

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

**МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННОЙ
РЕГРЕССИОННОЙ ЗАВИСИМОСТИ С РАСШИРЕННЫМ
НАБОРОМ СОСТАВЛЯЮЩИХ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 412 группы

направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Леднова Александра Петровича

Научный руководитель

ст. преподаватель

А. Д. Луньков

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2023

ВВЕДЕНИЕ

С учетом современного административно-территориального устройства России пространственная неоднородность и межрегиональное неравенство являются значимыми факторами при оценке социально-экономических показателей субъектов страны. Для оценивания различных показателей регионов с учетом этих факторов используются модели пространственной эконометрики, построенные на основе панельных данных.

В основе моделей пространственной эконометрики лежит идея о том что, при моделировании каких-либо показателей региона нужно учитывать не только влияние социально-экономических факторов внутри этого региона, но и значения этих же показателей в других регионах. Реализация этой идеи возможна за счет использования панельных данных, которые благодаря своей структуре позволяют строить модели с учетом индивидуальных отличий и пространственных взаимодействий между экономическими единицами.

Актуальность работы. Задача анализа зависимости между значениями социально-экономических показателей с пространственным влиянием является важной для больших стран со сложным административно-территориальным делением. Определение подобных зависимостей позволяет прогнозировать как изменение одних социально-экономических показателей повлияет на другие, что может быть важно, например, для планирования бюджета страны.

Целью бакалаврской работы является изучение моделей пространственной эконометрики и их построение для оценивания значений социально-экономических показателей субъектов России.

Для достижения данной цели можно выделить ряд **задач**:

- изучение способов оценивания неизвестных параметров регрессионных моделей;
- изучение структуры панельных данных и моделей, позволяющих учитывать индивидуальные различия между объектами;
- изучение моделей, позволяющих учитывать пространственные взаимодействия;
- изучение моделей, позволяющих одновременно учитывать пространст-

венные взаимодействия и индивидуальные различия между объектами;
— построение модели регрессионной зависимости для оценки стоимости жилья на основе социально-экономических показателей регионов России.

В качестве **объекта исследования** выступает задача анализа зависимостей в пространственно варьирующихся социально-экономических показателях.

Предметом исследования являются модели пространственной эконометрики.

В ходе исследования применялись следующие методы: анализ, моделирование, изучение научной литературы и ее обобщение.

Структура и содержание бакалаврской работы. Работа состоит из введения, пяти разделов, заключения, списка использованных источников, содержащего 20 наименований и двух приложений. *В первом разделе* приведена информация о методах оценивания неизвестных параметров регрессионных моделей. *Во втором и третьем разделах* описаны модели, позволяющие учитывать индивидуальные отличия и пространственные взаимодействия между отдельными объектами наблюдения соответственно. *Четвертый раздел* дает представление о моделях, в которых одновременно учитываются пространственные взаимодействия и индивидуальные отличия между объектами наблюдения. *Пятый раздел* посвящен построению модели для оценки стоимости жилья в России на основе социально-экономических показателей регионов.

Основное содержание работы

В первом разделе приведена информация о методах оценивания неизвестных параметров регрессионных моделей.

Модель множественной регрессии имеет вид: $y = X\beta + \varepsilon$.

Такая модель может быть оценена одним из методов оценивания неизвестных параметров: метод наименьших квадратов, обобщенный метод наименьших квадратов, метод максимального правдоподобия.

Во втором разделе приведена информация о панельных данных и описаны модели, позволяющие учитывать индивидуальные отличия между отдельными объектами наблюдения.

Панельные данные состоят из наблюдений одних и тех же экономических единиц или объектов, которые осуществляются в последовательные периоды времени, и представляют собой двумерные массивы, одна из размерностей которых имеет временную интерпретацию, а другая – пространственную. Благодаря своей структуре панельные данные позволяют строить более гибкие и содержательные модели. В частности, возникает возможность учитывать и анализировать индивидуальные отличия и пространственные взаимодействия между экономическими единицами.

Простейшая модель линейной регрессии построенной на основе панельных данных имеет вид:

$$y_{it} = x'_{it}\beta + \varepsilon_{it}. \quad (1)$$

Простейшая модель не учитывает панельную структуру данных. Для того, чтобы это исправить, нужно добавить в модель составляющую, позволяющую учитывать индивидуальные различия между экономическими единицами. Модель, полученная за счет добавления такой составляющей, имеет вид:

$$y_{it} = x'_{it}\beta + \mu_i + \varepsilon_{it}, \quad (2)$$

где величина μ_i выражает индивидуальный эффект объекта i , не зависящий от времени t , при этом регрессоры x_{it} не содержат константу.

В зависимости от предположений относительно характера величины μ_i рассматриваются две модели.

Первая из рассматриваемых моделей это *модель с фиксированным эф-*

фактом (fixed effect model). В этой модели предполагается, что в уравнении (2) величины μ_i являются неизвестными параметрами.

Вторая из рассматриваемых моделей это *модель со случайным эффектом* (random effect model). В этой модели предполагается, что в уравнении (2) $\mu_i = \alpha + u_i$ где α — параметр, общий для всех единиц во все моменты времени, а u_i — ошибки, некоррелированные с ε_{it} и некоррелированные при разных i .

В третьем разделе приведена информация о моделях пространственной эконометрики.

Миграция, технологические эффекты и торговые связи могут связывать экономические системы регионов, несмотря на политические границы между территориями. Для отражения в эмпирических моделях этих пространственных связей применяются модели пространственной эконометрики. В основе пространственной методологии лежит *матрица пространственных весов*, которая описывает систему связей между регионами и их интенсивность.

Предполагается, что матрица пространственных весов W остается постоянной во времени, а панель данных является сбалансированной, то есть представляет собой набор данных, в котором каждая экономическая единица непрерывно наблюдается каждый отведенный промежуток времени. Элемент матрицы w_{ij} отражает влияние региона i на регион j . Все матрицы пространственных весов — квадратные. Так как в матрице исключается возможность влияния региона на себя, то по диагонали матрицы стоят нули.

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

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

Использование матриц соседства не позволяет оценить интенсивность взаимосвязей между регионами. Для добавления такой возможности можно

использовать матрицу расстояний:

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

где d_{ij} – мера расстояния между i и j , γ - определяет интенсивность взаимосвязи, а $D(q)$ – квартили расстояний, $q = 1, 2, 3, 4$. Параметры γ и q считаются постоянными. В качестве меры расстояния может использоваться как фактическое расстояние между столицами изучаемых регионов, так и время, необходимое для преодоления этого расстояния.

Для пространственных связей можно выделить две ситуации: пространственная автокорреляция и пространственная неоднородность. Пространственная корреляция является следствием зависимости экономических объектов друг от друга. Пространственная неоднородность является следствием недиагональной структуры ковариационной матрицы ошибок.

Для моделирования пространственной автокорреляции используется *модель пространственного лага* (SAR – Spatial Autoregressive Model), которая имеет вид:

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

Для моделирования пространственной неоднородности используется *модель пространственной зависимости в ошибки* (SEM – Spatial Error Model), которая имеет вид:

$$y_{it} = x'_{it} \beta + u_{it}, \text{ где } u_{it} = \lambda \sum_{j=1}^n w_{ij} u_{jt} + \varepsilon_{it}.$$

В четвертом разделе представлены модели, в которых одновременно учитываются пространственные взаимодействия и индивидуальные отличия между объектами наблюдения.

Для получения моделей, позволяющих отразить реальные взаимодействия между различными экономическими единицами, нужно при составле-

нии регрессионной модели учитывать как пространственные взаимодействия между объектами наблюдения, так и индивидуальные отличия между ними. Такие модели одновременно включают в себя матрицу пространственных весов W и индивидуальный эффект объекта μ_i .

Модель пространственного лага, позволяющая учитывать индивидуальные отличия между объектами наблюдения имеет вид:

$$y_{it} = \rho \sum_{j=1}^n w_{ij} y_{jt} + x'_{it} \beta + \mu_i + \varepsilon_{it}. \quad (3)$$

Модель пространственной зависимости в ошибке, позволяющая учитывать индивидуальные отличия между объектами наблюдения имеет вид:

$$y_{it} = x'_{it} \beta + \mu_i + u_{it}, \text{ где } u_{it} = \lambda \sum_{j=1}^n w_{ij} u_{jt} + \varepsilon_{it}. \quad (4)$$

В зависимости от вида μ_i моделях (3) и (4) они могут учитывать как фиксированные, так и случайные индивидуальные эффекты.

Пятый раздел посвящен построению модели для оценки стоимости жилья в России на основе социально-экономических показателей регионов.

Для демонстрации возможности моделей пространственной эконометрики отражать в эмпирических моделях пространственные связи между экономическими единицами построена модель для оценки стоимости жилья в России на основе социально-экономических показателей регионов. Для моделирования этой зависимости была выбрана модель пространственного лага с фиксированным эффектом.

Для построения моделей были выбраны следующие социально-экономические показателя:

1. средние цены на вторичном рынке жилья (на конец года; рублей за квадратный метр общей площади);
2. среднедушевые денежные доходы населения (в месяц; рублей);
3. численность персонала, занятого исследованиями и разработками (количество человек на 100 тысяч человек населения);
4. число зарегистрированных преступлений (на 100 тысяч человек насе-

ления).

Данные о рассматриваемых показателях взяты с сайта федеральной службы государственной статистики (Росстат) из приложения к ежегодному сборнику «Регионы России. Социально-экономические показатели». Все расчеты проводятся по данным за 2000–2021 года для 76 регионов России. Данные для остальных регионов не были взяты в рассмотрение, так как, из-за отсутствия информации о социально-экономических показателях за различные периоды времени, содержат пропуски. В качестве матрицы пространственных весов выбраны матрица соседства и матрица обратных расстояний для регионов России.

Для выявления пространственных эффектов можно использовать *индексы Морана*. Индекс Морана для некоторого регионального показателя x и матрицы пространственных весов W может быть вычислен по формуле:

$$I(X, W) = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left[\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right] \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]}, \quad (5)$$

где n — число регионов, \bar{x} — среднее значение показателя x по регионам, w_{ij} — элементы матрицы матрицы пространственных весов.

Значения, которые может принимать индекс Морана, лежат в диапазоне от -1 до 1, где значение 1 говорит о наличии прямой пространственной зависимости, -1 о наличии обратной зависимости, а 0 указывает на отсутствие пространственной зависимости. Выявление положительного пространственного эффекта для какого-либо социально-экономического показателя соответствует тому, что регион, в котором этот показатель растет, способствует росту этого показателя в соседних регионах. В то же время отрицательный пространственный эффект соответствует тому, что регион с растущим показателем, мешает росту этого показателя в соседних регионах.

Значения индексов Морана и Z-статистика для средней стоимости квадратного метра на вторичном рынке жилья с матрицами соседства и обратных расстояний для каждого года приведены в таблице 1.

Вычисления индексов Морана и построение моделей производились средствами языка программирования Python.

Для наглядной демонстрации эффективности модели, учитывающей

Таблица 1 – Значения индексов Морана для средней стоимости квадратного метра на вторичном рынке жилья

Год	Матрица соседства	Z-оценка коэффициента Морана(МС)	Матрица обратных расстояний	Z-оценка коэффициента Морана(МОР)
2000	0.87231***	12.9926	0.88838***	12.9632
2001	0.86356***	12.7873	0.88085***	12.7769
2002	0.85761***	13.5812	0.87425***	13.5543
2003	0.85729***	13.4522	0.87504***	13.4421
2004	0.84596***	11.1972	0.86457***	11.2213
2005	0.82571***	10.6046	0.84201***	10.6029
2006	0.83919***	37.8662	0.84791***	37.2381
2007	0.43465***	16.4289	0.43555***	15.9602
2008	0.54182*	0.16970	0.55868*	0.17095
2009	0.26882*	0.16188	0.28932*	0.16752
2010	0.33977	0.01738	0.39939	0.01996
2011	0.55914***	9.55530	0.59102***	9.9676
2012	0.77987***	23.3792	0.79962***	23.5162
2013	0.86332***	37.7276	0.87496***	37.2995
2014	0.82856***	36.5528	0.84126***	36.0926
2015	0.80425***	24.8744	0.82158***	24.7600
2016	0.80552***	15.2699	0.82065***	15.2258
2017	0.79563***	13.4794	0.79359***	13.1605
2018	0.86608***	14.3012	0.87528***	14.1565
2019	0.86648***	8.71412	0.86366***	8.51415
2020	0.79459***	6.19051	0.80375***	6.14529
2021	0.83050***	11.6317	0.83883***	11.4952

индивидуальные и пространственные эффекты, кроме модели пространственного лага с фиксированным эффектом (3), можно рассмотреть простейшую модель (1), параметры которой оцениваются методом наименьших квадратов.

Оценки для параметров моделей (1) и (3) с матрицами соседстве и обратных расстояний приведены в таблице 2.

Значения коэффициентов, полученные в результате построения моделей, позволяют сделать вывод о том, что рост среднедушевых доходов населения и количества людей, занятых исследованиями или разработками в

Таблица 2 – Оценки параметров моделей

Модель		Коэффициенты при регрессорах			
		Пространственный коэффициент	Среднедушевые доходы населения	Численность высококвалифицированного персонала	Количество преступлений
Простейшая модель	коэф.		1.6809***	12.825***	-2.114***
	<i>t</i> - тест		84.310	16.673	-9.965
Модель пространственного лага с индивидуальным эффектом (МС)	коэф.	0.0708***	1.6350***	35.113***	-5.347***
	<i>t</i> - тест	6.2405	99.763	27.451	-24.94
Модель пространственного лага с индивидуальным эффектом (МОР)	коэф.	0.0719***	1.6339***	35.177***	-5.330***
	<i>t</i> - тест	6.0378	97.648	27.279	-24.16

регионе, приведет к росту средней стоимости квадратного метра жилплощади, в то время как увеличения количества преступлений окажет негативный эффект. Также, как и предполагалось, в моделях пространственного лага с фиксированным эффектом существует положительная пространственный эффект, то есть рост цен на жилплощадь в одном регионе приводит к их росту и в соседних регионах.

Для оценки качества рассматриваемых регрессионных моделей можно использовать среднюю квадратичную ошибку (MSE), среднюю абсолютную ошибку (MAE) и среднюю относительную ошибку (MAPE), задающиеся формулами:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2,$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|,$$

$$MAPE = 100 * \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i},$$

где $N = n * T$ — число наблюдений набора данных, y_i — фактическое значение зависимой переменной для i -го наблюдения, \hat{y}_i — значение, предсказанное моделью для i -го наблюдения. Значения перечисленных показателей для рас-

смаатриваемых моделей приведены в таблице 3.

Модель	MSE	MAE	MAPE
Простейшая модель	137110899.524	7848.0259	33.0809
Модель пространственного лага с индивидуальным эффектом (матрица соседства)	73849154.7999	6274.1475	27.6648
Модель пространственного лага с индивидуальным эффектом (матрица обратных расстояний)	73690659.1079	6273.3738	27.5123

Таблица 3 – Сравнение точности моделей

Сравнивая значения различных показатели качества моделей можно сделать вывод о том что модель, учитывающая пространственные и индивидуальные эффекты, лучше прогнозирует стоимость жилплощади. Такой результат свидетельствует о наличии значительного влияния пространственных взаимосвязей и индивидуальных отличий между регионами.

Модель пространственного лага с фиксированным эффектом, в которой в качестве матрицы пространственных весов использовалась матрица обратных расстояний, дает лучший прогноз, чем аналогичная модель с матрицей соседства, такой результат объясняется тем, что матрица обратных расстояний дает более точное представление о пространственной структуре регионов.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены структура панельных данных и модели, которые могут быть построены на их основе. Такие модели дают возможность учитывать индивидуальные отличия и пространственные взаимодействия между экономическими единицами. Эта возможность достигается за счет включения в классическую модель множественной регрессии дополнительных составляющих.

В качестве демонстрации эффективности моделей с дополнительными составляющими были построены модели для предсказания стоимости квадратного метра на вторичном рынке жилья на основе социально-экономических показателей регионов. В качестве демонстрационной модели была выбрана модель пространственного лага с фиксированным эффектом, использующая в качестве матрицы пространственных весов матрицу соседства и матрицу обратных расстояний, а для сравнений использовалась простейшая модель на основе панельных данных. Наиболее точной оказалась модель пространственного лага с фиксированным эффектом, использующая матрицы обратных расстояний. Такой результат позволяет утверждать, что на стоимость жилплощади оказывают влияния пространственные взаимосвязей и индивидуальных отличий между регионами, для учета которых в модель требуется добавлять дополнительные составляющие.