УДК 330.4+316.4+51-77

DOI: [10.14530/se.2017.3.019-039](https://doi.org/10.14530/se.2017.3.019-039)

ОСОБЕННОСТИ НЕОДНОРОДНОСТИ СТРАН ПО ЭТНИЧЕСКОМУ РАЗНООБРАЗИЮ: СЛУЧАЙ ШРИ-ЛАНКИ, США И КНР

Д.С. Васькин, А.Б. Шаповал

Васькин Дмитрий Сергеевич – магистрант. Национальный исследовательский университет Высшая школа экономики, ул. Мясницкая, 20, Москва, Россия, 101000. E-mail: vaskind@yandex.ru.

Шаповал Александр Борисович – доктор физико-математических наук, профессор. Национальный исследовательский университет Высшая школа экономики, ул. Мясницкая, 20, Москва, Россия, 101000; Лаборатория исследования социальных отношений и многообразия общества, Российская экономическая школа, ул. Новая, 100а, Сколково, Москва, Россия, 143026. E-mail: abshapoval@gmail.com.
ORCID: 0000-0001-5340-1930

Аннотация. В работе исследована неоднородность стран по этническому разнообразию на примере Шри-Ланки, КНР и США. Этническое разнообразие оценивается с помощью энтропийных индексов, индекса Джини – Симпсона и индекса фрагментации M , основанного на доле доминирующего этноса в регионе. Показано, что этническое разнообразие регионов связано с их географическим положением. Обосновано, что упорядочение регионов по их этническому разнообразию нечувствительно к выбору конкретной меры разнообразия. Неоднородность стран описана с помощью эмпирической функции распределения регионов. Выбрав в качестве меры разнообразия индекс M , показано, что соответствующее ранговое распределение регионов близко к линейному в Шри-Ланке и США, в отличие от КНР, где уровень этнического разнообразия низок, и в трети регионов доминирует один из этносов. Авторами предложена теоретико-вероятностная модель для объяснения найденных эмпирических закономерностей.

Ключевые слова: индекс этнического разнообразия, вероятностная модель расселения, функция распределения регионов, Шри-Ланка, США, КНР

Для цитирования: Васькин Д.С., Шаповал А.Б. Особенности неоднородности стран по этническому разнообразию: случай Шри-Ланки, США и КНР // Пространственная экономика. 2017. № 3. С. 19–39. DOI: [10.14530/se.2017.3.019-039](https://doi.org/10.14530/se.2017.3.019-039).

For citation: Vaskin D.S., Shapoval A.B. Features of Countries' Heterogeneity over Ethnic Diversity: Cases of Sri-Lanka, USA, PRC. *Prostranstvennaya Ekonomika = Spatial Economics*, 2017, no. 3, pp. 19–39. DOI: [10.14530/se.2017.3.019-039](https://doi.org/10.14530/se.2017.3.019-039).

© Васькин Д.С., Шаповал А.Б., 2017

Работа выполнена при поддержке Министерства образования и науки Российской Федерации, грант Правительства РФ, договор № 14.U04.31.0002.

ВВЕДЕНИЕ

Этническое, культурное и религиозное разнообразие населения неоднозначно влияет на экономические и социальные аспекты развития государства (Alesina, La Ferrara, 2005). С одной стороны, показано, что на возникновение и развитие городов положительно влияло этническое и культурное разнообразие их жителей (Sassen, 2011; Florida, 2002). С другой стороны, чрезмерное разнообразие приносит вред обществу. Как показано в работе (Easterly, Levine, 1997), тяжелая экономическая ситуация в ряде африканских стран связана со слишком высоким уровнем этнического разнообразия. Низкая эффективность и коррумпированность правительств объясняется высоким уровнем разнообразия общества (La Porta et al., 1999). Этот результат уточняется в работе (Alesina, Zhuravskaya, 2011), где показано, что этническое и лингвистическое, но не религиозное разнообразие является одним из факторов, который ухудшает качество правительства.

Этническое разнообразие нередко является причиной конфликтов в обществе (Бутаева и др., 2016). Природа таких конфликтов может быть различной: от мирных соревнований этнических групп до насильственных столкновений и гражданских войн (Козырев, 2001; Welsh, 1993). Предполагается, что зависимость между этническим разнообразием и конфликтами является немонотонной: как малое, так и большое этническое разнообразие может увеличивать вероятность конфликтов (Horowitz, 1985).

Эмпирические выводы о статистически значимом влиянии неоднородности на экономическое развитие стран или крупных регионов обычно связаны с локальными особенностями регионов и поэтому слабо переносимы от страны к стране. Наиболее детальная информация об анализируемых объектах предоставляется на уровне регионов. Из-за их неполноты и отсутствия межстрановых стандартов сбора данных использование этой информации затруднительно при сопоставлении выводов о различных странах. Тем не менее подробное изучение отдельных факторов региональной неоднородности позволяет глубже исследовать феномен разнообразия в контексте экономического развития.

Центральным инструментом анализа роли неоднородности в экономическом развитии является индекс разнообразия (Davydov, Weber, 2016). В данной работе оценивается этническое разнообразие Шри-Ланки, США и КНР путем вычисления и последующего анализа индексов фрагментации со следующей целью: выделить закономерности сосуществования этносов и предложить возможные объяснения найденных закономерностей. Для анализа используются энтропийные индексы разнообразия, индекс Джини – Симпсона и их модификации. Неоднородность регионов описывается с помощью эмпирической функции распределения.

В статье под распределением регионов понимается статистический тер-

мин, означающий последовательность регионов в соответствии с некоторым признаком (здесь – с уровнем разнообразия). Установлено, что распределение регионов по разнообразию этносов практически нечувствительно к выбору исследуемых индексов. Поскольку матрица расстояний между этносами до сих пор не построена, используется простой и одновременно эффективный способ учитывать близость между отдельными этносами при исследовании распределения регионов. Мы объединяем произвольно выбранную пару этносов в один, что приводит к новому распределению регионов по разнообразию этносов. При повторении этой процедуры для всевозможных пар этносов получается семейство распределений регионов. Установлено, что все попарные коэффициенты корреляции Спирмена между представителями полученного семейства близки к единице. Показано, что разнообразие регионов связано с их географическим положением: регионы с близкими характеристиками разнообразия соседствуют друг с другом. В статье предложена теоретико-вероятностная урновая модель расселения, которая демонстрирует распределение регионов по индексам разнообразия, согласующееся с наблюдениями в той части распределения, которое описывает регионы с одной доминирующей группой жителей. Теоретико-вероятностная модель расселения построена на основе модели сегрегации Шеллинга (Schelling, 1971). При определении модели предполагается, что индивид предпочитает жить среди представителей своего этноса. Результаты модели получены методом Монте – Карло. Компьютерные вычисления проводятся с помощью пакета прикладных программ Матлаб (MatLab).

Наилучшая согласованность между моделью и наблюдениями достигается на примере Шри-Ланки. Модель имитирует процессы рождаемости и миграции при условии высокой сегрегации этносов. Мы полагаем, что среди трех рассматриваемых кейсов именно в Шри-Ланке наблюдается наибольшая фрагментация этносов, в связи с чем модель демонстрирует наилучшее согласие с данными. В целом, более точная согласованность модели наблюдениям ожидается в стране, в которой при значительной фрагментации населения зафиксированы конфликты между этническими группами или имеет место иммиграция, при которой иммигранты предпочитают жить среди своей этнической группы.

Для анализа использованы данные из базы «Этнолог» (Ethnologue, 2009) и последних переписей населения, содержащие информацию о количестве людей, принадлежащих к каждой этнической группе в административных округах Шри-Ланки, Китая и США. Новые результаты в этой статье достигнуты за счет: а) использования современных данных, б) обоснования устойчивости распределения регионов к вариациям в определении этносов, в) построения вероятностной модели, объясняющей выявленные закономерности.

Необходимость принятия во внимание расстояния между этническими группами при оценке разнообразия стран обсуждается в ряде работ (Fearon, Laitin, 2003). Авторы использовали три различных определения расстояния, основанные на дереве языковых семейств, и показали, что каждый из трех построенных индексов фрагментации сильно коррелирует с индексом, не учитывающим расстояние между группами. Мы отказываемся от этого способа, оставаясь в рамках этнического, а не языкового разнообразия. Альтернативный способ измерения расстояния между группами людей используют меры генетической близости (Spolaore, Wacziarg, 2009). Однако для этносов, обсуждаемых в данной работе, генетическая близость до сих пор не оценена.

Выбор Шри-Ланки в качестве одной из исследуемых стран интересен с учетом недавней гражданской войны, продолжавшейся почти 30 лет и унесшей жизни почти 100 тыс. человек. Фрагментация населения, имеющая исторические корни, уходящее в колониальное прошлое страны, видимо взаимосвязана с прошедшим конфликтом. Универсальность эмпирических закономерностей, первоначально выявленная на данных Шри-Ланки, обсуждается в контексте анализа данных КНР, представляющую страны с низким уровнем разнообразия, и США, где разнообразие этносов достаточно велико, но в течение долгого времени отсутствовали войны внутри страны.

Работа организована следующим образом. В разделе 1 вводятся индексы разнообразия и обсуждаются их свойства. В разделе 2 анализируются особенности распределений регионов, построенных на основе различных индексов разнообразия. Влияние географического положения региона на величину разнообразия исследуется в разделе 3. В разделе 4 вводится и тестируется вероятностная модель расселения жителей по регионам. В заключительном разделе содержатся основные выводы проведенного исследования.

1. ФОРМАЛЬНЫЕ ОПРЕДЕЛЕНИЯ

Индекс разнообразия – это безразмерный показатель, который оценивает отклонение распределения исследуемой выборки от равномерного распределения. Несмотря на то, что появление индексов разнообразия связано с биологическими системами, в настоящее время эти индексы успешно применяются для оценки разнообразия, возникающего в ряде общественных категорий, включающих в себя этнос и религию.

В данной работе анализируется этническое разнообразие трех многонациональных стран: Шри-Ланки, КНР и США. Для анализа использованы данные по распределению людей из каждого административного округа этих стран по этническим группам. Таким образом, рассматривается страна, разделенная на J регионов (штатов), в которой проживают I не-

пересекающихся этнических групп (т. е. для простоты предполагается, что каждый человек принадлежит только к одному этносу). Пусть в округе j , $j = 1, \dots, J$, проживают N_j жителей. Через n_{ji} , $i = 1, \dots, I$, обозначается численность i -й этнической группы в j -м регионе. На самом деле количество I этнических групп, вообще говоря, меняется от региона к региону. Для удобства обозначений будем считать, что I постоянно, но часть этих групп может быть пустой, то есть $n_{ji} \in [0, N_j]$. Из сделанных обозначений следует, что $\sum_{i=1}^I n_{ji} = N_j$. Через $p_{ji} = \frac{n_{ji}}{N_j}$ обозначается доля i -го этноса в населении j -го округа. Тогда вектор $p_j = (p_{j1}, p_{j2}, \dots, p_{jI})$ задает распределение жителей по этническим группам в регионе j . Индексы разнообразия будут вычислены по векторам p_j .

Введем индексы разнообразия, которые используются в статье. Для упрощения обозначений временно опустим индекс вектора долей, который соответствует номеру региона, то есть вместо вектора $p_j = (p_{j1}, p_{j2}, \dots, p_{jI})$ рассматривается вектор $p = (p_1, p_2, \dots, p_I)$. Индекс $E(p)$ этнолингвистической фрагментированности Джини – Симпсона определяется формулой:

$$E(p) = 1 - \sum_{i=1}^I p_i^2.$$

Индекс Джини – Симпсона является оценкой вероятности того, что два случайно выбранных жителя региона принадлежат к разным этническим группам. Заметим, что при доминировании одной из групп населения (существовании координаты p_i , близкой к единице) индекс принимает значения, близкие к нулю. В этом случае индекс E фактически определяется размером доминирующей группы. Если только доминирующая группа представляет интерес, то, следуя Познеру (Posner, 2004), естественно ввести следующий индекс M ,

$$M(p) = 1 - \max_i p_i^2 - \left(1 - \max_i p_i\right)^2,$$

который зависит именно от наибольшей координаты в векторе p , игнорируя распределение остальных координат¹.

Другая мера разнообразия, используемая в данной работе, – однопараметрическое семейство индексов Реньи, которое было введено в (Renyi, 1961) для решения других задач. Индексы семейства с параметром $\alpha \in (0, +\infty)$, $\alpha \neq 1$, вычисляются по формуле:

¹ Заметим, что такой подход приписывают Нюю: когда он выбирал «каждой твари по паре», то его интересовало только количество видов. На самом деле, согласно Ветхому Завету, Ной действовал несколько сложнее, разделяя животных на чистых и нечистых. Нечистых животных он действительно брал по паре, а чистых – 7 пар. Впрочем, и при таком подходе несомненный приоритет Ноя – количество различных особей.

$$R_{\alpha}(p) = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^I p_i^{\alpha} \right),$$

где слагаемые вида $0 \log_2 0$ заменяются на 0. Заметим, что при исключении из рассмотрения нулевых координат вектора p значение индекса не изменится.

Индекс Реньи чувствителен к изменению количества (непустых) групп и величины отдельных групп. Роль этих двух показателей регулируется параметром α . Например, при малых значениях α индекс отражает в большей мере количество групп (ненулевых координат в векторе p), чем величину отдельных значений, а в пределе $\alpha \rightarrow 0$ значение индекса $R_0(p) = \log_2(I)$ зависит только от количества групп. Напротив, при больших значениях α значение индекса полностью характеризуется значением наибольшей координаты, поскольку $R_{\infty}(p) = \log_2 \left(\frac{1}{\max_i p_i} \right)$. Таким образом, при увеличении параметра α от нуля до $+\infty$ индекс Реньи осуществляет переход от измерения разнообразия с помощью количества групп до измерения разнообразия с помощью размера наибольшей группы. Последнее соответствует индексу M .

Самостоятельное значение имеет «промежуточный» индекс Реньи, который возникает при α , стремящемся к единице. Это – хорошо известная энтропия $H(p)$, (Shannon, Weaver, 1949), определяемая как

$$H(p) = - \sum_{i=1}^I p_i \log_2 p_i.$$

С увеличением видового разнообразия сообщества значение индекса $H(p)$ увеличивается и достигает максимального значения при равенстве долей всех видов в сообществе, т. е. при симметричном представительстве всех видов.

В качестве «промежуточных» значений параметра α мы рассмотрим α , равные 1 (т. е. энтропию в качестве индекса) и $1/2$.

2. ОСОБЕННОСТИ РАСПРЕДЕЛЕНИЙ ПО ИНДЕКСАМ

Распределение регионов по индексам

Распределение регионов Шри-Ланки по индексам E и M , измеряющим этническое разнообразие регионов, показано на рисунке 1. Формально пусть X – это индекс E или M , $F(x)$ – это эмпирическая функция распределения, определяемая как доля тех регионов j , для которых $X(p_j) < x$:

$$F(x) = \frac{\#j \in \{1, 2, \dots, J\} : X(p_j) < x}{J},$$

$p_j = (p_{j1}, p_{j2}, \dots, p_{ji})$ – вектор долей в регионе j , $F_c(x) = 1 - F(x)$ – дополнительная функция распределения.

Распределения, вычисленные по индексам E и M , имеют похожую структуру: график функции $F_c(x)$ состоит из пяти квазилинейных «ступенек». Можно говорить о шестой «ступеньке» в правой части графика, построенного по индексу M , но три правых точки, построенных по индексу E , не формируют «ступеньку». Оба феномена: квазилинейность функции распределения и группируемость регионов – представляют несомненный интерес.

Рис. 1. Графики дополнительной функции распределения $F_c(x)$ индексов $E(p)$ и $M(p)$

Примечание. Col – Colombo (6), Gam – Gampaha (3), Kal – Kalutara (4), Kan – Kandy (6), Mat – Matale (5), Nuw – Nuwara Eliya (7), Gal – Galle (2), Mat – Matara (2), Ham – Hambantota (1), Jaf – Jaffna (1), Man – Mannar (5), Vav – Vavuniya (6), Mul – Mullaitivu (4), Kil – Kilinochchi (1), Bat – Batticaloa (6), Amp – Ampara (7), Tri – Trincomalee (7), Kur – Kurunegala (3), Put – Puttalam (6), Anu – Anuradhapura (3), Pol – Polonnaruwa (3), Bad – Badulla (6), Mon – Monaragala (2), Rat – Ratnapura (4), Keg – Kegalla (4).

Общая структура графиков $F_c(x)$ для индексов E и M не случайна: порядок следования регионов в соответствии со значениями этих двух индексов почти полностью идентичен. Близость двух распределений в левой части графика на рисунке 1 объясняется тем, что малые значения индекса E возникают при доминировании одной из рассматриваемых групп; в этом случае оба индекса примерно равны $1 - \max_i p_{ji}^2$.

Мы проверили сохранение порядка в распределении регионов при использовании всех приведенных выше индексов с помощью рангового коэффициента корреляции Спирмена. Для каждой пары индексов (X, Y) , $X, Y \in \{E, M, H, R_{12}\}$ вычислен и записан в таблицу ранговый коэффициент корреляции

$$\rho_{XY} = 1 - \frac{6 \sum_{j=1}^J (r_{X_j} - r_{Y_j})^2}{J(J^2 - 1)},$$

где r_{X_j} и r_{Y_j} – это ранги (номера по порядку в распределении регионов) региона j при использовании индексов X и Y соответственно.

Таблица

Коэффициенты корреляции между индексами

Индекс	Е	R_{12}	Н	М
E	1	0,9338	0,9931	0,9908
R_{12}	0,9338	1	0,9562	0,9054
H	0,9931	0,9562	1	0,9769
M	0,9908	0,9054	0,9769	1

Согласно таблице, наименьший коэффициент корреляции между индексами равен 0,9054. Как известно, значения коэффициента корреляции ρ_{XY} лежат на отрезке $(-1, 1)$, причем чем ближе коэффициент к единице, тем точнее сохраняется порядок следования элементов выборки. Поэтому найденные коэффициенты ранговой корреляции свидетельствуют об устойчивости рангового распределения регионов по разнообразию к выбору индексов из указанного выше множества индексов.

Проверка устойчивости сохранения порядка

В предыдущем разделе было установлено сохранение порядка следования регионов при переходе от одного индекса разнообразия к другому, не принимая во внимание близость / удаленность отдельных этносов друг от друга. Сейчас этот пробел будет устранен.

Как правило, при решении аналогичных задач близость между разными группами индивидов оценивается показателем, лежащим между нулем и единицей. Мы предлагаем использовать иной подход, приписывая расстоянию между группами только значения 0 или 1. Более точно, мы выбираем произвольную пару (i_1, i_2) этносов и объединяем их в один, считая расстояние между ними нулевым (формально заменяем $p_{ji_1} \rightarrow p_{ji_1} + p_{ji_2} \rightarrow 0$ для всех $j \in 1, 2, \dots, J$), и определяем новый порядок следования регионов с помощью каждого из индексов: E, M, H и R_{12} . Далее эта операция повторя-

ется для всех остальных пар этнических групп. В результате возникает $4J(J-1)/2$, вообще говоря, различных ранговых распределений регионов. Для всех пар полученных распределений вычисляется коэффициент корреляции Спирмена.

Ясно, что большая часть возникших распределений бессмысленна, поскольку она получена путем объединения непохожих этносов. Важно, однако, что наименьший коэффициент корреляции Спирмена из полученных указанным выше способом коэффициентов равен 0,89, а нижний пятипроцентный квантиль этих коэффициентов равен 0,91. Поскольку при каждом попарном объединении разных этнических групп в одну ранговое распределение почти не изменяется, то корреляция распределений, полученных путем объединения похожих этносов, также велика. Наибольшая близость ранговых распределений наблюдается по индексам H и M . Наименьший коэффициент корреляции на возникающих парах ранговых распределений равен 0,97. Подчеркнем, что результат, подтверждающий устойчивость рангового распределения регионов к объединению близких этносов, установлен без выявления тех этносов, которые на самом деле являются близкими.

Степенные распределения регионов

Вопрос о существовании степенных закономерностей в ранговых распределениях регионов обсуждается в литературе (Гелашвили и др., 2004; Гелашвили и др., 2005; Zipf, 1949), причем обсуждение природы степенных распределений выходит за рамки экономических приложений (Gabaix, 1999; Shapoval, Shnirman, 2012; Aschwanden et al., 2016). С одной стороны, исследователи сообщают о степенных законах, выявленных эмпирически. С другой стороны, пока нет общепризнанной простой теоретической модели, которая бы определила механизм (то есть микрооснования), лежащий в основе группирования индивидов в соответствии со степенными законами.

Оставляя в стороне статистическую проверку согласия наблюдаемых распределений с теоретическими (усеченными) степенными случайными величинами, мы проводим визуальный анализ рассматриваемых распределений, полагая, что уже на этом шаге будут раскрыты эмпирические закономерности, теоретическое объяснение которых лежит за пределами данной работы.

Графики дополнительной функции распределения $F_c(x)$ в двойных логарифмических осях для индексов разнообразия H и $R_{1/2}$ представлены на рисунке 2. Ранговое распределение регионов по индексу H аккуратно аппроксимируется степенной функцией в своей центральной части, в которую

входит до 15 (т. е. чуть меньше половины) регионов. Ранговое распределение регионов, построенное по индексу $R_{1/2}$, аппроксимируется несколько менее аккуратно (из-за «провала») в окрестности точки $(2^{0,5}, 2^{-1,5})$. Отметим, что степенная функция с той же степенью аппроксимирует также пять точек (шесть для индекса H), лежащих в окрестности точки $(1, 1/2)$. Поэтому «следы» степенных законов несомненно прослеживаются в построенных распределениях. Однако источник их неясен; для его поиска требуется более подробный эмпирический анализ, подтверждающийся теоретическим моделированием.

Рис. 2. Графики дополнительной функции распределения H и $R_{1/2}$ в логарифмических координатах и прямые линии для иллюстрации степенного закона

3. ГЕОГРАФИЯ ИМЕЕТ ЗНАЧЕНИЕ

В этом разделе мы покажем, что распределение регионов по их разнообразию соотносится с их географическим положением. Поскольку порядок следования регионов устойчиво определяется наиболее простым индексом M , то мы используем именно этот индекс при сопоставлении распределения регионов с их географическим положением. Как показано на рисунке 3, каждой из семи групп, выделенных с помощью рангового распределения регионов по индексу M , присваивается номер из интервала $[1, 7]$.

Рис. 3. Дополнительная функция распределения регионов по индексу M в Шри-Ланке

Далее, на рисунке 4 приведена карта Шри-Ланки, на которой регионам присвоены те же номера, что и на рисунке 3. Цвет / штриховка выбраны в зависимости от того, какой этнос доминирует в регионе (см. легенду к рисунку 4).

Сопоставление карты с делением регионов на группы свидетельствует о том, что 7 регионов из центральной части распределения, которым приспаны номера 3 и 4 с доминирующим этносом Синхалезе (Sinhalese), составляют связную область. Области карты под номерами 2 и 6 с тем же доминирующим этносом несвязны. Переход от одной части этих областей к другой происходит через область 4 (или 1), которые относятся к близким группам. Регионы, обозначенные числом 7, являются особыми как на рисунке 3, так и на рисунке 4. Север страны с доминирующим этносом шри-ланкийских тамиллов (Sri Lanka Tamil) демонстрирует монотонное увеличение разнообразия, чему, видимо, способствует соседство с регионами, в которых доминирует другой этнос. Север и юг страны обладают наи-

большим разнообразием, тогда как центр страны характеризуется большим перемешиванием этносов. Таким образом, география, несомненно, влияет на разнообразие регионов.

Рис. 4. Карта округов Шри-Ланки

Примечание. Цифры указывают на группы рисунка 3, к которой относится регион; цвет / штриховка выбраны в зависимости от доминирующего этноса.

4. ИМИТАЦИОННАЯ МОДЕЛЬ

В этом разделе мы предложим модель разделения индивидов на группы, которая порождает квазилинейное распределение регионов по индексу M , согласующееся с эмпирическим распределением. Наша модель реализует

предположение, что индивид, выбирая место жительства, останется в регионе с вероятностью, равной доле его этноса в этом регионе (другими словами, индивид выберет регион для проживания, если наугад встреченный житель принадлежит к его этносу, но уйдет искать другое место жительства в противном случае).

Технически расселение жителей по регионам происходит один за другим. Каждый индивид сначала выбирает наугад (т. е. с равной вероятностью) регион $j \in 1, 2, \dots, J$. Пусть рассматриваемый индивид принадлежит к этносу i , а p_{ji} – это текущая доля этноса i в регионе j . Тогда мы полагаем, что индивид останется в регионе j , если наугад встреченный в регионе житель также принадлежит к этносу i (т. е. с вероятностью p_{ji}). В противном случае индивид возвращается к начальному шагу действий, т. е. вновь наугад выбирает регион (из всех J регионов, включая тот, в котором он не остался на предыдущем шаге). Процедура выбора региона продолжается до тех пор, пока на очередном шаге регион не будет выбран. Более формально обозначим через Y_k событие, означающее, что место жительства не выбрано за k шагов, и предположим, что событие Y_k произошло. Тогда процедура завершится выбором региона j на шаге $k + 1$ с вероятностью $1/J \cdot p_{ji}$, где $1/J$ – это вероятность выбора региона j , а p_{ji} – вероятность того, что наугад встреченный индивид принадлежит к этносу i . По формуле полной вероятности, суммируя вероятности того, что регион j выбран на шаге $k = 1, 2, \dots$ с весами $P\{Y_{k-1}\}$, где $P\{Y_0\} = 1$ мы получаем, что (полная) вероятность выбора региона j очередным индивидом, относящимся к этносу i , равна $p_{ji}/(p_{1i} + p_{2i} + \dots + p_{ji})$. Разумеется, доли p_{ji} уточняются после того, как очередной индивид найдет себе место жительства; для простоты эта зависимость игнорируется в обозначениях.

Начальные условия в процедуре расселения заданы следующим образом. Процедура расселяет N индивидов; число N может, например, совпадать с наблюдаемым количеством проживающих в стране. Выбор очередного индивида, который выбирает регион проживания указанным выше способом, происходит наугад (без возвращения, т. е. выбор региона осуществляется раз и навсегда). Пусть p_{ji} – это наблюдаемые доли этносов $i = 1, \dots, I$ в регионе j ; $i_0(j)$ – это номер этноса, который доминирует в регионе j (равенство доминирующих долей не наблюдается, поэтому этот исключительный случай здесь опущен). Предполагается, что первые J человек расселены так, что начальные доли в модели определяются формулами:

$$p_{i_0(j),j} = \beta \hat{p}_{i_0(j),j} + \frac{1 - \beta}{J}, \quad p_{i,j} = \frac{1 - p_{i_0(j),j}}{J - 1}, \quad i \neq i_0(j).$$

Число $\beta \in (0, 1)$ – это единственный параметр модели. Например, при

$\beta = 1$ предполагается, что начальное значение доминирующей доли совпадает с наблюдаемым. Напротив, при $\beta = 0$ начальное распределение долей равномерно. За счет выбора параметра β мы приближаем модельное распределение регионов по разнообразию к наблюдаемому. Начальные доли, скорее всего, соответствуют нецелому количеству индивидов каждого этноса, расселенных по округам. Это – технический аспект модели, не влияющий на ее интерпретацию: мы изучаем количество жителей, которые расселены в конце процедуры, когда возникающие числа будут округлены до ближайшего целого.

Мы покажем эмпирически, что доли p_{ji} стабилизируются в модели (по крайней мере при корректном выборе параметра β). Установлено, что при β , равном или близком единице, доминирующий этнос группируется в отдельных регионах слишком быстро, в результате чего в левой части распределения регионов по индексу M значения смещаются слишком сильно влево по сравнению с той же частью распределения, построенного по реальным данным.

Для каждого фиксированного N численный эксперимент по расселению N индивидов проводился 100 раз. Для всех экспериментов вычислена функция распределения $F_{c,N}$ регионов по индексу M . Мы изучаем среднее арифметическое найденных ста функций распределений и стандартное отклонение, обозначаемые $\mu_{c,N}$ и σ_N соответственно, найденное в каждой точке области значений распределения (т. е. для каждого региона). Стабилизация функции среднего $\mu_{c,N}$ дополнительных функций распределения достигается уже при $N = 20\,000$ (подтверждающий график не приводим).

Построенный график $\mu_{c,N}$ с $N = 20\,000$ далее заменяет предельный случай $N = \infty$. Рисунок 5 показывает соответствие средних распределений регионов $\mu_{c,N} \pm \sigma_N$ и распределения F_c по индексу M , построенных по модельным (с $N = 20\,000$) и реальным данным. Согласие между теоретической моделью и наблюдениями имеет место для всех регионов, кроме 3–4 точек с $M \in (0,36, 0,4)$, причем только две точки, построенные по реальным данным, отклоняются от модельного распределения больше, чем на стандартное отклонение. Средняя функция распределения $\mu_{c,N}$, построенная по модельным данным, в отличие от F_c , не имеет «ступенек». Возможно, близость распределения регионов к линейной функции является эмпирической закономерностью, а отклонения от линейности, связанные в том числе с группируемостью, имеет смысл использовать при оценке влияния разнообразия регионов на экономические и социальные процессы.

Мы обращаем внимание, что стандартное отклонение σ_N как функция индекса принимает наименьшие значения по краям распределения, что свидетельствует о более быстрой стабилизации динамики расселения в

регионах с большой и малой доминирующими долями. Выбор параметра β осуществлен таким образом, чтобы график $F_{c,\infty}$, построенный по модельным данным, совпадал с наблюдаемым распределением в левой части графика.

Рис. 5. Среднее $\mu_{c,N}$ (черные круги) дополнительных функций распределения $F_{c,N}$ по 100 экспериментам и данным Шри-Ланки при параметре $\beta = 0,7$; средние плюс-минус среднеквадратические отклонения (серые круги); дополнительная функция распределения F_c (звезды) регионов Шри-Ланки; $N = 20\,000$

Теперь мы проверим близость модельного и наблюдаемого распределений регионов по индексу M на примере КНР и США. Распределение регионов КНР по индексу M не имеет ничего общего с линейной функцией. Доминирующая группа в большинстве регионов Китая велика, поэтому в одиннадцати регионах значение индекса M близко к нулю. Модельное распределение достаточно аккуратно аппроксимирует F_c на этих регионах, где оба распределения близки к линейному (рис. 6). Правое модельное распределение также отклоняется от линейной функции, но уже не соответствует наблюдениям.

Третий пример – Соединенные Штаты Америки – представляет страну с большим разнообразием этносов. Как и в Шри-Ланке, распределение регионов по индексу M допускает линейную аппроксимацию. Группируемость регионов по разнообразию проявляется в меньшей степени, чем в Шри-Ланке. Используя начальные данные, характеризующие США, мы построили распределение регионов по индексу M в модели и получили функцию, близкую к линейной (рис. 7).

Рис. 6. Среднее $\mu_{c,N}$ (круги) дополнительных функций распределения $F_{c,N}$ по 100 экспериментам и данным КНР при параметре $\beta = 0,9$; средние плюс-минус среднеквадратические отклонения (точки); дополнительная функция распределения F_c (звезды) регионов КНР; $N = 20\,000$

Рис. 7. Среднее $\mu_{c,N}$ (круги) дополнительных функций распределения $F_{c,N}$ по 100 экспериментам и данным США при параметре $\beta = 0,3$, средние плюс-минус среднеквадратические отклонения (точки) и дополнительная функция распределения F_c (звезды) регионов США; $N = 20\,000$

Распределение, построенное по модельным данным, также близко к линейной функции, но с незначительно отличающимся наклоном. На самом деле модельный график можно условно разделить на две части с точкой деления примерно в $M \approx 0,28$. Наклон левой части графика оказы-

вается больше, чем правой. Поэтому за счет выбора параметра β удастся приблизить наблюдения за моделью в левой части графика (примерно в половине точек), тогда как в правой части графика наблюдается некоторое отклонение распределения, построенного по модельным данным, от наблюдаемого.

Таким образом, рассмотренные примеры свидетельствуют, что модельное распределение $F_{c,\infty}$ согласуется с распределением регионов F_c в тех регионах, где доля доминирующего этноса велика. Если же разнообразие регионов велико и распределение F_c близко к линейному, то и модельное распределение $F_{c,\infty}$ близко к линейному и оно согласуется с F_c для большинства регионов.

Естественно считать, что модельные правила имитируют процессы рождения и отчасти внутривострановой миграции, если индивиды образуют семьи внутри этноса и перемещаются с места на место, как правило, не покидая представителей своего этноса. В самом деле, тогда вероятность рождения ребенка, принадлежащего к какому-нибудь этносу, в регионе j (в первом приближении) пропорциональна доле жителей этого этноса в регионе j . Аналогичные рассуждения применимы к иммиграции (возможно, также к эмиграции, но при инвертировании модельной процедуры добавления жителей), если приток населения не зависит от этноса, и отчасти к внутривострановой миграции при тех же предположениях. Мы полагаем, что предположения модели согласуются с особенностями расселения жителей в Шри-Ланке и США. В Шри-Ланке проявляется определенная сегрегация населения по этносам, тогда как в Соединенных Штатах Америки наблюдается определенная группируемость мигрантов при выборе места проживания.

ЗАКЛЮЧЕНИЕ

В статье исследовано распределение регионов по индексам разнообразия для Шри-Ланки, КНР и США. В качестве меры разнообразия использованы индекс Джини – Симпсона, индекс M , основанный на доле доминирующего этноса, и энтропийные индексы. Установлено, что выбор конкретного индекса из указанной выше группы практически не влияет на распределение регионов по разнообразию. Распределение регионов (штатов) по индексу M близко к линейному в Шри-Ланке и США. Этот вывод не переносится на КНР, что, по-видимому, связано с низким уровнем разнообразия в этой стране.

Мы построили теоретико-вероятностную модель расселения индивидов по регионам. Распределение регионов в модели соответствует наблюдаемому для тех регионов, в которых доля доминирующего этноса велика. Модель, построенная по данным Шри-Ланки и США, демонстрирует согла-

сие с наблюдениями для всех регионов. Модель, построенная по данным КНР, оказывается точной только для тех регионов Китая, в которых доля доминирующего этноса близка к единице. Обращаем внимание, что квазилинейность распределения регионов, построенного по реальным данным, индуцирует квазилинейность модельного распределения регионов. Идентификация микрооснований квазилинейности может стать темой дальнейших исследований.

Интересно понять, являются ли отклонения от линейного распределения, в случае Шри-Ланки выраженные в группируемости регионов, характеристиками, связанными с экономическими показателями регионов или с уровнем конфликтов.

БЛАГОДАРНОСТИ

А. Шаповал благодарен К. Бутаевой за предоставленные данные, Д. Давыдову – за полезную критику первого варианта статьи и С. Измалкову – за обсуждение способов проверки устойчивости полученных эмпирических результатов.

СПИСОК ЛИТЕРАТУРЫ

- Бутаева К.О., Вебер Ш., Давыдов Д.В. Язык, культуры, миграция, конфликты: экономическая проекция // Вестник Московского университета. Серия 6. Экономика. 2016. № 1. С. 3–21.
- Гелашивили Д.Б., Иудин Д.И., Розенберг Г.С., Якимов В.Н., Шурганова Г.В. Степенной закон и принцип самоподобия в описании видовой структуры сообществ // Поволжский экологический журнал. 2004. № 3. С. 227–245.
- Гелашивили Д.Б., Чупрунов Е.В., Иудин Д.И. Методологические аспекты пангеометризма в популяционной биологии // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Биология. 2005. № 1. С. 39–56.
- Козырев Г.И. Введение в конфликтологию. М.: Владос, 2001. 176 с.
- Alesina A., La Ferrara E. Ethnic Diversity and Economic Performance // Journal of Economic Literature. 2005. Vol. 43. Pp. 762–800. DOI: 10.1257/002205105774431243.
- Alesina A., Zhuravskaya E. Segregation and the Quality of Government in a Cross Section of Countries // The American Economic Review. 2011. Vol. 101. No. 5. Pp. 1872–1911. DOI: 10.1257/aer.101.5.1872.
- Aschwanden M., Crosby N.B., Dimitropoulou M., Georgoulis M.K., Hergarten S., McAteer J., Milovanov A.V., Mineshige S., Morales L., Nishizuka N., Pruessner G., Sanchez R., Sharma A.S., Strugarek A., Uritsky U. 25 Years of Self-Organized Criticality: Solar and Astrophysics. Space Sci Rev. 2016. Vol. 198. Pp. 47–166. DOI: 10.1007/s11214-014-0054-6.
- Davydov D., Weber S. A Simple Characterization of the Family of Diversity Indices // Economics Letters. 2016. Vol. 147. Pp. 121–123. DOI: 10.1016/j.econlet.2016.08.036.
- Easterly W., Levine R. Africa's Growth Tragedy: Policies and Ethnic Divisions // The

- Quarterly Journal of Economics. 1997. Vol. 112. Issue 4. Pp. 1203–1250. DOI: 10.1162/003355300555466.
- Fearon J.D., Laitin D.D.* Ethnicity, Insurgency, and Civil War // *American Political Science Review*. 2003. Vol. 97. Issue 1. Pp. 75–90. DOI: 10.1017/S0003055403000534.
- Florida R.* The Rise of the Creative Class: And how it's Transforming Work, Leisure, Community and Every day Life. New York: Basic Books, 2002. 434 p.
- Gabaix X.* Zipf's Law for Cities: An Explanation // *Quarterly Journal of Economics*. 1999. Vol. 114. Issue 3. Pp. 739–767. DOI: 10.1162/003355399556133.
- Horowitz D.L.* Ethnic Groups in Conflict. University of California Press, 1985. 697 p.
- La Porta R., Lopez-de Silanes F., Shleifer A., Vishny R.* The Quality of Government // *The Journal of Law, Economics and Organization*. 1999. Vol. 15. Issue 1. Pp. 222–279.
- Posner D.N.* Measuring Ethnic Fractionalization in Africa // *American Journal of Political Science*. 2004. Vol. 48. Issue 4. Pp. 849–865. DOI: 10.1111/j.0092-5853.2004.00105.x.
- Renyi A.* On Measures of Entropy and Information // *Proceedings of the 4th Berkeley Symposium on Mathematics Statistics and Probability*. 1961. Vol. 1. Pp. 547–561.
- Sassen S.* Cities in a World Economy. Sage Publications, 2011. 424 p.
- Schelling T.* Dynamic Models of Segregation // *The Journal of Mathematical Sociology*. 1971. Vol. 1. Issue 2. Pp. 143–186. DOI: 10.1080/0022250X.1971.9989794.
- Shannon C.E., Weaver W.* The Mathematical Theory of Communication. The University of Illinois Press, 1949. 125 p.
- Shapoval A.B., Shnirman M.* The BTW Mechanism on a Self-Similar Image of a Square: A Path to Unexpected Exponents // *Physica A: Statistical Mechanics and its Applications*. 2012. Vol. 391. No. 1–2. Pp. 15–20. DOI: 10.1016/j.physa. 2011.08.020.
- Spolaore E., Wacziarg R.* The Diffusion of Development // *The Quarterly Journal of Economics*. 2009. Vol. 124. Issue 2. Pp. 469–529. DOI: 10.1162/qjec.2009.124.2.469.
- Welsh D.* Domestic Politics and Ethnic Conflict // *Ethnic Conflict and International Security* / Edited by M.E. Brown. Princeton: Princeton University Press, 1993. Pp. 43–60.
- Zipf G.K.* Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge, 1949. 573 p.

FEATURES OF COUNTRIES' HETEROGENEITY OVER ETHNIC DIVERSITY: CASES OF SRI-LANKA, USA, PRC

D.S. Vaskin, A.B. Shapoval

Dmitry Sergeevich Vaskin – Master's Degree Student. National Research University Higher School of Economics, 20 Myasnitskaya Street, Moscow, Russia, 101000. E-mail: vaskind@yandex.ru.

Alexander Borisovich Shapoval – Doctor of Physic-Mathematical Sciences, Professor. National Research University Higher School of Economics, 20 Myasnitskaya Street, Moscow, Russia, 101000; Laboratory for Research of Social Relations and Diversity, New Economic School, 100a Novaya Street, Room 246, Skolkovo, Moscow, Russia, 143026. E-mail: abshapoval@gmail.com.

ORCID: 0000-0001-5340-1930

A. Shapoval wishes to acknowledge the support of the Ministry of Education and Science of the Russian Federation, grant No. 14.U04.31.0002.

Abstract. We explore countries' heterogeneity over ethnic diversity on the example of Sri-Lanka, PRC, and USA. The ethnic diversity is estimated with entropy-like indices, the Gini-Simpson index, and a fragmentation index M that is based on the share of the dominating ethnos. We posit that ethnic diversity of regions is related to their geographical locations. We claim that ordering of regions with respect to their ethnic diversity is not sensitive to the choice of specific measure of diversity. Heterogeneity of countries is described with empirical cumulative distribution function of regions. Choosing index M as a measure of diversity, we exhibit that the corresponding cumulative distribution function is close to linear in Sri-Lanka and USA, unlike in PRC where the level of ethnic diversity is low and single ethnos dominates in the third of all regions. We introduce a probabilistic model in order to explain empirical findings. The model is based on the assumption that individuals prefer to live with people of their own ethnos. The model distribution of regions matches the distribution computed with real data.

Keywords: index of ethnic diversity; probabilistic model of allocation; cumulative distribution function of regions, Sri Lanka, USA, PRC

REFERENCES

- Butaeva K.O., Weber S., Davydov D.V. Language, Culture, Migration, and Conflicts: Projection into Economic Field. *Vestnik Moskovskogo Universiteta. Seriya 6. Ekonomika* [Vestnik of Moscow University. Series 6. Economy], 2016, no. 1, pp. 3–21. (In Russian).
- Gelashvili D.B., Iudin D.I., Rosenberg G.S., Yakimov V.N., Shurganova G.V. Power Law and Self-Similarity Principle in Describing the Species Structure of Communities. *Povolzhskiy Ekologicheskii Zhurnal = Povolzhskiy Journal of Ecology*, 2004, no. 3, pp. 227–245. (In Russian).
- Gelashvili D.B., Chuprunov E.V., Iudin D.I. Methodological Aspects of Pangeometry in Population Biology. *Vestnik Nizhegorodskogo Universiteta im. N.I. Lobachevskogo. Seriya: Biologiya = Vestnik of Lobachevsky University of Nizhni Novgorod. Series: Biology*, 2005, no. 1, pp. 39–56. (In Russian).
- Kozyrev G.I. *Introduction to Conflictology*. Moscow, 2001, 176 p. (In Russian).
- Alesina A., La Ferrara E. Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 2005, vol. 43, pp. 762–800. DOI: 10.1257/002205105774431243.
- Alesina A., Zhuravskaya E. Segregation and the Quality of Government in a Cross Section of Countries. *The American Economic Review*, 2011, vol. 101, no. 5, pp. 1872–1911. DOI: 10.1257/aer.101.5.1872.
- Aschwanden M., Crosby N.B., Dimitropoulou M., Georgoulis M.K., Hergarten S., McAteer J., Milovanov A.V., Mineshige S., Morales L., Nishizuka N., Pruessner G., Sanchez R., Sharma A.S., Strugarek A., Uritsky U. 25 Years of Self-Organized Criticality: Solar and Astrophysics. *Space Sci Rev*, 2016, vol. 198, pp. 47–166. DOI: 10.1007/s11214-014-0054-6.
- Davydov D., Weber S. A Simple Characterization of the Family of Diversity Indices. *Economics Letters*, 2016, vol. 147, pp. 121–123. DOI: 10.1016/j.econlet.2016.08.036.
- Easterly W., Levine R. Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics*, 1997, vol. 112, issue 4, pp. 1203–1250. DOI: 10.1162/003355300555466.
- Fearon J.D., Laitin D.D. Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 2003, vol. 97, issue 1, pp. 75–90. DOI: 10.1017/S0003055403000534.
- Florida R. *The Rise of the Creative Class: And how it's Transforming Work, Leisure, Community and Every day Life*. New York: Basic Books, 2002, 434 p.

- Gabaix X. Zipf's Law for Cities: An Explanation. *Quarterly Journal of Economics*, 1999, vol. 114, issue 3, pp. 739–767. DOI: 10.1162/003355399556133.
- Horowitz D.L. *Ethnic Groups in Conflict*. University of California Press, 1985, 697 p.
- La Porta R., Lopez-de Silanes F., Shleifer A., Vishny R. The Quality of Government. *The Journal of Law, Economics and Organization*, 1999, vol. 15, issue 1, pp. 222–279.
- Posner D.N. Measuring Ethnic Fractionalization in Africa. *American Journal of Political Science*, 2004, vol. 48, issue 4, pp. 849–865. DOI: 10.1111/j.0092-5853.2004.00105.x.
- Renyi A. On Measures of Entropy and Information. *Proceedings of the 4th Berkeley Symposium on Mathematics Statistics and Probability*, 1961, vol. 1, pp. 547–561.
- Sassen S. *Cities in a World Economy*. Sage Publications, 2011, 424 p.
- Schelling T. Dynamic Models of Segregation. *The Journal of Mathematical Sociology*, 1971, vol. 1, issue 2, pp. 143–186. DOI: 10.1080/0022250X.1971.9989794.
- Shannon C.E., Weaver W. *The Mathematical Theory of Communication*. The University of Illinois Press, 1949, 125 p.
- Shapoval A.B., Shnirman M. The BTW Mechanism on a Self-Similar Image of a Square: A Path to Unexpected Exponents. *Physica A: Statistical Mechanics and its Applications*, 2012, vol. 391, no. 1–2, pp. 15–20. DOI: 10.1016/j.physa.2011.08.020.
- Spolaore E., Wacziarg R. The Diffusion of Development. *The Quarterly Journal of Economics*, 2009, vol. 124, issue 2, pp. 469–529. DOI: 10.1162/qjec.2009.124.2.469.
- Welsh D. Domestic Politics and Ethnic Conflict. *Ethnic Conflict and International Security*. Edited by M.E. Brown. Princeton: Princeton University Press, 1993, pp. 43–60.
- Zipf G.K. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, 1949, 573 p.