровне.

Источник: Расчеты авторов.

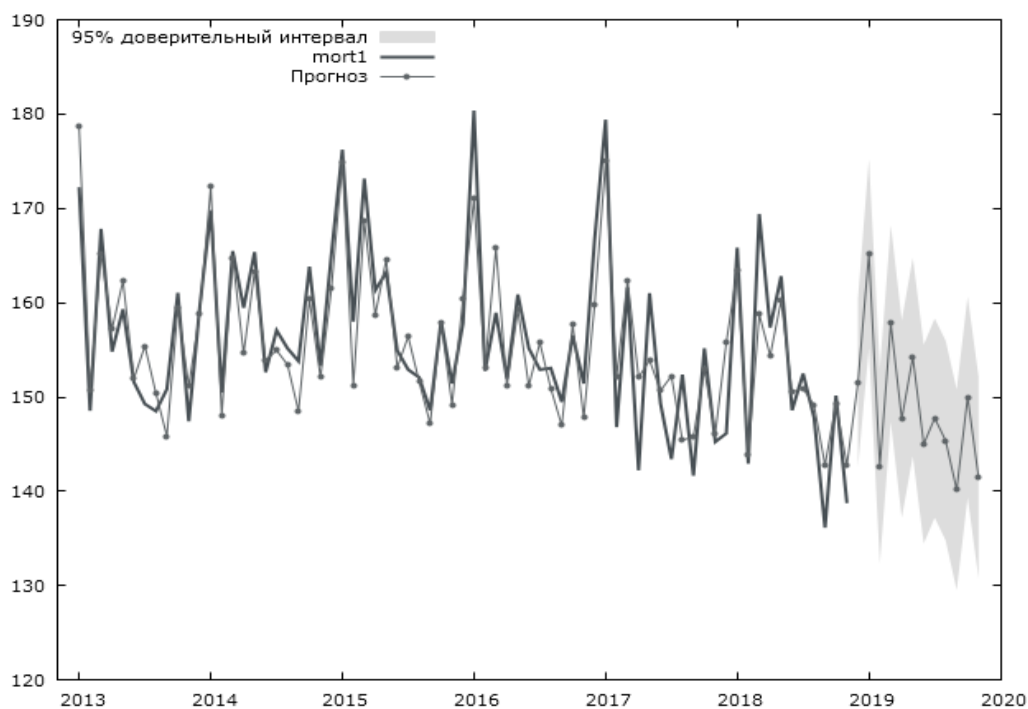


Рисунок 10. Число умерших в России: наблюдаемые и предсказанные значения на основе ARIMA-модели с 95%-ным доверительным интервалом, тыс.

Источник: Расчеты авторов.

На рисунке 10 представлены наблюдаемые значения числа умерших по месяцам (тыс. человек) в России и предсказанные по ARIMA-модели, а также построен 95%-ный доверительный интервал прогноза до декабря 2019 г. В таблице 7 приведены прогнозные значения *mort* на основе полученной ARIMA-модели со значениями 95%-ных доверительных интервалов, а также рассчитанные на основе прогнозных значений средние сезонные приросты.

Таблица 7. Прогнозные значения числа умерших в России на основе ARIMA-модели

Дата	Прогноз, тыс.	Ст. ошибка	95%-ный доверительный интервал прогноза	Сезонные приросты (к соответствующему месяцу предыдущего года)	
				тыс.	%
Декабрь 2018	151,532	4,531	(142,652, 160,411)	5,395	3,69
Январь 2019	165,218	5,087	(155,248, 175,187)	-0,618	-0,37
Февраль 2019	142,672	5,245	(132,393, 152,951)	-0,259	-0,18
Март 2019	157,800	5,304	(147,403, 168,196)	-11,608	-6,85
Апрель 2019	147,669	5,335	(137,213, 158,125)	-9,699	-6,16
Май 2019	154,240	5,355	(143,745, 164,736)	-8,565	-5,26
Июнь 2019	145,004	5,371	(134,477, 155,531)	-3,586	-2,41
Июль 2019	147,761	5,385	(137,206, 158,317)	-4,717	-3,09
Август 2019	145,375	5,399	(134,794, 155,957)	-2,447	-1,66
Сентябрь 2019	140,155	5,412	(129,547, 150,762)	4,015	2,95
Октябрь 2019	149,989	5,425	(139,357, 160,622)	-0,154	-0,10
Ноябрь 2019	141,454	5,438	(130,796, 152,112)	2,705	1,95

Источник: Расчеты авторов.

Прогнозные значения показывают, что на протяжении 2019 г. в России будет наблюдаться снижение числа умерших по месяцам, за исключением декабря 2018 г., сентября, ноября 2019 г., где наблюдается положительный прирост для показателя *mort*. Наибольший отрицательный прирост (-6,85%) ожидается в марте 2019 г. Наибольшее число умерших по-прежнему приходится на январь 2019 г. (165,218 тыс.), но с отрицательным приростом 0,37%, наименьшее – на сентябрь 2019 г. (140,155 тыс.) с положительным приростом 2,95%.

Прогнозирование числа зарегистрированных браков. При моделировании показателя *mar* была взята одна обычная и одна сезонная разность. Удовлетворительной оказалась модель SARIMA(2,1,0)(3,1,0)₁₂ в виде (6). Оценки модели (6) приведены в таблице 8.

$$(1 + \alpha_1 L + \alpha_2 L^2)(1 + \alpha_{12} L^{12} + \alpha_{24} L^{24} + \alpha_{36} L^{36}) \Delta \Delta_{12} y_t = \theta_0 + \varepsilon_t. \quad (6)$$

Таблица 8. Оценки SARIMA-модели (6) для *mar*

	Коэффициент	Ст. ошибка	z
Зависимая переменная: $\Delta \Delta_{12} mar$			
θ_0	-0,117	0,173	-0,6761
α_1	-0,966	0,071	-13,605
α_2	-0,585	0,070	-8,3131
α_{12}	-0,658	0,085	-7,6944
α_{24}	-0,567	0,090	-6,2793
α_{36}	-0,444	0,086	-5,1431
$\sigma = 11,84$; AIC = 1039,487; BIC = 1059,560, Q(20) = 19,23 (p-значение = 0,2); $\chi^2(2) = 1,906$ (p-значение 0,38)			

Примечание: Все коэффициенты, кроме константы, статистически значимы на 1%-ном уровне.

Источник: Расчеты авторов.

На основании теста Льюинга-Бокса для 20 лагов (Q-статистика = 19,23; p-значение = 0,2) был сделан вывод, что остатки модели обладают свойствами белого шума, коррелограмма остатков модели до 36 лага приведена в Приложении 4. Характеристики точности прогноза (RMSE = 12,65; MPE = -0,068; MAPE = 11,49) свидетельствуют о хороших прогностических свойствах полученной SARIMA-модели. Отрицательная

характеристика MPE свидетельствует о том, что прогноз чуть занижен. По характеристике MAPE прогноз для *mar* оказался самым «плохим». Результаты показывают, что в целом сезонный характер показателя на будущий год останется примерно таким же, как и в предыдущие годы.

На рисунке 11 представлены наблюдаемые значения числа зарегистрированных браков по месяцам регистрации в России и предсказанные по SARIMA-модели, а также построен 95%-ный доверительный интервал прогноза до декабря 2019 г. В таблице 9 приведены прогнозные значения *mont* на основе полученной ARIMA-модели со значениями 95%-ных доверительных интервалов, а также рассчитанные на основе прогнозных значений средние сезонные приросты.

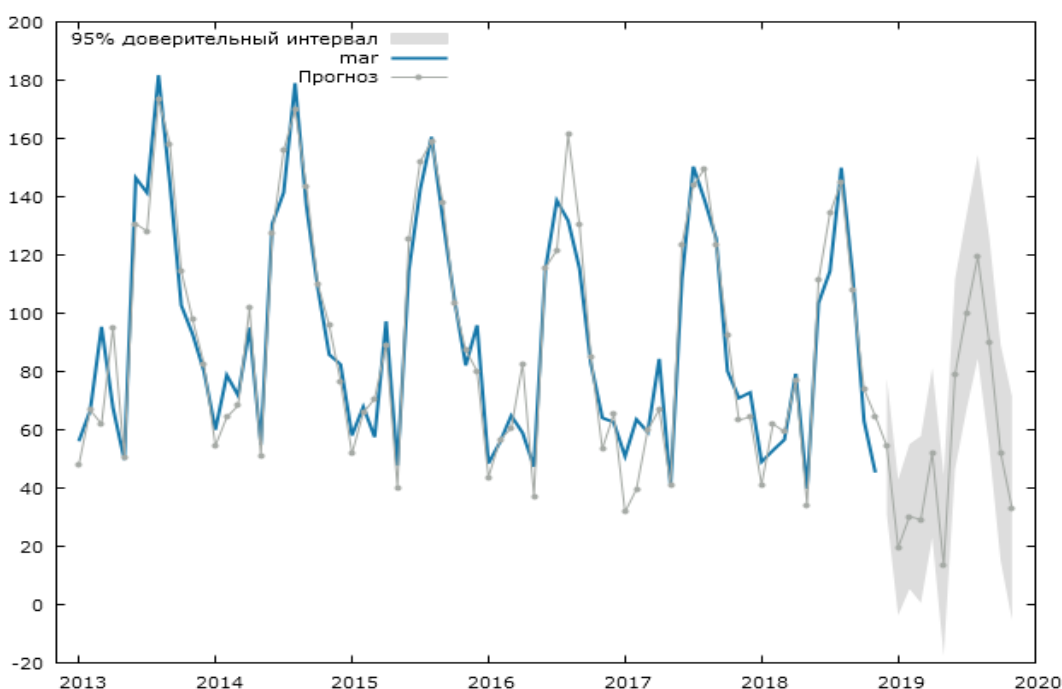


Рисунок 11. Число зарегистрированных браков в России: наблюдаемые и предсказанные значения на основе SARIMA-модели с 95%-ным доверительным интервалом, тыс.

Источник: Расчеты авторов.

Результаты моделирования указывают на значительное снижение числа зарегистрированных браков в 2019 г., однако надо отметить большую вариацию данного признака (в тыс.; среднее 97,209; минимум 39,331; максимум 192,59; стандартное отклонение 37,951) и как следствие большую стандартную ошибку прогноза и широкие границы 95%-ного доверительного интервала (для января, мая и ноября были получены отрицательные значения нижних границ доверительного интервала), что говорит о том, что, несмотря на хорошие прогностические свойства полученной модели, результаты стоит рассматривать аккуратно и привлекать экспертные оценки демографов. В целом в исходных данных также наблюдается резкое сокращение числа браков, так, например, в июле 2018 г. по отношению к июлю 2017 г. – на 35,952 тыс. и полученная модель оказалась очень «чувствительной» к данным скачкам. Снижение числа заключаемых и расторгаемых браков

демографы связывают с соответствующим движением «демографической волны»: сейчас в брачный возраст вступают малочисленные поколения родившихся в 1990-х годах, а также увеличивается доля пар, не регистрирующих брак (Щербакова 2018а).

Таблица 9. Прогнозные значения числа зарегистрированных браков в России на основе SARIMA-модели

Дата	Прогноз, тыс.	Ст. ошибка	95%-ный доверительный интервал прогноза	Сезонные приросты (к соответствующему месяцу предыдущего года)	
				тыс.	%
Декабрь 2018	54,337	11,835	(31,141, 77,533)	-18,508	-25,41
Январь 2019	19,633	11,842	(-3,577, 42,843)	-29,372	-59,94
Февраль 2019	30,285	12,673	(5,446, 55,124)	-22,687	-42,83
Март 2019	29,230	14,594	(0,626, 57,834)	-27,407	-48,39
Апрель 2019	52,154	14,760	(23,226, 81,083)	-27,104	-34,20
Май 2019	13,423	15,741	(-17,428, 44,275)	-26,333	-66,24
Июнь 2019	79,126	16,597	(46,596, 111,655)	-24,326	-23,51
Июль 2019	100,176	16,973	(66,9099, 133,442)	-14,240	-12,45
Август 2019	119,448	17,789	(84,5828, 154,313)	-30,518	-20,35
Сентябрь 2019	90,078	18,377	(54,059, 126,097)	-23,100	-20,41
Октябрь 2019	52,225	18,865	(15,250, 89,200)	-10,910	-17,28
Ноябрь 2019	33,184	19,521	(-5,0773, 71,444)	-12,197	-26,88

Источник: Расчеты авторов.

Сравнение прогностических свойств моделей SARIMA и Хольта-Винтерса. Проведенный анализ показал, что модели SARIMA хорошо описывают динамику выбранных для анализа демографических показателей. В заключение приведем сравнение прогнозов по характеристикам качества прогноза моделей SARIMA и Хольта-Винтерса с сезонностью, свидетельствующее в пользу SARIMA для всех рассматриваемых показателей. Результаты сравнения по RMSE, MAPE, MAE по тестовой части выборки приведены в таблице 10.

Таблица 10. Сравнение качества прогнозов по моделям SARIMA и Хольта-Винтерса (HW), по тестовой части выборки (T = 20)

Показатель	SARIMA			HW			Статистика Диболда-Мариано (p-значение), MSE
	RMSE	MAPE	MAE	RMSE	MAPE	MAE	
birth	3,59	0,02	2,86	5,84	0,035	4,98	S(1) = 5,26 (p-знач. = 0,0)
mort	4,58	0,026	3,88	5,56	0,032	4,88	S(1) = 5,02 (p-знач. = 0,0)
mar	9,92	0,1165	8,63	12,90	0,117	9,02	S(1) = 1,46 (p-знач. = 0,14)

Примечание: Нулевая гипотеза в тесте Диболда-Мариано: «нет различий в прогнозах».

Все рассчитанные характеристики по всем показателям принимают наименьшие значения для моделей SARIMA, что свидетельствует в пользу их хороших прогностических свойств. Результаты теста Диболда-Мариано указывают на то, что оценки по моделям на 5%-ном уровне значимости статистически не различаются лишь для показателя mar, а для birth и mort следует отдать предпочтение оценкам SARIMA-моделей.

ЗАКЛЮЧЕНИЕ

В работе были рассмотрены несколько примеров моделирования и прогнозирования демографических процессов с помощью SARIMA-моделей. Было выявлено, что исследуемые в работе демографические процессы имеют свои особенности: ряд числа рождений имеет второй порядок интегрируемости и содержит детерминированную сезонность, которая моделируется посредством введения в уравнение ARIMA-модели сезонных дамми-переменных, ряд числа умерших имеет первый порядок интегрируемости и детерминированную сезонность, ряд числа заключенных браков имеет первый порядок обычной и сезонной интегрируемости. Для ряда младенческой смертности наличие сезонности на основе анализа автокорреляционной функции и периодограммы не выявлено, что подтверждает выводы демографов об отсутствии сезонного фактора для данного показателя в последние годы. Все полученные SARIMA-модели оказались адекватными, а характеристики точности прогноза показали достаточно высокие прогностические свойства полученных моделей⁹. Для сравнения качества прогнозирования также были оценены модели Хольта-Винтерса с сезонностью. Результаты показали, что SARIMA-модели обладают лучшими характеристиками качества прогноза.

В дальнейшем было бы интересно проанализировать устойчивость полученных оценок моделей с учетом данных за более длительный период, когда возможно проявление эффектов «длинной памяти» (Granger 1980). Рассмотренный в работе статистический подход к прогнозированию демографических процессов может быть использован как один из возможных способов демографического прогнозирования, которое на сегодняшний день приобретает особую значимость во внутренней политике, при социально-экономическом планировании развития страны в целом или отдельных субъектов, при мониторинге реализации демографической программы в рамках майских указов Президента.

ЛИТЕРАТУРА

- Авдеев А., Блюм А., Троицкая И. (2002). Сезонный фактор в демографии российского крестьянства в первой половине 19 века: брачность, рождаемость, младенческая смертность. *Российский демографический журнал*, 1, 35-45. URL: <http://www.demoscope.ru/weekly/2003/0135/analit04.php>
- Айвазян С.А. (2010). *Методы эконометрики*. М.: Инфра-М.
- Баранов А.А., Намазова-Баранова Л.С., Альбицкий В.Ю., Терлецкая Р.Н. (2014). Тенденции младенческой и детской смертности в РФ в 1990-2012 гг. *Вестник Российской академии медицинских наук*, 69(11-12), 31-38. doi:10.15690/vramn.v69i11-12.1180
- Винник М.В. (2012). Сезонность демографических процессов (на примере метрических книг Покровского прихода г. Барнаул, 1877-1886 гг.). В кн. М.Б. Денисенко (Ред.),

⁹ Пока работа находилась на рецензировании, Росстат опубликовал оперативные данные по анализируемым показателям за декабрь 2018 – март 2019 г. Все опубликованные значения показателей (число родившихся, умерших, число браков) лежат в 95%-ном доверительном интервале, за исключением значений количества браков в марте 2019г. (см. Приложение 5).

- Демографические аспекты социально-экономического развития. Вып. (с. 251-267). М.: МАКС Пресс.
- Гиляровский Ф.В. (1866). *Исследования о рождении и смертности детей в Новгородской губернии*. Записки Императорского Русского Географического общества по отделению статистики. Т.1. Санкт-Петербург: LV-LVI.
- Ершов С.М. (1888). *Материалы для санитарной статистики Свияжского уезда. Опыт сравнительной демографии русской и татарской народностей*. (Дисс. д-ра медицины). Санкт-Петербург: Императорская Военно-медицинская академия.
- Канторович Г.Г. (2002) Анализ временных рядов (курс лекций). *Экономический журнал ВШЭ*, 6(3), 379–401. URL: <https://ej.hse.ru/2002-6-3/26547293.html>
- Кваша Е.А. (2003). Младенческая смертность в России в XX веке. *Социологические исследования*, 6, 47-55. URL: <http://ecsocman.hse.ru/socis/msg/18565164.html>
- Кендалл М., Стюарт А. (1976). *Многомерный статистический анализ и временные ряды*. М.: Наука.
- Махрова А.Г., Бочкарев А.Н. (2017). Маятниковая миграция в Московском регионе: новые данные. *Демоскоп Weekly*, 727-728. URL: <http://www.demoscope.ru/weekly/2017/0727/tema04.php>
- Турунцева М.Ю. (2011). Оценка качества прогнозов: простейшие методы. *Российское предпринимательство*, 12(8), 50-56. URL: <https://creativeconomy.ru/lib/6937>
- Щербакова Е.М. (2016). Россия: предварительные демографические итоги 2016 года (часть II). *Демоскоп Weekly*, 717-718. URL: <http://demoscope.ru/weekly/2017/0717/barom01.php>
- Щербакова Е.М. (2017) Россия: демографические итоги I полугодия 2017 года (часть II). *Демоскоп Weekly*, 741-742. URL: <http://demoscope.ru/weekly/2017/0741/barom01.php>
- Щербакова Е.М. (2018a). Россия: предварительные демографические итоги 2017 года (часть I). *Демоскоп Weekly*, 759-760. URL: <http://www.demoscope.ru/weekly/2018/0759/barom04.php>
- Щербакова Е.М. (2018b). Россия: предварительные демографические итоги 2017 года (часть II). *Демоскоп Weekly*, 761-762. URL: <http://demoscope.ru/weekly/2017/0761/barom01.php>
- Щербакова Е.М. (2019). Россия: предварительные демографические итоги 2018 года (часть II). *Демоскоп Weekly*, 803-804. URL: <http://demoscope.ru/weekly/2019/0803/barom01.php>
- Alho J. M., Spencer B. D. (2005). *Statistical demography and forecasting*. Springer.
- Beaulieu J.J., Miron J.A. (1993). Seasonal Unit Roots in Aggregate U.S. Data. *Journal of Econometrics*, 50(1), 305-328. doi:10.1016/0304-4076(93)90018-Z
- Booth H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3), 547–581. doi:10.1016/j.ijforecast.2006.04.001
- Box G.P., Jenkins G.M. (1970). *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- Cancho-Candel R., Llan J., Ardura-Fernánde J. (2007). Decline and loss of birth seasonality in Spain: Analysis of 33 421 731 births over 60 years. *Journal of Epidemiology and Community Health*, 61(8), 713-718. doi:10.1136/jech.2006.050211

- Dickey D.A., Fuller W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431. doi: 10.2307/2286348
- Dickey D.A., Hasza D.P., Fuller W.A. (1984). Testing for Unit Root in Seasonal Time Series. *Journal of the American Statistical Association*, 79(386), 355-367. doi: 10.2307/2288276
- Diebold F., Mariano R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253–263. doi:10.2307/1392185
- Eriksson A., Fellma J., Jord L., Pitkane K. (2008). Temporal, Seasonal, and Regional Differences in Births and Deaths in Åland (Finland). *Human Biology*, 80(2), 125-140. Retrieved from <https://www.jstor.org/stable/41466530>
- Feinstein C. (2002). Seasonality of Deaths in the U.S. by Age and Cause. *Demographic Research*, 6, 469-486. doi:10.4054/DemRes.2002.6.17
- Findley D.F., Monsell B.C., Bell W.R., Otto M.C., Chen B.-C. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business and Economic Statistics*, 16(2), 127–152. doi:10.2307/1392565
- Gomez V., Maravall A. (1996). Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User, WP 9628, Research Department, Banco de Espana.
- Granger C.W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238. doi:10.1016/0304-4076(80)90092-5
- Hiorns R.W. (1972). *Mathematical Models in Demography. The Structure of Human Populations*. Oxford: Clarendon Press.
- Holt C.C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 5–10. doi:10.1016/j.ijforecast.2003.09.015
- Hylleberg S., Engle R.F., Granger C. W.J., Yoo B.S. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics*, 44, 215-238. doi:10.1016/0304-4076(90)90080-D
- Hyndman R.J., Koehler A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688. doi:10.1016/j.ijforecast.2006.03.001
- Kale P., Andreozzi V., Nobre F. (2004). Time Series Analysis of Deaths Due to Diarrhoea in Children in Rio de Janeiro, Brazil, 1980-1998. *Journal of Health, Population and Nutrition*, 22(1), 27-33. URL: <https://www.jstor.org/stable/23499006>
- Kashyap R.L., Rao A.R. (1976). *Dynamic Stochastic Models from Empirical Data*. Academic Press, New York, San Francisco. London.
- Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54 (1–3), 159–178. doi:10.1016/0304-4076(92)90104-Y
- Lee R. (1974). Forecasting Births in Post-Transition Populations: Stochastic Renewal with Serially Correlated Fertility. *Journal of the American Statistical Association*, 69(347), 607–617. doi:10.2307/2285990
- Nakaji S., Parodi S., Fontana V., Umeda T., Suzuki K., Sakamoto J., Sugawara K. (2004). Seasonal Changes in Mortality Rates from Main Causes of Death in Japan (1970-1999). *European Journal of Epidemiology*, 19(10), 905-913. doi:10.1007/s10654-004-4695-8

- Phillips P.C.B., Perron P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75 (2), 335-346. doi:10.12691/ijefm-2-6-4
- Pollard J.H. (1970). On Simple Approximate Calculations Appropriate to Populations with Random Growth Rates. *Theoretical Population Biology*, 1, 208–218. doi:10.1016/0040-5809(70)90035-3
- Saboia J.L.M. (1974). Modeling and Forecasting Populations by Time Series - The Swedish Case. *Demography*, 11, 483–492. doi:10.2307/2060440
- Saboia J.L.M. (1977). Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting. *Journal of the American Statistical Association*, 72(358), 264–270. doi:10.2307/2286787
- Stafoggia M., Forastiere F., Michelozzi P., Perucci C. (2009). Summer Temperature-related Mortality: Effect Modification by Previous Winter Mortality. *Epidemiolog*, 20(4), 575-583. doi:10.1097/EDE.0b013e31819ecdf0
- Sun L., Klein E., Laxminarayan R. (2012). Seasonality and Temporal Correlation between Community Antibiotic Use and Resistance in the United States. *Clinical Infectious Diseases*, 55(5): 687-694. doi:10.1093/cid/cis509
- Torri T., Vaupel J.W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28, 519–531. doi:10.1016/j.ijforecast.2011.01.009
- Zhang Y. Bi P., Hiller J. (2008). Climate variations and salmonellosis transmission in adelaide, south australia: A comparison between regression models. *International Journal of Biometeorology*, 52(3), 179-187. doi:10.1007/s00484-007-0109-4

ПРИЛОЖЕНИЕ

Приложение 1. Основные понятия и формулы, используемые в работе, для анализа сезонности во временных рядах

Выборочная автокорреляционная функция $\rho(k)$ в зависимости от лага k определяется по формуле:

$$\hat{\rho}(k) = \text{corr}(y_t, y_{t-k}) = \frac{\frac{1}{T-k} \sum_{t=k}^T (y_t - \hat{\mu})(y_{t-k} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2},$$

где T – длина временного ряда, μ – выборочная оценка среднего.

$\rho(k)$ представляет собой последовательность коэффициентов корреляции между y_t и y_{t-k} .

Характеристическое уравнение ARMA процесса – это алгебраическое уравнение, на основе которого анализируется стационарность и обратимость процесса, описываемого уравнением вида:

$$\alpha_p(L)y_t = \theta_q(L)\varepsilon_t,$$

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

Для анализа стационарности составляется характеристическое уравнение AR части вида:

$$1 - \alpha_1 z - \alpha_2 z^2 \dots - \alpha_p z^p = 0$$

Для анализа обратимости процесса составляется характеристическое уравнение MA части вида:

$$1 + \theta_1 z + \theta_2 z^2 \dots + \theta_q z^q = 0$$

Если корни соответствующих характеристических уравнений по модулю больше единицы, то говорят о стационарности и обратимости процессов соответственно.

Единичный корень. Понятие единичного корня возникает при анализе характеристического уравнения процесса. Если один из корней характеристического уравнения равен единице, то анализируемый процесс будет нестационарным. Например, нестационарный процесс случайного блуждания $y_t = y_{t-1} + \varepsilon_t$ имеет характеристическое уравнение $1 - z = 0$, решением которого является единичный корень $z = 1$.

DF-тест Дики-Фуллера (Dickey, Fuller 1979) – статистический тест, позволяющий выявить (не)стационарность временного ряда. Другое название – тест *единичного корня*. В основе критерия лежит тестовое уравнение $y_t = a y_{t-1} + \varepsilon_t$. Если $a = 1$, то анализируется процесс случайного блуждания, который по определению является нестационарным. Если $a < 1$, то это соответствует стационарному AR процессу. Таким образом, $H_0: a = 1$

(нестационарный процесс), $H_1: a < 1$ (стационарный процесс). Далее переходят к уравнению $\Delta y_t = \beta y_{t-1} + \varepsilon_t$, где $\beta = a-1$, а соответствующие гипотезы имеют вид: $H_0: \beta = 0$ (нестационарный процесс), $H_1: \beta < 0$ (стационарный процесс). С помощью МНК рассчитывается оценка коэффициента и его стандартная ошибка, вычисляется $t_{набл.}$

$$\hat{t}_{набл.} = \hat{\tau} = \frac{\hat{\beta}}{S(\hat{\beta})} \sim DF$$

Распределение t -статистики отличается от классического распределения t -статистики, ее значения зависят от спецификации тестового уравнения, которое может содержать константу/тренд. Значения затабулированы в работе авторов теста (Dickey, Fuller 1979).

Порядок интегрируемости ряда y_t определяется исходя из того, сколько последовательных разностей $\Delta y_t = y_t - y_{t-1}$ надо применить к исходному нестационарному ряду, чтобы привести его к стационарному виду. Обозначается $I(d)$. Так, $I(0)$ – стационарный процесс, $I(1)$ -процесс будет стационарным после взятия 1-й разности, $I(2)$ -процесс станет стационарным после взятия 2-й разности и т.д. Аналогично вводится понятие *сезонной интегрируемости* ряда $I_s(d)$, и применяются сезонные разности $\Delta_{12}y_t = y_t - y_{t-12}$.

Модель Хольта-Винтерса с сезонностью (Holt-Winters Model) рассматривалась в виде:

$$\hat{y}_{t+\tau} = \{\hat{a}_t + \tau \cdot \hat{b}_t\} \cdot \hat{s}_{t-L+\tau}$$

$$\hat{a}_t = \alpha \frac{y_t}{\hat{S}(t-L)} + (1-\alpha)\{\hat{a}_{t-1} + \hat{b}_{t-1}\}$$

$$\hat{b}_t = \beta\{\hat{a}_t - \hat{a}_{t-1}\} + (1-\beta)\hat{b}_{t-1};$$

$$\hat{S}_t = \gamma \frac{y_t}{\hat{a}_t} + (1-\gamma)\hat{S}(t-L).$$

где a_t, b_t – параметры линейного тренда, S_t – сезонная компонента, $0 < \alpha, \beta, \gamma < 1$ – параметры адаптации, $L = 12$ – период сезонности для ежемесячных данных, τ – горизонт прогнозирования.

Таблица П1. Параметры адаптации в моделях Хольта-Винтерса

Показатель	Параметры адаптации		
	α	β	γ
birth	0,376	0,037	0,246
mort	0,405	0,000	0,011
mar	0,121	0,024	0,040

Примечание. Параметры адаптации подбирались автоматически в Stata на основе минимизации RMSE.

Приложение 2. Анализ периодической зависимости в анализируемых показателях

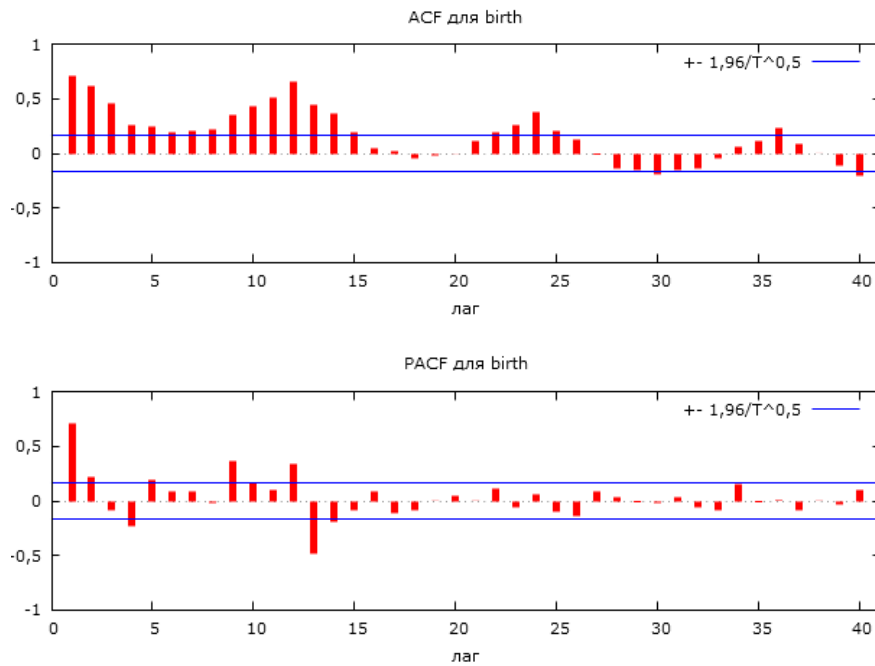


Рисунок П2.1. Коррелограмма для ряда birth

Источник: Расчеты авторов.

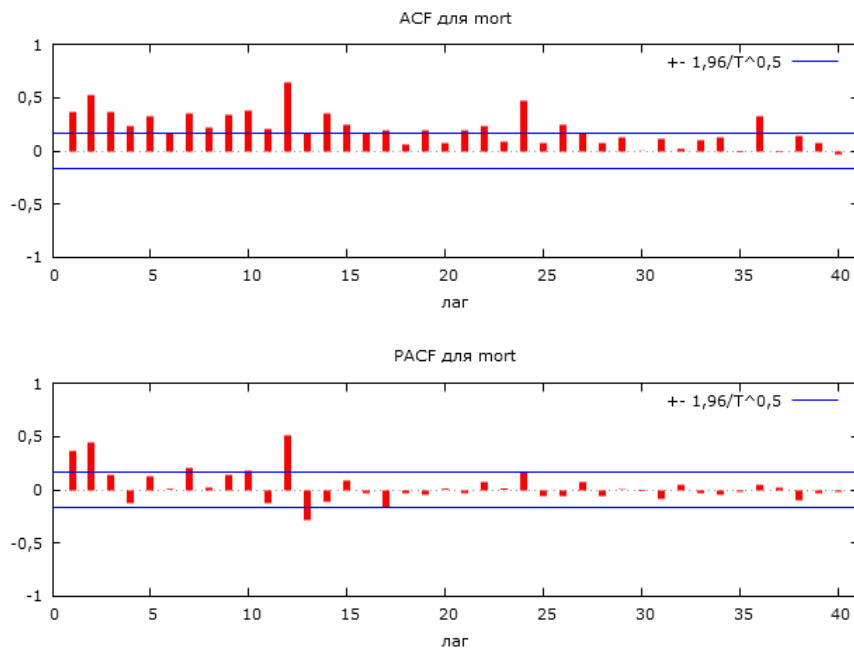


Рисунок П2.2. Коррелограмма для ряда mort

Источник: Расчеты авторов.

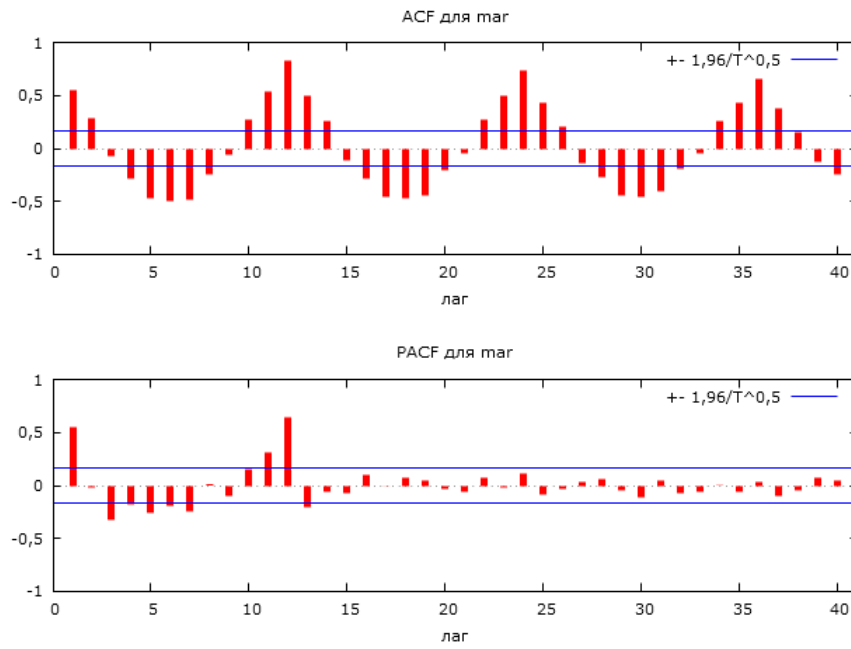


Рисунок П2.3. Коррелограмма для ряда mag

Источник: Расчеты авторов.

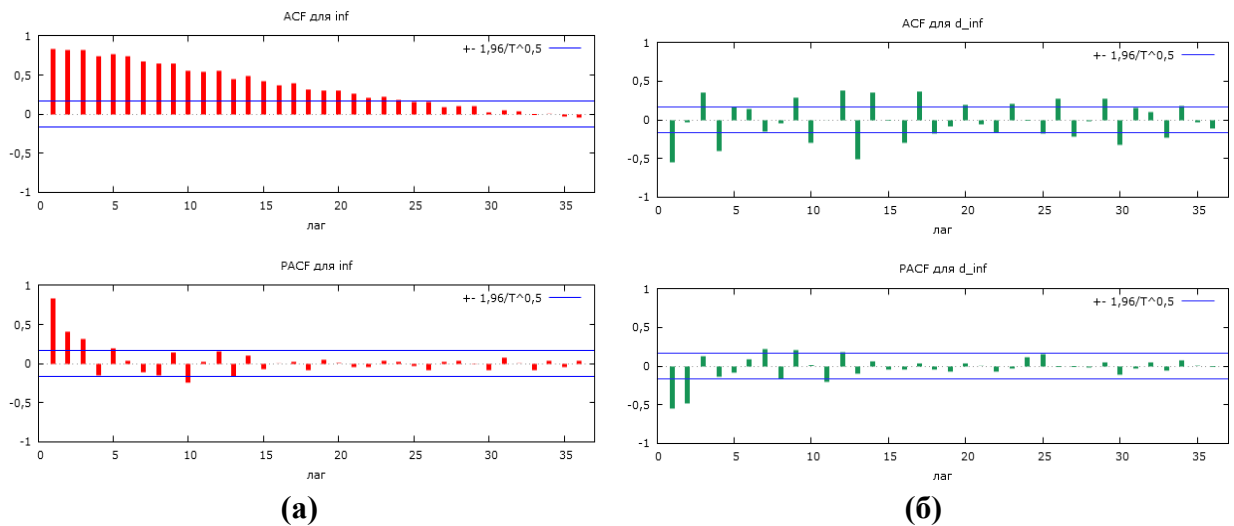


Рисунок П2.4. Коррелограмма для ряда inf (а) и Δ inf (б)

Источник: Расчеты авторов.

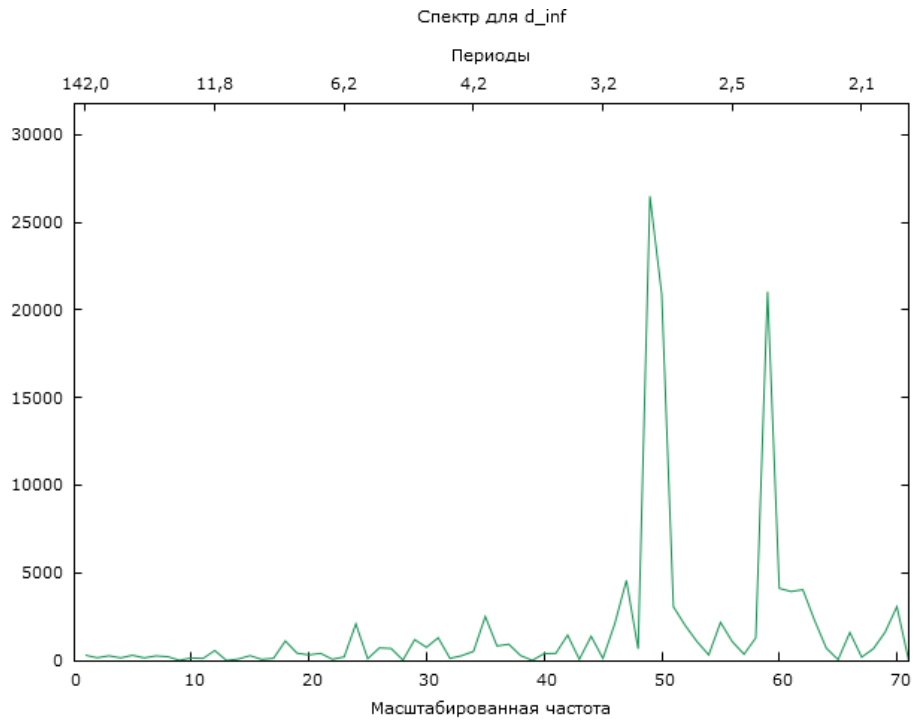


Рисунок П2.5. Периодограмма для Δinf

Источник: Расчеты авторов.

Приложение 3. Результаты тестирования рядов на стационарность

Таблица П3.1. Результаты тестирования ряда birth на стационарность

Временной ряд	Тест	Спецификация теста*	Значение статистик	P-значение
birth	ADF	1	-0,5714	0,874
		2	0,7403	0,999
		3	0,1030	0,997
		4	-1,3049	0,968
PP	1	-4,5054	0,0003	
	2	-4,4668	0,003	
KPSS	KPSS	1	0,6059	0,029
		2	0,4563	0,001
		3	0,7375	0,001
$\Delta birth$	ADF	1	-2,6835	0,077
		2	-3,0886	0,040
		3	-4,4996	0,002
	PP	1	-17,2985	0,000
		2	-17,4104	0,000
	KPSS	KPSS	1	0,0849
2			0,017	> 0,10
3			0,0423	> 0,10
$\Delta^2 birth$	ADF	1	-11,3973	0,000
	PP	1	-41,409	0,000
	KPSS	1	0,0673	> 0,10

Примечание: *: 1 – с константой, 2 – с константой и трендом, 3 – с константой, трендом и сезонными фиктивными переменными, 4 – с константой и квадратичным трендом.

Источник: Расчеты авторов.

Таблица ПЗ.2. Результаты тестирования рядов *mort* и *mar* на стационарность

Временной ряд	Тест	Спецификация теста*	Значение статистик	P-значение
<i>mort</i>	ADF	1	-1,2296	0,6639
		2	-2,556	0,301
		3	-1,109	0,7146
	PP	1	-8,4559	0,000
		2	-11,8998	0,000
		3	2,0854	0,000
	KPSS	1	2,0854	0,000
		2	0,1346	0,074
		3	2,2947	0,000
<i>mar</i>	ADF	1	0,627763	0,9905
		2	-0,631348	0,9768
		3	-0,834457	0,8088
	PP	1	-6,36873	0,0000
		2	-6,44458	0,0000
		3	0,1778	> 0,10
	KPSS	1	0,1778	> 0,10
		2	0,0438	> 0,10
		3	1,5795	0,000

Примечание: *: 1 – с константой, 2 – с константой и трендом, 3 – с константой и сезонными фиктивными переменными.

Источник: Расчеты авторов.

Таблица ПЗ.3. Результаты HEGY-теста для *birth* на сезонные единичные корни

Временной ряд	Модификация теста	Значение статистик	P-значение
<i>birth</i>	С константой и трендом	t1 = 0,09	0,99586
		t2 = -1,26	0,16791
		F1 = 0,16	0,83998
Δ <i>birth</i>	С константой и трендом	t1 = -3,18	0,08981
		t2 = -1,26	0,17097
		F1 = 0,17	0,83304
Δ^2 <i>birth</i>	С константой	t1 = -11,40	0,00000
		t2 = -1,43	0,13382
		F1 = 0,05	0,94426

Источник: Расчеты авторов.

Приложение 4. Коррелограммы остатков моделей

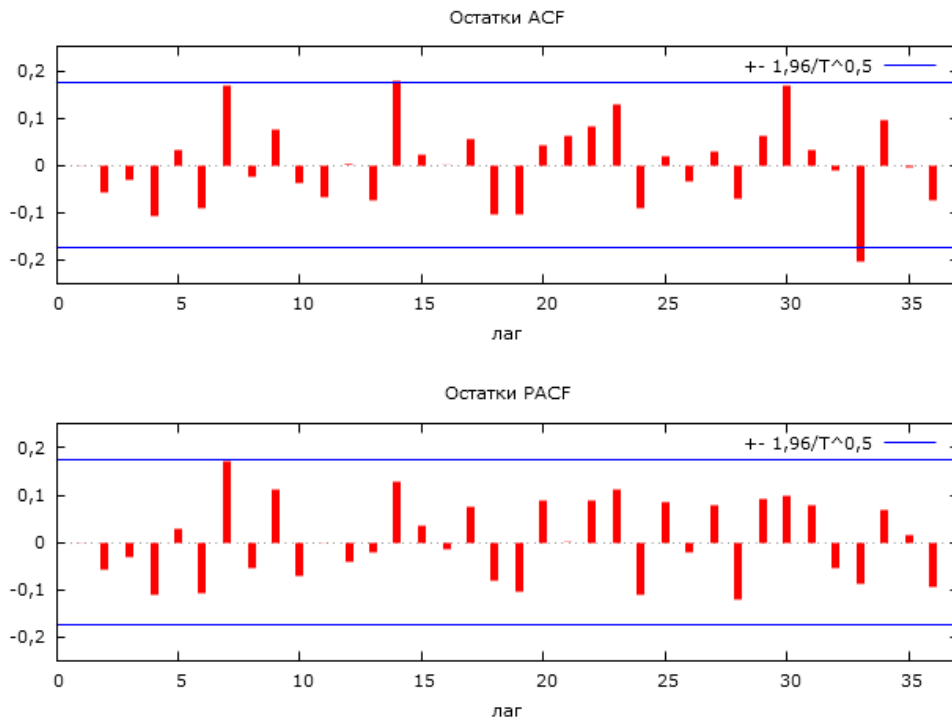


Рисунок П4.1. Коррелограмма остатков модели для birth

Источник: Расчеты авторов.

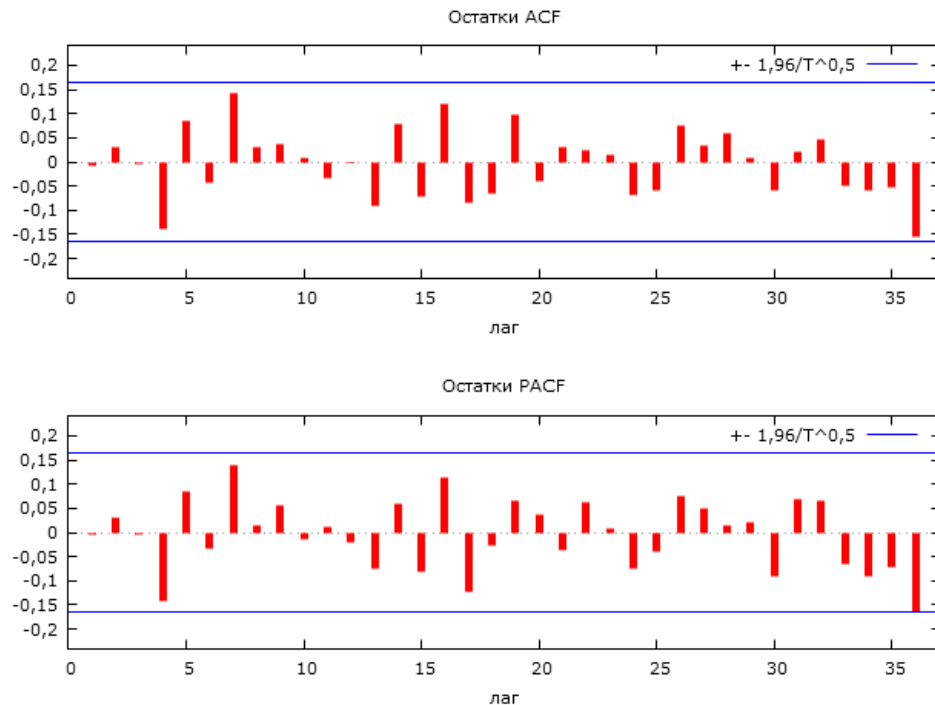


Рисунок П4.2. Коррелограмма остатков модели для mort

Источник: Расчеты авторов.

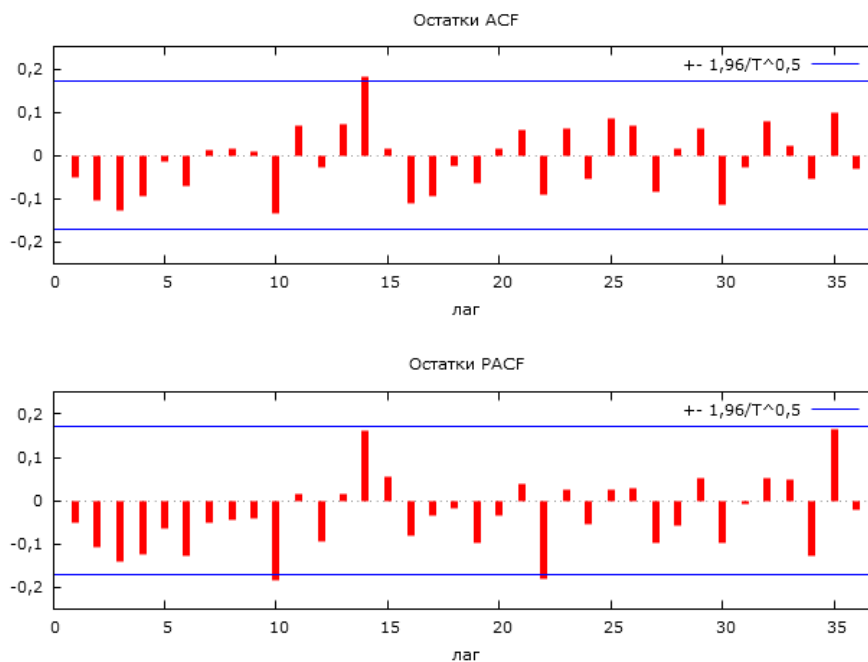


Рисунок П4.3. Коррелограмма остатков модели для lag

Источник: Расчеты авторов.

Приложение 5. Сравнение прогнозных значений и опубликованных данных

Таблица П5. Сравнение прогнозных значений на основе SARIMA-моделей и опубликованных данных

Дата	Birth, тыс.чел.		Mort, тыс.чел.		Mag, тыс.	
	прогноз*	опубл. данные	прогноз*	опубл. данные	прогноз*	опубл. данные
Декабрь 2018	(118,953, 132,343)	120,133	(142,652, 160,411)	145,440	(31,141, 77,533)	49,778
Январь 2019	(125,098, 139,834)	125,65	(155,248, 175,187)	168,495	(-3,577, 42,843)	32,224
Февраль 2019	(105,764, 123,365)	111,739	(132,393, 152,951)	142,287	(5,446, 55,124)	29,923
Март 2019	(115,626, 136,908)	117,793	(147,403, 168,196)	151,076	(0,626, 57,834)	61,731

Примечание: * – 95%-ный доверительный интервал прогноза на основе SARIMA-моделей.

STATISTICAL APPROACHES TO ANALYSIS AND MODELING OF SEASONALITY IN DEMOGRAPHIC DATA

LILIA RODIONOVA, ELENA KOPNOVA

According to the May Presidential Decree (2018), one of the national goals and strategic objectives of the development of the Russian Federation for the period up to 2024 is “ensuring sustainable natural growth in the population of the Russian Federation and increasing life expectancy to 78 years”. Thus, the increased need to monitor the current demographic situation, the study of the structure of demographic indicators, and the close attention of the community to the realization of national goals led to the choice of the topic of this study.

The paper studies the problems of modeling the seasonality of demographic indicators in the Russian Federation (the number of births, the number of deaths, infant mortality, the number of marriages) according to monthly data of Rosstat for the period 2007-2018. Foreign studies have shown that, along with traditional demographic methods, ARIMA models give good results in forecasting of demographic indicators (population size, birth and death rates, life expectancy). Using the approach based on SARIMA models in this work allowed us to obtain adequate models with good statistical and prognostic properties. The stationarity of processes was analyzed on the basis of the HEGY test. The indicators studied in the work had a number of features that must be taken into account when modeling. The series of the number of births and the number of deaths had second and first integration orders respectively and contained deterministic seasonality, the series of the number of marriages had the first integration order and seasonal integration, and the infant mortality series did not contain seasonality, which was confirmed based on the analysis of the autocorrelation function and periodogram. Point and interval estimates of the forecast for 2019 were built for all indicators here studied. To compare the quality of forecasting SARIMA-models, seasonal Holt-Winters models were also evaluated.

Key words: forecasting, time series analysis, ARIMA models, SARIMA models, HEGY test, statistical methods.

LILIA RODIONOVA (lrodionova@hse.ru), NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, RUSSIA.

ELENA KOPNOVA (ekopnova@hse.ru), NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS, RUSSIA.

DATE RECEIVED : MARCH 2019.

REFERENCES

- Aivazian S.A. (2010). *Metody ekonometriki* [Econometrics methods]. M.: Infra-M. (In Russ.)
- Alho J. M., Spencer B. D. (2005). *Statistical demography and forecasting*. Springer.
- Avdeev A., Blum A., Troitskaia I. (2002). Sezonnny faktor v demografii rossiyskogo krest'yanstva v pervoy polovine 19 veka: brachnost', rozhdayemost', mladencheskaya smertnost'. *Rossiyskiy demograficheskiy zhurnal*, 1, 35-45. (In Russ.) Retrieved from <http://www.demoscope.ru/weekly/2003/0135/analit04.php>
- Baranov A.A., Namazova-Baranova L.S., Al'bitskii V.Y., Terletskaya R.N. (2014). Trends of infant and child mortality in the Russian Federation in the period of 1990–2012. *Annals of the Russian academy of medical sciences*, 69(11-12), 31-38. (In Russ.) doi:10.15690/vramn.v69i11-12.1180

- Beaulieu J.J., Miron J.A. (1993). Seasonal Unit Roots in Aggregate U.S. Data. *Journal of Econometrics*, 50(1), 305-328. doi:10.1016/0304-4076(93)90018-Z
- Booth H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3), 547-581. doi:10.1016/j.ijforecast.2006.04.001
- Box G.P., Jenkins G.M. (1970). *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- Cancho-Candel R., Llan J., Ardura-Fernández J. (2007). Decline and loss of birth seasonality in Spain: Analysis of 33 421 731 births over 60 years. *Journal of Epidemiology and Community Health*, 61(8), 713-718. doi:10.1136/jech.2006.050211
- Dickey D.A., Fuller W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431. doi: 10.2307/2286348
- Dickey D.A., Hasza D.P., Fuller W.A. (1984). Testing for Unit Root in Seasonal Time Series. *Journal of the American Statistical Association*, 79(386), 355-367. doi: 10.2307/2288276
- Diebold F., Mariano R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253-263. doi:10.2307/1392185
- Eriksson A., Fellma J., Jord L., Pitkane K. (2008). Temporal, Seasonal, and Regional Differences in Births and Deaths in Åland (Finland). *Human Biology*, 80(2), 125-140. Retrieved from <https://www.jstor.org/stable/41466530>
- Ershov S.M. (1888). *Materialy dlya sanitarnoy statistiki Sviyazhskogo uyezda. Opyt sravnitel'noy demografii russkoy i tatarskoy narodnostey*. (PhD Thesis). Saint-Petersburg: Imperatorskaya Voyenno-meditsinskaya akademiya. (In Russ.)
- Feinstein C. (2002). Seasonality of Deaths in the U.S. by Age and Cause. *Demographic Research*, 6, 469-486. doi:10.4054/DemRes.2002.6.17
- Findley D.F., Monsell B.C., Bell W.R., Otto M.C., Chen B.-C. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business and Economic Statistics*, 16(2), 127-152. doi:10.2307/1392565
- Gilyarovskiy F.V. (1866). *Issledovaniya o rozhdenii i smertnosti detey v Novgorodskoy gubernii. Zapiski Imperatorskogo Russkogo Geograficheskogo obshchestva po otdeleniyu statistiki*. V.1. Saint-Petersburg: LV-LVI. (In Russ.)
- Gomez V., Maravall A. (1996). Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User, WP 9628, Research Department, Banco de Espana.
- Granger C.W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227-238. doi:10.1016/0304-4076(80)90092-5
- Hiorns R.W. (1972). *Mathematical Models in Demography. The Structure of Human Populations*. Oxford: Clarendon Press.
- Holt C.C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 5-10. doi:10.1016/j.ijforecast.2003.09.015
- Hylleberg S., Engle R.F., Granger C. W.J., Yoo B.S. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics*, 44, 215-238. doi:10.1016/0304-4076(90)90080-D
- Hyndman R.J., Koehler A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688. doi:10.1016/j.ijforecast.2006.03.001

- Kale P., Andreozzi V., Nobre F. (2004). Time Series Analysis of Deaths Due to Diarrhoea in Children in Rio de Janeiro, Brazil, 1980-1998. *Journal of Health, Population and Nutrition*, 22(1), 27-33. URL: <https://www.jstor.org/stable/23499006>
- Kantorovich G.G. (2002). Analiz vremennykh ryadov (kurs lektsiy). The *HSE Economic Journal*, 6(3), 379-401 (In Russ.). Retrieved from <https://ej.hse.ru/2002-6-3/26547293.html>
- Kashyap R.L., Rao A.R. (1976). *Dynamic Stochastic Models from Empirical Data*. Academic Press, New York, San Francisco. London.
- Kendall M., Stùart A. (1976). *Mnogomernyj statističeskij analiz i vremennye rãdy* [Multivariate statistical analysis and time series]. M.: Nauka. (In Russ.)
- Kvasha E. A. (2003). Mladencheskaya smertnost' v Rossii v XX veke. *Sotsiologicheskie Issledovaniia*, 6, 47-55. (In Russ.) Retrieved from <http://ecsocman.hse.ru/socis/msg/18565164.html>
- Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54 (1-3), 159-178. doi:10.1016/0304-4076(92)90104-Y
- Lee R. (1974). Forecasting Births in Post-Transition Populations: Stochastic Renewal with Serially Correlated Fertility. *Journal of the American Statistical Association*, 69(347), 607-617. doi:10.2307/2285990
- Mahrova A.G., Bočkarev A.N. (2017). Maâtnikovaâ migraciâ v Moskovskom regione: novye dannye. *Demoskop Weekly*, 727-728. (In Russ.). Retrieved from <http://www.demoscope.ru/weekly/2017/0727/tema04.php>
- Nakaji S., Parodi S., Fontana V., Umeda T., Suzuki K., Sakamoto J., Sugawara K. (2004). Seasonal Changes in Mortality Rates from Main Causes of Death in Japan (1970-1999). *European Journal of Epidemiology*, 19(10), 905-913. doi:10.1007/s10654-004-4695-8
- Phillips P.C.B., Perron P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75 (2), 335-346. doi:10.12691/ijefm-2-6-4
- Pollard J.H. (1970). On Simple Approximate Calculations Appropriate to Populations with Random Growth Rates. *Theoretical Population Biology*, 1, 208-218. doi:10.1016/0040-5809(70)90035-3
- Saboia J.L.M. (1974). Modeling and Forecasting Populations by Time Series - The Swedish Case. *Demography*, 11, 483-492. doi:10.2307/2060440
- Saboia J.L.M. (1977). Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting. *Journal of the American Statistical Association*, 72(358), 264-270. doi:10.2307/2286787
- Scherbakova E. M. (2016). Rossiya: predvaritel'nyye demograficheskiye itogi 2016 goda (chast' II). *Demoskop Weekly*, 717-718. (In Russ.). Retrieved from <http://demoscope.ru/weekly/2017/0717/barom01.php>
- Scherbakova E. M. (2017). Rossiya: demograficheskiye itogi I polugodiya 2017 goda (chast' II). *Demoskop Weekly*, 741-742. (In Russ.). Retrieved from <http://demoscope.ru/weekly/2017/0741/barom01.php>
- Scherbakova E. M. (2018a). Rossiya: predvaritel'nyye demograficheskiye itogi 2017 goda (chast' I). *Demoskop Weekly*, 759-760. (In Russ.). Retrieved from <http://www.demoscope.ru/weekly/2018/0759/barom04.php>

- Scherbakova E. M. (2018b). Rossiya: predvaritel'nyye demograficheskiye itogi 2017 goda (chast' II). Demoskop Weekly, 761-762. (In Russ.). Retrieved from <http://www.demoscope.ru/weekly/2018/0759/barom04.php>
- Scherbakova E. M. (2019). Rossiya: predvaritel'nyye demograficheskiye itogi 2018 goda (chast' II). Demoskop Weekly, 803-804. (In Russ.). Retrieved from <http://demoscope.ru/weekly/2019/0803/barom01.php>
- Stafoggia M., Forastiere F., Michelozzi P., Perucci C. (2009). Summer Temperature-related Mortality: Effect Modification by Previous Winter Mortality. *Epidemiolog*, 20(4), 575-583. doi:10.1097/EDE.0b013e31819ecdf0
- Sun L., Klein E., Laxminarayan R. (2012). Seasonality and Temporal Correlation between Community Antibiotic Use and Resistance in the United States. *Clinical Infectious Diseases*, 55(5): 687-694. doi:10.1093/cid/cis509
- Torri T., Vaupel J.W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28, 519–531. doi:10.1016/j.ijforecast.2011.01.009
- Turuntseva M. Yu. (2011). Assessment of Forecast Quality: the Simplest Methods. *Russian Journal of Entrepreneurship*, 12(8), 50-56. (In Russ.). Retrieved from <https://en.creativeconomy.ru/lib/6937>
- Vinnik M.V. (2012). Sezonnost' demograficheskikh protsessov (na primere metriceskikh knig Pokrovskogo prikhoda g. Barnaul, 1877-1886 gg.). V M.B. Denisenko (Ed.), *Demograficheskiye aspekty sotsial'no-ekonomicheskogo razvitiya*. Vyp. (pp. 251-267). M.: MAKSS Press. (In Russ.)
- Zhang Y. Bi P., Hiller J. (2008). Climate variations and salmonellosis transmission in Adelaide, South Australia: A comparison between regression models. *International Journal of Biometeorology*, 52(3), 179-187. doi:10.1007/s00484-007-0109-4