Дифференциальное отклонение и коэффициент Джини

Павлов О.И., к.ф.-м.н., доцент кафедры экономико-математического моделирования, РУДН, Москва, Россия

Павлова О.Ю., к.ф.-м.н., доцент кафедры высшей математики, статистики и информатики, АТиСО, Москва, Россия

Аннотация. В статье найден диапазон возможных значений дифференциального отклонения (равного определённому интегралу от квадрата производной функции Лоренца) при фиксированном значении коэффициента Джини и наоборот для двухзвенных кривых Лоренца. Найдены параметры эллипса, точки которого являются средними вершинами двухзвенных кривых Лоренца, имеющих одно и то же дифференциальное отклонение. Рассмотрены подходы к дифференциальному аналогу математического определения среднего класса в смысле Геворкяна-Малыхина.

Ключевые слова: показатель социальной стратификации, дифференциальное отклонение, коэффициент Джини, кривая Лоренца, индекс Гувера.

Differential deviation and the Gini coefficient

Pavlov O.I., PhD, Dr. SC. (Phys.-Math.), Associate Professor at the Department of Economic and Mathematical Modelling, RUDN University

Pavlova O.Yu., Dr. SC. (Phys.-Math.), Associate Professor at the Department of Higher Mathematics, Statistics and Computer Science

Annotation. Established is the range of the differential deviation (that is equal to the definite integral of the square of the derivative) given Lorenz coefficient and vice versa

for two-link polygonal curves. Described is the ellipse formed by the middle vertices of two-link polygonal curves having the same differential deviation. Discussed is a differential analog of the middle class mathematical definition due to P.S. Gevorgyan and V.I. Malykhin.

Keywords: income inequality measure, differential deviation, Gini coefficient, Lorenz curve, Hoover index.

Введение. Дифференциальное отклонение и другие показатели социальной стратификации

При исследовании социальной стратификации общества широко используются индексы социального неравенства. Наиболее часто применяют функции и показатели, определённые с помощью функции Лоренца $L(x)$.
Напомним, что эта функция определена на отрезке $[0; 1]$. $L(x)$ равно доле богатства (или дохода), которой обладают самые бедные члены данного общества, составляющие долю x от его численности [1]. График функции Лоренца называется кривой Лоренца (рис. 1).

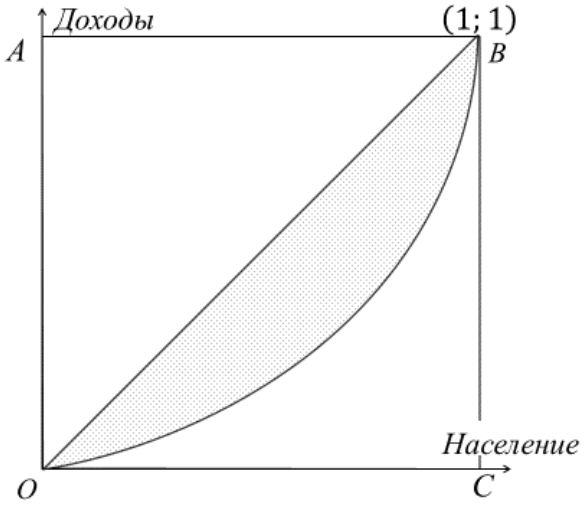

Рис. 1 – Кривая Лоренца

Отношение площади заштрихованной области (см. рис. 1) к половине площади квадрата (равной удвоенной площади заштрихованной области[1]) называется коэффициентом Джини данного общества. Мы будем обозначать коэффициент Джини, соответствующий функции Лоренца $L(x)$, через $J(L)$ или просто J.

Диагональ квадрата соответствует абсолютному равенству в обществе (все члены общества имеют один и тот же доход); чем далее расположена кривая Лоренца от диагонали, тем более неравномерно распределение доходов в обществе.

Например, для функции Лоренца $L(x) = x^2$,

$$J = 2\left(1 - \int_0^1 x^2 dx\right) = 2\left(1 - \frac{x^3}{3}\Big|_0^1\right) = 2\left(1 - \frac{1}{3}\right) = \frac{4}{3}$$

Не претендуя на отображение всех нюансов социальной стратификации, коэффициент Джини позволяет сравнивать равномерность распределения доходов между различными совокупностями (например, странами) или группами населения. Так, по уточнённым данным ФГТС за 2016 год [2], наибольшим неравенством по доходам в РФ отличались Ямало-Ненецкий автономный округ, г. Москва и Республика Башкортостан (значения коэффициента Джини равны 0,427, 0,421 и 0,419 соответственно); в целом по стране коэффициент составил 0,412. Сходная картина наблюдается и по предварительным данным за 2017 год [2].

Другим показателем социальной стратификации является индекс Гувера, который равен максимальной длине вертикального отрезка между кривой Лоренца и линией полного равенства $y = x$.

[1] Некоторые авторы понимают коэффициент Джини просто как площадь области между кривой Лоренца и диагональю квадрата (без удвоения).

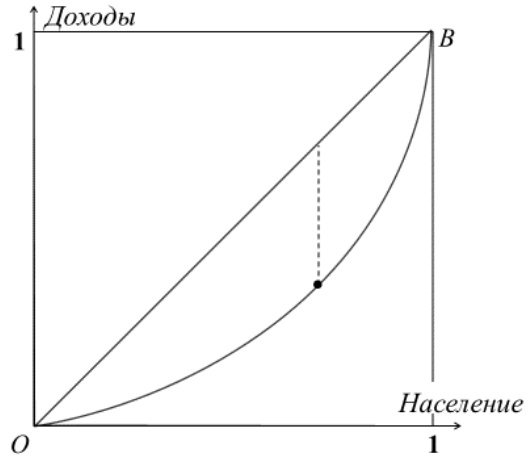

Рис. 2 – Индекс Гувера (индекс Робин Гуда)

Индекс Гувера также называют индексом Робин Гуда, так как он равен той доле дохода общества, которую необходимо перераспределить (передать от более богатой половины общества к более бедной) для достижения полного равенства.

В простейшем случае когда кривая Лоренца является двухзвенной ломаной линией с вершинами в точках $(0; 0)$, $(x; y)$ и $(1; 1)$ (что соответствует делению общества на две группы, в каждой из которых наблюдается полное равенство), коэффициент Джини и индекс Робин Гуда совпадают и равны $x - y$ (см. [3]). В частности, если $y = 0$, то $J = x$.

Так как кривая Лоренца обязательно имеет выпуклость, направленную вниз, производная такой функции является неубывающей на своей области определения. В случае абсолютного равенства в обществе, выполняется $L(x) = x$, следовательно $L'(x) = 1$ на всём отрезке [0; 1]. Если же кривая Лоренца далека от совпадения с диагональю, то неизбежно, она будет содержать участки, на которых производная далека от единицы. Так, возле начала координат кривая Лоренца почти горизонтальна, поэтому её производная близка к нулю, а возле точки $(1; 1)$ кривая Лоренца почти вертикальна, следовательно её производная имеет очень большую величину.

Поэтому, в качестве меры социального неравенства в обществе можно взять любой функционал, который принимает тем большее значение на производной

функции Лоренца, чем более эта производная отличается от константы 1. В качестве такого функционала в [4] был рассмотрен интеграл (2), названый *дифференциальным отклонением*.

$$\int_0^1 \left(L'(x)\right)^2 dx \tag{1}$$

Так как $\int_0^1 \left(L'(x)\right)^2 dx = \int_0^1 (L'(x) - 1)^2 dx + 1$, наименьшее значение дифференциального отклонения достигается на функции Лоренца абсолютного равенства $L(x) = x$ и равно 1. Для любой другой функции Лоренца оно будет строго больше единицы. Будем обозначать дифференциальное отклонение кривой Лоренца $L(x)$ через $D(L)$ или просто D.

В данной заметке мы продолжаем изучать дифференциальное отклонение и его взаимосвязь с коэффициентом Джини и другими индексами социального неравенства.

Возможные значения дифференциального отклонения при фиксированном значении коэффициента Джини

Пусть x_0 обозначает первую координату средней вершины двухзвенной кривой Лоренца. Фиксируем значение a коэффициента Джини, тогда вторая координата равна $x_0 - a$:

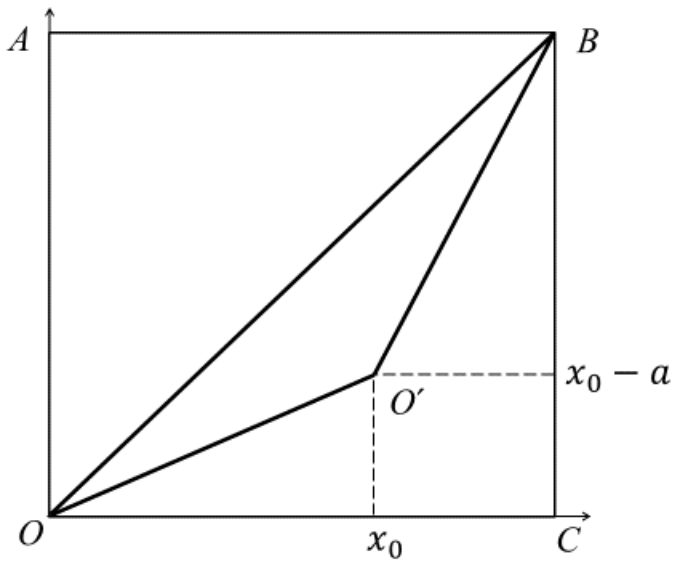

Рис. 3

В [4] показано, что для двухзвенной кривой Лоренца с вершинами $(0;0)$, $(x;y)$ и $(1;1)$, дифференциальное отклонение равно

$$D = \frac{y^2}{x} + \frac{(1-y)^2}{1-x} \tag{3}$$

Заменяя в формуле (3) x и y на x_0 и $x_0 - a$ соответственно, получаем

$$D = \frac{{y_0}^2}{x_0} + \frac{(1-y_0)^2}{1-x_0} = \frac{{y_0}^2(1-x_0) + (1-y_0)^2 x_0}{x_0(1-x_0)} =$$

$$= \frac{{y_0}^2 + x_0 - 2x_0y_0}{x_0(1-x_0)} = \frac{(y_0-x_0)^2 + x_0 - {x_0}^2}{x_0(1-x_0)} = \frac{(y_0-x_0)^2}{x_0(1-x_0)} + 1 =$$

$$= \frac{a^2}{x_0(1-x_0)} + 1 = a^2\left(\frac{1}{x_0} + \frac{1}{1-x_0}\right) + 1$$

Известно (см. [5]), что при $0 < x < 1$, $\frac{1}{x} + \frac{1}{1-x} \geq 4$, причём равенство достигается при $x = \frac{1}{2}$. Поэтому $D \geq 4a^2 + 1 = 4J^2 + 1$. Это неравенство нельзя улучшить при $a \leq \frac{1}{2}$ (так как в этом случае точка $P\left(\frac{1}{2}; \frac{1}{2} - a\right)$ лежит в треугольнике OBC ниже стороны OB; для двухзвенной кривой Лоренца OPB, коэффициент

Джини равен $\frac{1}{2}-\left(\frac{1}{2}-a\right)=a$, в то время как дифференциальное отклонение в точности равно $4a^2+1$ согласно (3)).

Но если $a>\frac{1}{2}$, то минимальная первая координата средней вершины двухзвенной ломаной с $J=a$ есть $x_0=a$ (сама вершина суть $(a;0)$). Поэтому, в силу монотонного возрастания функции $\frac{1}{x}+\frac{1}{1-x}$ на интервале $\left(\frac{1}{2};1\right)$,

$$D=a^2\left(\frac{1}{x_0}+\frac{1}{1-x_0}\right)+1\geq a^2\left(\frac{1}{a}+\frac{1}{1-a}\right)+1=$$

$$=a^2\left(\frac{1}{a(1-a)}\right)+1=\frac{a}{1-a}+1=\frac{1}{1-a}.$$

Следовательно, $D\geq\frac{1}{1-a}=\frac{1}{1-J}$ при $J>\frac{1}{2}$ и эта оценка неулучшаема. Таким образом, доказана

Теорема 1. $D\geq\begin{cases}4J^2+1, \text{если } 0<J\leq\frac{1}{2}\\ \frac{1}{1-J}, \text{если } \frac{1}{2}\leq J<1\end{cases}$.

Отметим, что угловые коэффициенты прямых OO' и $O'B$ (содержащих первое и второе звено ломаной соответственно, и поэтому равные производной функции Лоренца на интервалах $(0;x_0)$ и $(x_0;1)$ соответственно) равны $\frac{x_0-a}{x_0}$ и $\frac{1-x_0+a}{1-x_0}$ соответственно. При $x_0=\frac{1}{2}$ (возможно только если $a\leq\frac{1}{2}$) они равны $1-2a$ и $1+2a$ соответственно. Тогда, после вычитания единиц в формуле

$$D=\int_0^1(L'(x)-1)^2dx+1,$$

получаем, что оба звена ломаной вносят одинаковый вклад в указанный определённый интеграл.

Возможные значения коэффициента Джини при фиксированном значении дифференциального отклонения

Приравняв правую часть формулы (3) к константе K, получим геометрическое место точек, в которых дифференциальное отклонение равно этой константе. В [4] это равенство было приведено к виду

$$Kx^2 - 2xy + y^2 + x(1 - K) = 0 \qquad (4)$$

и показано, что это уравнение задаёт эллипс. Исследуем свойства этого эллипса. Сделаем замену $x = x' + \frac{1}{2}$, $y = y' + \frac{1}{2}$:

$$K\left(x' + \frac{1}{2}\right)^2 - 2\left(x' + \frac{1}{2}\right)\left(y' + \frac{1}{2}\right) + \left(y' + \frac{1}{2}\right)^2 + \left(y' + \frac{1}{2}\right)(1 - K) = 0,$$

откуда, после раскрытия скобок и приведения подобных, получаем уравнение в новых координатах

$$K(x')^2 - 2x'y' + (y')^2 + \frac{1 - K}{4} = 0.$$

Так как точки $(x'; y')$ и $(-x'; -y')$ одновременно принадлежат или не принадлежат эллипсу, делаем вывод, что он симметричен относительно нового центра координат, т.е. точки $\left(\frac{1}{2}; \frac{1}{2}\right)$.

Найдём точки пересечения эллипса с единичным квадратом. Если $x = 0$, то согласно (4), $y^2 = 0$, т.е. $y = 0$. Единственность точки пересечения эллипса с прямой $x = 0$ (являющейся осью ординат) означает касание этой прямой эллипса в точке $(0; 0)$. Из симметрии следует и касание прямой $x = 1$ в точке $(1; 1)$. Так как обе точки касания лежат на диагонали квадрата OB, им соответствуют не двухзвенные ломаные линии, а однозвенная ломаная – сама диагональ OB. Поэтому мы исключим точки касания из дальнейшего рассмотрения и под “эллипсом” будем понимать часть эллипса, лежащую в треугольнике OBC ниже диагонали OB. Эта часть эллипса задаётся явной функцией переменной x, определённой на интервале $(0; 1)$.

При $y = 0$, из (4) следует уравнение $Kx^2 + x(1 - K) = 0$, имеющее ненулевое решение $x = \frac{K-1}{K}$. Получаем точку пересечения $\left(\frac{K-1}{K}; 0\right)$. Симметричной будет точка пересечения $\left(\frac{1}{K}; 1\right)$.

Какая точка эллипса (4) является средней вершиной ломаной с наименьшими и наибольшими значением коэффициента Джини и индекса Гувера? Обозначим эти значения J_{min} и J_{max} соответственно. Эллипс содержит точки, сколь угодно близкие к диагонали OB. У двухзвенных ломаных со средними вершинами в этих точках значения коэффициента Джини и индекса Гувера сколь угодно близки к нулю, но не равны нулю. Поэтому J_{min} не достигается. Для нахождения J_{max} можно использовать геометрический подход. В зависимости от углового коэффициента k касательной к эллипсу в точке пересечения $\left(\frac{K-1}{K}; 0\right)$ (получаются два основных случая: $k > 1$ и $k \leq 1$), точка эллипса $(x; y)$, являющаяся средней вершиной двухзвенной ломаной с наибольшим значением коэффициент Джини, либо совпадает с $\left(\frac{K-1}{K}; 0\right)$, либо является той точкой эллипса, в которой касательная параллельна диагонали OB. Но проще обратить Теорему 1. Функция, определённая в правой части неравенства в условии этой теоремы – монотонно возрастающая на интервале $(0; 1)$ и имеет обратную, определённую на интервале $[1; +\infty)$. Эта обратная функция является неулучшаемой оценкой сверху для значения коэффициента Джини:

Теорема 2. $J \leq \begin{cases} \frac{\sqrt{D-1}}{2}, \text{если } 1 < D < 2 \\ \frac{D-1}{D}, \text{если } D \geq 2 \end{cases}$.

Дифференциальное отклонение и средний класс

П.С. Геворкян и В.И. Малыхин дали в [6] следующее определение: в обществе есть средний класс, если средняя его половина (в смысле богатства) обладает не

менее чем половиной всего богатства общества[2]. Это определение легко формулируется с помощью функции Лоренца. Как наличие среднего класса влияет на показатели социальной стратификации? Этот вопрос рассматривался в публикациях [7], [8], [9] в первую очередь применительно к коэффициенту Джини. Не совсем ясно как сформулировать определение аналогичное среднему классу в терминах дифференциального отклонения. В [4] показано, что неравенству

$$\int_{1/4}^{3/4} (L'(x) - 1)^2 dx \geq \frac{1}{2} \int_{0}^{1} (L'(x) - 1)^2 dx \qquad (5)$$

не удовлетворяет никакая двухзвенная кривая Лоренца, что делает это неравенство не очень пригодным для определения дифференциального отклонения.

Вопрос 1. Удовлетворяет ли неравенству (5) какая-либо кривая Лоренца кроме линии полного равенства $L(x) = x$?

Очевидно, при ответе на Вопрос 1 можно ограничиться кривыми Лоренца для которых распределение доходов внутри наиболее богатой и наиболее бедной четвертях населения является равномерным (равносильно тому, что самая левая и самая правая четверти кривой Лоренца являются отрезками).

Библиографический список

1. Lorenz M. O. Methods of Measuring the Concentration of Wealth // Publications of the American Statistical Association. – 1905. – Vol. 9. N 70. – P. 209-219.

2. Федеральная служба государственной статистики Российской Федерации. [Электронный ресурс]. URL: http://www.gks.ru/ (дата обращения: 25.12.2018).

3. Kämpke T., Radermacher F.J. Income Modeling and Balancing. A Rigorous Treatment of Distribution Patterns // Lecture Notes in Economics and Mathematical Systems. – Vol. 679, – 2015.

[2] Аналогично можно определить средний класс по доходам

4. Павлов О.И., Павлова О.Ю. Индекс социального неравенства, определяемый угловым коэффициентом кривой Лоренца // Управление экономическими системами (электронный журнал). – 2017. – №12. URL: http://uecs.ru/uecs-94-942016/item/4239-2016-12-24-07-45-16

5. Зорич В.А. Математический анализ. Часть I. М.: МЦМНО, – 2017.

6. Геворкян П.С., Малыхин В.И. Распределение богатства в обществе и средний класс // Труд и социальные отношения. – 2010. – № 12. – С. 90-97.

7. А. Н. Асташенко, В. И. Малыхин. Показатели социальной дифференциации // LAP LAMBERT Academic Publishing. – 2012.

8. Павлов О.И., Павлова О.Ю. Коэффициент Джини и математическое определение среднего класса // Инновации в жизнь. – 2016. – № 4. – С. 153-161.

9. Павлов О.И., Павлова О.Ю. Кривая Лоренца и математическое определение среднего класса // Управление экономическими системами (электронный журнал). – 2016. – №12. URL: http://uecs.ru/uecs-94-942016/item/4239-2016-12-24-07-45-16

References

1. Lorenz M. O. Methods of Measuring the Concentration of Wealth // Publications of the American Statistical Association. – 1905. – Vol. 9. N 70. – P. 209-219.

2. Federal State Statistics Service. [electronic resource]. URL: http://www.gks.ru/ (retrived: 25.12.2018).

3. Kämpke T., Radermacher F.J. Income Modeling and Balancing. A Rigorous Treatment of Distribution Patterns // Lecture Notes in Economics and Mathematical Systems. – Vol. 679, – 2015.

4. Pavlov O.I., Pavlova O.Yu. Income inequality measures defined by means of the Lorenz curve slope // Management of economic systems (scientific electronic journal). – 2017. – №12. URL: http://uecs.ru/uecs-94-942016/item/4239-2016-12-24-07-45-16

5. Zorich V.A. Mathematical Analysis. Part I. M.: MCCME, – 2017.

6. Gevorgyan P.S., Malykhin V.I. The distribution of wealth in society and the middle class // Labour and social relations. – 2010. – № 12. – P. 90-97.

7. A.N. Astashenko, V.I. Malykhin. Income inequality measures // LAP LAMBERT Academic Publishing. – 2012.

8. Pavlov O.I., Pavlova O.Yu. The Gini coefficient and a mathematical definition of the middle class // Innovations in life. – 2016. – № 4. – P. 153-161.

9. Pavlov O.I., Pavlova O.Yu. The Lorenz curve and a mathematical definition of the middle class // Management of economic systems (scientific electronic journal). – 2016. – №12. URL: http://uecs.ru/uecs-94-942016/item/4239-2016-12-24-07-45-16