

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**МОДЕЛЬ ДАРВИНА ДЛЯ РОССИЙСКИХ  
ПРОСТРАНСТВЕННЫХ ДАННЫХ**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 412 группы

направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Калиниченко Ильи Максимовича

Научный руководитель

старший преподаватель

\_\_\_\_\_

А. Д. Луньков

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2022

## ВВЕДЕНИЕ

**Актуальность темы.** В нынешних экономических условиях, когда открыты границы для импорта и экспорта товаров и ресурсов, и производственные технологии стали трансграничными, глобальными стали и вопросы охраны здоровья населения. Пандемия COVID-19, которая достаточно быстро распространилась по всему миру, парализовала нормальное функционирование большинства секторов экономики. Системы здравоохранения многих стран были не готовы к масштабным последствиям распространения коронавирусной инфекции, что потребовало колоссальных вложений ресурсов как человеческих, так и материальных для восстановления и сохранения уровня экономики.

В условиях неоднородности российских регионов и конкурентности распределения бюджетных средств важным является выявление оптимальной доли расходов на здравоохранение в ВРП для достижения лучшего экономического эффекта. Самыми важными критериями оценки эффективности расходов на здравоохранение могут служить их влияние на ожидаемую продолжительность жизни в конкретном регионе и темпы экономического роста.

**Цель бакалаврской работы** - эконометрическое исследование на определение средней оптимальной доли расходов на здравоохранение и спорт в ВРП, при которой темп среднего экономического роста регионов будет максимальным.

**Объект исследования** - Регрессионная модель, описывающая государственные расходы на здравоохранение и экономический рост в России.

**Предмет исследования** - Проверка основной гипотезы о существовании оптимального уровня расходов на здравоохранение и спорт, при котором положительное влияние на темпы роста ВРП является максимальным.

Для достижения поставленной цели необходимо решить **задачи**:

- Рассмотреть основные понятия, относящиеся к панельным данным;
- Рассмотреть основные положения пространственной эконометрики;
- Рассмотреть динамику и структуру расходов на здравоохранение в России;
- Рассмотреть гипотезу о существовании оптимального уровня расходов

на здравоохранение и спорт, при котором положительное влияние на темпы роста ВРП является максимальным;

— Проверить гипотезу с помощью пространственной модели Дарбина.

**Практическая значимость** проводимого исследования состоит в том, что на основании построенной пространственной модели Дарбина можно проанализировать региональные данные для определения средней оптимальной доли расходов на здравоохранение и спорт в ВРП для всех регионов России.

**Структура и содержание работы.** Работа состоит из введения, четырех разделов, заключения, списка использованных источников, содержащего 21 наименование, и трех приложений. Общий объем работы составляет 47 страниц.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы работы, формулируется цель работы и решаемые задачи, отмечается практическая значимость полученных результатов.

В **первом** разделе приводятся основные понятия панельных данных и описание модели с фиксированным эффектом.

Модель с фиксированным эффектом (fixed effect model) описывается регрессионной зависимостью вида  $y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}$ , в котором эффекты  $i$ -го региона  $\alpha_i$  являются неизвестными параметрами, которые подлежат оцениванию.

Панельные данные состоят из наблюдений одних и тех же экономических единиц или объектов, которые осуществляются в последовательные периоды времени. Если ввести фиктивные переменные для каждой экономической единицы:  $d_{ij} = 1$ , если  $i = j$ , и  $d_{ij} = 0$ , если  $i \neq j$ , то зависимость вида  $y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}$  может быть переписана в более привычном виде линейной регрессии

$$y_{it} = \sum_{j=1}^n \alpha_j d_{ij} + x'_{it}\beta + \varepsilon_{it}.$$

Для записи данного уравнения в матричной форме можно все фиктив-

ные переменные объединить в одну матрицу:

$$y = D\alpha + X\beta + \varepsilon.$$

Это соотношение можно рассматривать как стандартную модель регрессии и получать оценки параметров  $\alpha, \beta$  с помощью обычного метода наименьших квадратов. В панельных данных рост числа наблюдений может происходить как за счет увеличения количества экономических единиц  $n$ , так и за счет увеличения длительности наблюдений  $T$ . В первом случае происходит рост числа оцениваемых параметров, и гарантировать состоятельность, по крайней мере для оценок параметров  $\alpha$ , нельзя. Во втором случае МНК-оценки состоятельны, но большие временные интервалы при небольшом числе экономических единиц в панельных данных встречаются редко.

В панельных данных среди независимых переменных  $x_{it}$  могут быть такие, которые не меняются во времени для каждой экономической единицы. Например, при анализе заработной платы в число объясняющих факторов включают пол и/или расовую принадлежность индивидуума. Модель с фиксированным эффектом не позволяет идентифицировать соответствующие таким переменным коэффициенты.

Во **втором** разделе рассматриваются основные положения пространственной эконометрики.

В основе пространственной методологии лежит **матрица пространственных весов**, по значениям элементов которой можно судить о наличии связей между регионами и их интенсивности.

Элемент матрицы  $w_{ij}$  характеризует влияние некоторого процесса региона  $j$  на тот же процесс регион  $i$ . Все матрицы пространственных весов – квадратные. По диагонали матрицы стоят нули, так как в матрице исключается возможность влияния региона на себя.

В данной работе рассматриваются следующие типы матриц:

1. Матрица соседства.

Простейшим способом учета пространственных связей является бинарная матрица граничных соседей – предполагается, что на экономику объекта могут влиять только граничащие с ним регионы:

$$w_{ij} = \begin{cases} 1, & \text{если } j \text{ граничит с } i; \\ 0, & \text{если } j \text{ не граничит с } i. \end{cases}$$

Другим видом матриц соседства является матрица  $k$  ближайших соседей. Для каждого региона  $i$  рассчитываются расстояния до всех регионов  $d_{ij}$ . Затем из них выбираются  $k$  наименьших расстояний. Вычисляется величина  $d_i(k)$  такая, что  $d_{ij} \leq d_i(k)$ , для  $j$ , входящих в ближайшие  $k$  соседей региона  $i$ . Подразумевается, что на каждый объект значимо влияют только  $k$  ближайших соседей. Влияние остальных регионов незначимо:

$$w_{ij} = \begin{cases} 1, & \text{если } d_{ij} \leq d_i(k); \\ 0, & \text{если } d_{ij} > d_i(k). \end{cases}$$

## 2. Матрица расстояний.

Использование матриц соседства не позволяет оценить интенсивность взаимосвязей между регионами. Была разработана матрица расстояний, как альтернатива:

$$w_{ij} = \begin{cases} 0, & \text{если } i = j; \\ \frac{1}{d_{ij}^q}, & \text{если } d_{ij} \leq D(q); \\ 0, & \text{если } d_{ij} > D(q). \end{cases}$$

где  $d_{ij}$  – мера расстояния между  $i$  и  $j$ , а  $D(q)$  – квартили расстояний,  $q = 1, 2, 3, 4$ . В качестве меры расстояния может применяться как фактическое расстояние между столицами изучаемых регионов, так и время, которое необходимо для преодоления этого расстояния.

Матрица расстояний имеет свои модификации, которые также широко используются в пространственной эконометрике.

Матрица граничных потенциалов:

$$w_{ij} = \begin{cases} 0, & \text{если } i = j; \\ \frac{a_i b_{ij}^q}{d_{ij}^q}, & \text{если } d_{ij} \leq D(q); \\ 0, & \text{если } d_{ij} > D(q). \end{cases}$$

где  $b_{ij}$  - доля длины общей границы  $i$ -ого и  $j$ -ого регионов в длине границы  $i$ ,  $a_i$  - доля площади  $i$ -ого региона в площади всей исследуемой территории.

Матрица рыночных потенциалов. Учитывает мощность соседних регионов:

$$w_{ij} = \begin{cases} 0, & \text{если } i = j; \\ \frac{A_j}{d_{ij}^\alpha}, & \text{если } d_{ij} \leq D(q); \\ 0, & \text{если } d_{ij} > D(q). \end{cases}$$

где  $A_j$  - показатель мощности экономики  $j$ -го региона. Может оцениваться по величине валового регионального продукта (ВРП), промышленного производства, конечного потребления и т.д.

В пространственной эконометрике выделяются пространственные связи двух типов: пространственная автокорреляция и пространственная неоднородность.

Для моделирования пространственной автокорреляции используется модель пространственного лага SAR (Spatial Autoregressive Model) [?]:

$$Y = \rho W_1 Y + X\beta + \varepsilon,$$

где случайные ошибки распределены по нормальному закону ( $\varepsilon \sim N(0, \sigma^2 I_n)$ ),  $X$  - матрица независимых (объясняющих) переменных,  $W_1$  - матрица пространственных весов,  $\rho$  - пространственный параметр, который отвечает за силу пространственной зависимости,  $\beta$  - вектор параметров.

Для моделирования пространственной неоднородности – модель с пространственной ошибкой, предусматривающая существование пространственной зависимости в остатках регрессии SEM (Spatial Error Model):

$$Y = X\beta + u, \quad u = \lambda W_2 u + \varepsilon.$$

Для проверки наличия пространственных внешних эффектов используют ряд тестов. При выявлении наличия корреляции между переменными используются LM-тесты: Lagrange Multiplier и Robust Lagrange Multiplier для пространственного лага. Для проверки наличия корреляции между ошибками используются статистика Moran's I и LM-тесты: Lagrange Multiplier и

Robust Lagrange Multiplier для пространственной ошибки.

Статистика Moran's I рассчитывается по следующей формуле:

$$I = \frac{N}{S} \frac{e^T W e}{e^T e}, \quad S = \sum_i \sum_j w_{ij}$$
$$E(I) = \frac{N}{S} \frac{\text{tr}(MW)}{n - k},$$
$$V(I) = \left(\frac{N}{S}\right)^2 \left[ \frac{\text{tr}(MW MW^T) + \text{tr}(MW)^2 + (\text{tr}(MW))^2}{d - E(I)^2} \right],$$
$$d = (n - k)(n - k - 2),$$
$$Z = \frac{I - E(I)}{\sqrt{V(I)}}.$$

Z-статистика имеет асимптотическое нормальное распределение с нулевым математическим ожиданием и единичной дисперсией.

**Третий** раздел посвящен анализу государственных расходов на здравоохранение и экономического роста в России. В настоящее время отмечается замедление роста мировой экономики после динамичного выхода из рецессии, которую вызвала пандемия. После активного возобновления роста во второй половине 2020 года темпы восстановления мировой экономики снизились в первом полугодии 2021 года на фоне новых вспышек COVID-19. Рост международной торговли утратил динамику на фоне замедления роста мировой экономики и сохранения проблем с поставками.

В литературе существует несколько подходов к объяснению механизма влияния расходов на здравоохранение на экономический рост:

1. Положительное влияние системы мер по укреплению здоровья населения на продолжительность жизни и далее на совокупный доход экономики через производственную функцию как фактор, увеличивающий продолжительность жизни трудовых ресурсов;
2. Снижение издержек на восстановление трудоспособности и сглаживание негативных последствий в период эпидемий;
3. Рассмотрение инвестиций в здравоохранение как вложение в человеческий капитал.

Положительное влияние расходов на здравоохранение может быть объяснено следующими факторами: повышение данных расходов способствует большей социальной защищенности населения, улучшению физического здоровья и повышению производительности труда и его вовлеченности в экономические процессы.

Современная система распределения доходов по территории России привела к тому, что расходы на здравоохранение на душу населения 10 наиболее обеспеченных регионов в 2 раза выше, чем в самых малообеспеченных, что вызывает неравенство в доступности медицинских услуг для граждан, проживающих в разных субъектах РФ.

В работе рассмотрена **основная гипотеза**: Существует такой оптимальный уровень расходов на здравоохранение и спорт, при котором положительное влияние на темпы роста ВРП является максимальным.

Для проверки гипотезы используются панельные данные по 75 регионам Российской Федерации с 2015 по 2019 г., которые публикуются ФСГС ежегодно. В данной работе используется издание сборника «Регионы России» 2021 г., в связи с чем статистика по большинству используемых показателей ограничена 2019 г. Москва и Московская область объединены в один регион, потому что в 2012 г. границы Москвы были расширены за счет присоединения территорий Московской области. Из выборки также исключены субъекты РФ, по которым отсутствуют наблюдения или недостаточно данных, например, Республика Крым и Чеченская Республика.

В качестве зависимой переменной используется рост ВРП  $grp\_growth = \left(\frac{Y_{it+1}}{Y_{it}}\right)$  на душу населения, где  $i$  - номер региона, а  $t$  - наблюдаемый год.

В данной работе рассматриваются следующие категории расходов бюджетов субъектов РФ в ВРП:

- $shexp\_health$  – доля расходов на здравоохранение, физкультуру и спорт (основная переменная);
- $shexp\_gov$  – доля расходов на общегосударственные вопросы;
- $shexp\_econ$  – доля расходов на национальную экономику;
- $shexp\_housing$  – доля расходов на ЖКХ;
- $shexp\_edu$  – доля расходов на образование;



— *shexp\_socpol* – доля расходов на социальную политику.

Обозначения всех используемых переменных приведены в таблице 1.

Таблица 1 – Используемые обозначения переменных

Обозначение переменной	Полное название переменной
grp_growth	Рост ВРП
shexp_gov	Доля расходов бюджета субъекта РФ на общегосударственные вопросы в ВРП
shexp_econ	Доля расходов бюджета субъекта РФ на национальную экономику в ВРП
shexp_housing	Доля расходов бюджета субъекта РФ на жилищно-коммунальное хозяйство в ВРП
shexp_edu	Доля расходов на образование в бюджете субъекта РФ
shexp_health	Доля расходов бюджета субъекта РФ на здравоохранение, физическую культуру и спорт в ВРП
shexp_health2	Квадрат доли расходов бюджета субъекта РФ на здравоохранение, физическую культуру и спорт в ВРП
shexp_socpol	Доля расходов бюджета субъекта РФ на социальную политику в ВРП
grppercarppp	Валовой региональный продукт на душу населения
urbansharebig	Доля городского населения
invgdp	Доля инвестиций в основной капитал в ВРП
highed	Доля населения с высшим образованием
road	Плотность автодорог
import_grp	Доля импорта в ВРП
export_grp	Доля экспорта в ВРП

В теории государственные расходы влияют на экономический рост, но в то же время объем расходов может изменяться в зависимости от роста или снижения ВРП. Однозначно определить направление канала влияния затруднительно, из-за чего в модели может присутствовать эндогенность. Чтобы избежать проблемы несостоятельности оценок, вызванных эндогенностью, все значения регрессоров в модели используются с лагом в один год.

Так как основная гипотеза предполагает существование некоторой средней оптимальной доли расходов на здравоохранение в ВРП, после достижения которой дальнейшее увеличение объема затрат не дает максимального эффекта на экономический рост, то в модель включается не только доля расходов на здравоохранение (*shexp\_health*), но и ее квадрат (*shexp\_health<sup>2</sup>*).

Для проверки выдвигаемой гипотезы используется **пространственная модель Дарбина (SDM)**, которая имеет следующий вид:

$$Y_t = \alpha + \gamma_t + \delta_1 shexp\_health_t + \delta_2 shexp\_health_t^2 + \rho WY_t + X_t\beta + W\tilde{X}_t\theta + \varepsilon_t,$$

где  $Y_t = (Y_{1t}, \dots, Y_{Nt})'$  - вектор значений роста ВРП в  $N$  регионах в момент времени  $t (t = 1, \dots, T)$ ,  $X_t = (X_1, \dots, X_{12})$  - матрица независимых переменных,  $\tilde{X}_t = (shexp\_health_t, shexp\_health_t^2, X_1, \dots, X_{12})$  - матрица независимых переменных с уже включенными компонентами долей расходов на здравоохранение,  $\alpha = (\alpha_1, \dots, \alpha_N)'$  - вектор фиксированных эффектов,  $\gamma_t = (\gamma_1, \dots, \gamma_T)'$  - вектор временных эффектов,  $W$  - матрица пространственных весов,  $WY_t$  и  $WX_t$  - пространственные лаги зависимой и независимой переменных,  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$  - совместно нормально распределенные ошибки регрессии в момент времени  $t$ ,  $\delta_1$  и  $\delta_2$  - оцениваемые коэффициенты, а  $\beta, \theta$  и  $\rho$  - вектора оцениваемых параметров.

Традиционным показателем, отражающим наличие пространственных эффектов для рассматриваемой зависимой переменной, является индекс Морана:

$$I(y) = \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2},$$

где  $w_{ij}$  - элементы матрицы  $W$ ,  $y_i$  - значение роста ВРП на душу населения в  $i$ -м регионе,  $\bar{y}$  - среднее значение  $y$ .

Индекс Морана показывает, насколько сильно, в среднем, связаны значения показателей переменной  $y$  в рассматриваемом регионе и в соседних регионах. Положительная пространственная корреляция свидетельствует о том, что регионы с высоким и низкими темпами экономического роста стремятся образовывать соответствующие кластеры высокого и низкого экономического роста. В случае отрицательной пространственной корреляции такие кластеры не образуются, так как в этом случае регионы с высоким экономическим ростом окружены регионами с низким экономическим ростом, и наоборот.

В **четвертом** разделе приводятся оценки пространственной модели Дарбина с фиксированными и случайными эффектами, а также приводятся ре-

зультаты тестов Вальда.

В таблице 2 представлены результаты оценивания пространственной модели Дарбина.

Таблица 2 – Результаты оценивания пространственной модели Дарбина для российских регионов

Регрессоры	$\beta$	$\theta$	Прямой эффект	Косвенный эффект	Суммарный эффект
1	2	3	4	5	6
shexp_gov	0,979***	-1,204*	0,874***	-1,341	-0,467
shexp_econ	-0,0432	-0,495	-0,119	-0,847	-0,966
shexp_housing	0,121	1,000*	0,168	1,574**	1,742**
shexp_edu	0,342	-0,624	0,303	-1,213	-0,910
shexp_health	1,600***	1,306	1,982***	3,647***	5,629***
shexp_health2	-12,307***	-2,818	-14,97***	-18,13	-33,1**
shexp_socpol	-0,117	-0,201	-0,194	-0,504	-0,698
grppercapppp	7,78e-09	-3,3e-08	10,69e-09	-2,51e-08	-1,45e-08
urbansharebig	0,0189	-0,153	0,00255	-0,268	-0,265
invgdp	-0,0212	-0,369*	-0,11	-0,164*	-0,274
highed	0,00751	-0,479	-0,00612	-0,413	-0,407
road	-0,000301*	4,11e-05	-6,38e-05**	5,21e-05	-1,17e-05
import_grp	0,034***	-0,0159*	0,024	-0,0852	-0,061
export_grp	0,0057*	-0,0036*	0,0013	-0,0178	-0,0165
$\rho$	0,491***	0,491***	0,491***	0,491***	0,491***

Для поиска средней оптимальной доли расходов необходимо остановиться на прямом эффекте (реакции зависимой переменной региона с индексом  $i$  на изменение объясняющей переменной в том же регионе с индексом  $i$ ) от изменений переменной *shexp\_health*:

$$grp\_growth = -14,97 \cdot shexp\_health^2 + 1,982 \cdot shexp\_health + \dots$$

Продифференцировав равенство по *shexp\_health* и приравняв к 0, получаем, что максимальное значение среднего темпа роста ВРП достигается при средней доле расходов на здравоохранение в ВРП, равной 5,4%, то есть дальнейшее увеличение доли расходов будет оказывать меньший эффект на экономический рост.

В **заключении** приведены результаты бакалаврской работы.

### **Основные результаты**

- Рассмотрены основные понятия, относящиеся к панельным данным;
- Рассмотрены пространственно-эконометрические модели;
- Собраны панельные данные с сайта ФСГС по 75 регионам России;
- Написан код для оценивания пространственной модели Дарбина в программной среде STATA 16.0;
- Проанализировано, что для рассматриваемого временного периода 2015–2019 гг. средняя оптимальная доля государственных расходов на здравоохранение и спорт в ВРП составляет 5,4%.