

ИССЛЕДОВАТЕЛЬСКИЙ ПРОЕКТ ПО КУРСУ
«ЭКОНОМЕТРИКА-2»

**«С кем поведёшься, от того и наберёшься? или как
социальные связи помогают объяснять
предпочтения»**

Авторы

Семен Жижерин, Данила Кочнев

Содержание

1	Введение	1
2	Обзор литературы	3
3	Основная часть	7
3.1	Данные	7
3.2	Результаты оценивания	15
3.3	Устойчивость результатов	22
3.4	Ссылки	23
4	Заключение	24
A	Приложение: графики	25
	Литература	29

1 Введение

Настоящая работа потребовала вложения большого труда, и её написание было бы невозможно без помощи. Мы благодарим наших одноклассников Алину Куданову, Таисию Куликову и Николая Маркова за их вклад в работу над проектом. Мы также хотим отдельно поблагодарить Дмитрия Артамонова и Ольгу Сучкову за помощь и наставления по написанию работы.

Классической предпосылкой эконометрического анализа пространственной выборки является независимость наблюдений: в отличие от временных рядов или панельных данных, здесь сложно представить, как могут быть взаимосвязаны разные наблюдения, полученные за один и тот же период времени. Однако, как только речь заходит об исследовании различных предпочтений, становится понятно, что такое предположение нереалистично. Можно считать, что на политические и экономические взгляды, вкусы и ценности людей могут влиять мнения их друзей и знакомых – тех, с кем они чаще всего общаются. Сегодня, когда существенная доля социального взаимодействия перенесена в онлайн, особенно интересным становится изучение влияния социальных сетей на предпочтения их активных пользователей. Целью настоящей работы является выявление и последующая количественная оценка такого влияния на примере анализа предпочтений жителей европейских стран.

Существующие на данный момент исследования по теме сходятся на том, что социальные связи являются статистически значимым детерминантом самых разных предпочтений от туризма и отношения к работе до научного соавторства и евроскептицизма. Однако в большинстве работ влияние этого фактора оценивается включением некоторых прокси социальной связности в список регрессоров. Такой подход не учитывает пространственной структуры данных, то есть взаимозависимости наблюдений. Здесь кажется перспективным использование моделей пространственной эконометрики, которые бы учитывали влияние наблюдений в одном регионе на наблюдения в другом. Актуальность данной работы заключается в применении этих методов для уточнения полученных ранее оценок и проверки значимости описанного эффекта. Несмотря на то, что в литературе наиболее распространено использование пространственных моделей, понимающих близость стран в смысле географических расстояний, эта работа концентрируется на близости в смысле социальной связности. Как показано далее, такой подход в ряде случаев заметно выигрывает у географического в смысле точности оценивания.

В рамках настоящей работы анализируются различные предпочтения жителей Европы. Вначале исследуются политические взгляды и отношение к Евросоюзу. Мотивация заключается в том, что социальные связи способствуют распространению политической информации как через общение отдельных пользователей, так и в результате деятельности целых сообществ. Затем рассматриваются более общие предпочтения, имеющие отношение к жизненной позиции и ценностям. Вопросы толерантности и доверия как к отдельным социальным группам (допустим, лицам нетрадиционной сексуальной ориентации), так и к институтам (например, власти или церкви) также следует рассматривать через призму социальных связей. Мотивация здесь совершенно аналогичная: личное общение и деятельность групп в сети помогает распространять информацию, тем самым делая пространственные наблюдения взаимозависимыми.

2 Обзор литературы

За последний год было опубликовано большое число работ, посвящённых влиянию социальных связей на различные социо-экономические показатели. Существенная доля таких исследований опирается на серию статей М. Bailey и соавторов, где вводится показатель SCI (Social Connectedness Index), рассчитанный по базе активных летом 2019 г. пользователей социальной сети Facebook по формуле

$$SCI_{ij} = \frac{FB_Connections_{ij}}{FB_Users_i \times FB_Users_j} \quad (1)$$

Здесь $FB_Connections_{ij}$ означает число дружеских связей в сети Facebook между жителями регионов i и j , а FB_Users_i и FB_Users_j – количества различных пользователей этой сети в регионах i и j соответственно. В Bailey et al. (2020) и других работах этот показатель интерпретируется как вероятность дружбы между жителями регионов. В пользу такого понимания SCI_{ij} там же приводится несколько аргументов, включающих ограничение на максимальное число друзей в сети Facebook, «симметричность» дружеских связей (в отличие, например, от подписок в Twitter), а также результаты исследований, выявивших, что этой социальной сетью пользуются от 40% до 67% европейцев в зависимости от региона, причем зачастую дружбе в Facebook соответствует реальное знакомство.

В работе Bailey et al. (2020) проводится комплексный анализ детерминантов социальной связности. Авторы вычисляют попарные SCI между NUTS2 регионами стран Евросоюза¹ и проводят эконометрический анализ этих данных. Как показано в статье, социальная связность во многом определяется географической удалённостью, а также заметно возрастает, если жители регионов имеют общую религию или язык и схожие демографические показатели. Интересно и то, что SCI оказывается тем выше, чем больше различия между доходами регионов, что авторы связывают с миграционными потоками. Кроме того, большое влияние на величину социальной связности оказывают исторические границы стран и политических союзов, причём как сравнительно недавних (например, Чехословакии и Югославии), так и прекративших своё существование ещё в начале XX века (скажем, Австро-Венгрии). Поскольку значительная доля дисперсии SCI объясняется экзогенными факторами, можно считать эту переменную некоррелированной со случайными ошибками в регрессиях, описывающих различные предпочтения жителей Европы.

¹NUTS2 регионы – территориальные единицы стран Евросоюза с населением от 800 тысяч до 3 миллионов человек, зачастую совпадающие с внутристрановым административно-территориальным делением.

Наконец, в упомянутой статье Bailey et al. (2020) оценивается связь между SCI и политическими предпочтениями. Авторы используют два прокси евроскептицизма: результаты опроса «Евробарометр»², проведённого осенью 2018 г., и доли голосов за антиевропейские партии на парламентских выборах в период с 2000 до 2017 гг. На основе SCI авторы вычисляют доли социальных связей населения региона с жителями других стран Евросоюза – *EU_friends_abroad*, а затем с помощью МНК выясняют, что более высокие доли жителей других стран Евросоюза среди друзей в сети Facebook связаны с большим доверием Евросоюзу и меньшим евроскептицизмом. Однако, при добавлении к регрессии по второму прокси фиксированных эффектов стран оценка теряет статистическую значимость. Кажется, что эта связь всё-таки существует и достаточно сильна, поэтому можно предположить, что более продвинутые методы помогут чётче её выявить.

В более ранней статье Bailey et al. (2018) проводится похожий анализ детерминантов социальной связности, но на этот раз для США. Авторы приходят к выводу о том, что SCI оказывается тесно связан с торговыми потоками между штатами, более подробному анализу чего посвящена работа Bailey et al. (2021). При добавлении SCI в стандартную регрессию гравитационной модели международной торговли коэффициенты при географических расстояниях теряют значимость. Можно считать, что социальные связи не только тесно связаны с торговыми потоками, но также аккумулируют в себе всю информацию о расстоянии между странами. Несмотря на то, что авторы не называют выявленную связь причинно-следственной, в конце работы приводится набор аргументов в пользу этого. Иными словами, SCI_{ij} можно считать пропорциональным вероятности того, что между регионами i и j установятся торговые отношения.

SCI как прокси социальной связности также используется многими другими исследователями. Так, например, результаты, аналогичные анализу торговых потоков, получаются в Diemer and Regan (2020) при анализе цитирования патентов в США. При включении в регрессию одновременно социальной связности и географической удалённости последняя теряет значимость, в то время как SCI остаётся значим на 1%-уровне даже после введения контрольных переменных, отражающих профессиональные связи.

Опыт вышеуказанных исследований наталкивает на мысль о том, что предпочтения в одном регионе могут влиять на предпочтения в других, вызывая тем самым пространственную автокорреляцию. Для анализа таких данных могут при-

²Респондентам был задан вопрос: «I would like to ask you a question about how much trust you have in the European Union. Could you please tell me if you tend to trust it or tend not to trust it?».

меняться модели пространственной эконометрики, которые используют взвешивающую матрицу W . Интуиция заключается в том, что в модели указанная автокорреляция может быть связана как с зависимой переменной y (в некотором смысле аналог AR модели для временных рядов), так и со случайными ошибками (аналог MA модели). Иными словами, можно рассмотреть две спецификации:

$$y = \rho W y + X\beta + \varepsilon \quad (2)$$

$$y = X\beta + \varepsilon, \quad \varepsilon = \lambda W \varepsilon + \zeta \quad (3)$$

Где y – зависимая переменная, W – матрица пространственной связности (предполагается, что $w_{ii} = 0$), а X – матрица регрессоров. Следуя терминологии LeSage and Pace (2009) и Croissant and Millo (2019), будем называть модель (2) SAR (Spatial Autoregressive), а модель (3) – SEM (Spatial Error Model). Подобные модели успешно используются для построения эконометрических моделей с выраженной пространственной структурой. Так, в Elhorst et al. (2019) приводится пример применения таких моделей к анализу олигополизированных отраслевых рынков. Ещё один пример удачного использования методов пространственной эконометрики – исследование Иванова (2019), посвящённое экологической кривой Кузнеца (ЭКК), описывающей взаимосвязь между среднедушевым ВРП и выбросами в окружающую среду. Сравнивая три различных способа формирования матрицы W (на основе обратных географических расстояний, их квадратов, а также на основе соседства), автор подтверждает ключевую гипотезу об U-образном виде ЭКК для российских регионов. Для сравнения матриц весов в статье используется индекс пространственной автокорреляции Морана, рассчитываемый по формуле

$$I(W) = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N \sum_{j=1}^N w_{ij}} \quad (4)$$

Где y_i – исследуемый показатель. Индекс имеет асимптотически нормальное распределение, что позволяет проверять его на значимость.

Говоря непосредственно об оценке предпочтений, можно упомянуть работу Hsieh and van Kippersluis (2019), посвящённую оценке влияния дружеских эффектов на курение. Авторы располагают данными по учащимся нескольких школ и используют модифицированную SAR модель для анализа того, как на курение ученика влияет факт курения его друзей. Нововведением этой работы является

ся то, что авторы делят студентов на эмоционально устойчивых и неустойчивых (в зависимости от результатов школьного психологического теста) и предполагают, что дружеский эффект может различаться в зависимости от эмоциональной устойчивости как самого учащегося, так и его друзей. Таким образом, оцениваются сразу четыре параметра ρ , управляющие влиянием различных блоков матрицы W . Дополнительно авторы статьи устраняют очевидную эндогенность матрицы пространственной связности, используя логит-регрессию вероятности дружбы на подобранный набор инструментов. Результаты показывают, что эмоционально неустойчивые ученики действительно в существенно большей мере подвержены влиянию друзей в отношении курения, чем их более устойчивые сверстники.

Следуя вышеописанной мотивации, можно предположить, что модели пространственной эконометрики позволят лучше описать формирование предпочтений жителей европейских NUTS2 регионов, чем стандартные МНК регрессии. Разумно выдвинуть гипотезу о том, что SAR модель окажется более подходящим выбором, чем SEM, так как отражает непосредственное влияние наблюдений друг на друга. Наконец, в части политических предпочтений кажется, что введенный в Bailey et al. (2020) регрессор *EU_friends_abroad* не должен терять значимость в новой постановке, поскольку дружеские связи внутри европейских стран должны способствовать большему доверию Евросоюзу и меньшему евроскептицизму. Озвученные вопросы приводят к окончательному списку проверяемых гипотез:

1. модели пространственной эконометрики лучше подходят для описания предпочтений жителей Евросоюза, чем стандартный МНК;
2. параметр пространственного лага ρ в SAR постановке помогает объяснить дисперсию предпочтений и является статистически значимым;
3. предпочтения жителей Европы с точки зрения их пространственной структуры лучше описываются социальными связями, чем географическими расстояниями;
4. в части политических предпочтений большее число социальных связей внутри Еврозоны способствует большему доверию Евросоюзу и меньшему евроскептицизму.

3 Основная часть

3.1 Данные

Для проверки гипотез в первую очередь используются два набора пространственных данных, собранные Bailey et al. (2020) по NUTS2 регионам Европы, **trust_in_EU** и **anti_EU_votes**. В этих наборах данных есть значительное число контрольных переменных, которые используются для дальнейшего анализа. Помимо этих данных мы пользуемся агрегированными по NUTS2 регионам данными седьмой волны опроса World Values Survey (2017 – 2020) о предпочтениях и взглядах европейцев, **world_values_survey**. Для регрессий, где в качестве зависимой переменной выступает показатель из этого набора данных, в уравнение добавляются дополнительные контрольные переменные, характеризующие структуру принадлежности населения региона к различным объединениям. Полный список переменных представлен Таблице 1 ниже.

Таблица 1: Описание переменных

Переменная	Описание
Зависимые переменные	
<i>Trust_in_EU</i>	Доля опрошенных в регионе, ответивших «Tend to trust» на вопрос о доверии Евросоюзу
<i>Anti_EU_vote</i>	Доля голосов за антиевропеские партии в регионе
<i>Conf_in_civil_services</i>	Агрегированная переменная доверия государственным институтам
<i>Homo_neighbours</i>	Агрегированная переменная отношения к соседству с лицами нетрадиционной сексуальной ориентации
<i>Religion</i>	Агрегированная переменная важности религии в жизни
<i>Trust</i>	Агрегированная переменная доверия людям
Переменные интереса	
<i>EU_friends_abroad</i>	Доля социальных связей жителей региона с европейцами из других стран
$W \times y$	Пространственный лаг зависимой переменной

Переменная	Описание
Контрольные переменные	
<i>ED0_2</i>	Доля в населении региона людей с образованием ниже среднего
<i>ED3_4</i>	Доля в населении региона людей со средним образованием
<i>ED5_8</i>	Доля в населении региона людей с высшим образованием
<i>median_age</i>	Медианный возраст населения региона
<i>Y0_9</i>	Доля населения региона в возрасте 0-9 лет
<i>Y10_19</i>	Доля населения региона в возрасте 10-19 лет
<i>Y20_29</i>	Доля населения региона в возрасте 20-29 лет
<i>Y30_39</i>	Доля населения региона в возрасте 30-39 лет
<i>Y40_49</i>	Доля населения региона в возрасте 40-49 лет
<i>Y50_59</i>	Доля населения региона в возрасте 50-59 лет
<i>Y60_69</i>	Доля населения региона в возрасте 60-69 лет
<i>Y70_MAX</i>	Доля населения региона в возрасте старше 70 лет
<i>Health</i>	Субъективная оценка респондента для состояния его здоровья
<i>share_other_EU</i>	Доля жителей региона, родившихся в другой европейской стране
<i>unemp_rate</i>	Уровень безработицы в регионе в % к 2018 году
<i>B</i>	Доля жителей региона, занятых в добыче полезных ископаемых
<i>C</i>	Доля жителей региона, занятых в производстве
<i>D</i>	Доля жителей региона, занятых в электро-, газо- и пароснабжении
<i>E</i>	Доля жителей региона, занятых в водоснабжении, канализации, рекультивации
<i>F</i>	Доля жителей региона, занятых в строительстве
<i>G</i>	Доля жителей региона, занятых в торговле и ремонте транспорта
<i>I</i>	Доля жителей региона, занятых в грузоперевозках и хранении

Переменная	Описание
Контрольные переменные	
<i>J</i>	Доля жителей региона, занятых в размещении и сфере продовольственных услуг
<i>L</i>	Доля жителей региона, занятых в сфере информационных технологий
<i>M</i>	Доля жителей региона, занятых в сфере недвижимости
<i>N</i>	Доля жителей региона, занятых в научно-технической деятельности
<i>income_thous</i>	Средний доход жителей региона, тыс. евро
<i>Member_control_1</i>	Агрегированная переменная принадлежности к религиозным организациям
<i>Member_control_2</i>	Агрегированная переменная принадлежности к организациям, связанным с образованием или искусством
<i>Member_control_3</i>	Агрегированная переменная принадлежности к профсоюзным организациям
<i>Member_control_4</i>	Агрегированная переменная принадлежности к политическим партиям
<i>Member_control_5</i>	Агрегированная переменная принадлежности к экологическим организациям
<i>Member_control_6</i>	Агрегированная переменная принадлежности к профессиональным ассоциациям
<i>Member_control_7</i>	Агрегированная переменная принадлежности к организациям, связанным со спортом и развлечениями
<i>Member_control_8</i>	Агрегированная переменная принадлежности к потребительским кооперативам
<i>Member_control_9</i>	Агрегированная переменная принадлежности к благотворительным организациям
<i>Member_control_11</i>	Агрегированная переменная принадлежности к группам взаимопомощи

Поскольку для проверки гипотез применяются SAR и SEM модели, встаёт вопрос о создании пространственной матрицы W . Для каждого из используемых наборов данных мы конструируем три матрицы пространственной связности. $W^{(d)}$ и $W^{(d^2)}$ – матрицы на основе обратных величин географических расстояний и их квадратов, соответственно, а $W^{(SCI)}$ – матрица на основе SCI. На диагонали каждой из матриц – нули, производится нормировка³. Использование географических расстояний между центроидами регионов является одним из стандартных подходов к конструированию пространственной матрицы. У этого подхода есть существенное преимущество: пространственная матрица очевидным образом экзогенна. Однако, в статье Bailey et al. (2020) показано, что в некоторых случаях SCI лучше объясняет пространственные связи в данных, чем меры, основанные на расстояниях. В то же время, как уже отмечалось ранее, есть содержательные основания считать $W^{(SCI)}$ экзогенной. Различия между географическими расстояниями и SCI хорошо видны на графе социальной связности европейских стран (Рис. 1)⁴.

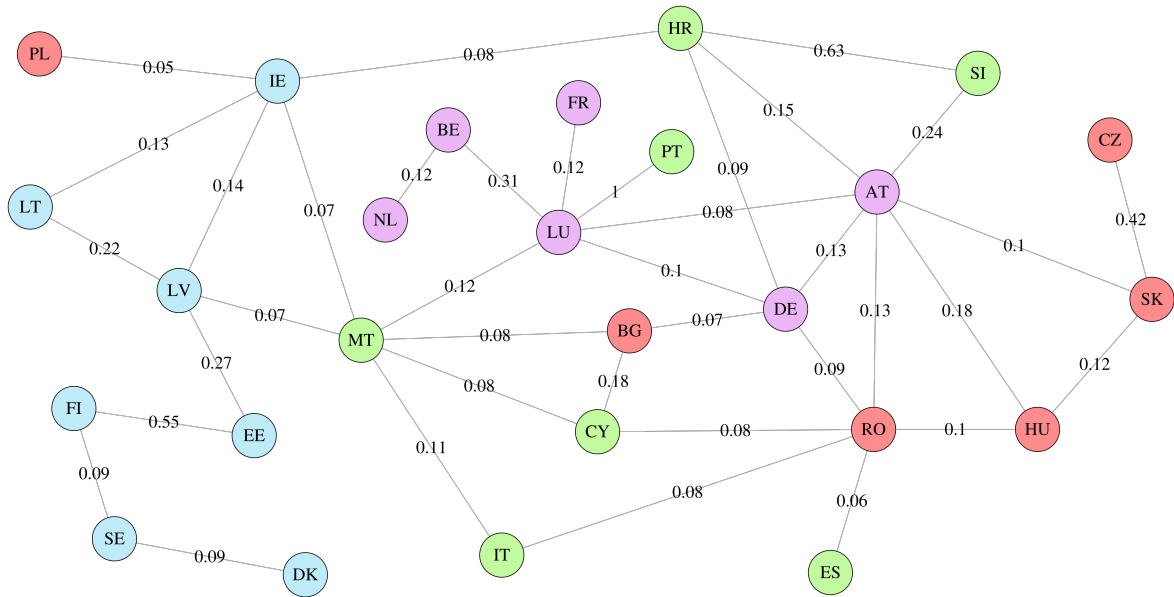


Рис. 1: Граф социальной связности европейских стран

³Нормировка производится так, чтобы сумма элементов каждой из строк матрицы была равна единице. Такой подход является общепринятым в литературе по теме, поскольку позволяет избежать проблем при оценивании модели и интерпретировать коэффициент при пространственном лаге.

⁴Наиболее слабые связи опущены для наглядности.

Цветом на графе выделены группы стран по географическому принципу: голубым – страны Северной Европы, лиловым – страны Западной Европы, зелёным – страны Южной Европы, красным – страны Восточной Европы. Веса дуг графа – SCI между странами, нормированный на величину SCI между Люксембургом и Португалией. Граф иллюстрирует важные особенности социальных связей в Европе. Во-первых, прослеживается явная корреляция между SCI и географической близостью: легко видеть, что, например, в странах Северной Европы внутригрупповые связи довольно сильны, в то время как внешние связи немногочисленны и относительно слабы. Во-вторых, очевидно, что структура SCI не повторяет географическую: наблюдается ряд сильных межгрупповых связей. Таким образом, можно сделать вывод о целесообразности использования $W^{(SCI)}$ для анализа предпочтений.

Важным элементом нашей эмпирической стратегии является добавление в модель фиксированных эффектов, основанное на интуиции Croissant and Millo (2019). У используемых наборов данных есть особенность, делающая затруднительным применение традиционных фиксированных эффектов: для ряда стран есть лишь одно или два наблюдения в силу их небольшой площади⁵. Тем самым, возникает ситуация, когда в уравнение добавляются переменные, объясняющие единственное наблюдение, что едва ли положительно сказывается на качестве модели. Чтобы избежать этой проблемы и в то же время ввести фиксированные эффекты, мы используем полученную Bailey et al. (2020) кластеризацию европейских регионов (Рис. 2)⁶. Число 19 выбрано по нескольким причинам. Во-первых, такое количество кластеров обеспечивает относительно равномерное распределение по ним регионов в используемых наборах данных. Во-вторых, при добавлении большего числа фиксированных эффектов оценки при переменных интереса в основных моделях меняются незначительно. В-третьих, такая степень дискретизации представляется разумной с точки зрения количества порождаемых фиктивных переменных. На карте кластеров также хорошо видны особенности социальных связей. Контуры кластеров часто совпадают с политическими границами стран, однако существуют кластеры, части которых не имеют общей сухопутной границы (например, кластер Португалии и Швейцарии). Мы полагаем, что в силу этих особенностей добавление в модель для объяснения предпочтений фиксированных эффектов кластеров регионов вместо фиксированных эффектов стран вполне целесообразно.

⁵Также для некоторых стран наблюдения имеются не по NUTS2 регионам, а по более крупным территориальным единицам – NUTS1.

⁶Авторы статьи применяют алгоритм hierarchical agglomerative linkage clustering.

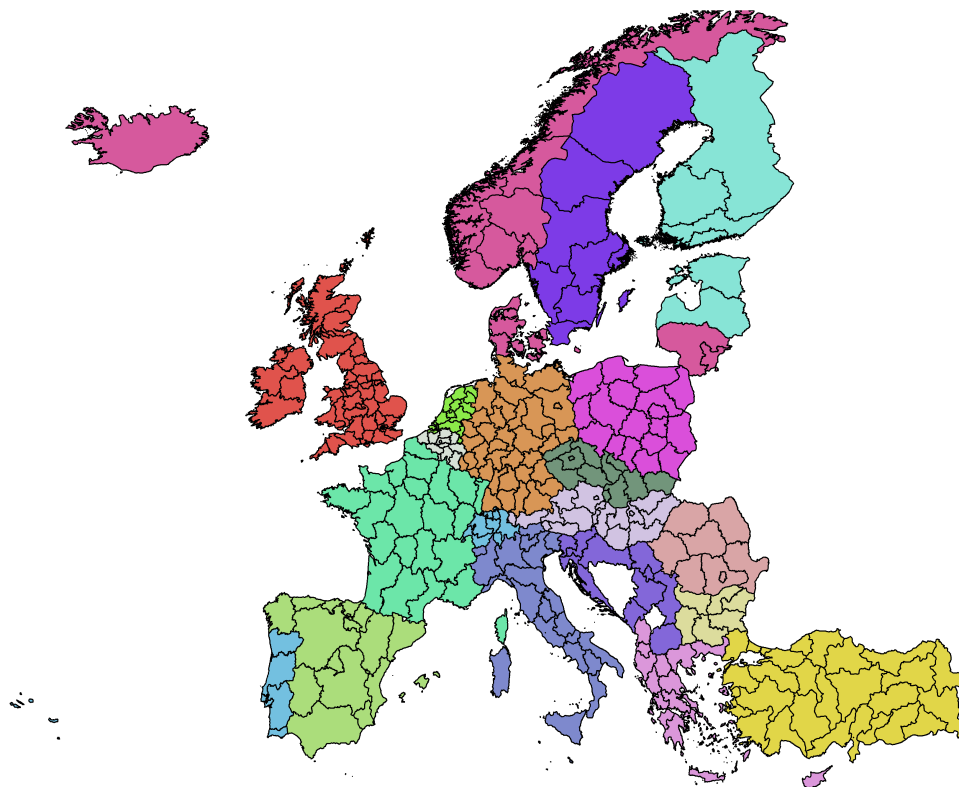


Рис. 2: 19 кластеров европейских регионов на основе SCI

Продолжая описание данных, уместно привести таблицу с индексами пространственной корреляции Морана для основных зависимых переменных⁷.

Таблица 2: Индексы пространственной корреляции

Переменная	Набор данных	N	$W^{(d)}$	$W^{(d^2)}$	$W^{(SCI)}$
<i>Trust_in_EU</i>	trust_in_EU	198	0,096	0,292	0,482
<i>Anti_EU_vote</i>	anti_EU_votes	215	0,062	0,201	0,459
<i>Conf_in_civil_services</i>	world_values_survey	190	0,117	0,321	0,443
<i>Homo_neighbours</i>	world_values_survey	190	0,259	0,549	0,553
<i>Religion</i>	world_values_survey	190	0,216	0,505	0,623
<i>Trust</i>	world_values_survey	190	0,217	0,510	0,643

⁷ N – число наблюдений в наборе данных.

Все индексы являются статистически значимыми на 1% уровне. Это позволяет говорить о пространственной коррелированности изучаемых предпочтений и свидетельствует в пользу целесообразности применения моделей пространственной эконометрики. Отметим также, что индекс Морана для $W^{(SCI)}$, как правило, оказывается больше индексов для матриц на основе расстояний. Такой результат можно интерпретировать как аргумент в пользу верности гипотезы 3 о том, что SCI лучше отражает пространственную структуру предпочтений, чем географические расстояния.

Поскольку данные имеют выраженную пространственную структуру, логично визуализировать их при помощи карты. Так, на рисунке 3 видны кластеры регионов со схожим уровнем доверия ЕС. Особенно хорошо прослеживается контраст между Великобританией и странами Северной и Центральной Европы.

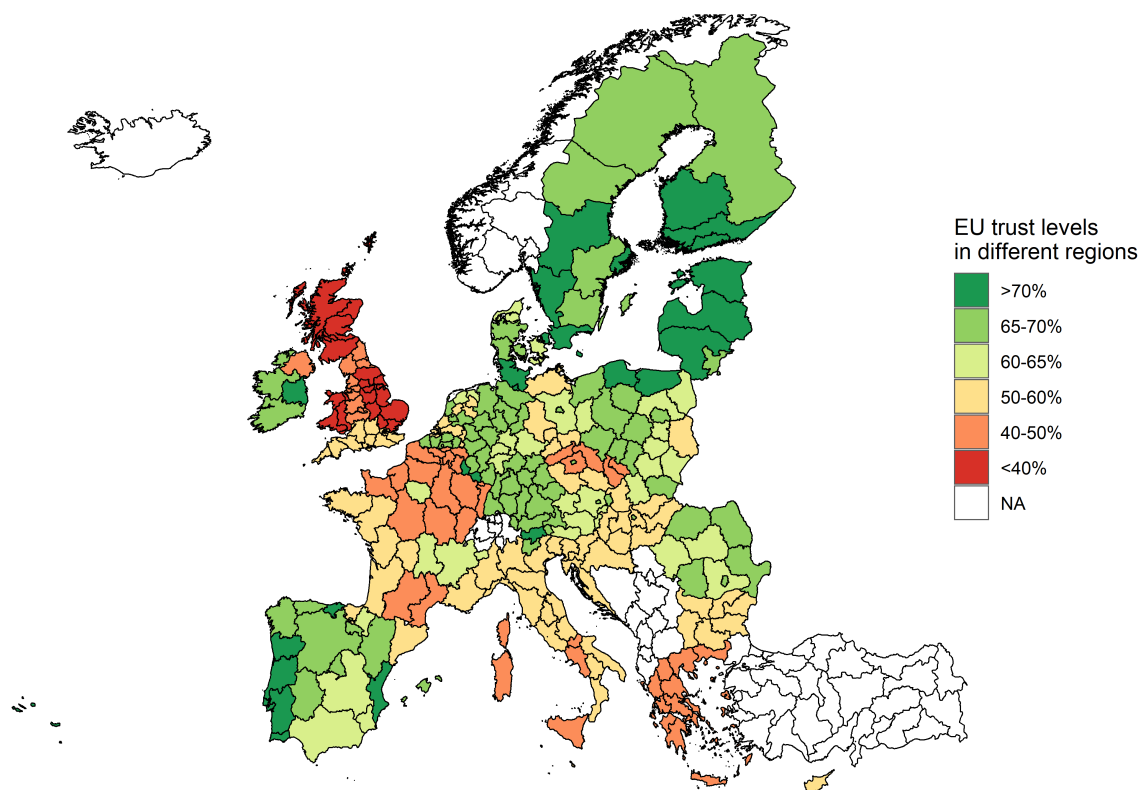


Рис. 3: Уровни доверия ЕС в регионах

Если обратиться к другому прокси евроскептицизма, доле голосов за партии евроскептиков, можно увидеть похожую картину (Рис. 4).

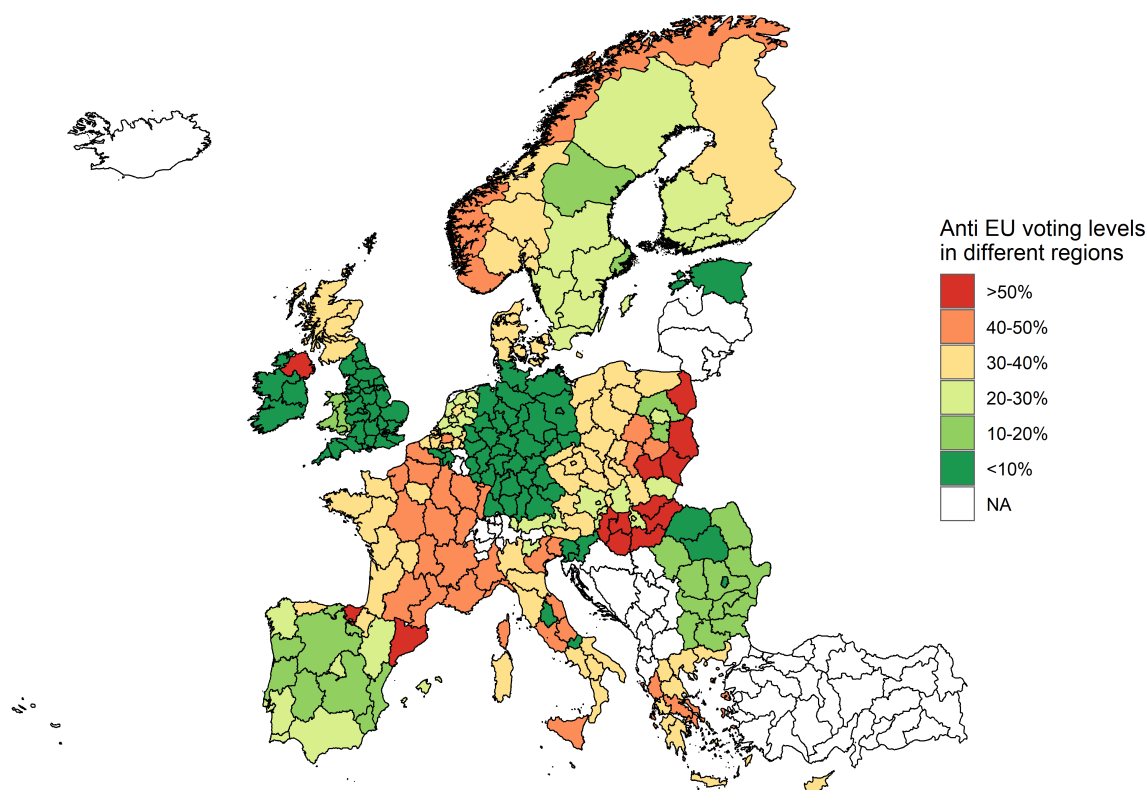


Рис. 4: Доли голосов за партии евроскептиков в регионах

Заметно, что существует отрицательная корреляция между двумя показателями. В то же время, показатели характеризуют качественно различные выражения евроскептицизма: если уровни доверия ЕС оценены на основе агрегированных опросных данных, то доли голосов за партии евроскептиков измерены непосредственно и являются хорошей оценкой процента активных противников ЕС. Тем самым, совместный анализ этих показателей крайне интересен, поскольку в случае получения сходных результатов можно говорить об их устойчивости и делать более уверенные выводы о степени влияния различных факторов на евроскептицизм.

Оценивание пространственных моделей проводилось методом квазимоксимального правдоподобия (QML). Работа с данными, оценка моделей и построение графиков выполнены в пакете R, R Core Team (2020), с помощью библиотеки `splm`, описанной в Millo and Piras (2012).

3.2 Результаты оценивания

Все регрессионные модели рассматривались в двух постановках: короткой и длинной. Оцениваемые уравнения имеют одинаковую структуру, однако короткая версия отличается от длинной отсутствием некоторых контрольных переменных, которые в большинстве случаев оказывались незначимы⁸. Стремление сократить количество регрессоров обусловлено довольно малым числом наблюдений в расчёте на оцениваемый параметр. Здесь и далее приводятся результаты оценивания для длинных постановок, поскольку выясняется, что они устойчивы относительно удаления части наблюдений из выборки (см. раздел об устойчивости). Это смягчает опасения насчёт излишней параметризованности.

Рекомендованная в Croissant and Millo (2019) и применённая в Иванова (2019) схема сравнения моделей заключается в следующем: на первом этапе тестирования проводятся два теста множителей Лагранжа для объединённой модели. Первый из них проверяет гипотезу $\lambda = 0$ при условии $\rho = 0$ (еггог-тест), а второй – в каком-то смысле симметричную гипотезу $\rho = 0$ при условии $\lambda = 0$ (lag-тест). Если ровно один из вышеуказанных тестов принимает нулевую гипотезу, то соответствующий параметр полагается равным нулю. Если оба теста отвергают нулевую гипотезу, проводятся два робастных теста с гипотезами $\rho = 0$ при условии $\lambda \neq 0$ и $\lambda = 0$ при условии $\rho \neq 0$ и решение о выборе модели принимается на их основе (в случае повторного отвержения обеих нулевых гипотез отбор производится по наименьшему р-значению).

Что касается меры качества подгонки, то для сравнения моделей используются критерий Акаике AIC и псевдо- R^2 Nagelkerke, вычисляемый по формуле

$$R^2 = \frac{1 - \left(\frac{L_0}{L}\right)^{2/N}}{1 - L_0^{2/N}} \quad (5)$$

Здесь L и L_0 – значения функции правдоподобия при оценке длинной модели и модели регрессии на константу соответственно, а N – размер выборки. Этот показатель совпадает со стандартным R^2 для МНК оценивания, так что его использование в целях сравнения вполне оправданно.

⁸Длинная регрессия включает все контрольные переменные, кроме *share_other_EU*, а также регрессоров, вызывающих чистую мультиколлинеарность. Короткая регрессия из образовательного контроля включает только *ED5_8*, из отраслевого контроля – только *E*, *G* и *L*. Также в короткой регрессии отсутствует *unemp_rate*.

Для набора данных **trust_in_EU** LM тесты (Таблица 3) указывают на то, что для матриц $W^{(SCI)}$ и $W^{(d^2)}$ SAR модель оказывается предпочтительной. Что касается $W^{(d)}$, то пространственные модели с ней проигрывают обычному МНК.

Таблица 3: Результаты LM тестов для *Trust_in_EU*

	$W^{(SCI)}$	$W^{(d)}$	$W^{(d^2)}$
	$p - value$	$p - value$	$p - value$
error	0,259	0,484	0,831
lag	0,001	0,416	0,056
error (robust)	0,005	0,017	0,069
lag (robust)	0,000	0,016	0,009

Для **anti_EU_votes** тесты (Таблица 4) указывают на предпочтительность пространственной модели только для матрицы $W^{(SCI)}$, причём в SAR версии. Остальные регрессии снова оказываются хуже стандартного МНК и поэтому не приводятся в Таблице 5, посвящённой сравнению моделей.

Таблица 4: Результаты LM тестов для *Anti_EU_vote*

	$W^{(SCI)}$	$W^{(d)}$	$W^{(d^2)}$
	$p - value$	$p - value$	$p - value$
error	0,386	0,787	0,197
lag	0,025	0,961	0,337
error (robust)	0,010	0,665	0,387
lag (robust)	0,001	0,732	0,940

Тот факт, что пространственная модель оказалась лучше простого МНК хотя бы для одной из рассматриваемых матриц, подтверждает первую из выдвинутых гипотез для обоих наборов данных **trust_in_EU** и **anti_EU_votes**: SAR модель действительно может лучше описывать политические предпочтения европейцев. Следует упомянуть, что, хотя при построении регрессий и использовались робастные стандартные ошибки, тест Бреуша – Пагана на гетероскедастичность принял нулевую гипотезу по обоим наборам данных. Это позволяет не сомневаться в эффективности полученных оценок и их значимости. Подробнее лучшие модели, касающиеся политических предпочтений, представлены в Таблице 5.

Таблица 5: Анализ евроскептицизма

	<i>Trust_in_EU</i>		<i>Anti_EU_vote</i>
	$W^{(d^2)}$	$W^{(SCI)}$	$W^{(SCI)}$
<i>EU_friends_abroad</i>	0,393*** (0,089)	0,380*** (0,086)	-0,098 (0,202)
ρ	0,254* (0,138)	0,469*** (0,106)	0,363*** (0,122)
Cluster Fixed Effects	Y	Y	Y
Socioeconomic controls	Y	Y	Y
Membership controls	-	-	-
Number of observations	198	198	215
R^2	0,798	0,807	0,700
AIC	-592,1	-601,5	-339,1

Note: * $p < 0, 1$; ** $p < 0, 05$; *** $p < 0, 01$

Пространственный лаг ρ оказывается значим в каждой из отобранных постановок. Это подтверждает вторую из выдвинутых гипотез. Наконец, поскольку модели с матрицами, построенными на основе географических расстояний, проигрывают даже обычному МНК в наборе **anti_EU_votes**, а для данных **trust_in_EU** имеют меньший R^2 и больший критерий Акаике AIC, можно считать подтверждённой и третью гипотезу. Социальные связи на самом деле лучше объясняют политические предпочтения, чем географическая удалённость.

Результаты анализа соответствуют выводам статьи Bailey et al. (2020): для набора данных **trust_in_EU** регрессор *EU_friends_abroad* оказывается статистически значим. Более того, численное значение коэффициента перед ним, оценённое в Bailey et al. (2020), равнялось 0,326, в то время как в SAR модели этот коэффициент оценён как 0,380. Аналогичной близости в оценках коэффициента при *EU_friends_abroad* для набора данных **anti_EU_votes** нет, однако это неудивительно ввиду его незначимости как в SAR постановке, так и в оригинальной работе Bailey et al. (2020). Здесь следует оговориться о том, что интерпретация коэффициентов в пространственных моделях отличается от стандартной интерпретации МНК, поэтому различие (пусть и небольшое) вполне естественно и не свидетельствует о пропуске существенных переменных или иной форме эндогенности.

В силу спецификации SAR модели коэффициенты при переменных в правой части нельзя интерпретировать так же, как в стандартной линейной регрессии. Результат изменения любого регрессора будет разлагаться на два эффекта – прямой и косвенный. Прямой эффект заключается в непосредственном влиянии изменения регрессора в регионе i на зависимую переменную в том же регионе, в то время как косвенный – во влиянии этого изменения на другой регион j . Проиллюстрировать это можно выкладками из LeSage and Pace (2009):

$$y = \rho W y + X\beta + \varepsilon \quad \Rightarrow \quad (I_n - \rho W)y = X\beta + \varepsilon$$

$$\frac{\partial y_i}{\partial x_{jr}} = [(I_n - \rho W)^{-1} I_n \beta_r]_{ij} \quad (6)$$

Здесь x_{jr} – наблюдение r -ого регрессора в регионе j . Из формулы видно, что прямые и косвенные эффекты правда отличаются, то есть $\frac{\partial y_i}{\partial x_{jr}}$ для различных j , вообще говоря, не совпадают. Для количественной интерпретации, следуя примеру Иванова (2019), используем усреднённое значение этих эффектов. В частности, для переменной *EU_friends_abroad* в наборе данных **trust_in_EU** прямой эффект составляет 0,390, косвенный – 0,325, а усреднённый – 0,357. Таким образом, увеличение доли европейских друзей из других стран на 1% будет сопряжено с ростом доверия Евросоюзу на 0,39% в данном регионе и на 0,33% в других регионах.

Все сказанное о длинной и короткой спецификациях остаётся верным и для регрессий по наборам данных, взятых из опроса World Values Survey⁹. Почти все данные здесь гомоскедастичны, о чём вновь говорит тест Бреуша – Пагана¹⁰. Спецификационные тесты для переменной *Religion* представлены ниже в Таблице 6.

Таблица 6: Результаты LM тестов для *Religion*

	$W^{(SCI)}$	$W^{(d)}$	$W^{(d^2)}$
	$p - value$	$p - value$	$p - value$
error	0,784	0,377	0,834
lag	0,009	0,213	0,073
error (robust)	0,000	0,027	0,096
lag (robust)	0,000	0,017	0,015

⁹За исключением того, что длинные регрессии включали дополнительно все контрольные переменные, связанные с участием в организациях, а короткие – только *Member_control_2*, *Member_control_4* и *Member_control_7*.

¹⁰Кроме переменной *Trust*, однако робастные стандартные ошибки должны это нивелировать.

Указанные результаты аналогичны результатам тестов для набора данных **trust_in_EU** и свидетельствуют о том, что наиболее предпочтительны SAR модели с матрицами $W^{(SCI)}$ и $W^{(d^2)}$. Для переменной *Conf_in_civil_services* на удивление лучшие в сравнении с МНК оценки позволяет получить только SAR модель с матрицей $W^{(d^2)}$ (см. р-значения тестов в Таблице 7).

Таблица 7: Результаты LM тестов для *Conf_in_civil_services*

	$W^{(SCI)}$	$W^{(d)}$	$W^{(d^2)}$
	$p - value$	$p - value$	$p - value$
error	0,868	0,363	0,752
lag	0,439	0,425	0,038
error (robust)	0,014	0,001	0,000
lag (robust)	0,010	0,001	0,000

Итоговое сравнение лучших моделей для переменных *Religion* и *Conf_in_civil_services* представлено в Таблице 8.

Таблица 8: Анализ отношения к религии и доверия государственным институтам

	<i>Religion</i>		<i>Conf_in_civil_services</i>
	$W^{(d^2)}$	$W^{(SCI)}$	$W^{(d^2)}$
<i>EU_friends_abroad</i>	-0,103 (0,140)	-0,013 (0,138)	0,340** (0,135)
ρ	0,254** (0,109)	0,416*** (0,109)	0,379*** (0,139)
Cluster Fixed Effects	Y	Y	Y
Socioeconomic controls	Y	Y	Y
Membership controls	Y	Y	Y
Number of observations	190	190	190
R^2	0,876	0,879	0,660
AIC	-531,1	-535,8	-543,7

Note: * $p < 0, 1$; ** $p < 0, 05$; *** $p < 0, 01$

Говоря о результатах исследования *Religion*, можно отметить, что вновь SAR модели оказываются лучше стандартного МНК, причём коэффициент ρ в них значим на уровне 1%. При этом R^2 в модели с матрицей $W^{(SCI)}$ выше. Итого по данной переменной подтверждаются все три основные гипотезы. Того же нельзя сказать о *Conf_in_civil_services*: здесь последнюю гипотезу принять не получается, поскольку лучшей из пространственных моделей является SAR модель с матрицей $W^{(d^2)}$. Выходит, дисперсию отношения к государственным институтам география объясняет лучше, чем социальная связность.

Интересные результаты получаются при анализе переменной *Trust*. Спецификационные LM тесты для всех трёх матриц указывают на превосходство SAR модели и потому не приводятся. При этом коэффициенты ρ оказываются значимыми на 5%-уровне и, что занятно, отрицательными (см. оценки моделей в Таблице 9). Интерпретация этого может быть следующей: менее доверчивые и более мнительные люди склонны поддерживать социальные связи только с более доверчивыми и наивными людьми. Это вполне естественно, ведь кажется, что сообщество, целиком состоящее из не доверяющих друг другу людей, не может быть устойчивым. Также интересно, что переменная *EU_friends_abroad* оказывает значимый положительный эффект на доверие. Получается, у более доверяющих людей больше доля друзей среди жителей других европейских регионов.

Таблица 9: Анализ уровня доверия людям

	<i>Trust</i>		
	$W^{(d)}$	$W^{(d^2)}$	$W^{(SCI)}$
<i>EU_friends_abroad</i>	0,319*** (0,111)	0,295*** (0,110)	0,375*** (0,105)
ρ	-1,005** (0,399)	-0,308** (0,147)	-0,687*** (0,138)
Cluster Fixed Effects	Y	Y	Y
Socioeconomic controls	Y	Y	Y
Membership controls	Y	Y	Y
Number of observations	190	190	190
R^2	0,886	0,885	0,895
AIC	-625,2	-623,2	-639,4

Note: * $p < 0, 1$; ** $p < 0, 05$; *** $p < 0, 01$

Что касается *Homo_neighbours*, здесь снова побеждают пространственные модели. Однако тут впервые тесты указывают на SEM модель, причём сразу для двух матриц – $W^{(SCI)}$ и $W^{(d^2)}$ (Таблица 10). В качестве интерпретации такого выбора в Croissant and Millo (2019) предлагается пропуск фактора, объясняющего пространственную автокорреляцию зависимой переменной. Подробнее оценки моделей приведены в Таблице 11.

Таблица 10: Результаты LM тестов для *Homo_neighbours*

	$W^{(SCI)}$	$W^{(d)}$	$W^{(d^2)}$
	<i>p – value</i>	<i>p – value</i>	<i>p – value</i>
error	0,014	0,097	0,070
lag	0,826	0,044	0,086
error (robust)	0,000	0,000	0,000
lag (robust)	0,003	0,000	0,000

Таблица 11: Анализ отношения к соседству с лицами нетрадиционной ориентации

	<i>Homo_neighbours</i>		
	$W^{(d)}$	$W^{(d^2)}$	$W^{(SCI)}$
<i>EU_friends_abroad</i>	0,478** (0,237)	0,549** (0,230)	0,483** (0,225)
ρ	0,658*** (0,200)	-	-
λ	-	-0,796*** (0,207)	-0,964*** (0,215)
Cluster Fixed Effects	Y	Y	Y
Socioeconomic controls	Y	Y	Y
Membership controls	Y	Y	Y
Number of observations	190	190	190
R^2	0,861	0,864	0,869
AIC	-330,1	-334,2	-341,2

Note: * $p < 0, 1$; ** $p < 0, 05$; *** $p < 0, 01$

3.3 Устойчивость результатов

В качестве проверки устойчивости результатов были проведены две серии дополнительных тестов.

Первая серия тестов была посвящена устойчивости к удалению части выборки. Для каждой постановки из полного набора данных были удалены 10 случайных наблюдений, после чего на оставшейся подвыборке оценивалась лучшая пространственная модель. По каждому набору данных проводилось по 1000 итераций, а затем составлялась эмпирическая плотность распределения $\hat{\rho}$ или $\hat{\lambda}$ (в зависимости от модели). Примеры полученных плотностей для **trust_in_EU** и **anti_EU_votes** представлены на Рисунках 5 и 6 (жирной чертой отмечена оценка параметра ρ по полной выборке, а пунктирами – 95%-доверительные интервалы). Аналогичные графики для других переменных находятся в Приложении А.

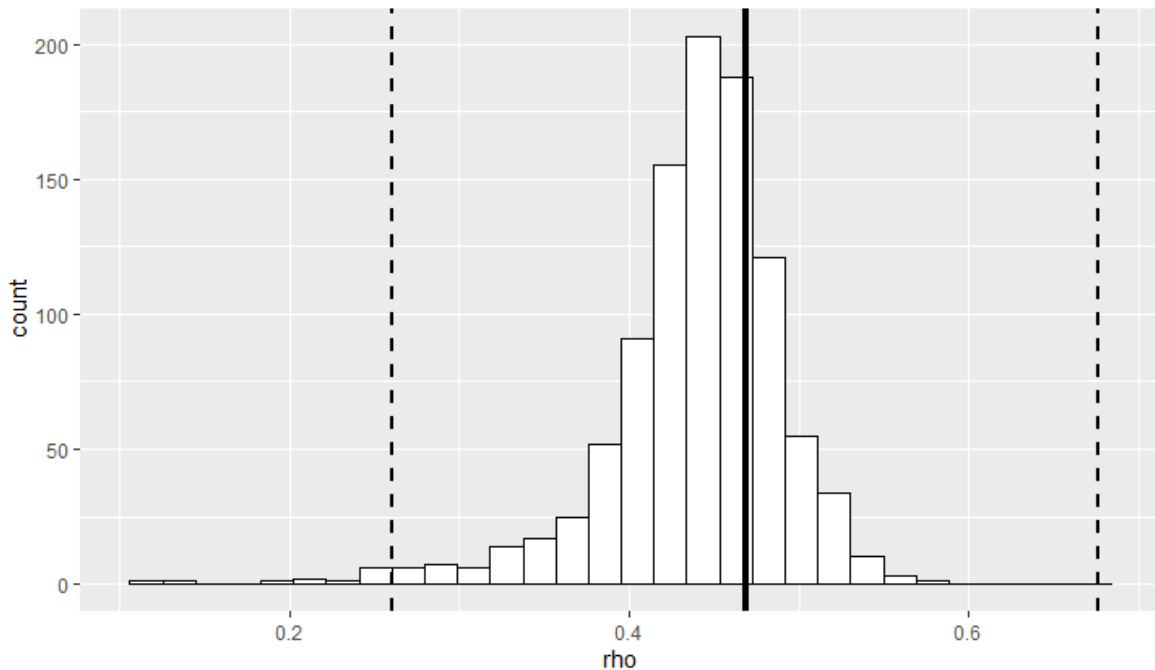


Рис. 5: Эмпирическая плотность распределения $\hat{\rho}$ в тестах на устойчивость по переменной *Trust_in_EU*

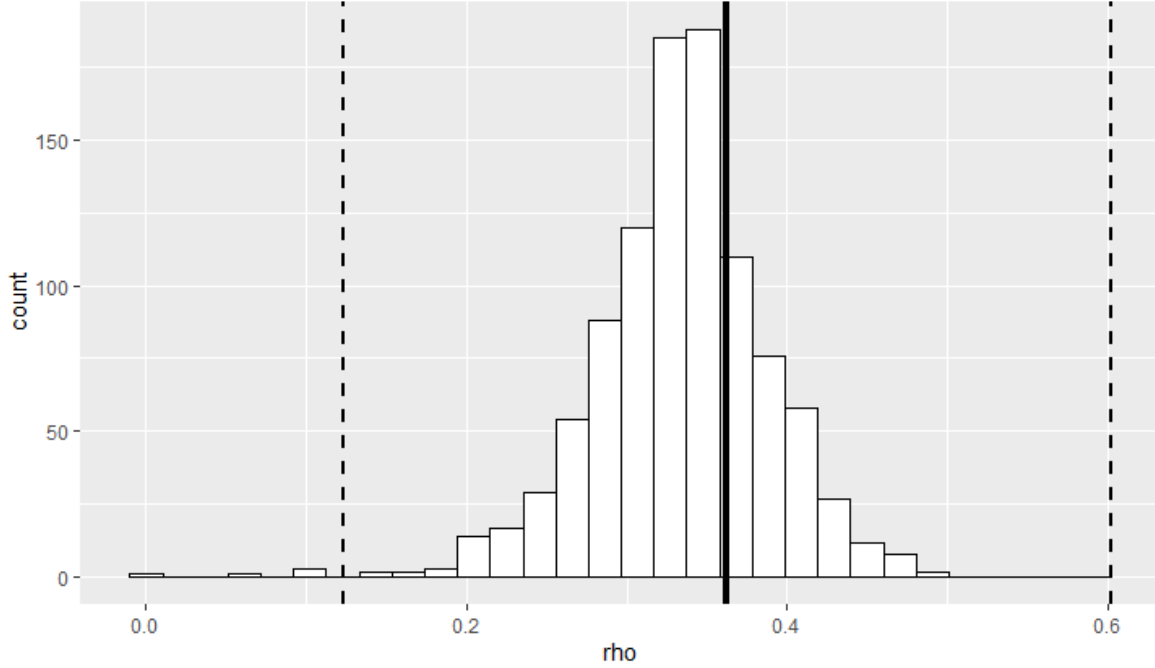


Рис. 6: Эмпирическая плотность распределения $\hat{\rho}$ в тестах на устойчивость по переменной *Anti_EU_vote*

Затем для всех моделей был проведён плацебо-тест, который заключался в проверке значимости коэффициента ρ для SAR модели (и, соответственно, коэффициента λ для SEM модели) при использовании случайной матрицы W . Случайность понималась в смысле равномерности распределения её элементов до нормировки. Иными словами, была построена модель, где $w_{ii} = 0$ и при $i \neq j$

$$w_{ij} = \frac{\xi_{ij}}{\sum_{j \neq i} \xi_{ij}}, \quad \xi_{ij} \sim U[0; 1] \quad (7)$$

Для всех приведённых выше регрессий параметры ρ и λ со случайной матрицей W незначимы и построенные на их основе модели проигрывают МНК.

3.4 Ссылки

Bailey et al. (2020) следуют принципам воспроизводимых научных исследований и публикуют свои расчёты в *github-репозитории*, что позволяет проверять и дополнять их результаты. Такой подход представляется нам правильным, поэтому мы также размещаем код, написанный нами для этого исследования, в репозитории. С нашими расчётами можно ознакомиться по *ссылке*.

4 Заключение

Подводя итог проведённому исследованию, следует отметить, что все предложенные гипотезы оказались в той или иной мере приняты. Во-первых, модели пространственной эконометрики, как показали LM тесты, для всех рассматриваемых предпочтений оказались лучше стандартного МНК хотя бы для одной матрицы весов из трёх предложенных. Получается, исследуемые зависимые переменные действительно характеризуются заметной пространственной автокорреляцией, игнорирование которой приводит к неэффективности оценок.

Во-вторых, во всех построенных регрессиях, за исключением постановок с матрицами $W^{(SCI)}$ и $W^{(d^2)}$ для переменной *Homo_neighbours*, параметр пространственного лага ρ оказался статистически значим. Здесь можно сделать вывод о том, что SAR модель заметно лучше подходит для описания предпочтений жителей европейских стран, чем альтернативная SEM модель.

В-третьих, как показывают рассчитанные для пространственных моделей коэффициенты R^2 и критерии Акаике AIC, матрица $W^{(SCI)}$, отражающая «близость» регионов в смысле социальной связности, в большинстве случаев лучше подходит для описания рассмотренных предпочтений, чем весовые матрицы, построенные на основе географических расстояний.

Наконец, результаты регрессий переменных, связанных с доверием Евросоюзу и евроскептицизмом, хорошо согласуются с результатами, полученными в Bailey et al. (2020). Более того, оценки коэффициентов при переменной *EU_friends_abroad* в тех уравнениях, где этот регрессор значим, довольно близки по численному значению. Это наводит на мысль о том, что доли друзей, проживающих в других европейских странах, схожим образом влияют на предпочтения людей и поэтому являются важным фактором при их анализе.

Настоящая работа использует как сравнительно новые для предметной области методы – модели пространственной эконометрики – так и свежие данные по SCI. Вместе, такое сочетание позволяет добиться интересных с академической точки зрения результатов и безусловно имеет большой потенциал для дальнейшего развития.

А Приложение: графики

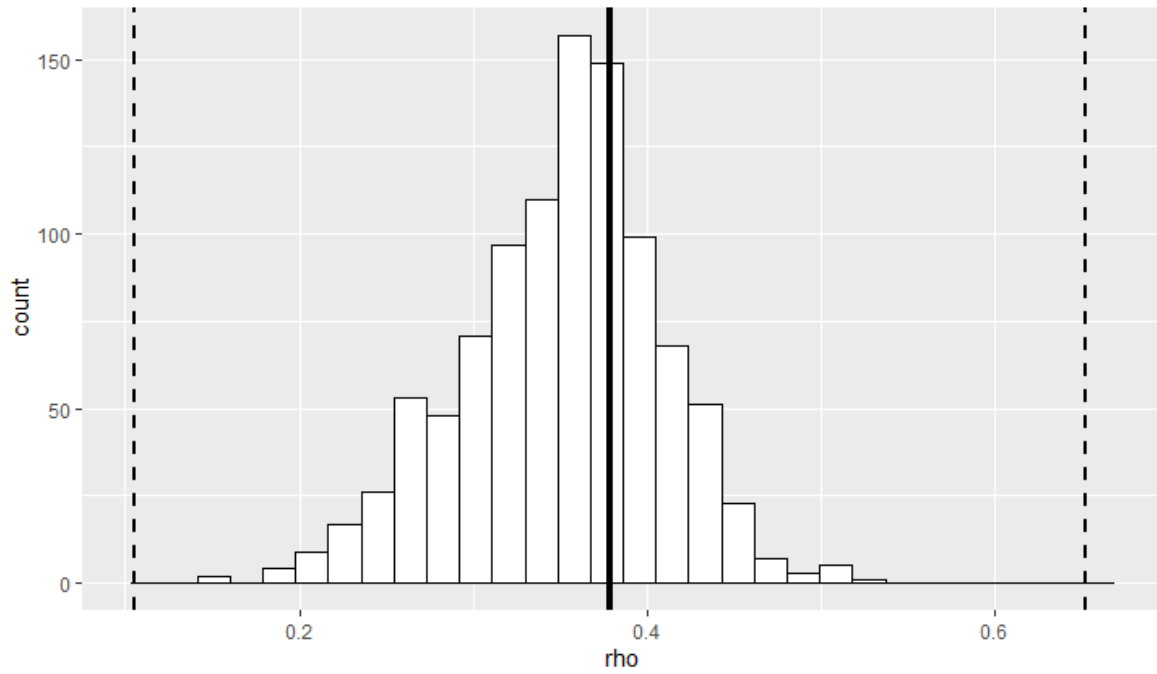


Рис. 7: Эмпирическая плотность распределения $\hat{\rho}$ в тестах на устойчивость по переменной *Conf_in_civil_services*

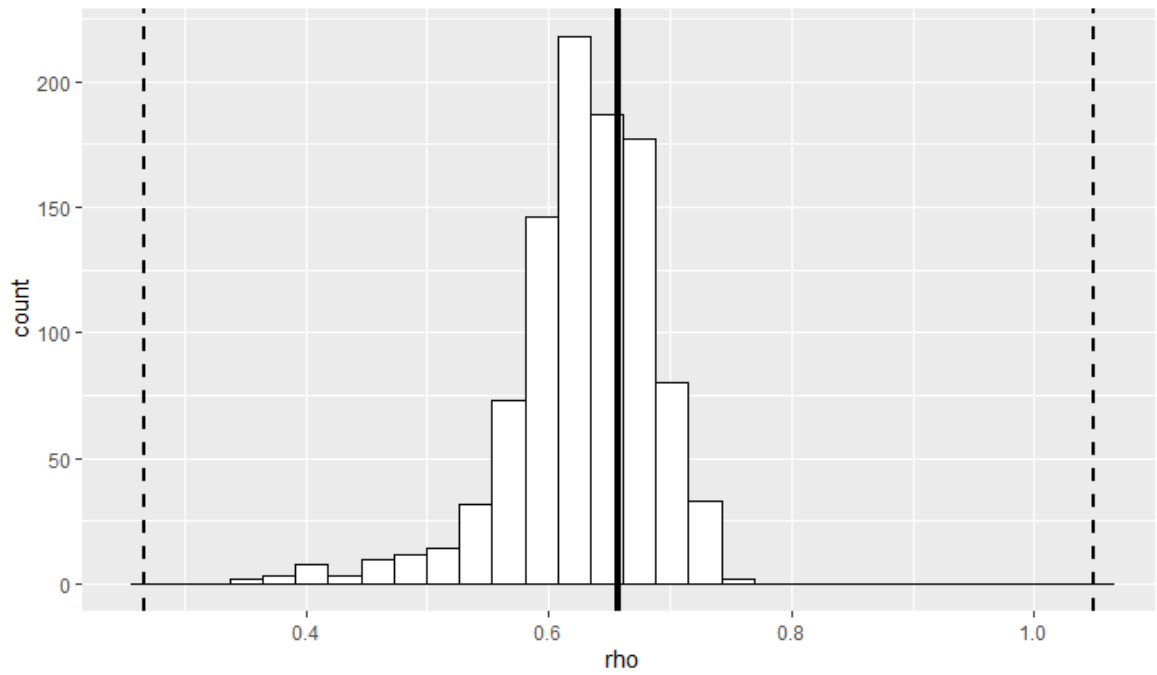


Рис. 8: Эмпирическая плотность распределения $\hat{\rho}$ в тестах на устойчивость по переменной *Homo_neighbours*

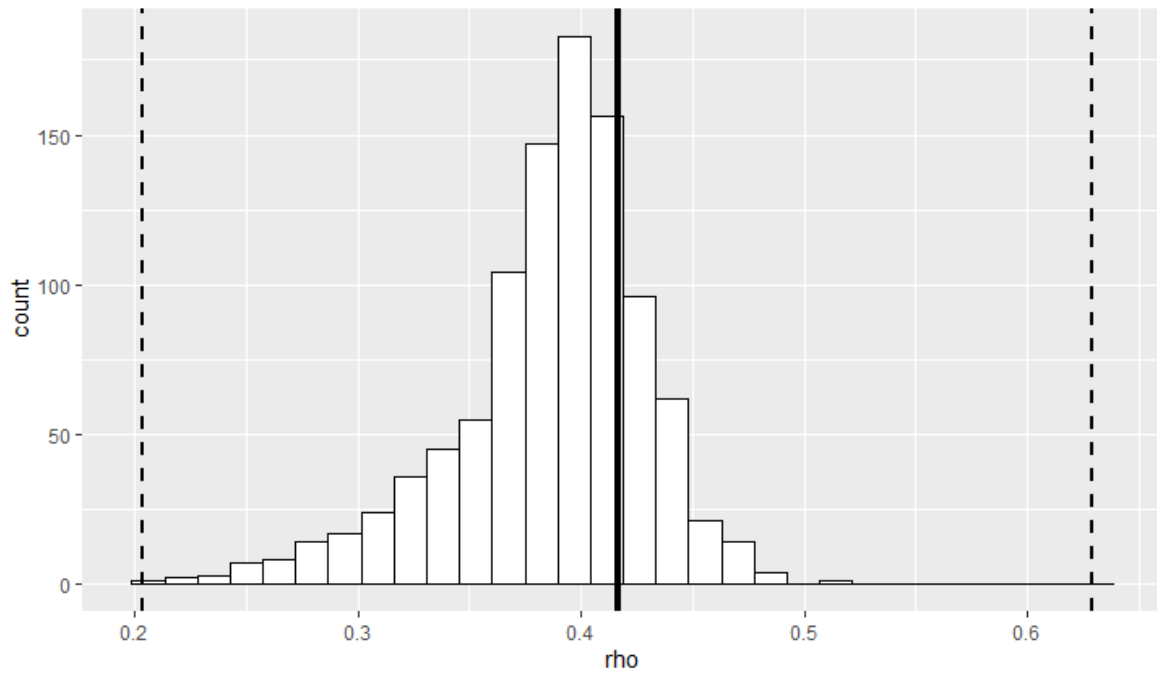


Рис. 9: Эмпирическая плотность распределения $\hat{\rho}$ в тестах на устойчивость по переменной *Religion*

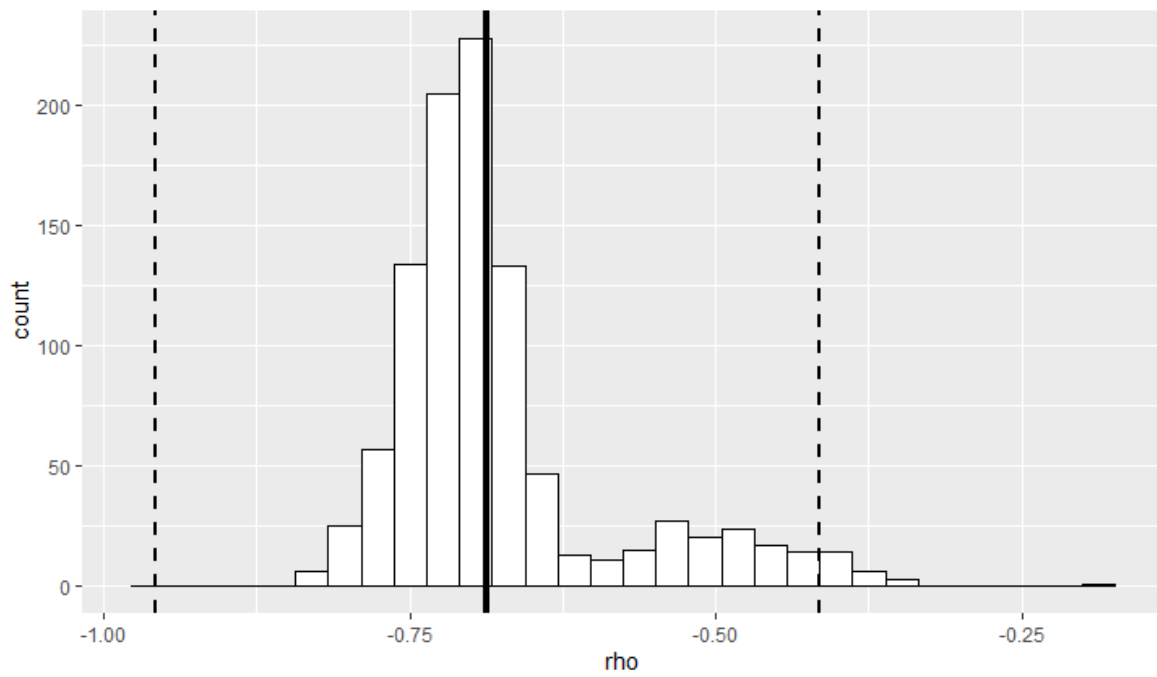


Рис. 10: Эмпирическая плотность распределения $\hat{\rho}$ в тестах на устойчивость по переменной *Trust*

Литература

- Bailey, M., Cao, R., T Kuchler, J. S., and Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(2):259–280.
- Bailey, M., Gupta, A., Hillenbrad, S., Kuchler, T., Richmond, R., and Stroebe, J. (2021). International trade and social connectedness. *Journal of International Economics*, 129(103418).
- Bailey, M., Johnston, D., Kuchler, T., Russel, D., State, B., and Stroebe, J. (2020). The determinants of social connectedness in europe. *Social Informatics. SocInfo 2020. Lecture Notes in Computer Science*, 12467(8310):1–14.
- Croissant, Y. and Milla, G. (2019). *Panel Data Econometrics with R*. John Wiley Sons Ltd.
- Diemer, A. and Regan, T. (2020). No inventor is an island: social connectedness and the geography of knowledge flows in the US. *CEP Discussion Papers, Centre for Economic Performance, LSE.*, (1731).
- Elhorst, P., Abreu, M., Amaral, P., Bhattacharjee, A., Corrado, L., Doran, J., Fuerst, F., Gallo, J. L., McCann, P., Monastiriotis, V., Quatraro, F., and Yu, J. (2019). Raising the bar (11). *Spatial Economic Analysis*, 14(2):129–132.
- Hsieh, C. and van Kippersluis, H. (2019). Smoking initiation: Peers and personality. *Quantitative Economics*, 9(2):825–863.
- LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman Hall/CRC.
- Milla, G. and Piras, G. (2012). splm: Spatial panel data models in R. *Journal of Statistical Software*, 47(1):1–38.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Иванова, (2019). ВРП и загрязнение окружающей среды в регионах России: пространственно-эконометрический анализ. *Квантиль*, (14):53–62.