К ВОПРОСУ О ПРИМЕНЕНИИ КОЭФФИЦИЕНТА ДЖИНИ И ДРУГИХ ПОКАЗАТЕЛЕЙ НЕРАВЕНСТВА

К.П. Глущенко

В статье обсуждается проблема, поднятая Г.Л. Громыко и И.Н. Матюхиной в публикации «Об использовании коэффициента Джини в экономико-статистических исследованиях» (журнал «Вопросы статистики», № 9 за 2015 г.), и излагается точка зрения по поводу применения средних, относительных и некоторых других величин для расчета коэффициента Джини.

Как указывает автор, коэффициент Джини лишь один из многих измерителей неравенства, и сказанное относительно коэффициента Джини в равной мере относится и к остальным, близким по содержанию показателям (например, к индексам Тейла, Аткинсона, Херфиналя-Хиршмана, Гувера и т. д.). Автор утверждает, что критика использования коэффициента Джини основана на ошибочных посылках, в первую очередь на неверной трактовке понятия распределения применительно к статистике: оно интерпретируется как способ разделения некоторого блага (или «антиблага») между определенными объектами (отдельными лицами, предприятиями, регионами и т. п.). Однако в статистике понятие распределения подразумевает эмпирическое распределение вероятностей. Кроме того, по мнению автора, в рассматриваемой публикации абсолютизируется понимание коэффициента Джини как «индекса концентрации» и его связь с кривой Лоренца.

В статье приводятся конкретные примеры, иллюстрирующие правомерность использования коэффициента Джини и сходных показателей, рассчитанных на основе средних и относительных величин, для измерения неравенства (дифференциации, неравномерности, различий и т. п.).

Ключевые слова: измерение неравенства, распределение, коэффициент Джини, кривая Лоренца, душевые показатели, группировка данных.

JEL: D31, D63.

Введение

Авторы недавно опубликованной в журнале «Вопросы статистики» статьи [1] о применении коэффициента Джини в экономико-статистических исследованиях приглашали к дискуссии по этому поводу. Как представляется, они ломятся в открытую дверь: в мире существует обширная литература, в которой данный вопрос изучен во всех тонкостях и подробностях. В опубликованном в 1990 г. обзоре [2] говорится о более чем 700 работах по проблематике коэффициента Джини и смежным вопросам (хотя в самом обзоре рассматриваются «только» 385 из них). За следующие 25 лет их число выросло еще больше, вероятно перевалив за тысячу. Если же говорить в целом об измерении неравенства, то соответствующая литература вообще необъятна. Отечественные экономисты и статистики остались в стороне от этих исследований - тема неравенства в советские времена была довольно скользкой, и теоретические изыскания на эту тему не были востребованы[1]. Так что вполне понятно, почему в статье [1] даны ссылки только на учебники и справочные издания по статистике.

Авторы утверждают, что коэффициент Джини «может ... использоваться *лишь для суммарных ... показателей*, которые могут быть реально подвергнуты *разбиению (распределению) по группам*... Что же касается средних показателей, то непосредственно для них, на наш взгляд, коэффициент Джини неприемлем, поскольку средние ... не являются *распределением*» [1, с. 65] (выделения оригинала). Согласно этому утверждению, некорректно оценивать степень межрегионального неравенства по душевым доходам, уровню

Глущенко Константин Павлович (glu@nsu.ru) - д-р экон. наук, ведущий научный сотрудник Института экономики и организации промышленного производства СО РАН (ИЭОПП СО РАН); профессор Новосибирского государственного университета (НГУ) (г. Новосибирск, Россия).

[1] Эмпирические работы, хотя и немногочисленные (и с очень осторожными формулировками), все же имелись. Можно назвать публикации П.П. Маслова, Б.В. Ракитского, Н.М. Римашевской, А.Н. Шапошникова, представителей новосибирской социологической школы и ряд других.

рождаемости, доле пахотных земель, уровню безработицы и т. д. Получается, что нельзя рассчитывать показатель неравенства в оплате труда представителей разных профессий - врачей, учителей, строителей, промышленных рабочих, работников сельского хозяйства и т. п., поскольку при этом неизбежно придется оперировать показателями средней заработной платы. Более того, из приведенной цитаты следует, что не всегда можно использовать коэффициент Джини, даже имея дело не со средними или относительными, а с первичными данными. Например, оценить степень дифференциации жителей страны по возрасту - ведь нет «суммарного возраста» всех жителей России, который был бы «разделен» между ними. То же можно сказать и применительно к росту учеников какой-нибудь школы.

Хотя в статье говорится только о коэффициенте Джини, утверждение авторов шире. Коэффициент Джини - лишь один из широкого класса измерителей неравенства, существо которых одно и то же, и сказанное относительно коэффициента Джини в равной мере относится к любому из них, например к индексам Тейла, Аткинсона, Херфиндаля-Хиршмана, Гувера и т. д.

Побудило авторов к написанию статьи то, «что в последнее время все чаще приходится встречаться с научными работами, в которых коэффициент Джини рассчитывается не для *суммарного показателя*, распределяемого по группам, а для *среднего показателя* или *относительного, рассчитанного на душу населения*» [1, с. 62] (курсив оригинала). Однако такие работы появились отнюдь не «в последнее время». В качестве примера можно привести публикацию 50-летней давности [3], в которой оценивается межрегиональное неравенство в разных странах по данным о доходах на душу населения регионов (правда, с применением коэффициента вариации, а не коэффициента Джини, но это значения не имеет). Такой подход к анализу межстрановых и межрегиональных различий используется в большом количестве исследовательских публикаций (из которых работа [3] далеко не первая) и в обзорах ОЭСР, Всемирного банка и других международных организаций; при этом неравенство оценивается не только по душевым доходам, но и по другим показателям в расчете на душу населения. Но, конечно, довод типа «все так делают, и уже давно» - не аргумент; заблуждения могут быть массовыми и живучими.

Аргумент же состоит в том, что в статье [1], в первую очередь, неверно трактуется понятие распределения применительно к статистике. Из-за этого и ошибочный вывод об области применения коэффициента Джини. Различных точек зрения по этому вопросу в литературе, вопреки утверждению авторов статьи, нет. Нет, собственно, и предмета для дискуссии, тем не менее приходится в нее вступить, чтобы показать, в чем заблуждаются авторы рассматриваемой публикации, поскольку из-за их немалого авторитета положения статьи могут внести смятение в умы исследователей.

Распределение и показатели неравенства

Центральное положение рассматриваемой статьи состоит в следующем: «Сам термин “распределение” применим лишь к объемным (суммарным) показателям, которые действительно могут быть разделены между отдельными группами единиц совокупности, и при этом количественно определена часть этого показателя у отдельных групп (в процентах)» [1, с. 58] (сказанное можно отнести и к самим единицам совокупности, ведь каждая группа может состоять и из одной единицы). Иными словами, распределение понимается в обыденном смысле: как способ дележа некоего блага (или «антиблага») между определенными объектами - отдельными лицами, предприятиями, регионами и т. п. (и, следовательно, нельзя сказать «распределение работников отрасли по возрасту», надо говорить «распределение возраста между работниками», а тогда, согласно приведенной цитате, такое распределение не имеет права на существование). Но статистика-то пользуется иным понятием распределения, подразумевающим эмпирическое распределение вероятностей.

Его можно построить для любой совокупности, состоящей из единиц $i = 1,...,N$, каждая из которых имеет некоторую числовую характеристику - значение признака x_i (будем считать, что единицы упорядочены по возрастанию этого значения: если $i < k$, то $x_i \leq x_k$, что дает вариационный ряд признака). Эмпирическим распределением является дискретное распределение, приписывающее каждому элементу вариационного ряда вероятность $p(x_i) = p_i = 1/N$. При этом природа признака не имеет никакого значения: это может быть, действительно, характеристика самой единицы (численность населения региона), среднее по элементам, которые могут быть выделены в единице (душевой доход в регионе), относительный показатель (доля пахотных земель в общей площади региона), или вообще некоторая величина, полученная любым путем (прожиточный минимум в регионе). Таково общепринятое понимание распределения в статистике (см., например, [4, с. 818]), которое, насколько известно, до сих пор не вызывало никаких возражений[2].

В социально-экономической статистике значение признака чаще всего не является случайной величиной, тем не менее при этом с эмпирическим распределением можно обращаться как с «обычным» распределением вероятностей. Кроме того, эмпирическому распределению легко придать вероятностную трактовку, интерпретируя p_i как вероятность того, что у случайно выбранной единицы значение признака окажется равным x_i. Если единицы объединены в группы с усреднением значения признака по группе (если оно не одинаково), то $p(x_j) = n_j/N$, где x_j - среднее по j-й группе, а n_j - число единиц в ней, не обязательно одинаковое для всех групп. Отсюда легко получить функцию распределения: $F(x_i) = \sum_{k=1}^{i} p_k = i/N$ или для сгруппированных данных - $F(x_j) = \sum_{k=1}^{j} p_k = \sum_{k=1}^{j} n_k / N$.

Разделив весь диапазон значений признака на несколько равных или неравных частей и подсчитав относительные частоты (или, как предпочитают говорить преподаватели статистики, частости) единиц, попадающих в каждый диапазон, получим гистограмму - оценку плотности распределения [4, с. 145]. И все показатели неравенства (а их имеется, наверное, несколько десятков), по сути, измеряют тем или иным способом «ширину» плотности распределения. Каждый из них представляет собой просто некоторую статистику - функцию от результатов наблюдений, каковую (в том числе и коэффициент Джини) можно рассчитать для любого распределения, независимо от способа его построения.

«Показатели неравенства» - наиболее распространенный термин, но их можно назвать и показателями дифференциации, неравномерности, различий, разброса, концентрации, в зависимости от содержания изучаемого явления. Они делятся на два класса. Один включает показатели, учитывающие только часть содержащейся в распределении информации: относительный размах вариации x_N/x_1, коэффициент фондов, межквартильное расстояние и т. п. Другой класс - показатели неравенства, которые используют информацию обо всем распределении, то есть рассчитываются по всем единицам наблюдения (или их группам). К их числу относятся: стандартное отклонение логарифмов значений признака (логарифм используется для того, чтобы сделать показатель неравенства безразмерным), коэффициент вариации, коэффициент Джини, индексы Тейла, Аткинсона, Гувера, Херфиндаля-Хиршмана и др. О них можно также сказать, что они измеряют среднее расстояние между единицами наблюдения на оси признака. А различие между ними обусловлено тем, как определено (в математическом смысле) это расстояние.

Коэффициент Джини[3] - всего лишь «первый среди равных». Его широкая популярность во многом обусловлена традицией. Есть у него и ряд достоинств. Он ограничен сверху, стремясь к единице по мере увеличения числа наблюдений - его максимальное значение равно $(N - 1)/N$, легко интерпретируется, допускает удобное наглядное представление с помощью кривой Лоренца. Но есть и недостатки: в частности, его невозможно декомпонировать, что не позволяет определить вклад той или иной единицы (группы единиц) в общую величину неравенства, или же разделить неравенство на внутригрупповое и межгрупповое (например, внутри федеральных округов и между ними). Индекс Херфиндаля-Хиршмана тоже ограничен сверху, его максимальное значение всегда равно единице, а минимальное равно $1/N$, стремясь к нулю с ростом числа наблюдений. Можно построить и «кривую Херфиндаля-Хиршмана» - аналог кривой Лоренца. Кроме того, он очень просто декомпонируется. Однако этот индекс применяется - опять же, скорее по традиции - при оценке степени концентрации экономической деятельности (хотя для этого можно использовать и другие измерители неравенства). Правда, у него есть свойство, которое зачастую оказывается большим недостатком: он очень сильно реагиру-

[2] Можно дать и строгое определение, однако чтобы не вдаваться в математические тонкости, делать этого не будем (такое определение приведено, например, в [5, с. 94]).

[3] Его обычно называют коэффициентом, когда он выражен в долях единицы, и индексом - когда в процентах. Но похоже, в современной литературе не очень строго придерживаются этого деления, нередко термин «индекс Джини» используют в обоих случаях.

ет на изменения в правой части распределения. Относительно экономического неравенства это означает, что его значение будет сильно меняться при изменении доходов самых богатых, что мало кому интересно (а вот при изучении, например, концентрации фирм в какой-нибудь отрасли такое свойство может быть полезно). Коэффициент Джини в наибольшей степени реагирует на изменения в средней части распределения. Это может оказаться неудобным при исследованиях бедности. В таких случаях более пригоден индекс Аткинсона, содержащий свободный параметр, меняя который можно «регулировать» чувствительность индекса к изменениям в разных частях распределения. Есть свои достоинства и у других измерителей неравенства.

В работе [1] абсолютизируется понимание коэффициента Джини как показателя концентрации, приводящее к выводу, что его расчет по средним и душевым показателям неприемлем, «так как противоречит содержанию показателя Джини как индекса концентрации» и «*средние* величины не характеризуют *распределение* и *концентрацию* (средоточие) исследуемого показателя по группам» [1, с. 62, 63] (курсив оригинала). Но как уже отмечалось, «индекс концентрации» - всего лишь одно из возможных названий, далеко не всегда соответствующее содержанию получаемой величины. Коррадо Джини, действительно, интересовал вопрос концентрации богатства в отдельных группах общества, и поэтому он назвал построенный им показатель «rapporto di concentrazione» - отношение концентрации (в англоязычной литературе - «concentration ratio *R*» или «Gini concentration ratio»[4]). Однако впоследствии было понято, что область применения этого показателя гораздо шире, чем только измерение степени концентрации чего-либо. Коэффициент Джини обладает многими интересными и полезными свойствами, что обусловило его широкое применение в разных, подчас весьма неожиданных, областях[5], где название «индекс концентрации» совершенно не отражает его суть. Так, коэффициент Джини, рассчитанный по душевым доходам в регионах, конечно, измеряет не степень концентрации доходов в регионах, а величину межрегионального неравенства по доходам. А если он рассчитан по ценам некоторого товара в регионах, то это показатель межрегионального разброса (или дифференциации) цен на данный товар[6].

Авторы рассматриваемой публикации исходят из неразрывной связи коэффициента Джини с кривой Лоренца и утверждают, что К. Джини «предложил использовать график Лоренца для оценки социального неравенства» [1, с. 57]. Это отнюдь не так. К. Джини разрабатывал свой измеритель неравенства *R* совсем из других соображений (в своем первоначальном виде он давно не используется) и лишь затем показал его связь с кривой Лоренца, а также со средним абсолютным различием доходов [7]. И хотя формально всегда можно построить кривую Лоренца, соответствующую данному коэффициенту Джини, она не во всех случаях имеет содержательный смысл. Можно, например, оценить с его помощью степень различия роста учеников школы. Необходимый для построения соответствующей кривой Лоренца «суммарный рост школьников» (и тем более «кумулятивная доля роста» в нем) представляется всего лишь странным. Но если коэффициентом Джини измерена межрегиональная дифференциация цены товара, то «суммарная цена» по всем регионам - вообще нонсенс, даже экономисты-первокурсники знают (вернее, должны знать), что цены складывать нельзя.

Применяемая в работе [1] формула коэффициента Джини исходит из геометрических соображений, связанных с графиком Лоренца. Приведем ее в более наглядном варианте, известном как формула Брауна [8]:

$$G = 1 - \sum_{i=1}^{N} p_i \left(\sum_{k=1}^{i-1} \frac{x_k}{X} + \sum_{k=1}^{i} \frac{x_k}{X}\right), \qquad (1)$$

где $X = x_1 + x_2 + \ldots + x_N$.

Это не лучшая формула, она требует утомительного вычисления промежуточных величин, которые могут быть использованы для построения кривой Лоренца (что делается нечасто), но совсем не обязательны для расчета коэффициента Джини. Единственное достоинство этой формулы в

[4] Термин «concentration ratio» без уточняющего определения имеет в настоящее время другой смысл.

[5] Например, в регрессионном анализе (минимизация, вместо суммы квадратов остатков регрессии, коэффициента Джини остатков, что дает так называемую «регрессию Джини»), для построения показателей мобильности по доходам и т. д. [6].

[6] Правда, для характеристики межрегиональных или межстрановых различий цен обычно применяется не коэффициент Джини, а стандартное отклонение логарифмов цен, но это обусловлено всего лишь традицией, сложившейся в данной области исследований.

том, что ее можно непосредственно применять в случае, когда наблюдения сгруппированы в разновеликие группы. Если группы одинаковы или расчет ведется по отдельным единицам, без их объединения в группы, то, напомним, все p_i одинаковы: $p_i = 1/N$ (N - число единиц либо групп), а если группы разновелики, то $p_i = n_i/M$ (тогда N будет представлять число групп, а M - общее количество единиц в рассматриваемой совокупности).

Кроме «геометрических», есть еще три группы формул коэффициента Джини: исходящие из среднего абсолютного различия значений признака, получаемые интегрированием функции распределения и основанные на ковариации значений признаков и их рангов [6, с. 12-13]. Они применимы для расчета непосредственно по всем наблюдениям (единицам совокупности) или по равновеликим группам наблюдений. Можно их использовать и для разновеликих групп, но тогда требуется модификация этих формул - включение в них взвешивающих величин p_i; для разных формул (и даже разновидностей одних и тех же) это делается разным образом.

Формула среднего абсолютного различия значений признака была предложена самим К. Джини. Она имеет вид:

$$G = \frac{\sum_{i=1}^{N}\sum_{k=1}^{N}\left|x_i - x_k\right|}{2N^2\overline{x}}. \qquad (2)$$

Эта формула показывает, что оценка неравенства с помощью коэффициента Джини основана на попарных сопоставлениях всех наблюдений (например, численности населения регионов, душевых доходов в них, средних доходов в группах населения и т. д.). Видно также, что расстояние между единицами (группами) i, k определено как $\left|x_i - x_k\right|$ (а, например, в коэффициенте вариации используется евклидово расстояние). Эта формула не требует никакой дополнительной обработки наблюдений, даже их упорядочения. Некоторое неудобство состоит в громоздкости вычислений: нужно рассчитать $(N-1)N$ разностей.

Использование интегралов функции распределения дает (в дискретном варианте) такую формулу коэффициента Джини:

$$G = \frac{N+1}{N} - \frac{2\sum_{i=1}^{N}(N+1-i)x_i}{N^2\overline{x}}. \qquad (3)$$

Здесь значения признака должны быть упорядочены по возрастанию.

И, наконец, ковариационная формула выглядит таким образом:

$$G = \frac{2\operatorname{cov}(x, r/N)}{\overline{x}}, \qquad (4)$$

где x - вариационный ряд, а r - ранги его элементов: $r_i = i$. Как видно, в эту формулу явным образом «встроена» функция распределения - r/N и есть эта функция: $r_i/N = i/N = F(x_i)$. Выглядит эта формула гораздо элегантнее, чем формула (3), а по трудоемкости вычислений не отличается от нее.

Как ни удивительным это может показаться, но все четыре формулы выводятся одна из другой. Они существуют во множестве разновидностей, но результат расчетов по ним одинаков, поэтому совсем не обязательно приводить «рабочую формулу» (ведь не приводят же в научных публикациях формулу расчета, скажем, дисперсии). Необходимость в этом возникает лишь в случаях, когда эти формулы каким-то образом модифицируются, например вводится взвешивание для учета разновеликости групп, в которые объединены наблюдения.

Статистические примеры

Рассмотрим теперь, к чему приведет следование рекомендациям, высказанным в статье [1]. Допустим, нас интересует дифференциация российских регионов по уровню рождаемости (числу родившихся на 1000 человек населения). Утверждается, что коэффициент Джини по этому показателю неприемлем, так как «признак рассчитан на человека, а распределяются *регионы*» [1, с. 61] (курсив оригинала). Пользуясь данными о численности населения регионов и уровне рождаемости[7], можно вычислить общее число рождений по регионам - тот самый результат «дележа» общего числа родившихся в стране детей между регионами[8]. Но, во-первых, предметом анализа тогда окажется различие вовсе не в уровнях рождаемости, а в общем числе родившихся.

[7] Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 39-40, 50-53.

[8] Здесь и далее, если не сказано иное, исходные данные для расчетов относятся к 2013 г.; расчеты проведены по 83 субъектам Российской Федерации, при этом данные по Архангельской и Тюменской областям - без входящих в них автономных округов (во избежание двойного счета); численность населения - среднегодовая.

Во-вторых, очевидно, что чем больше население региона, тем больше детей в нем рождается. И действительно, коэффициент корреляции между этими величинами равен 0,98 - практически функциональная зависимость. Коэффициент Джини для межрегиональных различий в численности населения составляет 0,448, а для различий в количестве рождений - 0,438, то есть они разнятся на 1 процентный пункт. В то же время коэффициент Джини для дифференциации регионов по уровню рождаемости гораздо меньше: 0,109. На рис. 1 приведены соответствующие кривые Лоренца.

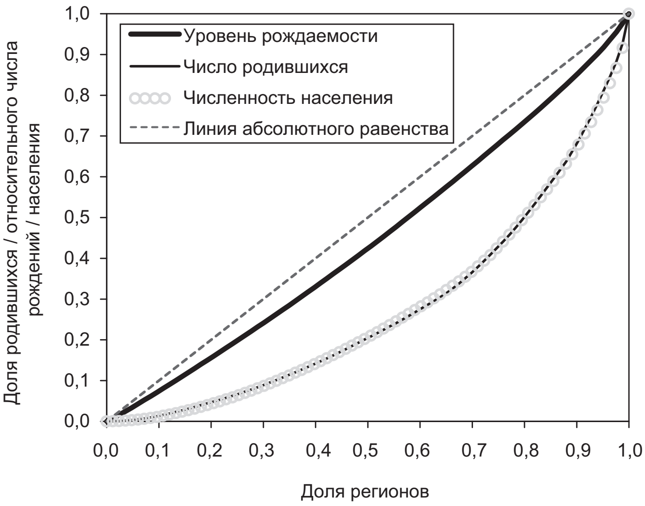

Рис. 1. **Кривые Лоренца для распределения регионов по уровню рождаемости, числу родившихся и численности населения**

Источник: расчеты автора по данным: Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 39-40, 50-53.

Как видно, кривые Лоренца для числа родившихся и численности населения близки настолько, что на глаз почти неразличимы. И какую новую информацию о различиях в рождаемости между регионами нам это дало? Аналогичным образом обстоит дело и с уровнем преступности в регионах. Коэффициент корреляции между численностью населения и общим числом преступлений составляет 0,94; коэффициент Джини по общему числу преступлений равен 0,460 (а по численности населения, напомним, 0,448), тогда как по уровню преступности - 0,191[9]. Кривые Лоренца для численности населения и общего числа преступлений хотя и различаются, но совсем ненамного, опять-таки не давая никакой информации о межрегиональных различиях в уровне преступности. Из-за схожести картины с той, которая приведена на рис. 1, графическую иллюстрацию давать не будем.

Обратимся теперь к межрегиональному неравенству по денежным доходам, используя данные о численности населения регионов и среднедушевых денежных доходах в регионах[10]. Здесь корреляция между численностью населения и общими объемами доходов, полученных жителями регионов, тоже довольно высока: 0,93. Но все же кривые Лоренца для этих двух показателей заметно различаются. Межрегиональное неравенство по общему объему дохода оказывается весьма значительным: коэффициент Джини равен 0,534, тогда как неравенство по доходам на душу населения составляет 0,190. Полученные результаты иллюстрирует рис. 2.

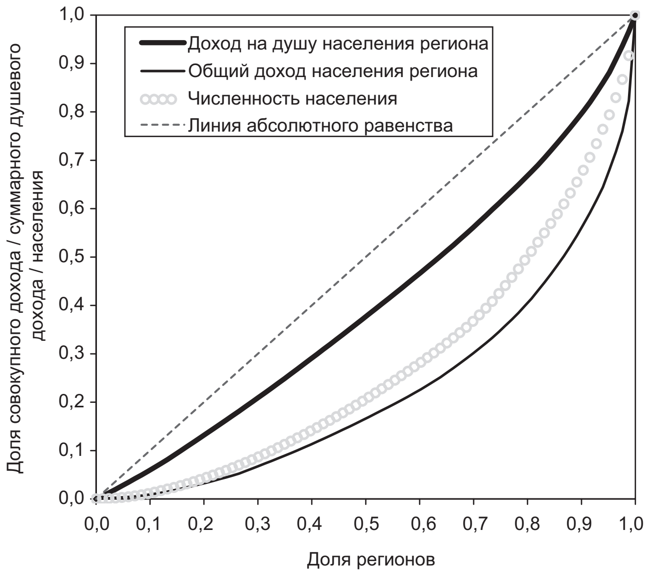

Рис. 2. **Кривые Лоренца для распределения регионов по душевому доходу, совокупному доходу жителей региона и численности населения**

Источник: расчеты автора по данным: Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 39-40, 132-133.

Немало внимания в [1] уделено группировке данных. Она действительно необходима, когда мы имеем дело с очень большим числом наблюдений

[9] Рассчитано по данным: Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 39-40, 326-327.

[10] Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 39-40, 132-133.

(например, всего населения, предприятий региона и т. п.). Однако когда число единиц наблюдения обозримо, как это имеет место в случае регионов, никакой нужды в группировке нет. Более того, она вносит искажения, занижая коэффициент Джини из-за того, что не учитывается внутригрупповая дифференциация. Так, сгруппировав данные о душевых доходах в регионах в соответствии с [1, таблица 3] (с тем отличием, что там автономные округа учтены в составе включающих их областей) и рассчитав коэффициент Джини по сгруппированным данным, получим величину 0,184. Правда, расхождение с точным значением (0,190) оказалось небольшим, составив 3,2%. А вот если сгруппировать регионы по федеральным округам, оно станет существенным: такая группировка дает величину коэффициента Джини, равную 0,106.

Рассмотрим содержательную интерпретацию оценки дифференциации по душевым показателям на примере межрегионального неравенства по доходам. Хотя при этом говорится о регионах, по сути рассматриваются индивидуумы: каждый регион представлен своим «средним» (то есть получающим доход, равный среднему по региону) жителем. И при оценке межрегионального неравенства мы сравниваем между собой представительных жителей разных регионов (подобно тому, как мы берем «среднего» учителя, рабочего и т. д., оценивая неравенство по заработной плате между профессиями, не заботясь о том, каковы доли совокупной заработной платы отдельных профессиональных групп в общем фонде заработной платы рассматриваемой совокупности профессий). Тогда понятно, что можно с полным правом суммировать душевые доходы (это будет совокупный доход всех представительных жителей регионов) и определять долю каждого региона (то есть его представительного жителя) в этом суммарном доходе. Аналогична интерпретация и в случае уровней рождаемости, преступности и т. д.

В качестве примера «некорректного» использования коэффициента Джини в статье [1] рассматриваются анализ неравенства регионов России по прямым иностранным инвестициям (ПИИ) в расчете на душу населения и его методика в информационно-аналитическом докладе [9]. «Трудно понять логику авторов данной методики и доклада», - отмечается в [1, с. 62]. Эта логика такая же, как в приводившихся выше примерах: резонно предположить, что чем крупнее экономика региона, тем, скорее всего, будет больше и объем ПИИ в нем, поэтому необходимо элиминировать фактор величины экономики региона. Ее индикатором может служить численность населения региона[11]. Действительно, величина коэффициента корреляции между численностью населения регионов и поступлением ПИИ в них оказывается немалой, равняясь 0,72 (расчеты относительно ПИИ основаны на данных Росстата[12] и [9, с. 25-26] за 2010 г.). Коэффициент Джини по душевым ПИИ составляет 0,876 (в [9, с. 6] приводится несколько иная величина - 0,872, расхождение обязано двойному счету в [9], где автономные округа учтены дважды - один раз как таковые, а второй - в составе включающих их областей).

Если же следуя [1], рассчитывать неравенство регионов по общему объему ПИИ, то получим заниженную оценку: коэффициент Джини составит 0,803. Расчеты неравенства регионов по общему объему ПИИ, приведенные в [1, с. 61-62], дают еще меньшую величину: 0,753. Причины тут две. Первая состоит в том, что из расчета почему-то исключены восемь регионов, куда ПИИ не поступали [1, табл. 4, 4а, строка 1]. Это, естественно, занизило величину неравенства (как если бы мы при анализе благосостояния населения проигнорировали всех лиц, не имеющих денежных доходов). Учтя эти регионы, получим значение коэффициента Джини, равное 0,777. Но и оно занижено, по второй причине: из-за расчета по сгруппированным данным вместо расчета непосредственно по всем наблюдениям.

Хотя корректность анализа в [9] не дает повода для сомнений (если не говорить о двойном счете, приведшем к небольшим искажениям), результаты относительно ПИИ в [1], в отличие от приведенных выше примеров с рождаемостью и преступностью, все же дают определенное представление о межрегиональных различиях. Но обусловлено это исключительно особенностями рассматриваемого распределения: весьма зна-

[11] Более точную характеристику величины экономики дает валовой региональный продукт (ВРП), и, казалось бы, лучше рассматривать отношения объемов ПИИ к ВРП. Этот показатель приведен в [9], но, на наш взгляд, авторы совершенно правильно не стали вдаваться в дальнейший его анализ, поскольку имеются немалые трудности с обеспечением межрегиональной сопоставимости ВРП из-за различия региональных уровней цен его составляющих [10].

[12] Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 39-40, 326-327.

чительными различиями регионов по объемам ПИИ, далеко не пропорциональным численности их населения, хотя и коррелирующим с ней.

Как было сказано выше, коэффициент Джини лишь один из многих измерителей неравенства. Посмотрим, какие результаты дает применение разных измерителей для анализа одного и того же явления: межрегионального неравенства по (душевым) доходам. На рис. 3 показана динамика этого неравенства, оценки которого получены с помощью коэффициента Джини по формуле (4), коэфициента вариации $V = \sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 / N} / \overline{x}$, индекса Тейла $T = \sum_{i=1}^{N}(x_i / \overline{x})\ln(x_i / \overline{x}) / N$, стандартного отклонения логарифмов доходов $\sigma = \sqrt{\sum_{i=1}^{N}(\ln x_i - \overline{\ln x})^2 / N}$ и индекса Херфиндаля-Хиршмана $H = \sum_{i=1}^{N}(x_i / X)^2 = (V^2 + 1) / N$ (значения последнего на рис. 3а отложены на правой вертикальной оси). Здесь использованы данные по 79 регионам: автономные округа, входящие в другие субъекты Российской Федерации, учитывались в их составе и отдельно не выделялись; исключена Чеченская Республика, данные о доходах в которой имеются только с 2010 г.

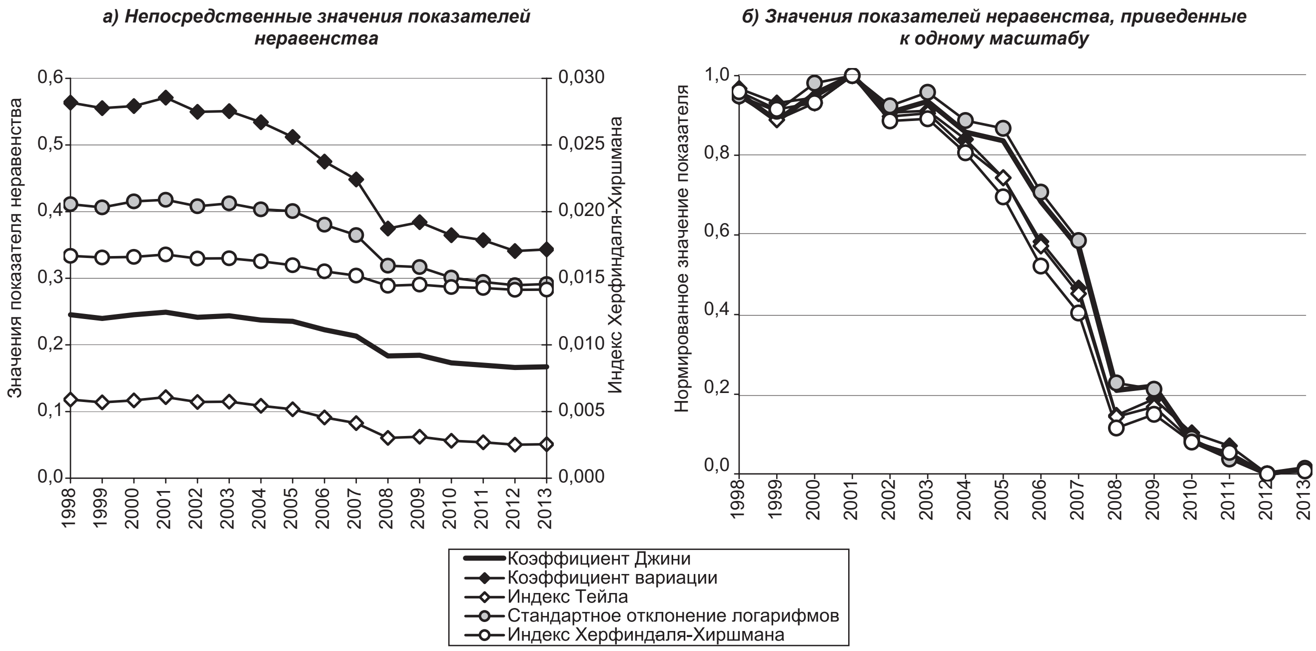

Рис. 3. **Динамика межрегионального неравенства по доходам в России**

Источник: расчеты автора по данным Росстата. URL: www.gks.ru/dbscripts/cbsd/DBInet.cgi?pl=2340019 (дата обращения: август 2015 г.).

На рис. 3а видно, что все пять показателей неравенства дают схожую качественную картину: характер динамики примерно один и тот же. Однако их трудно сравнивать из-за больших различий в значениях: в то время как коэффициент вариации в 1998 г. равен 0,564, величина индекса Херфиндаля-Хиршмана составляет 0,017. Для удобства сравнения показатели неравенства на рис. 3б приведены в сопоставимый вид нормировкой на диапазон их изменения в течение рассматриваемого периода: $Z'_t = (Z_t - \min\{Z_\tau\})/(\max\{Z_\tau\} - \min\{Z_\tau\})$, где Z_t - исходное значение показателя неравенства Z; Z'_t - его нормированная величина. Рис. 3б демонстрирует, что действительно разные показатели дают очень близкую картину динамики. Видны и различия между ними, особенно заметные на промежутке 2004-2009 гг. Они обусловлены разной чувствительностью того или иного показателя неравенства к изменениям в различных частях распределения.

В заключение приведем экзотическую, но тем не менее вполне правомерную оценку межрегионального неравенства - «неравенство по неравенству», которая с точки зрения положений рассматриваемой статьи должна представляться совершенно дикой. Зададимся таким вопросом: какова степень различия российских регионов по величине внутрирегионального неравенства по доходам? Гистограмма соответствующего распределения приведена на рис. 4.

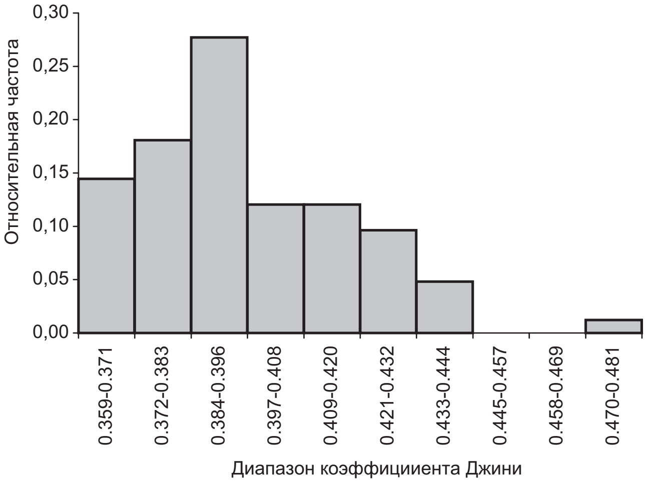

Рис. 4. **Распределение внутрирегионального неравенства по доходам**

Источник: расчеты автора по данным: Регионы России. Социально-экономические показатели. 2014: Стат. сб. / Росстат. - М., 2014. С. 132-133.

Сводную оценку степени различия регионов по внутрирегиональному неравенству дает коэффициент Джини, рассчитанный по коэффициентам Джини, которые характеризуют неравенство по доходам в регионах. Она оказывается низкой: 0,031. Таким образом, хотя неравенство по доходам внутри регионов довольно велико (в диапазоне от 0,359 до 0,481), между собой регионы в этом отношении *в целом* разнятся весьма мало.

* *
*

Критика в работе [1] общепринятого подхода к оценке неравенства (дифференциации, неравномерности и т. п.), в котором используются средние, относительные и другие величины, не являющиеся результатом деления некоторой величины между единицами (группами) наблюдения, исходит из неверных посылок. Это использование понятия распределения, не соответствующего принятому в статистике, абсолютизация понимания коэффициента Джини как «индекса концентрации» и его связи с кривой Лоренца. Таким образом, положения рассматриваемой работы не являются доказательством неправомерности указанного подхода, он является вполне корректным.

Литература

1. **Громыко Г.Л., Матюхина И.Н.** Об использовании коэффициента Джини в экономико-статистических исследованиях // Вопросы статистики. 2015. № 9. С. 56-66.

2. **Giorgi G.M.** Bibliographic portrait of the Gini concentration ratio // METRON - International Journal of Statistics. 1990. Vol. XLVIII. No. 1-4. P. 183-221.

3. **Williamson J.G.** Regional inequality and the process of national development: a description of patterns // Economic Development and Cultural Change. 1965. Vol. 13. No. 4. Part 2. P. 1-84.

4. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. - М.: Большая Российская энциклопедия, 1999.

5. **Cowell F.A.** Measurement of inequality // Handbook of Income Distribution. Vol. 1. Amsterdam: North Holland, 2000. P. 87-166.

6. **Yitzhaki S., Schechtman E.** The Gini methodology. A primer on a statistical methodology. - New York: Springer, 2013.

7. **Forcina A., Giorgi G.M.** Early Gini's contributions to inequality measurement and statistical inference // Journal Electronique d'Histoire des Probabilités et de la Statistique. 2005. Vol. 1. No. 1. P. 1-15.

8. **Brown M.C.** Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data // Social Science & Medicine. 1994. Vol. 38. No. 9. P. 1243-1256.

9. **Краснов С.М., Сайдуллаев Ф.С.** Мониторинг инвестиционной активности в регионах России. Прямые иностранные инвестиции в 2012 году. - М.: АНО «НИСИПП», 2013. URL: nisse.ru/upload/iblock/2fe/invest2012.pdf.

10. **Гранберг А.Г., Зайцева Ю.С.** Межрегиональные сопоставления валового регионального продукта в Российской Федерации: методологические подходы и экспериментальные расчеты // Вопросы статистики. 2003. № 2. С. 3-17.

ON THE ISSUE OF APPLICATION OF THE GINI COEFFICIENT AND OTHER INEQUALITY INDICES

Konstantin Gluschenko

Author affiliation: Institute of Economics and Industrial Engineering, Siberian Branch of the Russian Academy of Sciences, and Novosibirsk State University (Novosibirsk, Russia). E-mail: glu@nsu.ru

This article discusses a problem posed by G.L. Gromyko and I.L. Matyukhina in their publication «On the use of the Gini index in economic and statistical studies» (*Voprosy statistiki*, 2015, no. 9), and states a viewpoint regarding calculation of the Gini coefficient with the use of average, relative, and some other values.

As the author points out, the Gini coefficient is merely one of many inequality measures; thus the statements regarding the Gini coefficient equally well relate to other indices having the same essence (e.g. the Theil index, Atkinson index, Herfindahl-Hirschman index, Hoover index, etc.). The author states that the criticism of the application of the Gini coefficient bases on fallacious premises, first of all, on erroneous interpretation of the notion of distribution in statistics. The distribution is interpreted as a way of dividing a good (with positive or negative utility) between some units (individuals, firms, regions, etc.). However, the statistics uses a different notion of distribution, empirical probability distribution. Besides, in author's sight, the publication under consideration absolutizes the interpretation of the Gini coefficient as 'concentration ratio' and its connection with the Lorentz curve.

The article provides specific examples that illustrate the adequacy of the use of the Gini coefficient and similar indices calculated from average and relative values for measuring inequality (differentiation, unevenness, dispersion, etc.).

Keywords: inequality measurement, distribution, Gini index, Lorentz curve, per capita indicators, pooling of data.
JEL: D31, D63.

References

1. **Gromyko G.L., Matyukhina I.N.** Ob ispolzovaniyi koeffitsienta Dzhini v ekonomiko-statisticheskikh issledovaniyakh [On the use of the Gini index in economic and statistical studies]. *Voprosy statistiki*, 2015, no. 9, pp. 56-66. (In Russ.).
2. **Giorgi G.M.** Bibliographic portrait of the Gini concentration ratio. *METRON - International Journal of Statistics*, 1990, vol. XLVIII, no. 1-4, pp. 183-221.
3. **Williamson J.G.** Regional inequality and the process of national development: a description of patterns. *Economic Development and Cultural Change*, 1965, vol. 13, no. 4, part 2, pp. 1-84.
4. **Prokhorov Yu.V.** (ed. in chief). *Veroyatnost' i matematicheskaya statistika*: *Entsiklopediya* [Probability and mathematical statistics: Encyclopedia]. Moscow, Great Russian Encyclopedia Publ., 1999. (In Russ.).
5. **Cowell F.A.** Measurement of inequality. In: *Handbook of Income Distribution*, vol. 1. Amsterdam, North Holland, 2000, pp. 87-166.
6. **Yitzhaki S., Schechtman E.** *The Gini methodology. A primer on a statistical methodology*. New York, Springer, 2013.
7. **Forcina A., Giorgi G.M.** Early Gini's contributions to inequality measurement and statistical inference. *Journal Electronique d'Histoire des Probabilités et de la Statistique*, 2005, vol. 1, no. 1, pp. 1-15.
8. **Brown M.C.** Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data. *Social Science & Medicine*, 1994, vol. 38, no. 9, pp. 1243-1256.
9. **Krasnov S.M., Saidullaev F.S.** *Monitoring investitsionnoy aktivnosti v regionakh Rossii. Pryamyye inostrannyye investitsii v 2012 godu* [Monitoring of investment activity in the regions of Russia. Foreign direct investments in 2012]. Moscow, NFP «NISSE» Publ., 2013. (In Russ.). Available at: http://nisse.ru/upload/iblock/2fe/invest2012.pdf.
10. **Granberg A.G., Zaitseva Yu.S**. Mezhregioal'nyye sopostavleniya valovogo regional'nogo produkta v Rossiyskoy Federatsii: metodologichesiye podkhody i eksperimantal'nyye raschyoty [Inter-regional comparisons of gross regional product in the Russian Federation: Methodological approaches and experimental computations]. *Voprosy statistiki*, 2003, no. 2, pp. 3-17. (In Russ.).