

Институт естественных и гуманитарных наук
Сибирского федерального университета
пр. Свободный, 79, Красноярск, 660000, Россия
E-mail: artyem_s@list.ru

МЕТОДИКА ИССЛЕДОВАНИЯ СТРУКТУРЫ РЫНКА НА ОСНОВЕ НЕПОЛНЫХ ДАННЫХ, НА ПРИМЕРЕ РЕГИОНАЛЬНЫХ КНИГОТОРГОВЫХ РЫНКОВ

В статье проведено исследование книготоргового рынка с помощью метода Ципфа – Парето. Дан обзор состояния современного мирового книготоргового рынка. Исследована деятельность отдельной книготорговой фирмы, регионального рынка и книготоргового рынка РФ в целом. Выявлена интересная особенность формирования покупок «бюджетными» покупателями, что можно трактовать как новую теорию поведения покупателя на рынке. Также в статье содержится оригинальный математический аппарат исследования экстремумов ломанной кривой примененный впервые.

Ключевые слова: книготорговый рынок, метод Ципфа – Парето.

Закон Парето

В 1897 г. итальянский экономист Вильфредо Парето (1848–1923), ученик Леона Вальраса и его приемник на кафедре политэкономии Лозаннского университета, установил некоторую закономерность в распределении доходов в капиталистических странах, а также в странах, где преобладали феодальные и раннекапиталистические отношения. Опираясь на эту закономерность, Парето пытался сделать некоторые общие выводы экономического и социологического характера [3].

На материалах статистики различных стран В. Парето составлял кумулятивные ряды распределения, показывающие, сколько лиц имеет доход не ниже некоторых указанных в рядах величин. Затем он построил графики таких рядов распределения, откладывая на оси абсцисс величины доходов x , а на оси ординат число лиц, имеющих доход, который равен или больше x . В. Парето обнаружил, что в большинстве случаев кривые, отражающие исследуемые им распределения, похожи одна на другую – они являются гиперболоми, уравнение которых может быть выражено формулой $y = A / (x - a)^\beta$, где a – минимальный доход, от которого на графике начинается кривая; A и β – некоторые положительные параметры. Такая кривая, называемая кривой Парето, изображена на рис. 1.

Из уравнения кривой видно, что когда $x \rightarrow \infty$, то $y \rightarrow 0$; а $y \rightarrow \infty$, то $x \rightarrow a$; следовательно, кривая Парето имеет две асимптоты: $x = a$ и $y = 0$. Если ось y перенести в точку P , соответствующую минимальному из рассматриваемых нами доходов, то $a = 0$ и уравнение кривой Парето приобретает вид

$$y = A / x^a, \text{ или } y = Ax^{-a}.$$

Парето проводил наблюдения на статистике разных стран разных периодов: такие распределения он строил по данным по Англии, Пруссии и Саксонии в XIX в., во Флоренции эпохи возрождения, в средневековом Базеле, в Аугсбурге XV–XVI вв., в Перу в конце XVIII в. и т. д., и всегда он получал одни и те же характерные уравнения. Социальные системы менялись, а закон распределения доходов оставался прежним – так утверждал Парето.

Из эмпирически установленной им закономерности В. Парето пытался вывести некий общий социологический закон, который он считал естественным законом, действующим во все времена и при любом общественном строе. Отсюда следовало, что всякие социальные реформы, направленные на устранения в распределении национального дохода, заранее обречены на неудачу, так как естественный закон о распределении доходов действует при любых условиях и доходы будут распределяться в соответствии с приведенной им формулой. В. Парето выразил эту мысль следующим образом: «Поскольку существует тенденция к тому, что доходы среди населения распределяются по определенной форме, все измене-

ния в одной части кривой отразятся на других участках кривой; в конечном итоге общество возвращается к своей нормальной форме так же, как раствор данной соли всегда дает однородные кристаллы» [3].

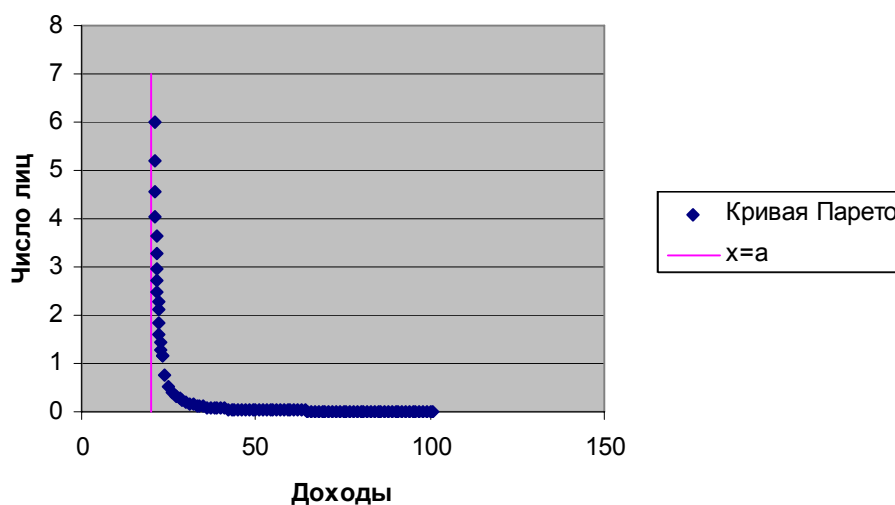


Рис. 1. Распределение Парето

Распределение Ципфа

Подобную зависимость нашел в лингвистике Джордж Кингсли Ципф (1902–1950). Его именем наравне с В. Парето и назван этот закон (закон Ципфа – Парето). Ципф брал выборку, состоящую из k слов, взятых из разных текстов одного автора, и упорядочивал все слова, фигурирующие там, по убыванию их частот. Слово, которое встречается чаще всего, получает ранг 1; слово ранга 2 – это следующее по частоте употребления слово.

Для понимания введем некоторые обозначения: обозначим через $W(r)$ слово, которое в данной классификации имеет ранг r . Продолжаем эту процедуру до тех пор, пока не закончится текст, при этом слово $W(r)$ встретиться нам $i(r, k)$ раз. Частота употребления слова $W(r)$ равна $i(r, k) / k$.

Как выясняется, частота $i(r, k) / k$ обратно пропорциональна рангу r , умноженному на десять, т. е. $i(r, k) / k = 1/10r$.

Необходимо подчеркнуть следующий факт: определение ранга предполагает, что частота $i(r, k) / k$ изменяется в направлении, обратном увеличению r , однако из этого определения никоим образом не следует, что она будет изменяться обратно пропорционально r , – это чисто эмпирическое открытие.

Применение закона Ципфа – Парето

Подобные закономерности впоследствии применялись практически во всех областях науки. Особенное распространение они получили в лингвистике, наукометрии [8; 9; 10], экологии. Выдвигалось множество трактовок этого закона, так, например, многие ученые [5; 6; 7] объясняют это так: данная закономерность носит общий характер и описывает ситуацию «свободной конкурентной борьбы» за распределение конечного количества каких-либо условных «благ». Оказывается, все мыслимое многообразие объектов, ситуаций и причинно-следственных связей не меняет характера этой зависимости: коль скоро имеется свободная конкуренция, ее результаты укладываются на «логарифмическую прямую» – меняются лишь константа и крутизна наклона прямой [4].

Кстати, у Парето коэффициент β (угол наклона прямой) имел значение характеристики общества 1,5. Считалось, что общество находится в равновесии, отклонение вызывало нарушения равновесия и приводило к революциям и прочим потрясениям.

Самое «удачное», на мой взгляд, определение дал известный русский ученый А. И. Яблонский, и ему же принадлежит одно из самых, пожалуй, подробных изучений исследований

этого закона. Вот что он писал: «Структурные особенности закона Ципфа – Парето отражают, по-видимому, глубинные механизмы формирования сложных систем, обуславливающих их устойчивость. Замкнутость систем, описываемых законом Ципфа – Парето, является достаточно универсальным свойством, характеризующим этот закон. Речь идет не только о научных дисциплинах, языковых или экономических системах. Например, при изучении миграции населения в качестве критерия замкнутости географического региона по закону Ципфа – Парето» [8].

Иными словами, «попадание» в распределение Ципфа – Парето говорит о сформированности системы, ее целостности, устойчивости, и, наоборот, если система существенно отклоняется от закона Ципфа – Парето, то она подвержена изменениям, неустойчива.

Исследование книготоргового рынка с помощью метода Ципфа – Парето

С помощью метода Ципфа – Парето было проведено исследование продаж одной фирмы разным группам покупателей: корпоративным и розничным.

Были взяты данные за несколько месяцев 2006 г. и построено на них распределение Парето [1], причем данные о покупках корпоративных клиентов и данные о покупках розничных покупателей были разделены.

Ордината точки распределения – логарифм величины (суммы) покупки одной отдельной взятой книги, например, учебника по философии автора Радугина. Абсцисса – логарифм ранга данной покупки во всей их совокупности за месяц.

Например:

Наименование	Выручка по отгрузке
Основы начертательной геометрии: Учебник для вузов	36 940,17
Практикум по кормлению сельскохозяйственных животных	35 569,80
История ветеринарии: Учеб. пособие	20 421,00

Учебник «Основы начертательной геометрии» был продан за месяц на сумму 36 940,17 – это самая большая сумма покупки, поэтому данная книга (покупка) получает ранг 1, и на график мы наносим точку $(\ln 1; \ln 36940,17)$. Практикум по кормлению сельскохозяйственных животных был продан на сумму 35 569,8, соответственно на график наносится точка $(\ln 2; \ln 35569,8)$ и т. д.

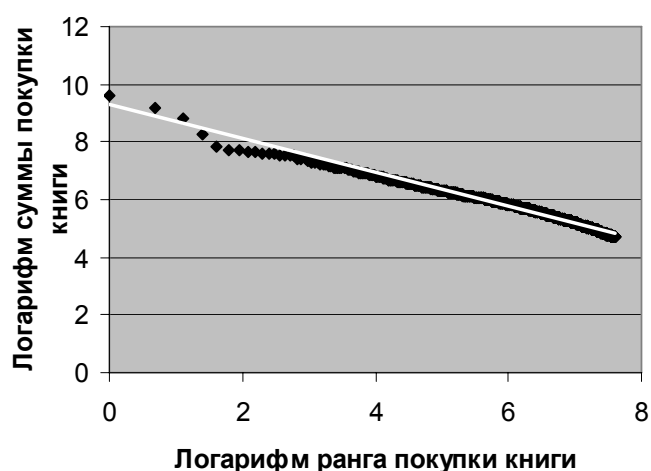


Рис. 2. Распределение Ципфа – Парето по суммам покупок книг

Распределение Ципфа – Парето по суммам покупок книг полностью совпадает с «рациональным» поведением (здесь и далее по тексту под рациональным будет пониматься поведение, которое описывается законом Ципфа – Парето), всегда точно ложится на прямую. В отличие от распределения «корпоративных продаж», которое всегда выглядит характерным образом (рис. 3.), т. е. отклоняется от «рационального» поведения (от прямой).

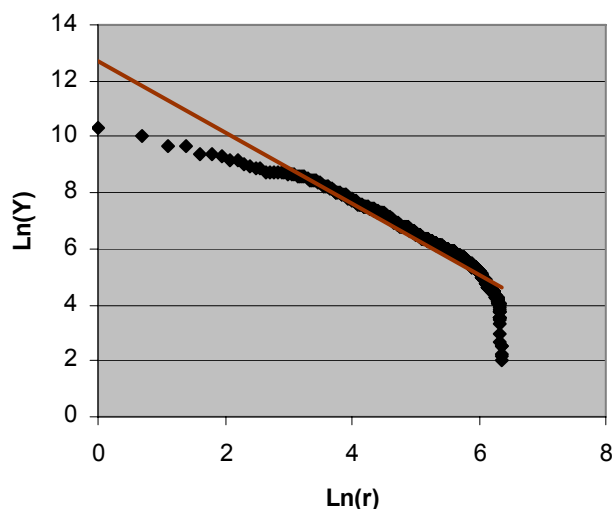


Рис. 3. Диаграмма Циффа – Парето корпоративные продажи

Вывод: розничный потребитель ведет себя на рынке «рационально» по Парето в отличие от корпоративного, поведение которого существенно отклоняется от «рационального» по каким-то причинам.

Причины различий в поведении корпоративных потребителей и розничных понять несложно. Дело в том, что розничный потребитель самостоятельно расходует свои деньги. Покупки же корпоративного потребителя происходят следующим образом: собственником, владельцем данной организации выделяется определенный бюджет и расходуются специальным человеком или отделом, отвечающим за это, т. е. может иметься несоответствие интересов между собственником средств и человеком или группой людей, которые их тратят. Собственник хочет, чтобы эти средства были потрачены с максимальной пользой, у субъекта, который их тратит, имеются, видимо, свои интересы, отсюда и отклонение от «рационального» поведения.

Оказывается, что распределение «корпоративных продаж» по Циффа – Парето более точно описывается не одной прямой, а тремя (рис. 4).

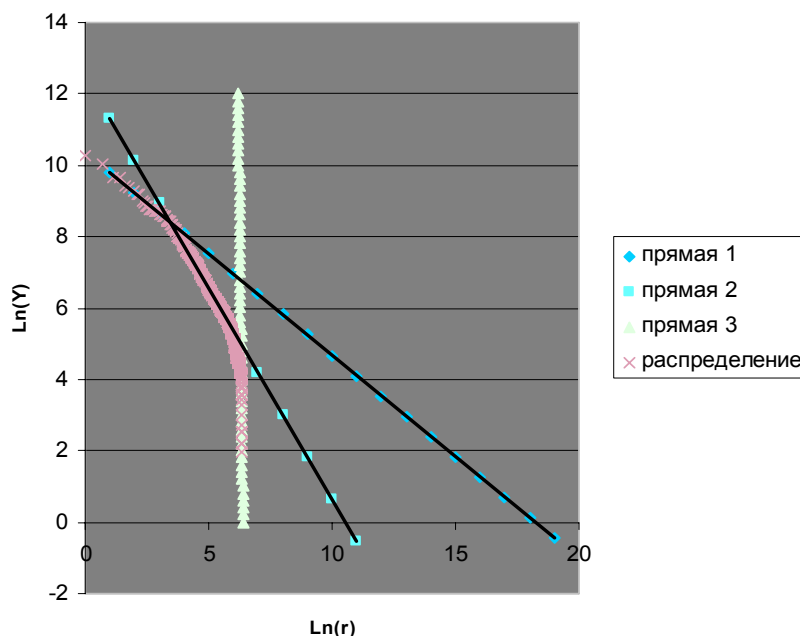


Рис. 4. Разложение распределения Циффа – Парето корпоративных продаж на три прямые

Розовым цветом выделено само распределение и даны три прямые (1, 2, 3), аппроксимирующие соответственно первую, вторую и третью области распределения, если общий коэффициент детерминации равен 0,8, то для каждой из трех областей он более 0,95. Таким образом, наше распределение Ципфа – Парето, построенное на корпоративных продажах, точно раскладывается в три прямые.

И так получается, что наша нецелостная, на первый взгляд, система раскладывается на три сложившихся целостных подсистемы.

Также было выяснено, что три участка распределения соответствуют трем группам товара, в зависимости от популярности: массового использования, более узкого назначения и очень специфичного назначения. Исследование проводилось на книготорговом рынке учебной литературы, поэтому были выделены: базовые учебники, специальные учебники и научные книги.

Библиотеки (а это и есть корпоративные потребители), как видно, по каким-то причинам уделяют большее внимание первой группе книг, которые им надо купить большим количеством и «непропорционально по Парето» большую часть своего бюджета тратят на эти книги, соответственно «непропорционально по Парето» меньшую часть своего бюджета они тратят на специальную литературу с меньшим количеством, и научную книгу финансируют по остаточному принципу.

Покупатель в данном случае действует «нерационально по Парето», но «рационально» со своей точки зрения. Интересно, что подобным образом, ведут себя абсолютно все корпоративные покупатели. Таким образом, можно говорить о новом типе рациональности: везде, где имеется несовпадение средств собственника и человека (группы людей), ответственного за использование этих средств, даже в условиях четкой регламентации действий последних, наблюдается подобное явление.

Как вывод для нас важно следующее – данный феномен характерен не только для книготоргового рынка, так как психология людей, занятых в книготорговой сфере, не отличается от психологии людей, занятых в других сферах. И поэтому подобное искривление будет наблюдаться и на других рынках, где присутствуют государственные закупки.

Теперь расширим наш анализ, исследуем книготорговые рынки регионов РФ и весь рынок РФ в целом.

Объект для анализа – объем печати литературы, с разбивкой по целевому назначению по регионам (субъектам) РФ. Взяты данные по производству литературы в разрезе следующих целевых групп [2]:

- учебные и методические издания;
- литературно-художественные издания;
- издания для детей и юношества;
- издания для широкого круга читателей;
- справочные издания;
- производственно-практические издания;
- научно-популярные издания;
- научные издания;
- официальные издания;
- информационно-рекламные издания.

Мы построим распределение Ципфа – Парето на этих данных по каждому региону. Так как поведение корпоративных (бюджетных) закупок и розничных покупателей нам известно, а на книготорговых рынках нет других потребителей¹, то мы можем ответить на вопрос, каких потребителей больше на рынке данного региона (важно понимать, что обе группы потребителей могут быть и из других регионов), т. е. как будет выглядеть распределение: ляжет ли оно в одну прямую – розничных, или разложится на области – корпоративных.

Для выделения областей используем следующий метод.

Для каждого распределения (всего у нас их 78) вычислим коэффициент детерминации нарастающим итогом. Например, возьмем данные по московской области (табл. 1).

¹ Здесь нужно отметить, что мы делаем допущение об отсутствии экспорта книг, поведение которого еще не изучали.

Таблица 1

Объем печати книг в тыс. экз. в разрезе по целевым группам

Раздел	Количество (тыс. экз)
учебные и методические издания	345,6
религиозные издания	229,9
издания для широкого круга читателей	220
справочные издания	193,6
литературно-художественные издания	153,4
научные издания	140,7
издания для детей и юношества	67,5
производственно-практические издания	54
официальные издания	53
научно-популярные издания	37
информационно-рекламные издания	15,7
нормативно-производственные издания	7

Проранжируем данные, возьмем натуральный логарифм от количества и ранга и получим распределение, где Y – количество; r – ранг (табл. 2).

Таблица 2

Натуральный логарифм от величины объема печати и ранга этой величины

LN(Y)	LN(r)
5,845282037	0
5,437644432	0,693147181
5,393627546	1,098612289
5,265794175	1,386294361
5,033048889	1,609437912
4,946629964	1,791759469
4,212127598	1,945910149
3,988984047	2,079441542
3,970291914	2,197224577
3,610917913	2,302585093
2,753660712	2,397895273
1,945910149	2,48490665

Вычислим коэффициент детерминации линейной функции для первых трех точек, далее для первых четырех и т. д. до всех точек распределения.

Получены следующие результаты: R^2

0,920919
 0,943078
 0,946641
 0,95816
 0,775607
 0,778945
 0,811018
 0,822959
 0,773132
 0,724999

Как мы видим, коэффициент детерминации (далее КД) растет с добавлением точек, но после седьмой точки он резко падает и больше не поднимается до прежнего значения. Это и есть излом на графике распределения. Таким образом, если в динамике КД максимальное

значение не последнее, то выделяются, области «более прямые», чем весь график, которые мы видели на данных одной фирмы.

Продолжаем исследовать наши данные начиная с точки, на которой был максимум КД. Запишем снова наши данные (табл. 3):

Таблица 3

Натуральный логарифм от величины объема печати и ранга этой величины

LN(Y)	LN(r)
4,946629964	1,791759469
4,212127598	1,945910149
3,988984047	2,079441542
3,970291914	2,197224577
3,610917913	2,302585093
2,753660712	2,397895273
1,945910149	2,48490665

И также вычислим КД, начиная с третьей точки.

Коэффициент детерминации нарастающим итогом R^2 : 0,935054
0,835607
0,894083
0,883119
0,872528

Как мы видим, максимальное значение первое, т. е. вторая область состоит из трех точек. Таким образом, КД первой области больше КД всего распределения, так же как КД второй и третьей области:

$$0,95 > 0,72;$$

$$0,93 > 0,72.$$

На графике распределения продаж в московской области выделяются области, которые лучше описываются законом Ципфа – Парето, чем все распределение, что мы и наблюдали на данных одной фирмы.

Вычислим все значения КД общих, первых и вторых и построим графики их распределения.

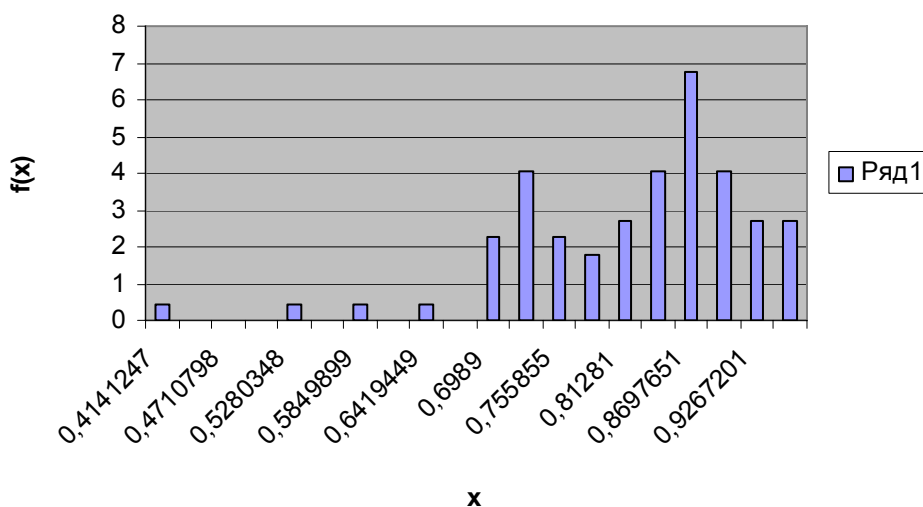


Рис. 4. Распределение общего коэффициента детерминации

На графике видно, что распределение общего КД имеет нормальный характер с центром на значении 0,9.

Поведение КД первой области похоже на экспоненциальное, и самое частое значение близко к единице (рис. 5).

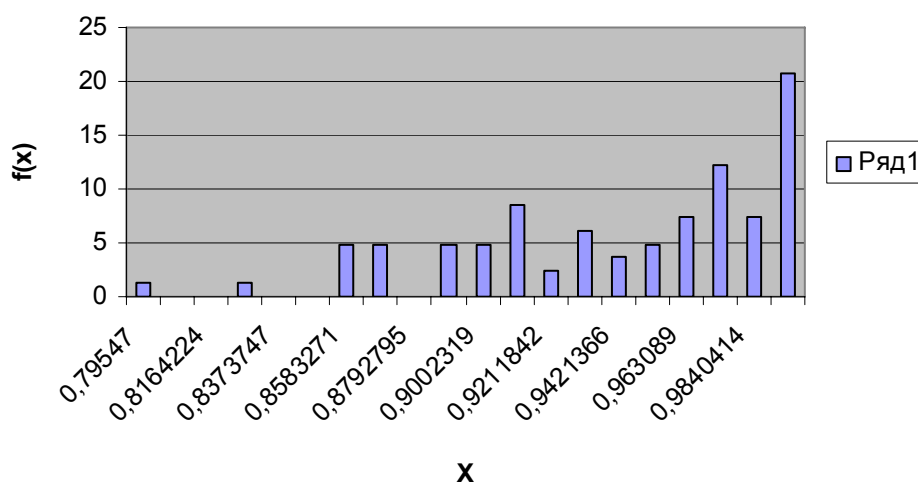


Рис. 5. Распределение коэффициента детерминации первой области

Распределение КД второй области похоже на первое, самое частое значение близко к единице (рис. 6).

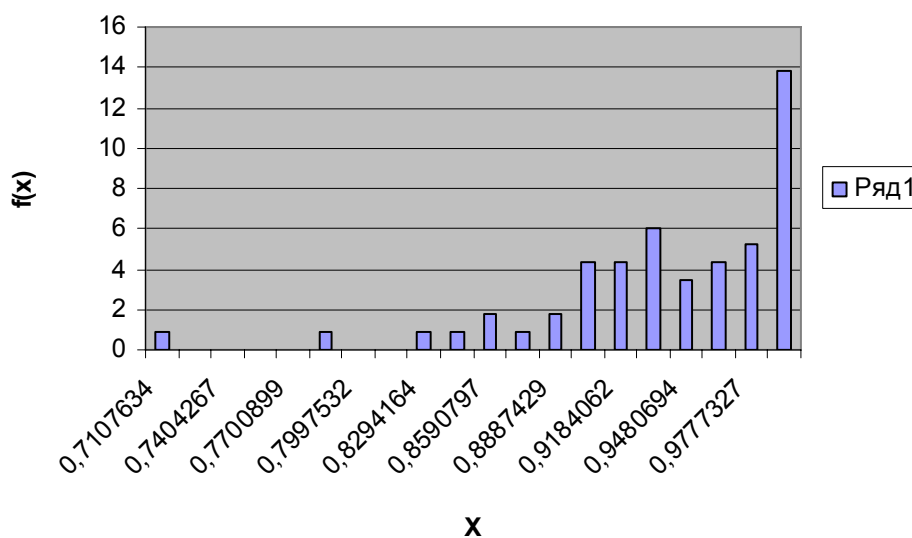


Рис. 6. Распределение коэффициента детерминации второй области

Доказательство значимости коэффициентов детерминации. Для доказательства значимости КД используем критерий Стьюдента: рассчитаем величину $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ и сравним с табличным значением $t_{\alpha, n-2}$, где α – уровень значимости, n – количество наблюдений. Коэффициент детерминации первой, второй области и общий значим при $\alpha = 0,05$.

Доказательство значимости различий коэффициентов прямой. Теперь докажем, что прямые, аппроксимирующие первую, вторую область и все распределение, значимо различаются. Для этого рассмотрим выборки угловых коэффициентов прямых соответственно первой, второй области и прямой, аппроксимирующей все распределение. Используем критерии Вальда – Вольфовица и Манна – Уитни для доказательства неоднородности рассматриваемых совокупностей. Результаты приведены в табл. 4.

Таблица 4

Значение статистических критериев по выборкам значений угловых коэффициентов прямых, аппроксимирующих первую, вторую область и все распределение Ципфа – Парето на данных печати книг по всем регионам РФ

группа 1	группа 2	N	N2	cp1	cp2	Вальд – Вольфовиц	ур. знач.	Манна – Уитни	ур. знач.
область 1	область 2	70	57	-1,22089	-3,63333	5,278103	< 0,001	6,679033	< 0,001
общее	область 1	78	70	-2,08452	-1,22089	1,948267	0,051384	-6,60190	0,000000
общее	область 2	78	57	-2,08452	-3,63333	5,200719	0,000000	5,194395	0,000000

Уровни значимости критериев внушают доверие, поэтому можно сделать вывод – региональные рынки книжной продукции «искривлены» по Ципфа – Парето, что означает превышение бюджетных закупок книг в регионах над розничными. Иными словами, население страны покупает книг меньше, чем бюджет. А следствие этого – бюджетные средства расходуются «нерационально», хотя выделяются значительные суммы на книгообеспеченность библиотек, в то же время уровень образованности (покупки книг гражданами) оставляет желать лучшего. На мой взгляд, правительству пора пересмотреть политику в этом направлении. Как доказательство последнего приведем распределение, построенное на продажах книг по всей РФ (рис. 7).

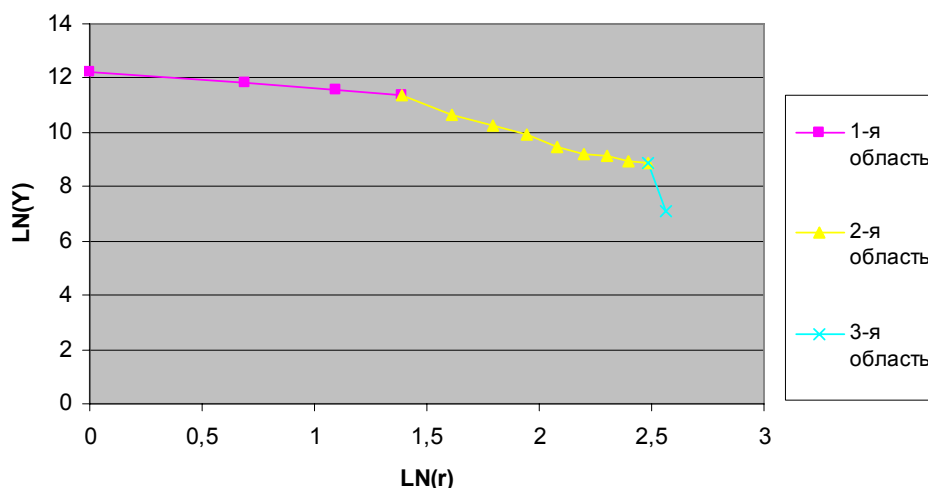


Рис. 7. Распределение Ципфа – Парето на данных печати книг по РФ

Как видно, четко выделяются три области на всем распределении.

Регионы, где не выделяются области. Надо отметить одну важную деталь: количество наблюдений в выборке КД общий 78, а выборке КД первой области уже 70, т. е. есть регионы, где не выделяются области на всём распределении. Они приведены в следующей таблице.

Таблица 5

Регионы РФ, где не выделяются области в распределении Ципфа – Парето по данным печати книг

Область	R^2 общий	T	$t_{\text{табличное}}, \lambda$ – уровень значимости
Магаданская область	0,96	13,79	5,9, при $\lambda = 0,001$
Курганская область	0,82	6,62	4,78, при $\lambda = 0,001$
Костромская область	0,94	10,39	5,9, при $\lambda = 0,001$
Кабардино-Балкарская республика	0,88	6,83	5,9, при $\lambda = 0,001$
Башкортостан	0,94	13,34	4,58, при $\lambda = 0,001$
Чеченская республика	0,85	5,53	4,032, при $\lambda = 0,01$
Карачаево-Черкесская республика	0,86	5	4,6, при $\lambda = 0,01$
Томская область	0,91	7,27	6,85, при $\lambda = 0,001$

Коэффициенты детерминации значимы при достаточном уровне значимости. Это означает, что если в остальных областях рынок искривлен, то в этих восьми розничные покупки превалируют над бюджетными. Стоит внимательно рассмотреть книготорговый рынок этих регионов, чтобы ответить на вопрос, почему так происходит? Ведь данная структура рынка является, на мой взгляд, более предпочтительной для нации и больше отвечает её интересам.

Таким образом, в исследовании разработан метод изучения структуры любого рынка, где присутствуют как розничные, так государственные закупки, на основе данных о совокупном производстве или потреблении какого-либо товара с разбивкой на группы. Это остро актуально для современной России, в период, когда роль государства во всех отраслях народного хозяйства увеличивается, причем в отдельных отраслях государство играет главенствующую роль. С помощью данного метода можно дать ответ на вопрос, «рационально» ли используются бюджетные средства (наличие искривления)? Какова доля закупок бюджета на данном рынке по сравнению с розничными покупателями? Как изменится рыночное распределение потребления товара под воздействием государственных структур? И в какой мере?

В заключение я бы хотел выразить свою благодарность людям, внесшим ценный вклад в написание этой статьи: О. П. Секретенко за помощь в математических исследованиях, В. Г. Суховольскому и Т. М. Овчинниковой за ценные коррективы, Р. Г. Хлебопросу за направление в исследовании.

Список литературы

1. Степанов А. Д., Ноздрина Е. В. Современное состояние рынка литературы Красноярского края. Изучение с помощью метода Ципфа – Парето // Вестн. КрасГАУ. 2007. С. 47–48.
2. Печать Российской Федерации в 2006 году. Статистический сборник. РКП, 2007.
3. Ланге О. Введение в эконометрику. М., 1964. С. 152.
4. Мандельброт Б. Фракталы, случай и финансы. М., 2001. С. 214.
5. Суховольский В. Г. Экономика живого Новосибирск: Наука, 2004. С. 60.
6. Овчинников А. С., Овчинникова Т. М., Суховольский В. Г. Как делятся ресурсы в экономике: анализ распределения фирм по размерам / Препринт 809Ф Красноярск, 2000. 15 с.
7. Khlebopros R. G., Okhonin V. A., Fet A. I. Catastrophes in nature and society. Singapore: World Scientific, 2002. С. 288–292.
8. Хайтун С. Д. Проблемы количественного анализа науки. М., 1989.
9. Яблонский А. И. Модели и методы исследования науки. М.: Прогресс-Традиция, 2001.
10. Яблонский А. И. Структурные закономерности науки. М., 1996.

Материал поступил в редколлегию 24.07.2008

A. D. Stepanov

METHODOLOGY OF STUDY OF MARKET STRUCTURE BASED ON INCOMPLETE INFORMATION, EXAMPLE OF BOOK REGIONAL MARKETS

The article studied booksales market using Zipfa – Pareto. Made review the status of the modern world booksales market. We investigate the activities of individual booksales firms, the regional market and booksales Russian market as a whole. Revealed an interesting feature of the formation of purchases «budgetary» buyers that could be interpreted as a new theory of behavior buyer in the market. The same article contains the original study mathematical apparatus extremes curve used for the first time.

Keywords: book market, method Zipfa – Pareto.