NATIONAL RESEARCH UNIVERSITY

Семинар НУГ «Оценка влияния макрошоков на социально-экономические процессы в регионах России»

Основные пространственно-эконометрические модели и их применение к российским данным

Демидова
Ольга Анатольевна
https://www.hse.ru/staff/demidova_olga
E-mail:demidova@hse.ru

План презентации

- Особенности пространственно-эконометрических моделей
- Понятие соседства
- Матрицы весов
- Пространственные лаги
- Выявление глобальной пространственной зависимости
- Глобальные индексы Морана, Гири, Гетиса-Орда
- График Морана
- Выявление локальной пространственной зависимости
- Локальные индексы пространственной автокорреляции (LISA): индексы Морана, Гири, Гетиса-Орда
- Базовые пространственные модели
- Интерпретация результатов оценивания пространственных моделей
- Выбор между пространственными моделями
- Пространственно-эконометрические исследования с использованием региональных российских данных

Все взаимосвязано

Уолдо Рудольф Тоблер (1930-2018)

(Tobler, 1970) Первый закон географии:

«… все связано со всем остальным, но близкое связано больше, чем отдаленное» (everything is related to everything else, but near things are more related than distant things)

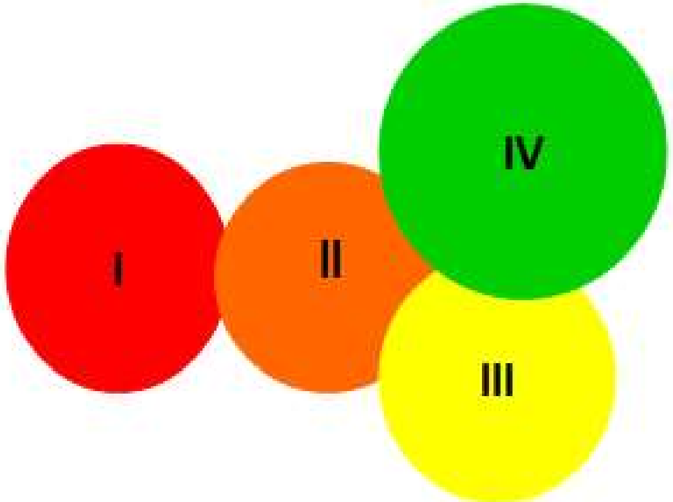
Tobler's Addendum

“Near can take on many meanings in different situations" Tobler (2004)

Пространство vs время

Характеристики	Время	Пространство
Размерность	dim = 1	dim = 2
Объекты	Моменты времени	Страны, регионы и т.п.
Направление влияния	Одностороннее	Двустороннее
Соседи	t-1, t, t+1	Описываются с помощью взвешивающей матрицы W
Лаги	Y_{t-1}	WY
Коэффициент корреляции	ρ – коэф. автокорреляции	ρ – коэф. пространственной корреляции

Пример взвешивающей матрицы



$$W_b = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Пример пространственного лага

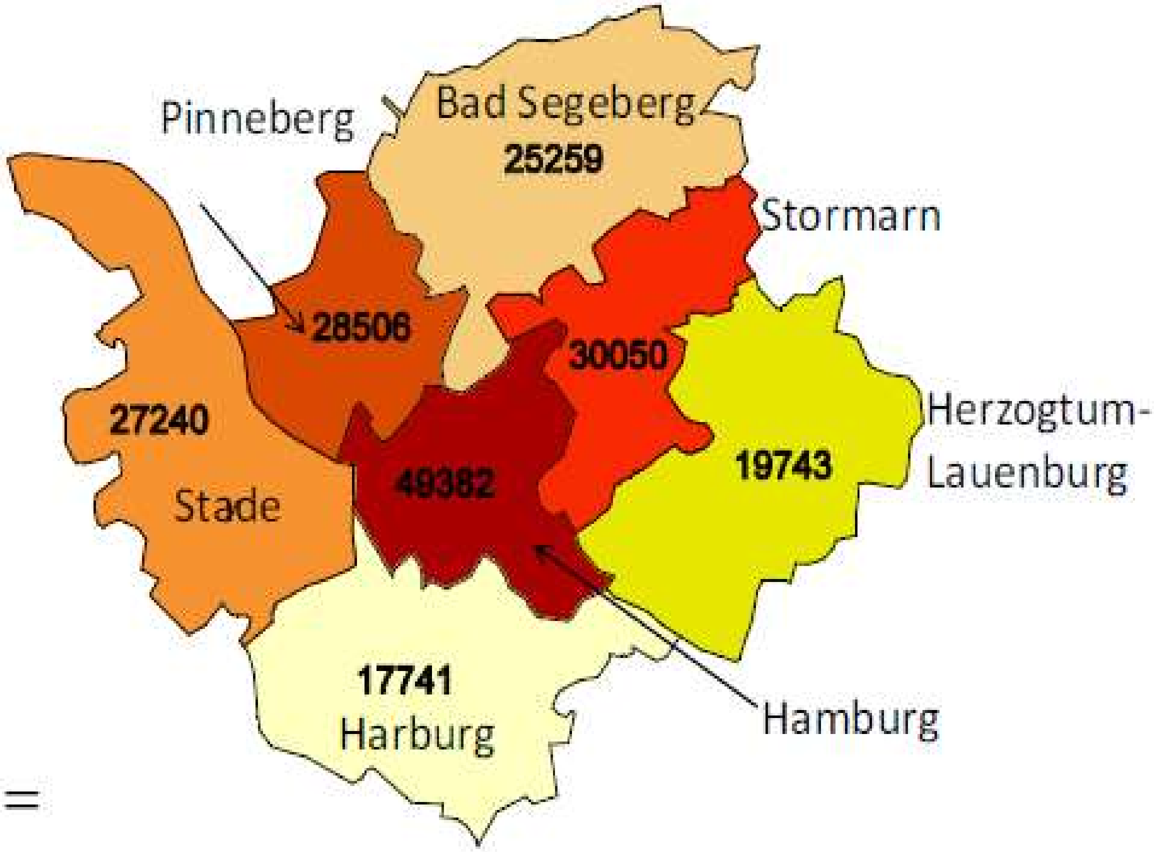
$$WY = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} Y_2 \\ 1/3(Y_1 + Y_3 + Y_4) \\ 1/2(Y_2 + Y_4) \\ 1/2(Y_2 + Y_3) \end{pmatrix}$$

Пример 1 пространственного лага

Spatial lag

GDP per capita 2008,

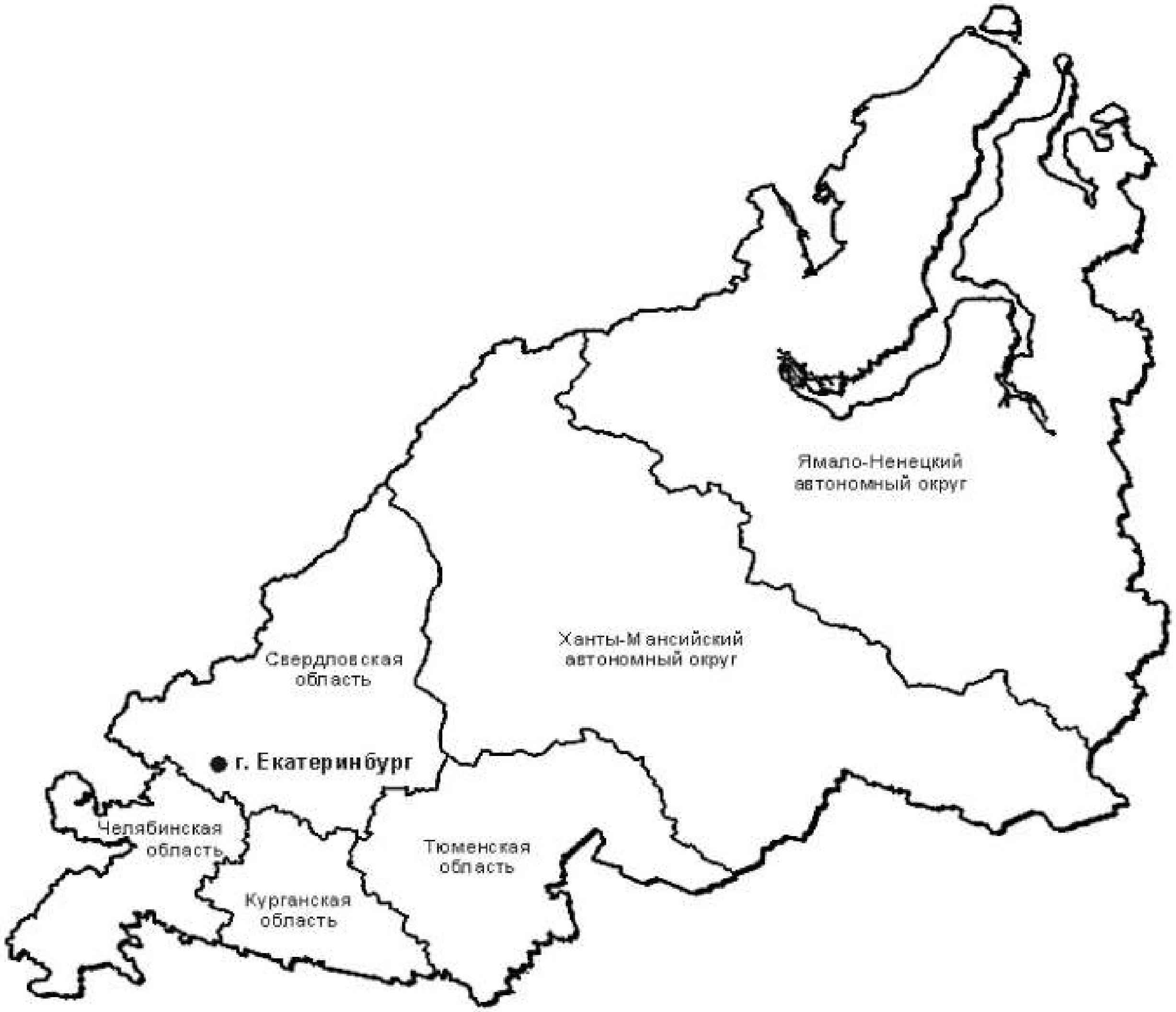
at current prices



$$[Wy]_{HAMBURG} = 24{,}756.5 =$$

$$\frac{1}{6}*17{,}741+\frac{1}{6}*19{,}743+\frac{1}{6}*25{,}259+\frac{1}{6}*27{,}240+\frac{1}{6}*28{,}506+\frac{1}{6}*30{,}050$$

Пример 2 пространственного лага



Карта регионов Уральского федерального округа

Пример 2 пространственного лага

1) Челябинская область 2) Курганская область,

3) Свердловская область, 4) Тюменская область (без АО),

5) Ханты-Мантийский АО, 6) Ямало-Ненецкий АО.

$$W = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Пример 2 пространственного лага

Y – уровень региональной безработицы, согласно данным Росстата (Регионы России, 2020), $Y = (5.1\ 7.8\ 4.2\ 4.1\ 2.5\ 1.9)'$ Пространственный лаг этой переменной $WY = (6\ 4.47\ 4.88\ 4.83\ 3.4\ 2.5)'$

$$WY = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 5.1 \\ 7.8 \\ 4.2 \\ 4.1 \\ 2.5 \\ 1.9 \end{pmatrix} = \begin{pmatrix} 6 \\ 4.47 \\ 4.88 \\ 4.83 \\ 3.4 \\ 2.5 \end{pmatrix}$$

Общий вид взвешивающей матрицы

$$W = \begin{pmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \dots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 0 \end{pmatrix}$$

Свойства взвешивающей матрицы:

1) $w_{ij} \geq 0,\ i, j = 1, \dots, n,$

2) $w_{ii} = 0,\ i, = 1, \dots, n$

3) $\sum_{j=1}^{n} w_{ij} = 1$

Виды взвешивающих матриц

- Основанные на географической близости регионов

- Основанные на социо-культурной близости регионов
  - Общий язык
  - Общая религия

- Основанные на экономической близости регионов
  - Величина торговых потоков
  - Технологические расстояния

  - Основанные на институциональной близости регионов

Часто взвешивающие матрицы нормируют по строке

Географические матрицы, основанные на соседстве регионов

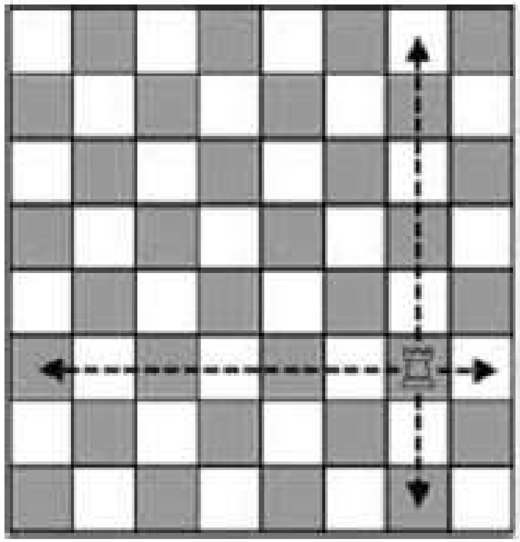
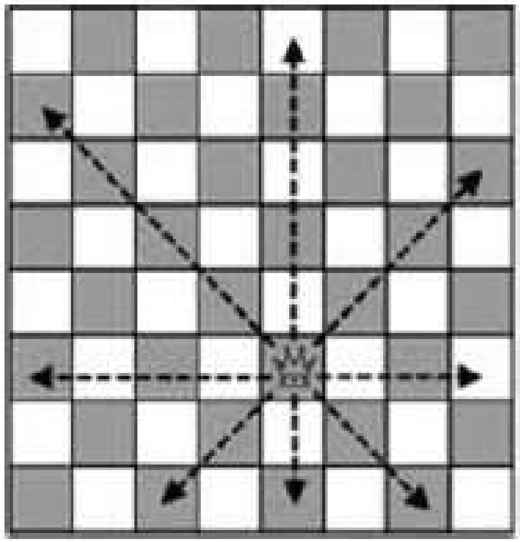
1) $w_{ij} = \frac{1}{n_i}$, если регион j является соседом региона i и 0 иначе,

где n_i – количество регионов – соседей региона i (это могут быть регионы, с которыми у региона i есть общая граница или регионы, расположенные не далее определенного расстояния d от региона i),

2) $w_{ij} = \frac{1}{k}$, если регион j входит в число ближайших k соседей региона i и 0 иначе.

Какой регион считать соседним?

- **Rook's** case: neighbours are those cells that share a common side
- **Queen's** case: neighbours are those cells that share a common edge or a common side

Какой регион считать соседним?

Разные типы «соседей»

- Соседи первого порядка
- Соседи второго порядка (соседи соседей)
- Соседи к-го порядка
- К ближайших соседей

Географические матрицы, основанные на расстояниях

1) $w_{ij} = \frac{1}{d_{ij}^{\theta}} / \sum_{j=1}^{n} d_{ij}^{\theta}$ или $w_{ij} = exp\{-\theta d_{ij}\} / \sum_{j=1}^{n} exp\{-\theta d_{ij}\}$

2) $w_{ij} \sim \frac{b_{ij}^{\alpha}}{d_{ij}^{\beta}}$,

где d_{ij} - расстояние между регионами i и j (существуют разные способы его измерения, например, как евклидово расстояние между центрами столиц регионов или между центрами полигонов, в которые вписаны соответствующие регионы или расстояние по автодорогам между столицами регионов, либо время в пути между столицами регионов и т.д. и т.п.), θ - некоторый дополнительный параметр (decay parameter) или некоторое конкретное число (наиболее популярные значения 1 и 2), b_{ij} - отношение общей границы регионов i и j к периметру региона i; α, β – дополнительные параметры.

+ и – введения взвешивающих матриц

+ Удается резко сократить количество оцениваемых параметров

- Неопределенность с выбором матрицы W

- Эти матрицы могут быть эндогенными

Чт о делать?

•Байесовский подход к выбору матриц (М.Фишер и Д.Лесаж)

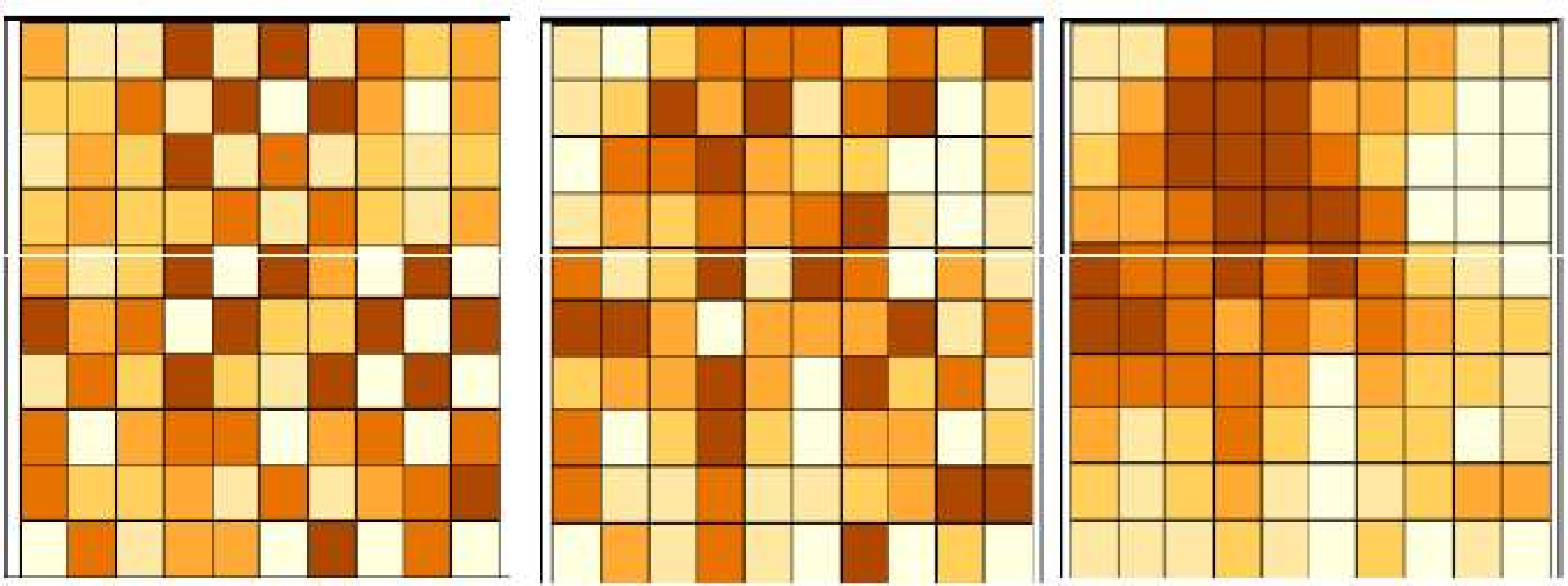
•Выпуклые комбинации взвешивающих матриц (Д.Лесаж)

•Специальная техника оценивания моделей с эндогенными взвешивающими матрицами (Д.Пирас)

•Оценка взвешивающей матрицы

Выявление пространственной зависимости

Пространственная корреляция



Отрицательная Случайная Положительная

Выявление пространственной зависимости

Глобальные индикаторы пространственной зависимости

- Глобальный индекс Морана I
- Глобальный индекс Гири, C
- Глобальный индекс Гетиса-Орда, G

Индекс Морана

Глобальный индекс Морана для выбранного показателя Y рассчитывается по формуле:

$$\mathrm{I}(Y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

и принимает значения -1 < I(Y) < 1.

Индекс Морана

H_0: I = 0 (расположение регионов является случайным)

H_1: I ≠ 0 (расположение регионов не является случайным)

$$z = \frac{I - E(I)}{\sqrt{\mathrm{var}(I)}} \sim N(0,1), \qquad E(I) = -\frac{1}{n-1} \approx 0 \text{ при больших } n$$

I значим и > 0 => положительная пространственная корреляция,

I значим и < 0 => отрицательная пространственная корреляция

Индекс Морана

I значим и > 0 => положительная пространственная корреляция (если выбранный регион имеет высокое (низкое) значение рассматриваемого показателя, то и в соседних регионах этот показатель велик (низок)),

I значим и < 0 => отрицательная пространственная корреляция (если выбранный регион имеет высокое (низкое) значение рассматриваемого показателя, то и в соседних регионах этот показатель низок (велик)).

Проверка значимости I равносильна проверке

значимости коэффициента наклона в регрессии

$WY = \beta_0 + \beta_1 Y + \varepsilon$.

Индекс Гири (Geary)

Индекс Гири для выбранного показателя Y рассчитывается по формуле:

$$C(Y) = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(Y_i - Y_j)^2}{2n\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

и принимает значения $0 < C(Y) < 2$,

$E(C) = 1$.

Индекс Гири (Geary)

H_0: C = 1 (расположение регионов является случайным)

H_1: C < 1 (положительная пространственная корреляция)
или H_1: C > 1 (отрицательная пространственная корреляция)

$$z = \frac{C - 1}{\sqrt{\mathrm{var}(C)}} \sim N(0,1),$$

z < 0 и p-v < α (уровень значимости) => положительная пространственная корреляция,
z > 0 и p-v < α => положительная пространственная корреляция

Индекс Гетиса-Орда

Индекс Гетиса-Орда для выбранного показателя Y рассчитывается по формуле:

$$G(Y) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} Y_i Y_j}{\sum_{i=1}^{n}\sum_{i=1}^{n} Y_i Y_j}$$

$E(G) = \frac{1}{n} \approx 0$ **для больших n.**

Индекс Гетиса-Орда

H_0: G = E(G) (расположение регионов является случайным или есть кластеры как с большими значениями показателя Y, так и с малыми значениями показателя Y)

H_1: G > E(G) (имеют место кластеры с большими значениями показателя Y),
или H_1: G < E(G) (имеют место кластеры с малыми значениями показателя Y)

$$z = \frac{\mathrm{G} - \mathrm{E(G)}}{\sqrt{\mathrm{var}(G)}} \sim N(0,1),$$

z > 0 и p-v < α (уровень значимости) => кластеры с большими значениями показателя Y,
z < 0 и p-v < α => кластеры с малыми значениями показателя Y.

Пример вычисления глобальных индексов пространственной зависимости

Example

1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8
5	6	7	8	9

	value	z	p-value	
Moran	0.84	2.35	0.0093	
Geary	0.12	2.35	0.0093	in the range (0;1)

9	3	8	2	7
3	8	2	7	2
8	2	7	2	6
2	7	2	6	1
7	2	6	1	5

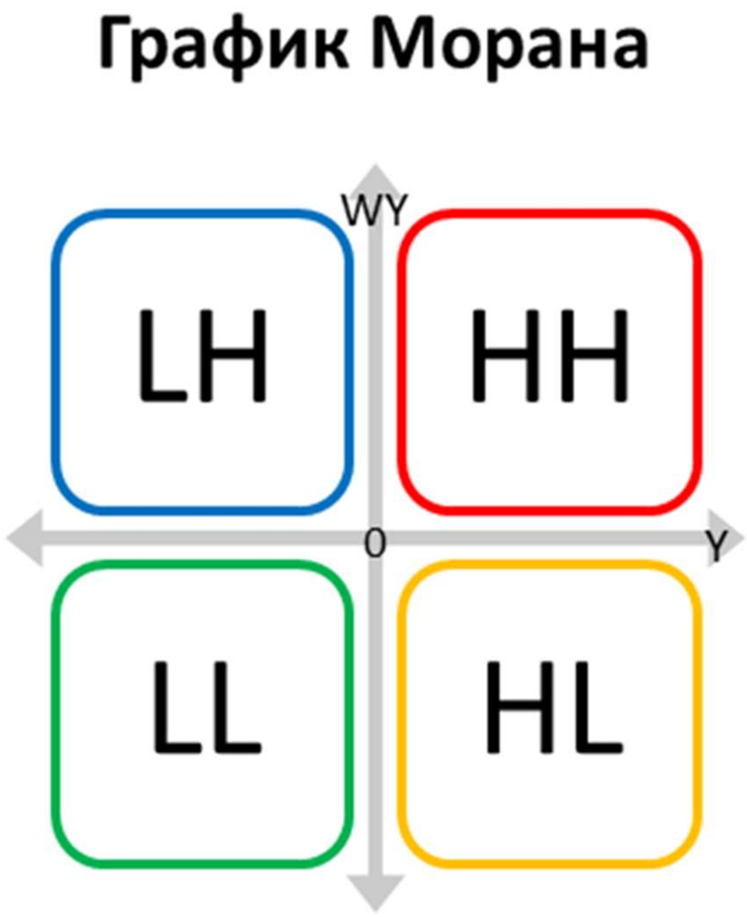
	value	z	p-value	
Moran	-0.89	1.37	0.085	
Geary	1.75	1.37	0.085	in the range (1;2)

График Морана

Весьма популярным способом выявления пространственной зависимости является график Морана, для каждого *i*-го региона по оси абсцисс откладывается значение рассматриваемого показателя Y_i, а по оси ординат – средневзвешенное значение этого же показателя WY_i (обычно соответствующие показатели предварительно центрируются и нормируются, т.е. преобразуются следующим образом:

$$\tilde{Y}_i = \frac{Y_i - \bar{Y}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}).$$

График Морана



Если большинство точек расположено в первом и третьем ортанте, то говорят о кластеризации, поскольку в этом случае большинство регионов окружено себе подобными.

Пример графика Морана

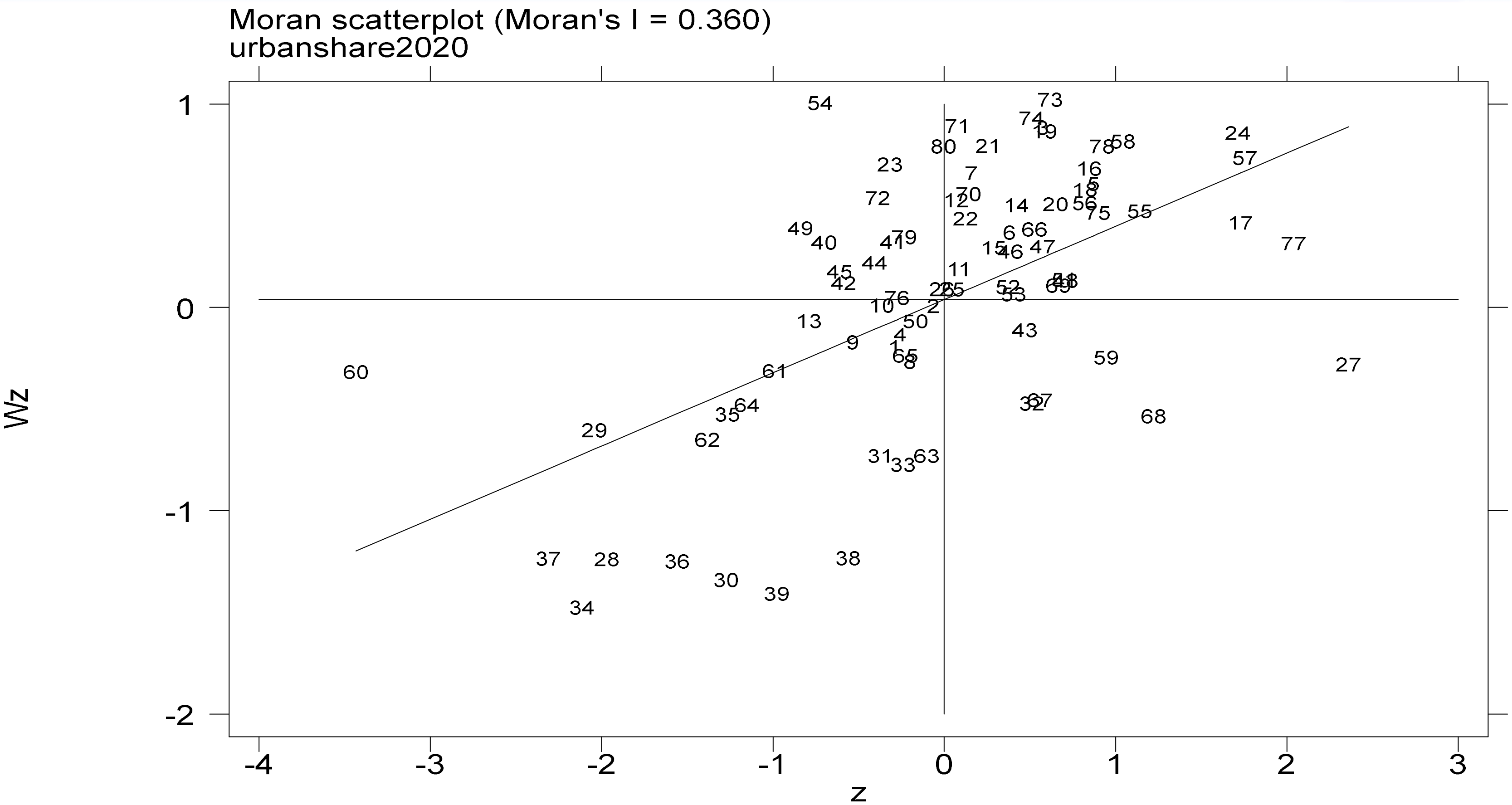


График Морана для доли городского населения в 2020. Наблюдения, «вылезающие» за 2, часто рассматривают как выбросы (outliers).

Локальные индексы пространственной зависимости

Кроме глобальных, рассчитывают также локальные значения перечисленных индексов (LISA – local indicators of spatial association) для каждого региона.

Основная информация об этих локальных индикаторах пространственной зависимости содержится в (Anselin L. Local Indicators of Spatial Association-LISA // Geographical Analysis. 1995. Vol. 27. No. 2. Pp. 93–115.).

См. также статью Подколзина Е. А., Демидова О. А., Кулецкая Л. Е. Пространственное моделирование электоральных предпочтений в Российской Федерации // Пространственная экономика. 2020. Т. 16. № 2. С. 70-100.

Локальный индекс Морана

Локальный индекс Морана для каждого из регионов $i = 1, \dots, n$ определяется следующим образом:

$$I_i(Y) = (Y_i - \bar{Y}_{J_i}) \sum_{j \in J_i}^{n} w_{ij} \, (Y_j - \bar{Y}_{J_i})^2 \, ,$$

где J_i обозначает множество соседей региона *i*,

$\bar{Y}_{J_i}$ – среднее значение рассматриваемого показателя именно для этих соседних для региона i наблюдений, n – число регионов.

Локальный индекс Морана

С его помощью можно проверить гипотезу

$H_0\colon I_i = \mathrm{E}(I_i)$ **(т.е. соседние с i-м регионы расположены случайным образом по рассматриваемому показателю) при альтернативной гипотезе**

1) $H_1\colon I_i > \mathrm{E}(I_i)$ **(т.е. регион** i **расположен в локальном кластере) или**
2) $H_1\colon I_i < \mathrm{E}(I_i)$ **(если у региона** i **наблюдается высокое значение рассматриваемого показателя Y, то у соседних регионов соответствующие показатели являются низкими, и наоборот).**

Локальный индекс Морана

Эти гипотезы проверяются с помощью тестовой статистики $z = \frac{I_i - E(I_i)}{\sqrt{\mathrm{Var}(I_i)}}$

Точный закон соответствующего распределения тестовой статистики неизвестен, его приходится восстанавливать с помощью симуляций.

Процедура выбора между основной и альтернативной гипотезами имеет некоторые особенности.

Правило выбора между основной и альтернативной гипотезой имеет вид: если $\mathrm{p-value} < \frac{\alpha}{n}$**, то гипотеза о случайном характере распределения регионов в окрестности *i*-го региона отвергается.**

Локальный индекс Гири

Локальный индекс Гири для каждого из регионов $i = 1, \dots, n$ определяется по формуле (Fischer, Wang, 2010, p.29)

$$c_i(Y) = \sum_{j \in J_i}^{n} w_{ij} \, (Y_i - Y_j)^2$$

и используется аналогично локальному индексу Морана.

Локальный индекс Гетиса-Орда

Локальные индексы Гетиса-Орда для каждого из регионов $i = 1, \dots, n$ и расстояния d определяются следующим образом (Fischer, Wang, 2010, p.27):

$$G_i(d,Y) = \frac{\sum_{j\neq i}^{n} w_{ij}(d)Y_j}{\sum_{j\neq i}^{n} Y_j}, G_i^*(d,Y) = \frac{\sum_{j=i}^{n} w_{ij}(d)Y_j}{\sum_{j=i}^{n} Y_j}$$

где $w_{ij}(d) = 0$, если расстояние между регионами i и j больше

d и $\dfrac{1}{\text{количество регионов, расстояние от которых до региона } i \text{ не превышает d}}$,

если расстояние между регионами i и j не превышает d. Отличие этих индексов состоит во включении или нет в суммы показателя Y_i для самого региона i.

Локальный индекс Гетиса-Орда

С помощью локальных индексов Гетиса-Орда для каждого региона i можно проверить гипотезу:
$H_0 \colon G_i(d) = \mathrm{E}(G_i(d))$ **(т.е. регион i не входит в локальный кластер по рассматриваемому показателю) при альтернативной гипотезе**
1) $H_1 \colon G_i(d) > \mathrm{E}(G_i(d))$ **(регион i входит в локальный кластер с большими значениями рассматриваемого показателя, их часто называют hot spots) или**
2) $H_1 \colon G_i(d) < \mathrm{E}(G_i(d))$ **(регион i входит в локальный кластер с малыми значениями рассматриваемого показателя, их часто называют cold spots).**
Алгоритм проверки соответствующей гипотезы с помощью z – статистики совпадает с описанным выше для локального индекса Морана (в этом случае также основная гипотеза отвергается, если $\mathrm{p-value} < \frac{\alpha}{n}$**).**
Локальные индикаторы пространственной зависимости могут быть использованы для выявления локальных очагов нестационарности.

Пример использования локальных индексов

Anselin L. Local Indicators of Spatial Association-LISA // Geographical Analysis. 1995. Vol. 27. No. 2. Pp. 93–115.

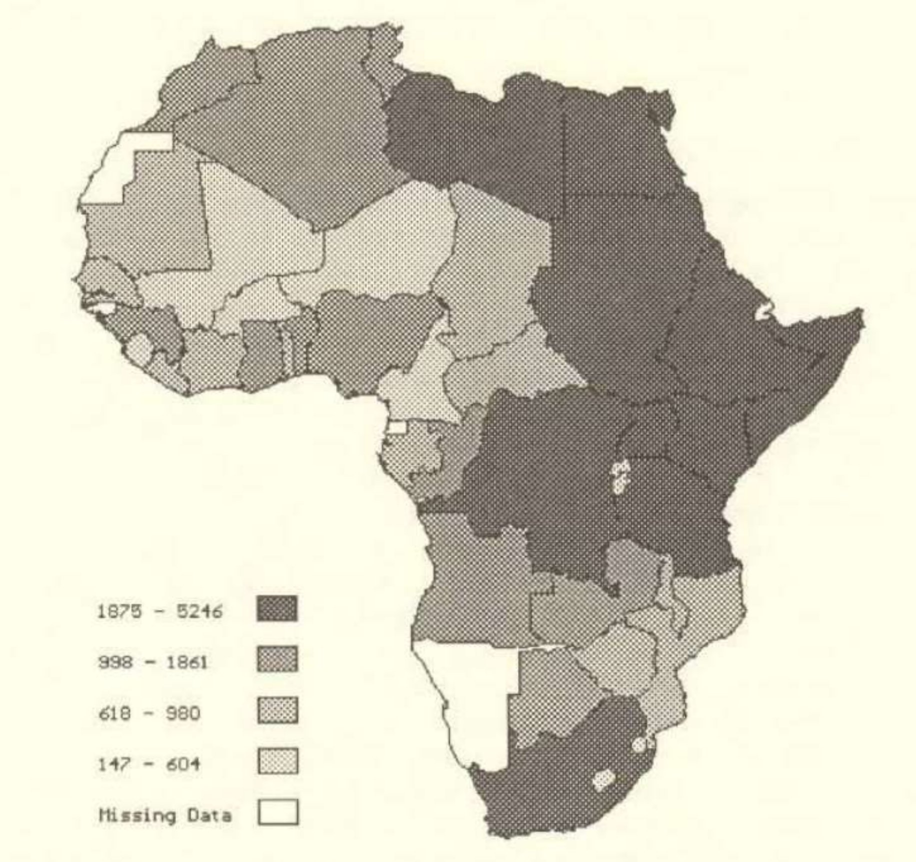
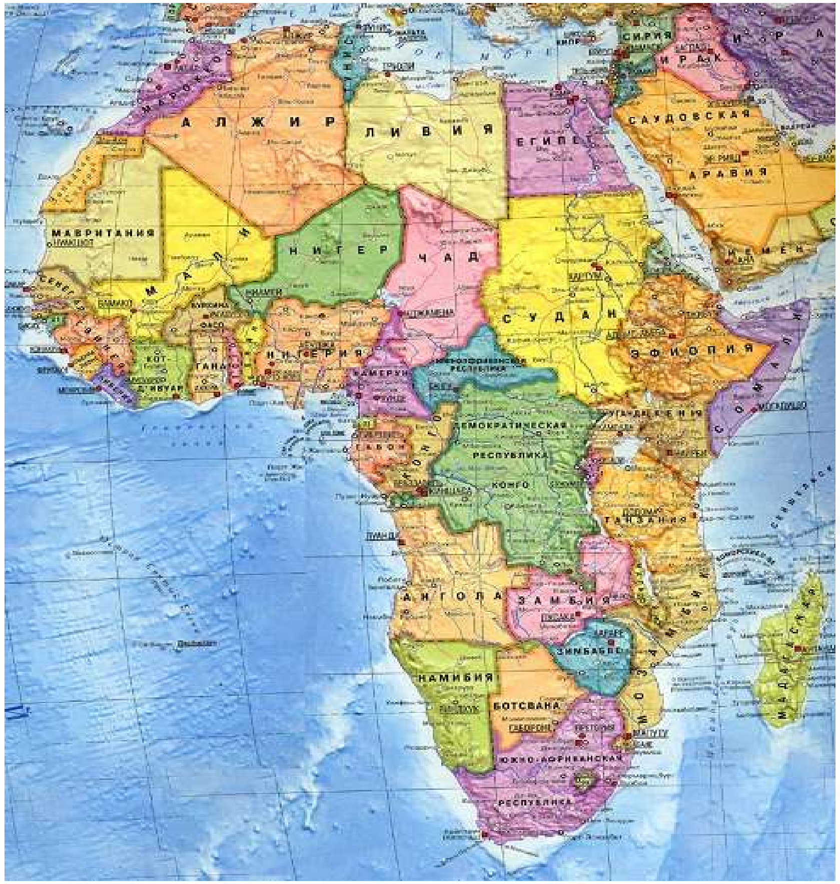


FIG. 1. Total Conflict Index for African Countries (1966–78)



Пример использования локальных индексов

Anselin L. Local Indicators of Spatial Association-LISA // Geographical Analysis. 1995. Vol. 27. No. 2. Pp. 93–115.

TABLE 1
Measures of Local Spatial Association

Id	Country	G_i^*	p	I_i	$z(I_i)$	p_n	p_r
1	Gambia	−0.984	0.1626	0.375	0.428	0.3342	0.4727
2	Mali	−1.699	0.0447	0.464	1.482	0.0692	0.0456
3	Senegal	−1.463	0.0717	0.257	0.623	0.2667	0.0270
4	Benin	−1.301	0.0966	0.194	0.484	0.3142	0.0612
5	Mauritania	−0.605	0.2726	0.097	0.269	0.3940	0.4111
6	Niger	−1.049	0.1471	0.231	0.774	0.2193	0.2404
7	Ivory Coast	−1.417	0.0782	0.290	0.788	0.2154	0.0611
8	Guinea	−1.449	0.0737	0.183	0.519	0.3020	0.0365
9	Burkina Faso	−1.751	0.0400	0.508	1.479	0.0695	0.0339
10	Liberia	−1.041	0.1490	0.186	0.398	0.3452	0.1333
11	Sierra Leone	−0.870	0.1921	0.265	0.444	0.3286	0.4006
12	Ghana	−1.103	0.1351	0.148	0.326	0.3721	0.0885
13	Togo	−0.991	0.1610	0.219	0.462	0.3219	0.1894
14	Cameroon	−1.133	0.1285	0.259	0.711	0.2387	0.1706
15	Nigeria	−1.173	0.1205	0.114	0.306	0.3798	0.0851
16	Gabon	−0.789	0.2150	0.204	0.349	0.3634	0.3139
17	CAR	1.174	0.1203	−0.442	−1.046	0.1477	0.0613
18	Chad	0.463	0.3218	−0.105	−0.225	0.4111	0.2125
19	Congo	−0.203	0.4198	0.011	0.079	0.4684	0.4734
20	Zaire	2.023	0.0216	0.710	2.591	0.0048	0.0404
21	Angola	1.235	0.1085	0.118	0.270	0.3936	0.0999
22	Uganda	3.336	**0.0004**	1.943	4.928	**0.0000**	0.0031
23	Kenya	3.503	**0.0002**	1.197	3.060	**0.0011**	0.0016
24	Tanzania	1.098	0.1360	0.272	0.973	0.1652	0.1898
25	Burundi	0.774	0.2194	−0.484	−0.872	0.1915	0.1040
26	Rwanda	1.457	0.0725	−0.752	−1.613	0.0534	0.0285
27	Somalia	1.183	0.1184	0.453	0.731	0.2324	0.1266
28	Ethiopia	2.627	0.0043	0.725	1.422	0.0775	0.0090
29	Zambia	0.753	0.2258	0.042	0.219	0.4134	0.1934
30	Zimbabwe	−0.200	0.4209	−0.010	0.033	0.4868	0.4041
31	Malawi	0.212	0.4161	−0.229	−0.388	0.3490	0.2088
32	Mozambique	−0.288	0.3868	0.017	0.114	0.4545	0.4728
33	South Africa	−0.868	0.1927	−0.183	−0.480	0.3156	0.1435
34	Lesotho	−0.298	0.3827	−0.419	−0.423	0.3361	0.2341
35	Botswana	0.041	0.4837	−0.004	0.039	0.4845	0.3691
36	Swaziland	−0.659	0.2548	0.017	0.063	0.4749	0.4128
37	Morocco	0.022	0.4913	−0.097	−0.111	0.4557	0.4995
38	Algeria	−0.363	0.3583	−0.010	0.040	0.4841	0.4139
39	Tunesia	0.579	0.2813	0.005	0.046	0.4818	0.1804
40	Libya	2.553	0.0053	0.804	2.300	0.0107	0.0133
41	Sudan	4.039	**0.0000**	2.988	9.898	**0.0000**	**0.0003**
42	Egypt	4.421	**0.0000**	6.947	10.679	**0.0000**	0.0058

They are Uganda (22), Kenya (23), Sudan (41), and Egypt (42), which themselves form a cluster in the northeast of Africa, part of the so-called Shatterbelt.

Пример использования индекса Морана

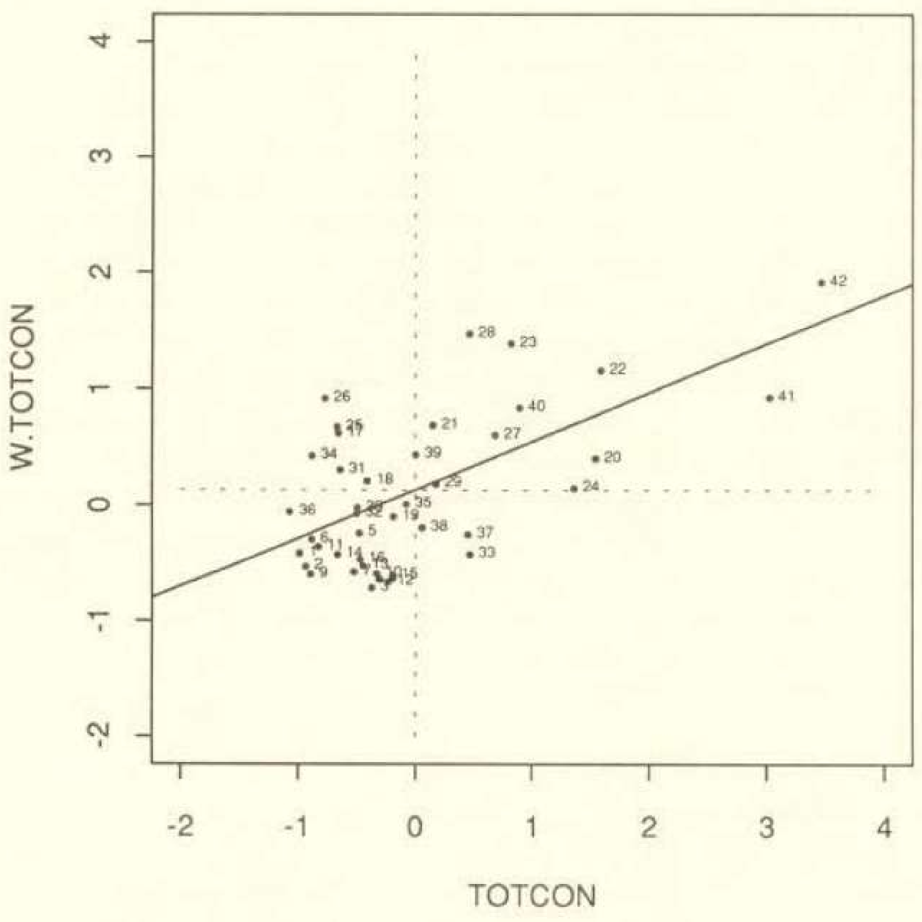


FIG. 3. Moran Scatterplot for Total Conflict ($I = 0.417$)

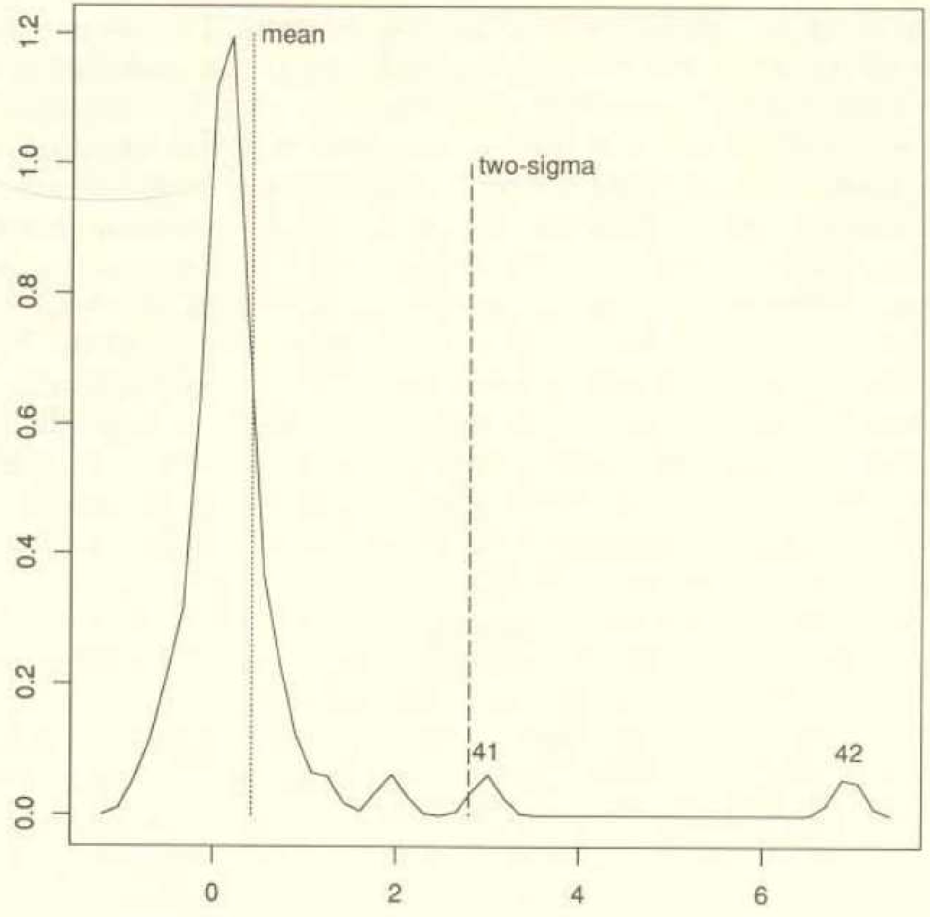


FIG. 4. Local Moran Outliers

Базовые пространственные модели

Базовые пространственные модели

- **Модель с пространственной зависимостью в ошибках (SEM)**
- **Модель с пространственной зависимостью в объясняющих переменных (SLX)**
- **Модель пространственной автокорреляции (SAR)**
- **Модель Дарбина (SDM)**

Модель с пространственной зависимостью в ошибках (SEM)

Когда в линейных регрессионных моделях в качестве объектов наблюдений используются страны, регионы и т.д. имеющие общие границы, потоки товаров и услуг и т.п., то предположение и независимости этих наблюдений, которое входит в условие теоремы Гаусса-Маркова, является нереалистичным. Если пространственная зависимость между факторами в моделях, оцениваемым по таким наблюдениям, не учтена, то она проявится в ошибках регрессии.

Модель с пространственной зависимостью в ошибках (SEM)

Пространственную зависимость шоков для соседних (в широком смысле) географических единиц можно учитывать с помощью модели с пространственной зависимостью в ошибках SEM (spatial error model):

$$Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u,$$

где n – число наблюдений, Y – зависимая переменная, i_n единичный n-мерный вектор, X – матрица объясняющих переменных, α, β – оцениваемые коэффициенты, ε – ошибки регрессии, W - пространственная матрица, λ – коэффициент пространственной корреляции ошибок регрессии, u – независимые одинаково распределенные случайные величины с дисперсией σ_u^2 (обычно делается предположение, что они нормально распределены).

Модель с пространственной зависимостью в ошибках (SEM)

$Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u,$

Для проверки, действительно ли имеет место пространственная зависимость в остатках, обычно проверяют гипотезу $H_0: \lambda = 0$

(при альтернативной гипотезе $H_1: \lambda \neq 0$)

с помощью тестовой статистики Лагранжа

$$LM = \left(\frac{e'We}{e'en^{-1}}\right)^2 \frac{1}{tr[W'W+W^2]}.$$

Если p-v < α (значимости), то H_0 отвергается (есть пространственная зависимость в ошибках).

Модель с пространственной зависимостью в ошибках (SEM)

$$Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u$$

Поскольку ковариационная матрица ошибок регрессии в случае их пространственной зависимости не пропорциональна единичной, а именно,

$$var[\varepsilon] = \sigma_u^2 (I - \lambda W)^{-1} (I - \lambda W')^{-1},$$

то оценки параметров α, β методом наименьших квадратов будут несмещенными, но их стандартные ошибки (а следовательно, и t-статистики для проверки значимости) будут вычислены неверно. Поэтому для оценки параметров обычно используют метод максимального правдоподобия.

Модель с пространственной зависимостью в ошибках (SEM)

$Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u$

Оценки коэффициентов β в модели с пространственной зависимостью в ошибках интерпретируются традиционным для линейных моделей образом: если коэффициент β_j при переменной X_j значим, то при увеличении X_j на одну единицу, зависимая переменная изменится на $\hat{\beta}_j$ единиц.

Модель с пространственной зависимостью в ошибках (SEM)

$Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u$

Для интерпретации коэффициента пространственной корреляции λ ошибок регрессии удобно использовать следующее представление ошибок регрессии:

$\varepsilon = (I - \lambda W)^{-1}u = (I + \lambda W + \lambda^2 W^2 + \cdots)u = u + \lambda W u + \lambda^2 W^2 u + \cdots,$

где $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$, $u = (u_1, \dots, u_n)', Wu = \left((Wu)_1, \dots, (Wu)_n\right)', (Wu)_i = \sum_{j=1}^{n} w_{ij}\, u_j$ **и т.д.**

В этом разложении u_i характеризует шоки рассматриваемого i-го региона, $(Wu)_i$ – средние шоки соседних с i-м регионов, $(W^2u)_i$ – шоки в соседних с соседними регионами и т.д. Чем меньше $|\lambda|$, тем быстрее затухает влияние шоков в соседних регионах.

Модель с пространственной зависимостью в объясняющих переменных (SLX)

Spatial lag model: $Y = \alpha i_n + X\beta + WX\theta + \varepsilon$,

где Y – зависимая переменная, n – число рассматриваемых объектов, i_n – единичный вектор длины n, $X = (X_1, \dots, X_k)$ – матрица объясняющих переменных, $W = (w_{ij})$ – нормированная по строкам пространственная матрица, WX – матрица пространственных лагов объясняющих переменных, ε – вектор ошибок регрессии.

Можно оценить модель с помощью МНК.

Модель с пространственной зависимостью в объясняющих переменных (SLX)

Более удобна следующая форма этого уравнения:

$$Y_i = \alpha + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} +$$

$$+ \theta_1 \sum_{j=1}^{n} w_{ij} X_{1j} + \cdots + \theta_k \sum_{j=1}^{n} w_{ij} X_{kj} + \varepsilon_i, i = 1, \dots, n,$$

из которой очевидно, что

1) при изменении переменной $X_m, m = 1, \dots, k$ **на одну единицу измерения в регионе** i **,** Y_i **изменится на** $\hat{\beta}_m$,

2) при изменении переменной $X_m, m = 1, \dots, k$ **на одну единицу измерения в каждом из соседних с i-м регионов,** Y_i **изменится на** $\hat{\theta}_m$, **а если только в регионе** j_0, **то на** $\hat{\theta}_m w_{i j_0}$ **. Это будут локальные спилловеры.**

Модель пространственной авторегрессии (SAR)

SAR (spatial autoregressive model): $Y = \alpha i_n + \rho WY + X\beta + \varepsilon$**,**

где Y – зависимая переменная, n – число рассматриваемых объектов, i_n **– единичный вектор длины n,** $X = (X_1, \dots, X_k)$ **– матрица объясняющих переменных,** WY **– пространственный лаг зависимой переменной, ε – вектор ошибок регрессии.**

WY **является эндогенной переменной, поэтому МНК применять нельзя, соответствующие оценки не будут состоятельными.**

Модель пространственной авторегрессии (SAR)

Модель

$$Y = \alpha i_n + \rho WY + X\beta + \varepsilon ,$$

можно переписать в следующем виде

$(I - \rho W)Y = \alpha i_n + X\beta + \varepsilon$ **или**

$$Y = (I - \rho W)^{-1}\alpha i_n + (I - \rho W)^{-1}X\beta + (I - \rho W)^{-1}\varepsilon$$

Сделав предположение о нормальности ε,

можно оценить параметры модели с помощью метода максимального правдоподобия.

Спилловеры в модели пространственной авторегрессии (SAR)

$Y = \alpha i_n + \rho W Y + X\beta + \varepsilon$,

$Y = (I - \rho W)^{-1} \alpha i_n + (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$

поскольку $(I - \rho W)^{-1} X\beta = (I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots) X\beta = X\beta + \rho W X\beta + \rho^2 W^2 X\beta + \rho^3 W^3 X\beta + \cdots$,

то $\rho W X\beta$ **характеризует влияние соседей первого порядка,**

$\rho^2 W^2 X\beta$ **– соседей второго порядка (или соседей соседей),**

$\rho^3 W^3 X\beta$ **– соседей третьего порядка и т.д.**

Еще один способ оценки модели пространственной авторегрессии (SAR)

$$Y = \alpha i_n + \rho WY + X\beta + \varepsilon\ ,$$

$$Y = (I - \rho W)^{-1}\alpha i_n + (I - \rho W)^{-1}X\beta + (I - \rho W)^{-1}\varepsilon$$

$$Y = (I - \rho W)^{-1}\alpha i_n + (I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots)X\beta +$$

$$+(I - \rho W)^{-1}\varepsilon + (I - \rho W)^{-1}\varepsilon$$

Для эндогенного пространственного лага WY используют инструменты: переменные, входящие в матрицы $X, WX, W^2\mathrm{X}$ (предложение Kelejian, Prucha).

Пространственная модель Дарбина (SDM)

Пространственная модель Дарбина (SDM (spatial Durbin model) – пространственная модель Дарбина) позволяет наиболее полно учесть пространственную зависимость между рассматриваемыми объектами (чаще всего странами или регионами), поскольку включает пространственные лаги как зависимой, так и объясняющих переменных.

NATIONAL RESEARCH UNIVERSITY

Пространственная модель Дарбина (SDM)

Функциональная форма этой модели, оцениваемой по наблюдениям за n объектами (например, регионами):

$$Y = \alpha i_n + X\beta + \rho WY + WX\theta + \varepsilon,$$

где Y – зависимая переменная, n – число рассматриваемых объектов, i_n – единичный вектор длины n, X – матрица объясняющих переменных $X_1, \dots, X_k$, WY – пространственный лаг зависимой переменной, WX – матрица пространственных лагов объясняющих переменных, ε – вектор ошибок, β, ρ, θ - оцениваемые параметры.

Пространственная модель Дарбина (SDM)

Модель

$$Y = \alpha i_n + \rho WY + X\beta + WX\theta + \varepsilon ,$$

можно переписать в следующем виде

$(I - \rho W)Y = \alpha i_n + X\beta + WX\theta + \varepsilon$ **или**

$$Y = (I - \rho W)^{-1}\alpha i_n + (I - \rho W)^{-1}X\beta +$$

$$+ (I - \rho W)^{-1}WX\theta + (I - \rho W)^{-1}\varepsilon$$

Сделав предположение о нормальности ε, **можно оценить параметры модели с помощью метода максимального правдоподобия.**

Интерпретация результатов оценивания пространственных моделей

Для интерпретации результатов оценивания пространственных моделей используются предельные эффекты. Продемонстрируем их вычисление на примере пространственной модели Дарбина (SDM).

$$Y = (I - \rho W)^{-1}(\alpha i_n + X\beta + WX\theta + \varepsilon)$$

$$\frac{\partial E(Y)}{\partial X_m} = \begin{pmatrix} \frac{\partial E(Y_1)}{\partial X_{m1}} & \cdots & \frac{\partial E(Y_1)}{\partial X_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(Y_n)}{\partial X_{m1}} & \cdots & \frac{\partial E(Y_n)}{\partial X_{mn}} \end{pmatrix} = (I - \rho W)^{-1}(\beta_m I + W\theta_m) =$$

$$= (I - \rho W)^{-1} \begin{pmatrix} \beta_m & w_{12}\theta_m & \cdots & w_{1n}\theta_m \\ w_{21}\theta_m & \beta_m & & \vdots \\ \vdots & & \ddots & \\ w_{n1}\theta_m & \cdots & & \beta_m \end{pmatrix} = S(X_m), m = 1, \dots, k.$$

Интерпретация результатов оценивания пространственных моделей

$$S(X_m) = (I - \rho W)^{-1} \begin{pmatrix} \beta_m & w_{12}\theta_m & \cdots & w_{1n}\theta_m \\ w_{21}\theta_m & \beta_m & & \vdots \\ \vdots & & \ddots & \\ w_{n1}\theta_m & \cdots & & \beta_m \end{pmatrix}, m = 1, \ldots, k.$$

Диагональные элементы этой матрицы называются прямыми предельными эффектами, элемент $s_{ii}(X_m)$ показывает, как изменится зависимая переменная Y в регионе i при изменении фактора X_m в этом же регионе на одну единицу измерения.

Внедиагональные элементы этой матрицы называются косвенными предельными эффектами (или *спилловерами*, эффектами перелива), элемент $s_{ij}(X_m)$, $i \neq j$ показывает, как изменится зависимая переменная Y в регионе i при изменении фактора X_m в другом регионе j на одну единицу измерения.

Интерпретация результатов оценивания пространственных моделей

Поскольку предельных эффектов очень много, n прямых и $n^2 - n$ косвенных, то обычно интерпретируют средние предельные эффекты:

- ADE – average direct effect, средний прямой эффект: $ADE(X_m) = \frac{1}{n}\sum_{i=1}^{n} s_{ii}$, который показывает, как изменение переменной X_m на одну единицу влияет на изменение переменной Y в этом же регионе (например, как инвестиции в выбранном регионе влияют на экономический рост в этом регионе),
- AIE – average indirect effect, средний косвенный (или спилловер) эффект: $AIE(X_m) = \frac{1}{n}\sum_{\substack{i=1 \\ i \neq j}}^{n} s_{ij}$, который показывает, как изменение переменной X_m на одну единицу в соседних регионах влияет на изменение переменной Y в выбранном регионе (например, как инвестиции в соседних регионах влияют на экономический рост в выбранном регионе),
- ATE – average total effect, средний общий эффект: $ATE(X_m) = \frac{1}{n}\sum_{i=1}^{n} s_{ij}$, который показывает, как изменение переменной X_m на одну единицу во всех регионах влияет на изменение переменной Y в выбранном регионе (например, как инвестиции в выбранном регионе и в соседних регионах в совокупности влияют на экономический рост).

Выбор между пространственными моделями

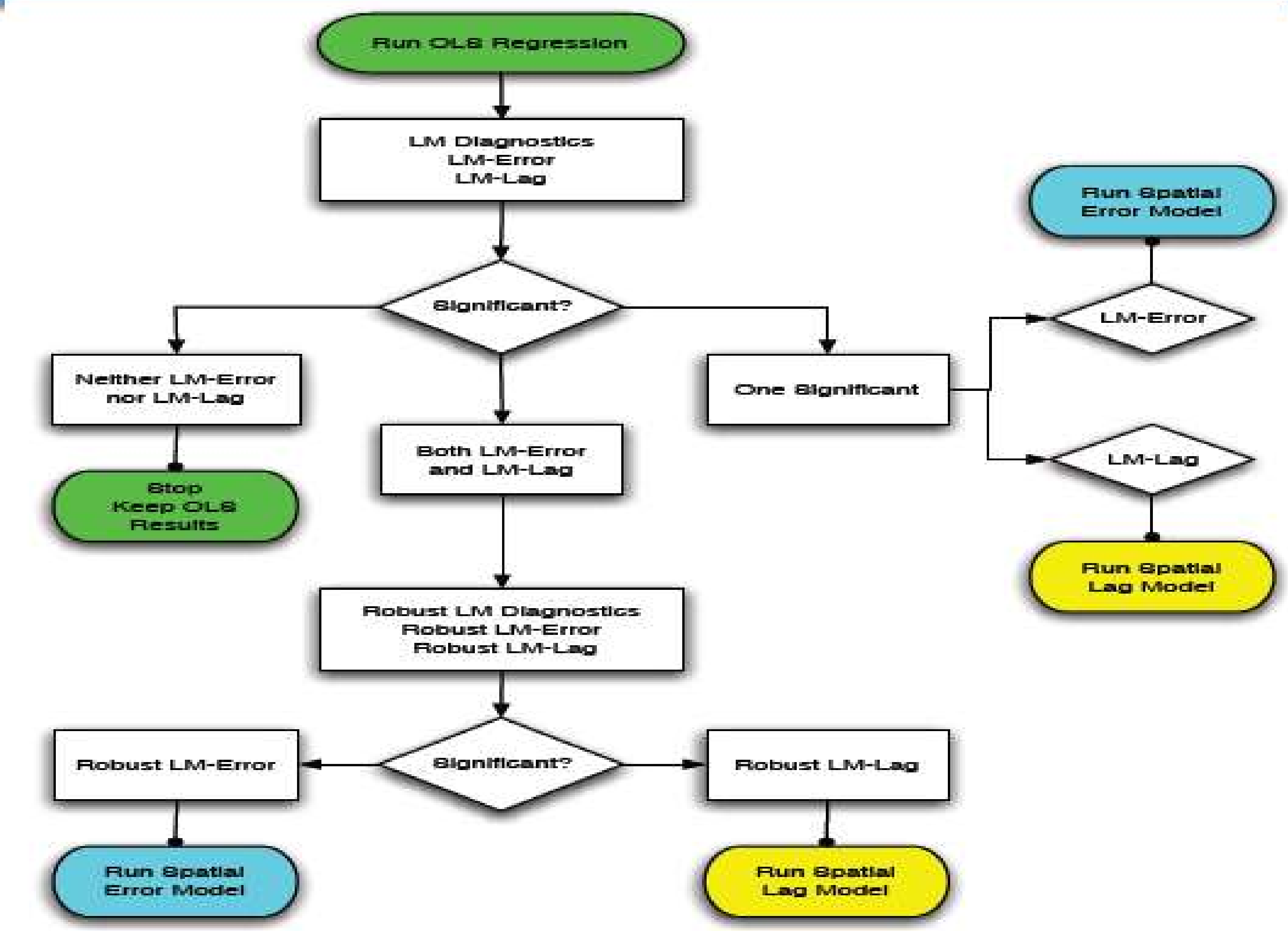
Два основных подхода:

1) От простых моделей к сложным (Anselin et al., 1996)

2) От сложных моделей к простым (Elhorst (2014), LeSage, Pace (2009), Kelejian, Prucha (2010))

Схема Л.Анселина

THE CHOICE OF THE MODEL



Первый подход к выбору моделей

Первый подход (Anselin et al., 1996) состоит в оценке линейной модели $Y = \alpha i_n + X\beta + \varepsilon$, сохранении остатков этой регрессии $e' = (e_1, \dots, e_n)$ и проверке гипотез

1) H_0: $Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u, \lambda = 0$, т.е. пространственная зависимость в ошибках регрессии отсутствует,

H_1: $Y = \alpha i_n + X\beta + \varepsilon, \varepsilon = \lambda W\varepsilon + u, \lambda \neq 0$, т.е. это модель SEM,

с помощью теста Лагранжа и тестовой статистики (Fischer, Wang, 2010, p.36):

$$LM(error) = \left(\frac{e'We}{e'en^{-1}}\right)^2 \frac{1}{tr[W'W + W^2]} \overset{H_0}{\sim} \chi^2(1),$$

2) H_0: $Y = \alpha i_n + \rho WY + X\beta + \varepsilon,\ \rho = 0$, т.е. пространственный лаг зависимой переменной отсутствует,

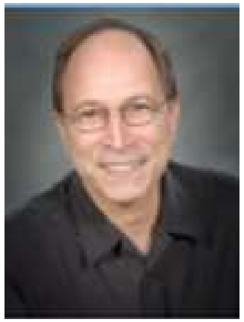
H_1: $Y = \alpha i_n + \rho WY + X\beta + \varepsilon, \rho \neq 0$, т.е. это модель SAR,

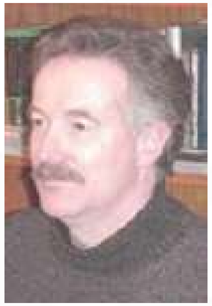
с помощью теста Лагранжа и тестовой статистики

$$LM(lag) = \left(\frac{e'WY}{e'en^{-1}}\right)^2 \frac{1}{H} \overset{H_0}{\sim} \chi^2(1),$$

$$\mathrm{H} = \{(\mathrm{WX}\hat{\beta})'[I - X(X'X)^{-1}X'](WX\hat{\beta})\hat{\sigma}_\varepsilon^{-2}\} + tr(W'W + W^2).$$

Второй подход к выбору моделей





При втором подходе (Elhorst (2014), LeSage, Pace (2009), Kelejian, Prucha (2010)) вначале оценивается наиболее общая модель (GNS – general nesting spatial model):

$$Y = \alpha i_n + X\beta + \rho WY + WX\theta + u, \varepsilon = \lambda W\varepsilon + u,$$

а потом с помощью теста Лагранжа или Вальда проверяются гипотезы:

1) $H_0: \lambda = 0$ (тогда модель сводится к SDM),
2) $H_0: \rho = 0, \lambda = 0$ (тогда модель сводится к SLX),
3) $H_0: \theta = 0, \lambda = 0$ (тогда модель сводится к SAR),
4) $H_0: \rho = 0, \theta = 0, \lambda = 0$ (тогда модель сводится к линейной регрессионной модели),
5) $H_0: \theta + \rho\beta = 0$ (тогда модель сводится к SEM).

Схема П.Элхорста

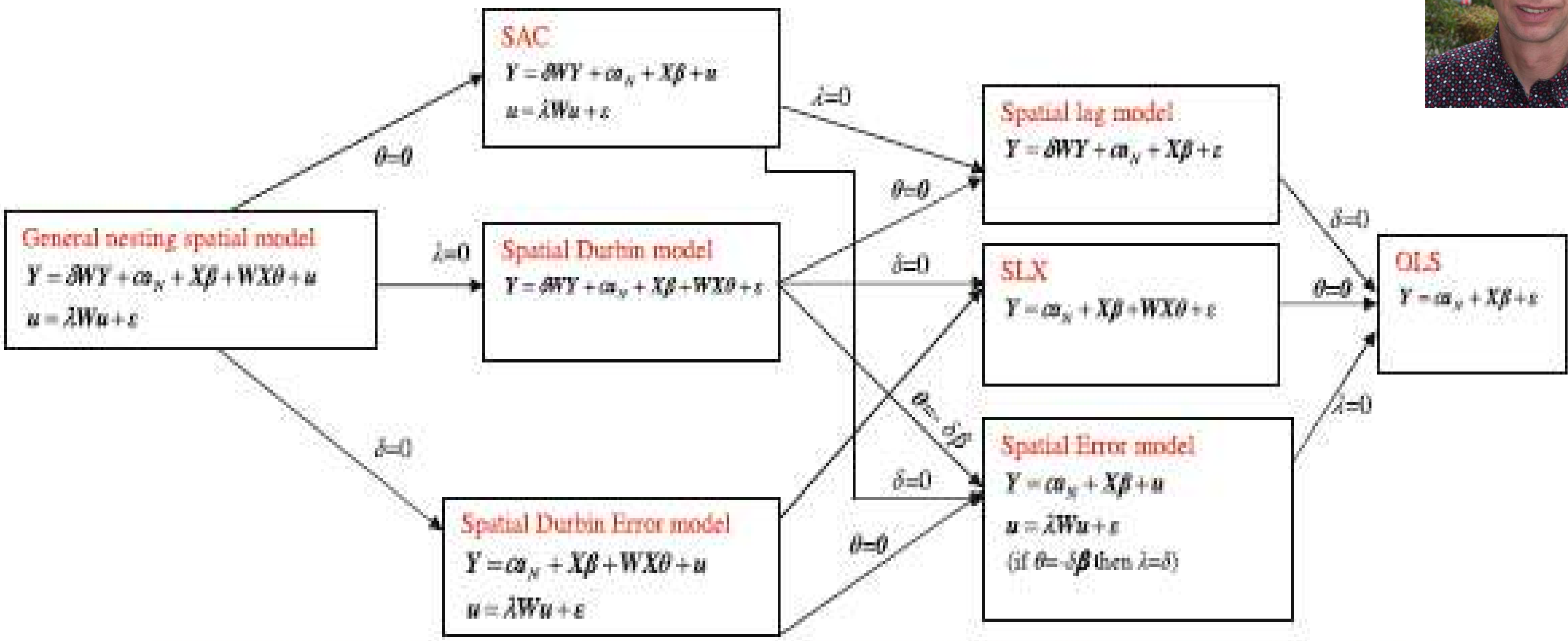


Fig. 2.1 The relationships between different spatial dependence models for cross-section data (*source* Halleck Vega and Elhorst 2012)

Пространственные модели панельных данных

Основные пространственные модели, оцениваемые по пространственным данным, аналогичны перечисленным ранее.

Например, пространственная модель Дарбина имеет следующую функциональную форму:

$$Y_t = \alpha + X_t\beta + \rho W Y_t + W X_t\theta + c_t + \varepsilon_t,$$

где $t = 1, \dots, T$ — рассматриваемые моменты времени, n — число регионов, $Y_t = (Y_{1t}, \dots, Y_{nt})'$ – зависимая переменная, X_t — матрица объясняющих переменных,
$\rho W Y_t$ - пространственный лаг зависимой переменной,
$W X_t$ - матрица пространственных лагов объясняющих перемнных,

$\alpha = (\alpha_1, \dots, \alpha_n)'$ – вектор фиксированных индивидуальных эффектов,
β, θ – векторы оцениваемых коэффициентов,
c_t — временные эффекты, ε_t – вектор ошибок.

Выбор пространственной модели панельных данных

Рекомендуется начинать с пространственной модели Дарбина :

$$Y_t = \alpha + X_t\beta + \rho W Y_t + W X_t \theta + c_t + \varepsilon_t,$$

После оценки модели рекомендуется проверить гипотезы об ограничениях на параметры:
H_0: $\theta = 0$ **(если эта гипотеза не отвергается, то модель сводится к модели пространственной авторегрессии SAR),**
H_0: $\theta = \rho = 0$ **(если эта гипотеза не отвергается, то модель сводится к линейной регрессионной модели, нет необходимости учитывать пространственную зависимость),**
H_0: $\theta + \rho\beta = 0$ **(если эта гипотеза не отвергается, то модель сводится к модели с пространственной зависимостью в ошибках SEM).**
Если значимы какие-нибудь из параметров θ**, то говорят о локальных (идущих от соседей) спилловерах. Если же значим коэффициент пространственной автокорреляции** ρ**, то имеют место глобальные спилловеры, отражающие влияние на регион не только непосредственных соседей, но и соседей второго, третьего и т.д. порядка.**
Интерпретируются также прямые, косвенные и общие предельные эффекты.

Пространственно-эконометрические исследования с использованием региональных российских данных

Михайлова Татьяна Николаевна,
РЭШ, РАНХиГС
A.Markevich, T.Mikhailova, Economic Geography of Russia// The Oxford Handbook of the Russian Economy, 2012

«Spatial structure of the economy is slow to change, especially in Russia, where, as we know from the evidence of the last 20 years, mobility of population is still rather limited. This means that in order to understand Russian spatial economy, we have to understand its evolution in historical perspective».

Пространственно-эконометрические исследования с использованием региональных российских данных

Луговой Олег Валерьевич, Институт Экономики Переходного Периода, РАНХиГС

Луговой О.В. и др. (2007). Экономико-географические и институциональные аспекты экономического роста в регионах, М., ИЭПП

"При статистическом анализе на региональных данных возникает ряд проблем, которые изучаются пространственной эконометрикой. В рамках обычной модели (безусловной или условной) конвергенции игнорируется возможность пространственного взаимодействия, поскольку неявно предполагается, что регионы в рассматриваемой экономической системе представляют собой независимые географические единицы. Такие факторы как мобильность капитала и трудовых ресурсов, распространение (диффузия) знаний и технологий, транспортные затраты – существенно влияют на межрегиональное взаимодействие, а значит и на основные показатели регионов и темпы их роста. Разумно предполагать, что регионы, ближе расположенные друг к другу, как правило, более интегрированы между собой, чем расположенные на значительном расстоянии. Основная предпосылка пространственной эконометрики состоит в том, что исследуемые показатели могут быть автокоррелированы в пространстве, т.е. наблюдения изучаемых показателей в пространстве и их динамика не случайны, а определяются региональной принадлежностью".

Пространственно-эконометрические исследования с использованием региональных российских данных

Константин Холодилин, Алексей Ощепков (НИУ ВШЭ), Борис Селиверстов

Kholodilin K., Oshchepkov A. Y., Siliverstovs B. The Russian regional convergence process: Where is it leading? // *Eastern European Economics*. 2012. Vol. 50. No. 3. P. 5-26

Разделение всех российских регионов на 4 группы с помощью графика Морана.

«we find a strong regional convergence among high-income regions located near other high-income regions. Our results indicate that estimating the speed of convergence using aggregate data may result in misleading conclusions regarding the nature of the convergence process among Russia's regions».

Пространственно-эконометрические исследования с использованием региональных российских данных

Коломак Евгения Анатольевна ИЭиОПП СО РАН, Новосибирский Государственный Университет

Пространственные экстерналии как ресурс экономического роста// Регион: Экономика и Социология, 2010, №4

«Проведенный анализ показал, что в России, несмотря на большие расстояния, относительно низкую плотность деловой активности и сравнительно высокие издержки межрегионального взаимодействия, работают импульсы и мультипликаторы экономического роста, которые не локализуются в границах региона, а распространяются на другие территории. Однако если в европейской части страны преобладают положительные экстерналии экономического роста, то в восточной части доминируют отрицательные внешние эффекты. Такие различия являются достаточно естественными и объясняются большими пространствами и дефицитом инфраструктуры транспорта и связи в регионах Сибири и Дальнего Востока».

Пространственно-эконометрические исследования с использованием региональных российских данных

Буфетова Анна Николаевна, ИЭиОПП СО РАН, Новосибирский Государственный Университет

Пространственные аспекты концентрации экономической активности в России // Пространственная экономика. – 2016. – № 3. – С. 38-56.

«Анализ пространственных эффектов показал, что их характер зависит от соотношения и степени различия уровня развития регионов и их ближайших соседей, а высокая дифференциация регионов по уровню экономической активности препятствует развитию отсталых регионов и способствует дальнейшей поляризации. В таких условиях более адекватной сложившейся ситуации представляется политика, направленная на сдерживание роста регионального неравенства и становление конкурентного сотрудничества регионов».

Пространственно-эконометрические исследования с использованием региональных российских данных

Демидова Ольга Анатольевна

НИУ ВШЭ

1. **Пространственно – авторегрессионная модель для двух групп взаимосвязанных регионов (на примере восточной и западной части России) // Прикладная эконометрика, 2014. Т. 34. № 2. С. 19—352.**
2. **Демидова О.А., Иванов Д.С., Модели экономического роста с неоднородными пространственными эффектами (на примере российских регионов) // Экономический журнал Высшей школы экономики, 2016. Т. 20. № 1. С. 52—75**
3. **Демидова О. А., Карнаухова Е. Е., Коршунов Д. А., Мясников А. А., Серегина С. Ф. Асимметричные эффекты денежно-кредитной политики в регионах России // Вопросы экономики. 2021. № 6. С. 77-102.**

Пространственно-эконометрические исследования с использованием региональных российских данных

Вакуленко Елена Сергеевна

НИУ ВШЭ

Анализ связи между региональными рынками труда в России с использованием закона Оукена //Прикладная эконометрика, 2015, 40, с.28-48

Оценивается широкий набор спецификаций пространственно-эконометрических моделей. В результате делается вывод, что оценки, построенные без учета пространственного взаимодействия, дают заниженные значения коэффициента Оукена.

Пространственно-эконометрические исследования с использованием региональных российских данных

Иванова Вера Ивановна

НИУ ВШЭ, Санкт-Петербург

1. Ivanova V. I. Spatial Convergence of Real Wages in Russian Cities // The Annals of Regional Science. 2017

2. Иванова В. И. Региональная конвергенция доходов населения: пространственный анализ // Пространственная экономика. 2014. № 4. С. 100-119.

Пространственно-эконометрические исследования с использованием региональных российских данных

Балаш Владимир Алексеевич,

Саратовский Государственный Университет

В.А.Балаш, О.С.Балаш, А.В.Харламов, Эконометрический анализ геокодированных данных о ценах на жилую недвижимость// Прикладная Эконометрика 22, 2011

Геокодированные данные существенно расширяют возможности экономического исследования пространственно распределенных явлений и процессов. Для моделирования цен на вторичном рынке жилья Саратова в работе был использован подход географически взвешенной регрессии. Переменные коэффициенты модели, плавно изменяющиеся по территории, позволяют в агрегированной форме отразить закономерности и локальные особенности ценообразования на вторичном рынке жилья, которые трудно воспроизвести стандартными методами.

Пространственно-эконометрические исследования с использованием региональных российских данных

Тимирьянова Венера Маратовна,

Башкирский Государственный Университет

Тимирьянова В.М., Зимин А.Ф., Юсупов К.Н. Экономическая активность территорий: сравнительный анализ способов оценки пространственных эффектов // Пространственная экономика. 2021. Т. 17. № 4. С. 41–68.

«В статье обсуждаются различные варианты учета пространственной зависимости данных в рамках иерархического и пространственного подходов. На основе литературного обзора определяются преимущества и недостатки каждого подхода и потенциал их сочетания. Проводится сравнение результатов расчетов моделей OLS, SAR, SEM, HLM, HSAR. При используемом наборе данных (2285 муниципальных образований в разрезе 85 субъектов РФ) акцент в работе сделан не на выявление связи между зависимой переменной и факторами, а на сравнение пространственных эффектов, которые могут быть выделены в рамках каждой из рассматриваемых моделей».

Рекомендуемые ресурсы по пространственной эконометрике

1) О. А. Демидова Методы пространственной эконометрики и оценка эффективности государственных программ // Прикладная эконометрика. 2021. № 4(64). С. 107-134.

2) Учебные материалы исследовательской рабочей группы «Центр пространственной эконометрики в прикладных макроэкономических исследованиях: https://economics.hse.ru/cseamr/materials

Спасибо за внимание!

demidova@hse.ru

http://www.hse.ru/org/persons/demidova_olga