МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

Оценивание параметров динамических регрессионных моделей с учётом пространственной взаимосвязи

наименование темы выпускной квалификационной работы

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 412 группы

направления 01.03.02 «Прикладная математика и информатика»

механико-математического факультета

Пескова Игоря Сергеевича

фамилия, имя, отчество

Научный руководитель: А. Д. Луньков

подпись, дата

Заведующий кафедрой,
д.ф.-м.н., доцент С. П. Сидоров

подпись, дата

Саратов 2023

Введение В данной работе рассмотрены методы оценивания параметров динамических регрессионных моделей с учетом пространственной взаимосвязи.

Актуальность Оценивание параметров динамических регрессионных моделей является одной из важнейших задач в эконометрике. При этом, учет пространственной взаимосвязи между наблюдениями является необходимым условием для получения более надежных регрессионных моделей. Оценивание параметров динамических регрессионных моделей часто применяется в различных областях, таких как прогнозирование экономических показателей, управление финансами, анализ рынка ценных бумаг, определение рисков и т.д.

Цель Целью работы является оценивание параметров динамических регрессионных моделей. Эти модели позволяют за счет учета пространственной взаимосвязи более качественно описывать изучаемые экономические процессы, получать более точные и надежные результаты за счёт пространственной взаимосвязи.

Для достижения этой цели проведен сбор и статистический анализ данных.

Структура бакалаврской работы

— В первом разделе даны теоретические основы временных рядов и динамических регрессионных моделей с учетом пространственной взаимосвязи.
— Во втором разделе даны теоретические основы динамических регрессионных моделей с учётом пространственной взаимосвязи.
— В третьем разделе описаны методы оценивания параметров этих моделей.
— В четвертом разделе построена регрессионная модель для прогнозирования стоимости минимального набора продуктов питания по регионам России средствами языка Python.

Основное содержание работы Динамические регрессионные модели с учетом пространственной взаимосвязи являются одним из классов моделей, используемых для анализа зависимостей между переменными во времени и пространстве. Они представляют собой расширение классических регрессионных моделей, не учитывающих пространственную взаимосвязь между наблюдениями.

Общий вид динамической регрессионной модели с учетом пространственной взаимосвязи следующий:

$$y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ijt} + \sum_{k=1}^{q} \gamma_k y_{i,t-k} + \varepsilon_{it},$$

где y_{it} - значение зависимой переменной y для i-го наблюдения в момент времени t, x_{ijt} - значение j-й независимой переменной для i-го наблюдения в момент времени t, β_j - коэффициенты регрессии для независимых переменных, γ_k - коэффициенты регрессии для лагов зависимой переменной, ε_{it} - случайная ошибка.

Динамический ряд отличается от обычной выборки данных, так как при анализе рассматривается зависимость параметров от времени, а не только взаимосвязь характеристик и их статистические особенности.

Динамические регрессионные модели с учетом пространственной взаимосвязи подразделяются на два типа: пространственно-временные и пространственно-динамические модели.

Пространственно-временные модели учитывают как временную, так и пространственную зависимость между переменными. Они описываются так:

$$y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ijt} + \sum_{l=1}^{m} \delta_l w_{il} + \varepsilon_{it},$$

где w_{il} - значение l-й пространственной переменной для i-го наблюдения.

Пространственно-динамические модели рассматривают только пространственную зависимость между переменными, но не учитывают временную зависимость. Они могут быть представлены зависимостью:

$$y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ijt} + \sum_{k=1}^{q} \gamma_k y_{i,t-k} + \sum_{l=1}^{m} \delta_l w_{il} + \varepsilon_{it}.$$

Для оценивания параметров динамических регрессионных моделей с учетом пространственной взаимосвязи используются различные методы, такие как метод максимального правдоподобия, метод инструментальных переменных, метод наименьших квадратов, обобщенный метод моментов.

Матрица пространственных весов W – положительно определенная матрица размера (NN), она задает круг «соседей» в географическом или ином смысле, влияющих на каждую единицу наблюдения. Ненулевые элементы матрицы w_{ij} отражают влияние региона j на регион i. Часто матрица нормируется по столбцам, то есть каждый элемент матрицы делится на сумму элементов того столбца, в котором он находится. Эта нормировка упрощает интерпретацию.

Не существует жестких норм, устанавливающих правила определения не только пространственных весов, но и круга «соседей». Зачастую выбор «соседей» происходит с точки зрения географической близости, наличия общей границы, ее протяженности. Определение пространственных весов не ограничивается только географическим соображениями, также широкое применение находит экономическое расстояние между агентами. Традиционно пространственные веса принимаются равными $1/d_{ij}$, где d_{ij} - мера расстояния; также используются более сложные формы функциональной зависимости – отрицательная экспоненциальная, похожая на формулу гравитационной силы. Чтобы получить хорошие асимптотические свойства оценок, матрицу пространственных весов подбирают так что она удовлетворяла ограничениям регулярности.

Так как в матрице исключается возможность влияния региона на себя, то по диагонали матрицы стоят нули. Как правило, матрица весов стандартизируется по столбцам для упрощения процедуры ее обращения и для улучшения качества оценивания пространственных эффектов.

Выделяют пространственные связи двух типов: пространственная автокорреляция и пространственная зависимость в ошибки. Пространственная корреляция является следствием зависимости экономических объектов друг от друга. Пространственная зависимость в ошибке влечет за собой недиагональную структуру ковариационной матрицы ошибок. Модели с такой матрицей сложнее оценивать. Рассмотрим две модели, которые описывают вышеупомянутые связи:

Для описания пространственной автокорреляции используется модель пространственного лага (SAR – Spatial Autoregressive Model):

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

$$\varepsilon \sim N(0, \sigma^2 I_n),$$

а для описания пространственной зависимости в ошибки - модель с тем же названием (SEM – Spatial Error Model):

$$Y = X\beta + u$$

$$u = \lambda W_2 u + \varepsilon$$

$$\varepsilon \sim N(0, \sigma I_n).$$

Индекс Морана используется для оценки пространственной автокорреляции данных в области, и вычисляется по формуле:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

где n - число единиц анализа (например, городов, регионов, населенных пунктов); x_i - значение переменной X в i-й единице анализа; $\overline{x}$ - среднее значение переменной X; w_{ij} - весовой коэффициент, который может быть например равен 1, если i и j соответствуют соседним единицам пространства, и 0 в противном случае; $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ (нормализующий коэффициент). Суммы берутся по i и j с условием $i \neq j$.

Значение индекса Морана принимает значения от -1 до 1. Значение близкое к 1 указывает на положительную пространственную автокорреляцию, то есть, наличие скоплений высоких или низких значений переменной X в пространстве. Значение близкое к -1 указывает на отрицательную автокорреляцию, что означает наличие скоплений высоких значений X рядом с скоплениями низких значений X. Значение близкое к нулю указывает на случайную расстановку значений переменной X в пространстве.

Рассмотрим более подробно методы для оценивания параметров динамических регрессионных моделей с учетом пространственной взаимосвязи используются различные методы. Рассмотрим некоторые из них:

1. Метод максимального правдоподобия (Maximum likelihood, ML). Для этого метода необходимо специфицировать функцию правдоподобия для модели. Она будет иметь вид:

$$L(\theta|y, X) = f(y|X, \theta),$$

где y - зависимая переменная, X - матрица регрессоров, θ - вектор параметров модели, f - функция плотности вероятности. Ищется максимум этой функции по θ. При наличии пространственной взаимосвязи используются модели с матрицей весов, которая учитывает соседство между наблюдениями.

Такие модели имеют вид:

$$y = X\beta + \rho Wy + \varepsilon,$$

где ε - случайная ошибка, учитывающая пропущенные факторы, а β - вектор коэффициентов регрессии.

2. Метод инструментальных переменных (Instrumental variables, IV). Он используется для оценивания параметров моделей, в которых присутствует эндогенность. Он заключается в использовании дополнительных переменных, которые влияют на эндогенную переменную, но не коррелируют с ошибкой.

Для моделей с матрицей весов используются инструменты, которые учитывают пространственную взаимосвязь между наблюдениями. Они могут иметь вид:

$$y = X\beta + \rho Wy + \varepsilon,$$

где ε - случайная ошибка, а β - коэффициенты регрессии, оцениваемые с помощью инструментов.

3. Байесовский подход (Bayesian approach). Заключается в описании априорного распределения параметров модели и обновлении его с помощью данных.

Для моделей с матрицей весов априорное распределение может иметь вид:

$$\beta \sim N(\mu, \Sigma),$$

где μ и Σ - параметры распределения.

Анализ стоимости минимального набора продуктов питания по регионам россии

В работе были рассмотрены пространственные динамические регрессионные модели для стоимости минимального набора продуктов питания. В состав набора включено 33 наименования продовольственных товаров. Данные о стоимости набора определяются в расчете на одного человека в месяц

Для построения моделей были выбраны следующие социально-экономические показатели:

— Стоимости минимального набора продуктов питания (в рублях).
— Посевные площади сельскохозяйственных культур (в хозяйствах всех категорий; тысяча гектаров).
— Поголовье крупного рогатого скота (в хозяйствах всех категорий; на конец года; тысяч голов).
— Доля уловленных и обезвреженных загрязняющих атмосферу веществ в общем количестве отходящих загрязняющих веществ от стационарных источников (в процентах).
— Индексы тарифов на грузовые перевозки (декабрь к декабрю предыдущего года; в процентах).

Данные о рассматриваемых показателях взяты с сайта федеральной службы государственной статистики (Росстат) из ежегодного сборника «Регионы России. Социально-экономические показатели».

Все расчеты проводятся по данным за 2000-2021 года для 76 регионов России.

В качестве матрицы пространственных весов выбрана матрица обратных расстояний для регионов России.

Для работы были использованы следующие библиотеки Python:

— Matplotlib - библиотека для визуализации данных.
— Scipy - библиотека, предназначенная для выполнения научных и инженерных вычислений.
— spreg - пакет для оценки моделей авторегрессии с учетом пространственной взаимосвязи.

Сначала проводится подготовка данных, а именно: проверка на наличие пропусков в данных, нормализация, запись в структуры данных python для работы.

Использование матриц соседства не позволяет оценить интенсивность взаимосвязей между регионами. Для добавления такой возможности можно использовать матрицу обратных расстояний:

$$w_{ij} = \begin{cases} 0, & \text{если } i = j \\ \frac{1}{d_{ij}}, & \text{если } i \neq j \end{cases},$$

где d_{ij} - мера расстояния между i и j. В качестве меры расстояния может использоваться как фактическое расстояние между центрами изучаемых регионов, так и время, необходимое для преодоления этого расстояния.

Далее загружаются данные для матрицы весов, а именно таблица расстояний между регионами России. Расстоянием между регионами считается длина дороги между областными центрами преодолеваемой на машине.

Составляем из загруженных данных матрицу весов обратных расстояний по закону выше.

Оценим параметры регрессионной модели следующего типа для нашей задачи:

$$y = \rho W y + X\beta + \epsilon.$$

Воспользуемся методом обобщенных моментов для оценивания параметров. Рассмотрим простейшие расчёты с помощью библиотеки spreg. Для этого создадим модель с лагом 1 степени с нужными параметрами:

```
model = spreg.GM_Lag(
y,
x,
w=e,
name_y="price",
name_x=x_names,
name_w="table.russia",
name_ds="data"
)
```

Получим следующие оценки показателей:

Variable	Coefficient	Std.Error	Z-Statistic	Probability
CONSTANT	1352.8101520	505.9190150	2.6739658	0.0074960
plant	−0.0016117	0.0011474	−1.4045992	0.1601405
animal	−0.0201171	0.0571180	−0.3522031	0.7246859
pollution	−0.3367946	0.4360970	−0.7722929	0.4399410
ride	0.0121734	0.8929588	0.0136327	0.9891230
lag1	0.8375942	0.1477729	5.6681199	0.0000000
W price	−0.0823092	0.0894363	−0.9203109	0.3574103

(1)

Для сравнения методов, рассмотрим оценивание этой модели с помощью метода наименьших квадратов (МНК).

```
model = spreg.OLS(
y,
x,
w=e,
name_y="price",
name_x=x_names,
name_w="table.russia",
name_ds="data"
)
print(model.summary)
```

Получим следующие данные:

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	910.6253509	167.9976016	5.4204664	0.0000009
plant	−0.0016343	0.0012165	−1.3434962	0.1837830
animal	−0.0139020	0.0601432	−0.2311490	0.8179252
pollution	−0.3307137	0.4623853	−0.7152341	0.4770257
ride	0.1424888	0.9349150	0.1524083	0.8793370
lag1	0.8359526	0.1566873	5.3351648	0.0000013
W price	−0.0923092	0.0892323	−0.9213405	0.3564072

(2)

Значения коэффициентов, полученные в результате построения моделей, позволяют сделать вывод о том, что рост поголовья крупного рогатого скота или посевных площадей сельскохозяйственных культур, приведет к снижению стоимости минимального набора продуктов питания, в то время как увеличения индексов тарифов на грузовые перевозки окажет негативный эффект. Также, как и предполагалось, в моделях пространственного лага существует отрицательный пространственный эффект, то есть снижение стоимости минимального набора продуктов питания в одном регионе приводит к их снижению и в соседних регионах.

Заключение По итогам работы можно сделать вывод о том, что учет пространственной взаимосвязи при оценивании параметров динамических регрессионных моделей является необходимым для получения более точных и надежных результатов. Важно правильно уметь строить матрицу обратных расстояний, нормализовать. Были рассмотрены различные методы оценивания, такие как метод инструментальных переменных, метод максимального правдоподобия и обобщенный метод моментов. Каждый из них имеет свои преимущества и недостатки, и выбор метода зависит от конкретной задачи и доступных данных.