

Моделирование демографического потенциала регионов России с учетом различий в уровне их социально-экономического положения

А.Г. Сукиасян,

РЭУ им. Г.В. Плеханова, Москва; e-mail: sukiasyan.ag@rea.ru

Работа выполнена в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования РФ на тему «Модели, методы и алгоритмы искусственного интеллекта в задачах экономики для анализа и стилизации многомерных данных, прогнозирования временных рядов и проектирования рекомендательных систем», номер проекта FSSW-2023-0004.

Аннотация. Одной из самых актуальных проблем в России и ее регионах остается проблема демографического спада, выражающаяся сохраняющимся на протяжении нескольких десятилетий низким уровнем рождаемости и усугубляющаяся высоким уровнем преждевременной смертности среди молодежи и населения трудоспособного возраста. Следствием продолжающегося с начала 90-х гг. XX в. демографического спада являются рост демографической нагрузки, старение населения, дисбаланс на рынке труда. Для решения этих проблем государством разрабатываются и принимаются меры, направленные, в частности, на повышение рождаемости, выражающиеся во введении социальных выплат, льгот, а также других мер поддержки семей с детьми. Однако в долгосрочном периоде реализуемые меры демографической политики не привели к существенным изменениям в тенденции происходящего в стране демографического процесса. Это свидетельствует о существовании факторов, оказывающих более значительное влияние на динамику демографического процесса в стране и ее регионах, и обуславливает необходимость их выявления с использованием современного статистического и математического аппарата. В статье приведены результаты построения с учетом специфики исходных данных нескольких вариантов моделей, описывающих зависимость демографического потенциала от различающихся условий социально-экономического развития регионов России. При этом в качестве характеристики демографического потенциала в статье использовался показатель нетто-коэффициента воспроизводства, который характеризует как специфику рождаемости, так и смертности населения регионов России. Приведен сравнительный анализ четырех моделей, учитывающих эффект мультиколлинеарности и наличие грубых ошибок. На основе отобранной в ходе сопоставления характеристик качества модели проведен сценарный анализ, подтверждающий высокую чувствительность демографического потенциала к изменениям социально-экономических условий в регионах страны.

Ключевые слова: демографический потенциал, нетто-коэффициент воспроизводства, регрессия, мультиколлинеарность, грубые ошибки, сценарный анализ.

Классификация JEL: C21, C51, C52, C82.

ВВЕДЕНИЕ

Несмотря на реализуемые в России меры, направленные на стимулирование рождаемости, социальную поддержку семей с детьми (Госпрограммы РФ, 2023б, Зверева, 2006), снижение смертности населения от основных и социально-значимых причин смерти (Госпрограммы РФ, 2023а), в стране сохраняется тенденция к сокращению численности населения, его старению (Яковец, Голубков, 2018). Это, в свою очередь, негативно сказывается на экономике (Никонова, Красильникова, 2011, Лившиц и др., 2011), рынке труда (Сукиасян, 2024) и промышленности и других сферах государственной деятельности, поскольку согласно современным взглядам ученых-экономистов, население формирует потенциал развития государства (Рыбаковский, 2023, Баева, Уразова, 2020). В регионах России демографическая

проблема обретает особую актуальность ввиду их значительной дифференциации по половозрастному составу и социально-экономическому положению (Сукиасян, 2023). Данная статья посвящена проблеме построения эконометрической модели, описывающей закономерности влияния индикаторов социально-экономического развития регионов России на уровень демографического потенциала.

В качестве индикатора демографического потенциала автором предлагается использовать нетто-коэффициент воспроизводства населения, который позволяет анализировать процесс смены поколений при устойчивом естественном воспроизводстве населения (Валентей, 2016, Макаров, 2019). Нетто-коэффициент учитывает смертность в различных возрастных группах и показывает, сколько из рожденных девочек, в среднем, доживают до возраста своей матери (Ростовская, Ситковский, 2024). По мнению специалистов в области демографии, именно нетто-коэффициент более точно характеризует закономерности воспроизводства населения (Тихомиров, Тихомирова, 2022, Balbo и др., 2013). Для увеличения численности населения необходимо, чтобы значения этих коэффициентов превышали единицу, что означает: в среднем на одну женщину приходится более одной дочери. Если же эти показатели не достигают единицы, это указывает на тенденцию к сокращению численности населения.

Оценка нетто-коэффициента воспроизводства населения осуществляется на основе следующих показателей по пятилетним возрастным группам:

- число родившихся детей на 1000 женщин пятилетней возрастной группы k в регионе j в году t , $f_k^j(t)$;
- число умерших женщин на 1000 человек пятилетней возрастной группы i в регионе j в году t , $m_i^j(t)$.

На основе имеющихся показателей рождаемости (Федеральная служба государственной статистики, 2025) рассчитываются возрастные коэффициенты рождаемости девочек по пятилетним возрастным группам матери согласно формуле (1):

$$b_k^j(t) = \frac{f_k^j(t)}{200} \cdot 0.488, \quad k = \overline{4, 10}, \quad (1)$$

где 0.488 – доля девочек среди новорожденных, k соответствует возрастной группе матери.

Показатели смертности женского населения (Центр демографических исследований, 2025, ЕМИСС, 2025) были использованы для получения коэффициентов дожития с использованием формулы (2):

$$p_i^j(t) = 1 - \frac{m_i^j(t)}{200}, \quad i = \overline{1, 10}, \quad (2)$$

где i соответствует возрастной группе матери.

Нетто-коэффициенты воспроизводства для каждого региона j в году t определяются согласно формуле 3 (Тихомиров, Тихомирова, 2023):

$$HKB^j(t) = \sum_{k=4}^{10} \left[b_k^j(t) \cdot \prod_{i=1}^{k-1} p_i \right], \quad i = \overline{1,10}; \quad k = \overline{4,10}. \quad (3)$$

Оценки нетто-коэффициента воспроизводства населения, полученные для каждого региона за период с 1995 по 2022 гг., характеризуются достаточно высокой динамичностью и дифференциацией по регионам России (Калабихина и др., 2022). В целом за период с 1995 по 2022 гг. усредненное по регионам России значение нетто-коэффициента воспроизводства населения возросло с 0,662 до 0,680, т. е. на 2,7% (табл. 1). При этом в динамике этого показателя наблюдается рост в период с 1999 по 2015 гг. до своего пикового значения 0,879, после чего следует снижение на 22,7% до уровня, практически сопоставимого с уровнем 1995 г. В свою очередь, дифференциация регионов по уровню нетто-коэффициента воспроизводства варьирует от 14,0% в 2003–2004 гг. до 18,9% и 20,9% в 2011 и 1995 гг. соответственно.

Все вышеизложенное, а также тот факт, что значение нетто-коэффициента воспроизводства населения в исследуемом периоде не превышает единицы, что свидетельствует о продолжающемся в стране процессе депопуляции, обуславливают необходимость выявления взаимосвязей между индикатором демографического потенциала и характеристиками социально-экономического положения региона (Айвазян и др., 2019) с целью выявления факторов, оказывающих наибольшее влияние на изменение значения нетто-коэффициента воспроизводства, что позволит применять полученные результаты для обоснования стратегий по выравниванию региональных различий и разработке мер, направленных на повышение уровня этого показателя, а следовательно, перехода от депопуляции к режиму расширенного воспроизводства (Сукиасян, 2015).

Таблица 1. Описательная статистика нетто-коэффициента воспроизводства по регионам России за период с 1995 по 2022 гг.

Год	Среднее арифметическое	Стандартное отклонение	Коэффициент вариации, %
1995	0.662	0.138	20.9
2007	0.709	0.116	16.4
2010	0.781	0.140	17.9
2015	0.879	0.128	14.6
2022	0.680	0.113	16.6

Источник: составлено автором

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

В качестве индикаторов социально-экономического положения регионов России были рассмотрены следующие показатели за период с 1995 по 2022 гг., характеризующие различные сферы уровня развития региона (Иванов и др., 2015):

- численность врачей на 10000 человек (X_1);
- количество больничных коек на 10000 человек (X_2);
- преступность на 10000 человек (X_3);
- общая площадь жилых помещений, приходящаяся на человека, м²/чел. (X_4);
- численность безработных человек на 10000 трудоспособных (X_5);
- коэффициенты миграционного прироста на 10000 человек (X_6);
- мощность амбулаторно-поликлинических организаций на 10000 человек (X_7);
- численность посещения музеев на 1000 человек (X_8);
- платные услуги медицины, рублей на человека (X_9);
- расходы на здравоохранение, рублей на человека (X_{10}).

С целью приведения к сопоставимому виду показатели, измеряемые в денежных единицах, были пересчитаны в ценах 2016 г.

На первом этапе с применением методов корреляционного анализа были выявлены взаимосвязи между результирующей переменной (нетто-коэффициента воспроизводства, Y) и вышеуказанными характеристиками, а также взаимные корреляции между объясняющими переменными. Коэффициенты парной корреляции Пирсона и все последующие этапы анализа данных и построения модели базировались на основе усредненных за рассматриваемый период времени значений результирующей и объясняющих переменных, что позволило нивелировать влияние на результат случайных флуктуаций и шумов.

Исходя из анализа значений матрицы парных коэффициентов корреляции были сделаны выводы о достаточно сильной взаимосвязи нетто-коэффициента воспроизводства (Y) населения и такими показателями, как численность врачей на 10000 человек, общая площадь жилых помещений на человека, численность безработных, а также между численностью врачей и общей площадью жилых помещений, мощность амбулаторно-поликлинических организаций и расходы на здравоохранение.

Наиболее сильная отрицательная корреляция наблюдается между нетто-коэффициента воспроизводства и соотношением разводов к бракам, где значение коэффициента составляет - 0.55. Ещё более выраженная отрицательная связь наблюдается между Y и площадью жилых

помещений, коэффициент составляет -0.71, обусловленная сохраняющимся сокращением численности населения при возрастающих темпах ввода в эксплуатацию жилых помещений.

Следует также обратить внимание на положительную взаимосвязь между нетто-коэффициентом воспроизводства и численностью безработных на 10000 трудоспособного населения, что объясняется слабой включенностью женщин, занимающихся уходом за детьми, в рынок труда (Архангельский и др., 2016, Лившиц и др., 2023).

С целью выявления скрытых взаимосвязей, обусловленных одновременным влиянием нескольких показателей на результирующую переменную, а также возможных взаимосвязей между объясняющими переменными были рассчитаны множественные коэффициенты корреляции, которые оказались значимыми на уровне значимости 5%, что означает, что все показатели зависят от других довольно сильно. Наиболее значимыми коэффициентами множественной корреляции оказались у НКВ и численность разводов, тем самым можно сделать вывод, что остальные показатели очень качественно объясняют их вариацию

Полученные результаты позволяют сделать вывод о необходимости проверки совокупности объясняющих переменных на наличие эффекта мультиколлинеарности. В рамках анализа использовались три подхода для оценки мультиколлинеарности: факторы инфляции дисперсии VIF, статистика Фишера и критерий χ^2 (см. табл. 2).

Таблица 2. Расчётные значения VIF и статистики Фишера для каждой объясняющей переменной

Показатели	VIF	F-статистика Фишера	Значимость
X_1	1.626	4.798	**
X_2	5.332	33.210	***
X_3	2.593	12.214	***
X_4	3.422	18.572	***
X_5	3.040	15.643	***
X_6	2.352	10.364	**
X_7	2.486	11.395	**
X_8	1.662	5.075	**
X_9	2.424	10.915	**
X_{10}	4.276	25.117	**

Источник: составлено автором

Несмотря на то, что статистика Фишера подтвердила наличие мультиколлинеарности –

почти все значения $F_{расч}$ оказались выше табличного значения 2.07, а полученное

расчётное значение $\chi^2_{расч} = 86.93$ оказалось выше критического табличного равного 50.99, что указало на необходимость отвергнуть нулевую гипотезу об отсутствии мультиколлинеарности, по результатам расчёта коэффициентов VIF, каждый из которых не превышает порогового значения, равного 10, можно сделать вывод о несущественности влияния мультиколлинеарности на свойство эффективности параметров модели множественной регрессии, построенной на основе имеющихся данных. Это, в свою очередь, позволяет сделать вывод о возможности построения классической линейной регрессионной модели зависимости демографического потенциала от рассматриваемых социально-экономических показателей регионов, учитывающей весь набор объясняющих переменных без необходимости введения дополнительных процедур, направленных на устранение мультиколлинеарности. Однако, в рамках данного исследования с целью проведения сопоставительного анализа и выбора наилучшей модели, была осуществлена процедура построения модели на главных компонентах, которая будет рассмотрена далее.

Таким образом можно сделать вывод о возможности построения классической регрессионной модели, характеризующей

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

Параметры линейного уравнения множественной регрессии с полным перечнем факторов могут быть записаны следующим образом (Сукиасян, 2022):

$$\hat{y} = a_0 + \sum_{p=1}^{10} a_p x_p. \quad (4)$$

На рис. 1 представлена ограниченная («короткая») модель множественной регрессии для моделирования нетто-коэффициента воспроизводства населения от сокращенного числа социально-экономических характеристик по регионам России в среднем за период с 1995 по 2022 гг. Модель включает в себя только те объясняющие переменные, которым соответствуют статистически значимые по тесту Стьюдента параметры модели.

Отбор факторов в регрессионную зависимость делался с помощью метода «a posteriori», который подразумевает под собой пошаговое удаление факторов из модели регрессии с наибольшим p -значением соответствующего параметра до тех пор, пока не останутся только факторы со значимыми коэффициентами регрессии. В модели осталось 5 значимых показателей: численность разводов x_2 , численность преступности x_3 , площадь жилых помещений на человека x_4 , платные медицинские услуги на человека x_9 , расходы на

здоровоохранение на человека X_{10} . Коэффициент детерминации полученной модели составляет 79.3%, а сама модель значима по тесту Фишера на уровне значимости 1%.

Однако для повышения объясняющей способности модели и улучшения качества прогноза целесообразно исключить из выборки регионы, которые характеризуются аномальными значениями показателей по сравнению других регионов или, другими словами, могут быть охарактеризованы как грубые ошибки.

Model 2: OLS, using observations 1-79
Dependent variable: y

	coefficient	std. error	t-ratio	p-value	
const	1.25278	0.0528024	23.73	4.80e-036	***
x2	-7.60332e-05	1.11299e-05	-6.831	2.14e-09	***
x3	0.000827997	0.000141787	5.840	1.35e-07	***
x4	-0.0104054	0.00319018	-3.262	0.0017	***
x9	-3.19364e-05	5.61257e-06	-5.690	2.48e-07	***
x10	1.24790e-05	1.57708e-06	7.913	2.05e-011	***
Mean dependent var	0.714460	S.D. dependent var	0.107065		
Sum squared resid	0.185162	S.E. of regression	0.050363		
R-squared	0.792909	Adjusted R-squared	0.778724		
F(5, 73)	55.90029	P-value (F)	1.36e-23		
Log-likelihood	127.1148	Akaike criterion	-242.2296		
Schwarz criterion	-228.0129	Hannan-Quinn	-236.5339		

Рис. 1. Ограниченная модель множественной регрессии зависимости демографического потенциала регионов от уровня их социально-экономического развития.

Источник: составлено автором

В процессе подготовки данных для построения регрессионной модели особое внимание было уделено анализу на наличие выбросов. Для этого использовался поэтапный подход, включающий сочетание одномерных и многомерных методов. На первом этапе применялся критерий Смирнова–Граббса, позволяющий обнаружить грубые ошибки в каждом из показателей по отдельности, по результатам которого были выявлены 8 регионов, характеризующихся аномальными значениями тех или иных признаков: г. Москва г. Санкт-Петербург, Магаданская область, Республики Дагестан, Ингушетия и Тыва, Сахалинская область и Чукотский автономный округ.

Для дополнительной проверки применялся дисперсионный критерий Граббса. Он сравнивает дисперсию выборки с усечённой дисперсией (то есть без потенциального выброса). Такой подход позволил подтвердить, что не все наблюдаемые отклонения статистически значимы, а некоторые могли быть следствием обычной вариативности данных. В результате было подтверждено наличие аномальных значений в данных трех регионов: город Санкт-Петербург, Республика Ингушетия, а также Чукотском автономном округе.

Однако поскольку одномерный анализ не позволяет выявить ситуации, когда регион существенно отличается от основного массива субъектов по совокупности факторов, с целью

более полного анализа исходной выборки для каждого из отмеченных регионов была рассчитана многомерная статистика Хотеллинга. Статистически значимыми выбросами по этому критерию оказались г. Санкт-Петербург ($T^2 = 56.35$), Республика Ингушетия ($T^2 = 60.32$) и Чукотский автономный округ ($T^2 = 44.91$), значения которых превышают критическое значение, равное 24.38. Соответственно данные регионы были исключены из выборки с целью получения модели с лучшими по сравнению с приведенной выше моделью характеристиками качества.

Такая модификация позволила сохранить необходимый уровень репрезентативности данных и обеспечить оптимальный баланс между коэффициентом детерминации и стандартной ошибкой в последующей регрессионной модели (см. рис. 2).

Model 2: OLS, using observations 1-79
Dependent variable: y

	coefficient	std. error	t-ratio	p-value	
const	1.25278	0.0528024	23.73	4.80e-036	***
x2	-7.60332e-05	1.11299e-05	-6.831	2.14e-09	***
x3	0.000827997	0.000141787	5.840	1.35e-07	***
x4	-0.0104054	0.00319018	-3.262	0.0017	***
x9	-3.19364e-05	5.61257e-06	-5.690	2.48e-07	***
x10	1.24790e-05	1.57708e-06	7.913	2.05e-011	***
Mean dependent var	0.714460	S.D. dependent var	0.107065		
Sum squared resid	0.185162	S.E. of regression	0.050363		
R-squared	0.792909	Adjusted R-squared	0.778724		
F(5, 73)	55.90029	P-value(F)	1.36e-23		
Log-likelihood	127.1148	Akaike criterion	-242.2296		
Schwarz criterion	-228.0129	Hannan-Quinn	-236.5339		

Рис. 2. Ограниченная модель множественной регрессии зависимости демографического потенциала регионов

Источник: составлено автором

Модель без выбросов демонстрирует более высокие показатели качества: значение

коэффициента детерминации (R^2) составляет 0.828 против 0.793 у второй модели, а нормированный коэффициент детерминации достигает 0.815, что также выше, чем 0.779 у модели с выбросами (см. табл. 3). Дополнительно стоит отметить снижение стандартной ошибки: с 0.050 до 0.046, что указывает на более устойчивые оценки параметров и меньшую дисперсию остатков. Эти различия не критичны по числовому выражению, но в совокупности говорят в пользу более высокой точности и интерпретируемости модели без выбросов.

Таблица 3. Сравнение регрессионных моделей зависимости демографического потенциала от уровня развития социально-экономического положения регионов со значимыми параметрами

Модель без выбросов	Модель с учетом выбросов
Регрессионная статистика	Регрессионная статистика

Множественный R	0,910	Множественный R	0,890
R-квадрат	0,828	R-квадрат	0,793
Нормированный R-квадрат	0,815	Нормированный R-квадрат	0,779
Стандартная ошибка	0,046	Стандартная ошибка	0,050

Источник: получено автором

Однако данную модель целесообразно также проверить на автокорреляцию остатков. С этой целью была рассчитана статистика Дарбина-Уотсона. Полученное значение составило $DW = 1.91$, что близко к пороговому значению 2. Это говорит об отсутствии выраженной положительной или отрицательной автокорреляции. Однако, поскольку тест Дарбина-Уотсона имеет зону неопределённости, важно учитывать критические границы, определяемые числом наблюдений и количеством регрессоров. При 76 наблюдениях и 5 объясняющих переменных нижняя критическая граница составляет приблизительно 1.50, а верхняя – 1.71. Таким образом, наше значение $DW = 1.91$ находится выше верхней границы. Это позволяет с уверенностью заключить, что автокорреляции первого порядка в модели нет.

Для подтверждения этого результата был применён тест Бройша-Годфри, который не имеет ограничений на включение лагов зависимой переменной в качестве факторов и является более универсальным при проверке автокорреляции различных порядков. Результаты теста Бреуша-Годфри для автокорреляции первого порядка: статистика $LM = 2.12$, p -значение (χ^2): 0.145; статистика Фишера $F = 1.84$, p -значение (F): 0.18.

Оба p -значения превышают стандартный уровень значимости 5%, что не позволяет отвергнуть нулевую гипотезу об отсутствии автокорреляции. Таким образом, оба теста согласованно показали, что автокорреляция в модели отсутствует. Это подтверждает корректность спецификации модели в части предпосылки независимости ошибок. Отсутствие автокорреляции также благоприятно сказывается на точности прогнозов и интерпретируемости результатов: не возникает необходимости в применении альтернативных методов оценки (например, обобщённых наименьших квадратов) или введении лагов.

Исходя из вышеизложенного, модель без выбросов и с отобранными значимыми факторами представляется более надёжной и уместной для дальнейшего анализа и прогнозирования. Очистка данных позволила улучшить статистические характеристики модели без излишнего усложнения её структуры.

С целью повышения качества построенных моделей была также предпринята попытка построения модели, в качестве объясняющих переменных в которой вместо исходных показателей использованы главные компоненты, что позволило устранить эффект мультиколлинеарности при сохранении существенной доли информации об исходных переменных.

В результате анализа собственных значений матрицы корреляций было выявлено, что первая главная компонента имеет наибольшее собственное значение, равное 3.759, что составляет 37.6% от общей дисперсии исходных данных. Вторая и третья главная компонента имеют 17.4% и 14.9% от общей дисперсии данных соответственно. Таким образом, первые три главные компоненты объясняют 69.9% дисперсии исходных данных. Собственные значения последующих главных компонент соответственно объясняют все меньшую часть дисперсии данных (Jindrová A. и др, 2013).

Значения элементов собственных векторов было решено использовать в качестве объясняющих переменных в моделях регрессии с целью сопоставления их с приведенными выше моделями для изучения степени изменения характеристик качества моделей после устранения эффекта мультиколлинеанности (Тихомиров, Тихомирова, 2024).

Построенная регрессионная модель на главных компонентах, приведенная на рис. 3 и не включающая объясняющие переменные, характеризующиеся статистически не значимыми параметрами, имеет высокий коэффициент детерминации, равный 0,819, а также является статистически значимой, на что указывает p -значение F -статистики Фишера.

Model 8: OLS, using observations 1-79
Dependent variable: s_y

	coefficient	std. error	t-ratio	p-value	
PC1	0.187352	0.0258638	7.244	3.92e-010	***
PC2	-0.524765	0.0380314	-13.80	6.62e-022	***
PC3	-0.244181	0.0411059	-5.940	9.25e-08	***
PC5	-0.267747	0.0645248	-4.150	9.01e-05	***
PC6	0.139910	0.0676246	2.069	0.0421	**
PC7	0.211049	0.0853987	2.471	0.0158	**
PC10	0.612180	0.137265	4.460	2.96e-05	***
Mean dependent var	0.000000	S.D. dependent var	1.000000		
Sum squared resid	14.12312	S.E. of regression	0.442893		
R-squared	0.818934	Adjusted R-squared	0.803846		
F(6, 72)	54.27430	P-value(F)	9.14e-25		
Log-likelihood	-44.09158	Akaike criterion	102.1832		
Schwarz criterion	118.7693	Hannan-Quinn	108.8281		

Рис. 3. Ограниченная модель множественной регрессии на главных компонентах.

Источник: составлено автором

Аналогичным образом были сконструированы главные компоненты после удаления из исходной выборки обнаруженных на основе статистических тестов выбросов. Сочетание метода главных компонент с робастной регрессией позволяет, с одной стороны, уменьшить размерность исходных данных за счёт перехода к новым, некоррелированным признакам, а с другой – повысить устойчивость модели к выбросам. Это особенно важно, когда анализ проводится по ограниченному числу наблюдений при высокой размерности пространства признаков (Тихомиров, Тихомирова, 2019, Jindrová A. et al., 2013).

В результате анализа собственных значений для матрицы корреляций после удаления выбросов выявлено, что первая компонента объясняет 36,3% дисперсии, вторая – 20%. Вместе они объясняют большую часть дисперсии в 56,3%. Собственные значения у последующих главных компонент уменьшаются и имеют менее 10% объясненной дисперсии.

Построенная регрессионная модель, приведенная на рис. 4, демонстрирует высокий уровень объясняющей способности: скорректированное значение коэффициента детерминации R^2

составляет 0.819. Высокое значение F -статистики и её значимость свидетельствуют о том, что модель в целом является статистически значимой.

Сравнительный анализ двух регрессионных моделей – одной, построенной на основе всех наблюдений, и другой, из которой исключены выбросы, – демонстрирует, что удаление аномальных регионов способствует улучшению качества модели. В частности, значение коэффициента детерминации увеличивается с 0.819 до 0.834, а стандартная ошибка снижается с 0.442 до 0.422. Эти показатели указывают на то, что обновлённая модель более точно описывает поведение зависимой переменной при меньшем уровне ошибок.

Model 11: OLS, using observations 1-76
Dependent variable: s_y

	coefficient	std. error	t-ratio	p-value	
PC1	0.208696	0.0255969	8.153	9.54e-012	***
PC2	0.471933	0.0345205	13.67	1.91e-021	***
PC3	0.392621	0.0490132	8.011	1.75e-011	***
PC7	0.187032	0.0741385	2.523	0.0139	**
PC9	0.464470	0.113349	4.098	0.0001	***
PC10	-0.398481	0.125819	-3.167	0.0023	***
Mean dependent var	0.000000	S.D. dependent var	1.000000		
Sum squared resid	12.47839	S.E. of regression	0.422212		
R-squared	0.833621	Adjusted R-squared	0.821737		
F(5, 70)	70.14548	P-value(F)	6.90e-26		
Log-likelihood	-39.18339	Akaike criterion	90.36679		
Schwarz criterion	104.3512	Hannan-Quinn	95.95563		

Рис. 4. Ограниченная робастная модель на главных компонентах.

Источник: составлено автором

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

В ходе исследования было построено четыре различные модели, отличающиеся по методике и составу данных. Две из них базировались на классическом множественном регрессионном анализе (МНК), две другие – на использовании главных компонент (МГК). При этом каждая из методик была реализована дважды: с полным набором наблюдений и с предварительным исключением выбросов (см. табл. 4).

Обычная модель без предварительной очистки данных показала приемлемое значение коэффициента детерминации равно 0.793 при стандартной ошибке 0.05. Тем не менее, присутствие выбросов отразилось на стабильности модели: *p*-значения некоторых факторов указывали на их статистическую незначимость, а прогнозные свойства модели вызывали сомнения.

После удаления аномальных наблюдений модель на тех же исходных признаках улучшилась: коэффициент детерминации вырос до 0.828, а ошибка уменьшилась до 0.046. Это указывает на улучшение качества аппроксимации наблюдаемых значений. Однако анализ значений коэффициентов выявил, что большая их часть близка к нулю, а значит, модель при прогнозировании будет слабо реагировать на изменения входных переменных.

Таблица 4. Сопоставительный анализ характеристик качества построенных в ходе исследования моделей

	Количество объясняющих переменных	Коэффициент детерминации	Стандартная ошибка	Значимость модели
Модель с выбросами	5	0.793	0.050	***
Модель без выбросов	5	0.828	0.046	***
МГК с выбросами	7	0.818	0.442	***
МГК без выброса	6	0.834	0.422	***

Источник: составлено автором

Применение метода главных компонент позволило сократить многомерность данных. Модель МГК с выбросами характеризуется коэффициентом детерминации, равным 0.818, но при этом имела стандартную ошибку 0.442 — это выше, чем у обычной модели. После

$$R^2$$

исключения выбросов модель по МГК показала лучший результат среди всех: коэффициент детерминации равен 0.834, стандартная ошибка — 0.422. Однако стандартная ошибка достаточно большая, что даст недостаточно точный прогноз при изменении показателей.

На основании проведённого сравнения моделей было принято решение использовать для прогнозирования модель, построенную без выбросов на основе классического метода наименьших квадратов. Она продемонстрировала наилучший баланс между высокой объясняющей способностью и невысокой стандартной ошибкой, сохранив при этом интерпретируемость коэффициентов и устойчивость структуры модели. Построенная модель адекватно отражает динамику зависимой переменной и может быть применена для прогнозирования с различными сценариями, включая ухудшение и улучшение условий. Её

структура позволяет провести варьирование значений факторов и оценить, как возможные изменения в социальной среде отразятся на демографической ситуации в стране.

Для демонстрации прикладной ценности модели был проведён сценарный анализ, отражающий потенциальные изменения демографической ситуации при корректировке социально-экономических показателей. Была смоделирована ситуация, характеризующаяся улучшением всех социально-экономических условий в рамках их фактической вариабельности. Рассматривался сценарий, при котором численность разводов на 10000 браков x_2 снижается на 15%, преступность на 10000 человек x_3 уменьшается на 1%, жилищные условия x_4 улучшаются на 3%, мощность амбулаторно-поликлинических организаций x_7 возрастает на 10%, расходы на здравоохранение x_{10} увеличиваются на 5%.

Данный оптимистичный сценарий продемонстрировал увеличение прогнозного значения нетто-коэффициента воспроизводства большинства регионов, общий прирост составил 8.7%. Больше всего замечается прирост в Магаданской, Ленинградской и Мурманской областях. Однако в некоторых регионах улучшение социальных факторов сказалось отрицательно на уровне нетто-коэффициента воспроизводства. Это относится к Республикам Ингушетия, Дагестан, а также Кабардино-Балкарской, что может быть обусловлено как спецификой региональной социально-экономической структуры, так и высокой чувствительностью этих субъектов к изменениям отдельных факторов. Это подчеркивает необходимость индивидуального подхода при разработке социальных мер и показывает, что универсальные улучшения не всегда эффективны в рамках всей страны.

Стоит рассмотреть также пессимистичный сценарий, при котором: численность разводов на 10000 браков x_2 увеличивается на 12%, преступность на 10000 человек x_3 увеличивается на 4%, жилищные условия x_4 ухудшаются на 7%, мощность амбулаторно-поликлинических организаций x_7 уменьшаются на 7%, расходы на здравоохранение x_{10} уменьшаются на 8%. Такой сценарий, как и ожидалось, негативно отразился на уровне демографического потенциала большинства регионов, в целом за все регионы упадок составил 10.4 процентных пункта. Больше всего данное снижение отразилось на Магаданской, Ленинградской и Мурманской областях. В свою очередь Республики Ингушетия, Дагестан, а также Кабардино-Балкарская продемонстрировали прирост, что подтверждает специфику этих регионов и необходимость реализации адресных мер, направленных на выравнивание демографической ситуации в регионах страны.

ЗАКЛЮЧЕНИЕ

В ходе исследования была построена регрессионная модель с целью выявления влияния социально-экономических факторов на уровень демографического потенциала в регионах страны, в частности на значение нетто-коэффициента воспроизводства. Основное внимание было уделено предварительному анализу массива данных на наличие грубых ошибок (выбросов), а также влияния эффекта мультиколлинеарности на эффективность оценок параметров и характеристик качества построенных моделей.

Дополнительно в исследовании была предпринята попытка построения моделей с использованием метода главных компонент (МГК), а также робастной модели на основе МГК. Однако применение этих подходов не дало существенного улучшения качества модели. Это указывает на то, что использование более сложных оценочных процедур не всегда приводит к повышению точности прогноза и должно быть обосновано спецификой данных и задачей анализа.

Проведённый сценарный анализ показал, что при улучшении значений социально значимых факторов большинство регионов демонстрируют положительную динамику модельных значений показателя. Однако не во всех субъектах наблюдается прирост: в некоторых республиках Северного Кавказа отмечено снижение прогнозируемого значения. Это объясняется спецификой региональных условий, традиционно высокими базовыми значениями рождаемости и культурными особенностями, слабо коррелирующими с рассматриваемыми социально-экономическими переменными.

Практическая значимость выполненной работы заключается в возможности применения полученной модели для оценки потенциального эффекта изменений в социальной политике. При корректной интерпретации коэффициентов можно формировать рекомендации для управления демографической ситуацией на региональном уровне, а также использовать подходы стандартизации переменных в аналогичных исследованиях.

Таким образом, по результатам проведенного исследования была успешно достигнута цель анализа влияния социально-экономических факторов на воспроизводственные процессы в России. Проведённое моделирование и анализ позволяют утверждать, что несмотря на отдельные отклонения, модель отражает общие закономерности и может служить основой для выработки эффективных решений в области демографической политики.

В качестве вопросов и перспектив для дальнейшего исследования автором предполагается постановка и попытка решения задач, связанных с построением модели с сохранением грубых ошибок в выборке, но их корректировкой методами устойчивого оценивания (например, винзорирование, модель Хубера), а также оценки модели на панельных данных с учетом специфики региональных различий.

Важно подчеркнуть, что Российская Федерация является большой, многонациональной страной с выраженными региональными различиями в социально-экономическом развитии, структуре населения и миграционных тенденциях, что нашло свое подтверждение в рамках проведенного исследования. Учитывая это, построение единой модели, одинаково хорошо отражающей реальность всех регионов, представляет собой весьма сложную задачу, поэтому в одним подходом к улучшению качества прогноза демографического потенциала и обоснования влияния на его уровень изменений социально-экономических условий может быть решение задачи построения кластерной модели, обобщающей в кластеры схожие по своим социально-экономическим и демографическим характеристикам регионы. Тем не менее, полученные результаты позволяют использовать модель в качестве инструмента для оценки эффективности мер, направленных на выравнивание и обеспечение поступательного развития демографического процесса.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Айвазян С.А., Афанасьев М.Ю., Кудров А.В.** (2019). Индикаторы основных направлений социально-экономического развития и их агрегаты в пространстве характеристик региональной дифференциации // *Прикладная эконометрика*. № 2(54), с. 51–69. [Ayvazyan S. A., Afanasyev M. Yu., Kudrov A.V. (2019). Indicators of the main directions of socio-economic development and their aggregates in the space of characteristics of regional differentiation. *Applied Econometrics*, 2(54), 51–69. (in Russian).].
- Архангельский В.Н., Иванова А.Е., Рыбаковский Л.Л.** (2016). Результативность демографической политики России. *Экон-Информ*. - 307 с. [Arkhangelsky V.N., Ivanova A.E., Rybakovsky L.L. (2016). The effectiveness of Russia's demographic policy. *Ekon-Inform* (in Russian).].
- Баева Ф.Г., Уразова Е.А.** (2020). Демографический потенциал как объект государственного управления // *Ученые заметки ТОГУ*. № 4, с. 339–342. [Baeva F.G., Urazova E.A. (2020). Demographic potential as an object of public administration. *Scientific Notes of TOGU*, 4, 339–342. (in Russian).].
- Валентей Д.И.** О системе демографических знаний (1973). (2016) // *Вестник Московского университета. Серия 6: Экономика*. №4, 134–148. [Valentey D.I. On the system of demographic knowledge (1973). (2016). *Bulletin of the Moscow University. Series 6: Economics*, 4, 134–148. (in Russian).].
- Госпрограммы РФ** (2023). Государственная программа «Развитие здравоохранения». Дата обращения 02.06.2025,

<https://programs.economy.gov.ru/gp/-/subject/-/direction/7/gp/1/gpVersion/10395> [State Programs of the Russian Federation (2023). The state program «Healthcare Development». Date of request 06.02.2025, <https://programs.economy.gov.ru/gp/-/subject/-/direction/7/gp/1/gpVersion/10395>. (in Russian).].

Госпрограммы РФ (2023). Государственная программа «Социальная поддержка». Дата обращения 02.06.2025, <https://programs.economy.gov.ru/gp/-/subject/-/direction/7/gp/18/gpVersion/10374> [State Programs of the Russian Federation (2023). The state program «Social support». Date of request 06.02.2025. <https://programs.economy.gov.ru/gp/-/subject/-/direction/7/gp/18/gpVersion/10374>. (in Russian).].

ЕМИСС (2025). Возрастные коэффициенты смертности. Дата обращения 22.06.2025 <https://www.fedstat.ru/indicator/30974> [ЕМИСС (2025). Age-related mortality rates. Date of request 22.06.2025. <https://www.fedstat.ru/indicator/30974>. (in Russian).].

Зверева Н.В. (2006). Экономика семьи и дети (по материалам социально-демографического исследования). Вестник Московского университета. Серия 6: Экономика. №2, 79–95. [Zvereva N.V. (2006). The economy of the family and children (based on the materials of socio-demographic research). *Bulletin of the Moscow University. Series 6: Economics*, 2, 79–95. (in Russian).]

Иванов В.Н., Овсиенко Ю.В., Сухова Н.Н. (2014). Социальная сфера России в 1990-2000-е годы (Демографические проблемы) // Экономика и математические методы. Т. 50, № 2, с. 3-15. [Ivanov V.N., Ovsienko Yu.V., Sukhova N.N. (2014). The social sphere of Russia in the 1990s and 2000s (Demographic problems). *Economics and mathematical methods*, Vol. 50, 2. p. 3-15. (in Russian).].

Калабихина И.Е., Казбекова З.Г., Клименко Г.А., Колотуша А.В. (2022). Демографический рейтинг регионов по активности публикаций СМИо материнском (семейном) капитале // Прикладная эконометрика, № 3(67), 46–73. [Kalabikhina I. E., Kazbekova Z. G., Klimenko G. A., Kolotusha A.V. (2022). Demographic rating of regions by activity in SMIo publications on maternal (family) capital. *Applied Econometrics*, 3(67), 46–73. (in Russian).]

Лившиц В.Н., Тищенко Т.И., Фролова М.П. (2011). Кризис в России - взгляд позавчера, вчера, сегодня и завтра // Экономика и математические методы. Т. 47, № 1. с. 66-81. [Livshits V.N., Tishchenko T.I., Frolova M.P. (2011). The crisis in Russia - a look at the day before yesterday, yesterday, today and tomorrow. *Economics and Mathematical Methods*, Vol. 47, 1, p. 66-81. (in Russian).].

- Лившиц В.Н., Шаталова О.М., Касаткина Е.В.** (2023). Межрегиональная дифференциация в РФ: эмпирический анализ влияния территориальной локализации отраслей на уровень экономической активности регионов // *Экономика и математические методы*. Т. 59, № 3. с. 77-90. [**Livshits V.N., Shatalova O.M., Kasatkina E.V.** (2023). Interregional differentiation in the Russian Federation: an empirical analysis of the impact of territorial localization of industries on the level of economic activity of regions. *Economics and Mathematical Methods*, Vol. 59, 3. p. 77-90. (in Russian).].
- Макаров П.Ю.** (2019). Социально-демографический потенциал региона: показатели и современные тенденции // *Ученые записки*, № S2, с. 6–10. [**Makarov P.Y.** (2019). Socio-demographic potential of the region: indicators and current trends. *Scientific Notes*, S2, 6–10. (in Russian).].
- Никонова А.А., Красильникова Е.В.** (2011). Кризис человеческого потенциала России и методы анализа // *Экономика и математические методы*, Т. 47, № 2. с. 84-95. [**Nikonova A.A., Krasilnikova E.V.** (2011). The crisis of Russia's human potential and methods of analysis. *Economics and Mathematical Methods*, Vol. 47, 2. p. 84-95.].
- Ростовская Т.К., Ситковский А.М.** (2024). Ресурсы демографического развития: к вопросу об унификации понятий в демографических исследованиях // *Экономические и социальные перемены: факты, тенденции, прогноз*, № 1, с. 178–200. [**Rostovskaya, T. K., Sitkovsky A.M.** (2024). Demographic development resources: on the issue of unification of concepts in demographic research. *Economic and social changes: facts, trends, forecast*, 1, 178-200. (in Russian).].
- Рыбаковски О.Л.** (2023). Демографический потенциал: сущность, структура и основные факторы. *Уровень жизни населения регионов России*, № 3, с. 319–326. [**Rybakovsky O.L.** (2023). Demographic potential: the essence, structure and main factors. *Standard of living of the population of the Russian regions*, 3, 319–326. (in Russian).].
- Сукиасян А.Г.** (2023). Многомерный статистический анализ преждевременной смертности населения регионов России от социально значимых причин. // В С.У.Увайсов (Ред.) *Инновационные, информационные и коммуникационные технологии : Сборник трудов XX Международной научно-практической конференции* (с. 169–174). М.: Ассоциация выпускников и сотрудников ВВИА им. проф. Жуковского. [**Sukiasyan A.G.** (2023). Multidimensional statistical analysis of premature mortality of the population of Russian regions from socially significant causes. In S.U.Uvaisov (Ed.) *Innovative, information and communication technologies : Proceedings of the XX International Scientific and Practical Conference* (pp. 169–174). М.: Association of Graduates and Staff of the VVIA named after Prof. Zhukovsky. (in Russian).].

- Сукиасян А.Г.** (2015). К вопросу оценки демографического потенциала населения регионов России // *Научные труды Вольного экономического общества России*. Т. 192, № 3. с. 410-424. [**Sukiasyan A.G.** (2015). On the issue of assessing the demographic potential of the population of the regions of Russia. *Scientific Papers of the Free Economic Society of Russia*, Vol. 192, 3, 410-424. (in Russian).].
- Сукиасян А.Г.** (2022). Модели управления потенциалом воспроизводства населения регионов России на основе методов машинного обучения // *Народонаселение*. Т. 25, № 4. с. 16-29. [**Sukiasyan A.G.** (2022). Models for managing the population reproduction potential of Russian regions based on machine learning methods. *Population*. Vol. 25, 4, 16-29. (in Russian).].
- Сукиасян А.Г.** (2024). Применение методов многомерного статистического анализа для исследования трудового потенциала регионов России // *Экономика строительства*, № 11, с. 405-409. [**Sukiasyan A.G.** (2024). Application of multidimensional statistical analysis methods to study the labor potential of Russian regions. *Economics of Construction*, 11, 405-409. (in Russian).].
- Тихомиров Н.П., Тихомирова Т.М.** (2022). Методы оценки и регулирования режима воспроизводства населения // *Вестник Российского экономического университета имени Г.В. Плеханова*, Т. 19, № 4(124), 7–15. [**Tikhomirov N.P., Tikhomirova T.M.** (2022). Methods of assessment and regulation of the reproduction regime of the population. *Bulletin of the Plekhanov Russian University of Economics*, vol. 19, 4(124), 7–15. (in Russian).].
- Тихомиров Н.П., Тихомирова Т.М.** (2024). Индикаторы режима воспроизводства населения // *Вестник Алтайской академии экономики и права*, № 5-1, с. 135-141. [**Tikhomirov N.P., Tikhomirova T.M.** (2024). Indicators of the population reproduction regime. *Bulletin of the Altai Academy of Economics and Law*, 5-1, 135-141. (in Russian).].
- Тихомиров Н.П., Тихомирова Т.М.** (2019). Оценка и управление потенциалом воспроизводства населения России // *Федерализм*, № 3(95), с. 51-71. [**Tikhomirov N.P., Tikhomirova T.M.** (2019). Assessment and management of the reproduction potential of the Russian population. *Federalism*, 3(95), 51-71. (in Russian).].
- Тихомиров Н.П., Тихомирова Т.М.** (2023). Эконометрические методы обоснования мер по переходу к режиму расширенного воспроизводства населения в России // *Вестник Российского экономического университета имени ГВ Плеханова*, 3, с. 18–28. [**Tikhomirov N.P., Tikhomirova T.M.** (2023). Econometric methods of substantiating measures for the transition to a regime of expanded reproduction of the population in Russia. *Bulletin of the Plekhanov Russian University of Economics*, 3, 18–28. (in Russian).].

- Федеральная служба государственной статистики** (2025). *Демографический ежегодник России*. Дата обращения 07.07.2025. <https://rosstat.gov.ru/folder/210/document/13207>. [Federal State Statistics Service (2025). *Demographic Yearbook of Russia*. Date of request: 07.07.2025. <https://rosstat.gov.ru/folder/210/document/13207>. (in Russian).].
- Центр демографических исследований Российской экономической школы** (2025). *Российская база данных по рождаемости и смертности (РосБРУС)*. Дата обращения 07.07.2025. <https://www.nes.ru/demogr-fermort-data?lang=ru>. [Center for Demographic Research of the Russian School of Economics (2025). *The Russian database on fertility and mortality (RosBRIS)*. Date of request: 07.07.2025. <https://www.nes.ru/demogr-fermort-data?lang=ru>. (in Russian).].
- Яковец Т.Ю., Голубков В.В.** (2018). Прогноз демографической ситуации в России до 2033 г. // *Экономика и математические методы*. Т. 54, №4. с. 71-87. [Yakovets T.Y., Golubkov V.V. (2018). The forecast of the demographic situation in Russia until 2033. *Economics and mathematical methods*, Vol. 54, 4, 71-87. (in Russian).].
- Balbo N., Billari F.C., Mills M.** (2013). Fertility in Advanced Societies: A Review of Research // *European Journal of Population*, Vol. 29, 1, p. 1-38.
- Jindrová A. et al.** (2013). Dimensionality reduction of quality of life indicators. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*. Vol. 60, 7, 147–154.

Modeling the demographic potential of Russian regions considering differences in their socio-economic state

© 2025 A.G. Sukiasyan

Received 00.00.2025

A.G. Sukiasyan

Plekhanov Russian University of Economics, Moscow, Russia; e-mail: sukiasyan.ag@rea.ru

This research was performed in the framework of the state task in the field of scientific activity of the Ministry of Science and Higher Education of the Russian Federation, project "Models, methods, and algorithms of artificial intelligence in the problems of economics for the analysis and style transfer of multidimensional datasets, time series forecasting, and recommendation systems design", grant no. FSSW-2023-0004.

Abstract. One of the key problems in Russia and its regions is the problem of demographic decline, expressed by low birth rate that has persisted for several decades and is aggravated by the high rate of premature mortality among young people and the working-age population. The demographic decline that has been going on since the early 90s of the 20th century is caused by an increase in the demographic burden, an aging population, and an imbalance in the labor market. To solve these problems, the state develops and takes measures aimed at increasing the birth rate, expressed in the introduction of social benefits, as well as other measures to support families with children. However, in the long term, the demographic policy measures implemented have not led to significant changes in the trend of the demographic process taking place in the country. This indicates the existence of factors that have a more significant impact on the dynamics of the demographic process in the country and its regions and necessitates their identification using modern statistical and mathematical tools. The article presents the results of constructing models, which considers the specifics of the initial data, several variants of models describing the dependence of demographic potential on the differing

conditions of socio-economic development of the regions of Russia. At the same time, the article used the net reproduction rate indicator as a characteristic of demographic potential, which characterizes both the specifics of fertility and mortality in the Russian regions. A comparative analysis of four models is presented, taking into account the effect of multicollinearity and the presence of gross errors. Based on the measures of quality of models characteristics selected during the comparison, a scenario analysis was carried out, confirming the high sensitivity of demographic potential to changes in socio-economic conditions in the country's regions.

Key words: demographic potential, net reproduction rate, regression, multicollinearity, outliers, scenario analysis.

JEL Classification: C21, C51, C52, C82.