

НЕЛИНЕЙНЫЕ МОДЕЛИ ПАРНОЙ РЕГРЕССИИ В КУРСЕ ЭКОНОМЕТРИКИ

Полежаев В.Д.¹, Полежаева Л.Н.²

¹Московский государственный технический университет им Н.Э. Баумана, Москва, e-mail: vpolej@gmail.com;

²Российский экономический университет имени Г.В. Плеханова, Москва, e-mail: Polezhaeva.LN@rea.ru

Определение вида функции, с помощью которой характеризуется зависимость между признаками, является важным этапом регрессионного анализа. В работе представлен методический подход, позволяющий выбрать лучшую аппроксимирующую функцию из класса монотонных двухпараметрических функций, значение каждой из которых в точке, соответствующей обобщенному среднему по аргументу x , равно обобщенному среднему по переменной y . Среди них присутствуют: линейная, гиперболическая, квадратичная, логарифмическая, показательная и другие функции, наиболее востребованные при обработке статистических данных. В отличие от других наборов эмпирических формул, традиционно предлагаемых в литературе по экономико-математическому моделированию для обработки результатов наблюдений, представленные функции объединены общим свойством, характеризующим логику их объединения для решения поставленной задачи. К тому же эта логика позволяет при необходимости неограниченно расширить предлагаемый перечень формул и показывает направление исследований по их модификации с целью внедрения третьего параметра. При этом учитывается не только статистическая, но и аналитическая информация об исследуемом явлении. Реализация предложенной методики была успешно осуществлена авторами при подготовке магистров по дисциплинам «Математическое моделирование» в МГТУ им. Н.Э. Баумана и «Эконометрика (продвинутый уровень)» в РЭУ им. Г.В. Плеханова.

Ключевые слова: аппроксимация, функция, эмпирическая формула, обобщенное среднее, зависимость, экономико-математическое моделирование.

NONLINEAR PAIRED REGRESSION MODELS IN THE ECONOMETRICS COURSE

Polezhaev V.D.¹, Polezhaeva L.N.²

¹Bauman Moscow State Technical University, Moscow, e-mail: vpolej@gmail.com;

²Plekhanov Russian University of Economics, Moscow, e-mail: Polezhaeva.LN@rea.ru

The determination of the type of function by which the relationship between the characteristics is characterized is an important step in regression analysis. The paper presents a methodical approach that allows choosing the best approximating function from the class of monotone two-parameter functions, the value of each of which at a point corresponding to the generalized mean with respect to the argument x is equal to the generalized mean with respect to the variable y . Among them there are: linear, hyperbolic, quadratic, logarithmic, exponential and other functions, most in demand when processing statistical data. Unlike other sets of empirical formulas traditionally proposed in the literature on economic and mathematical modeling for the processing of observation results, the functions presented are combined by a common property characterizing the logic of their unification for the solution of the task. In addition, this logic allows, if necessary, unlimitedly expand the proposed list of formulas and shows the direction of research on their modification to introduce the third parameter. This takes into account not only statistical, but also analytical information about the phenomenon being investigated. The implementation of the proposed methodology was successfully carried out by the authors in the preparation of masters in the disciplines «Mathematical Modeling» at Bauman Moscow State Technical University and «Econometrics (advanced level)» in the Plekhanov Russian University of Economics.

Keywords: approximation, function, empirical formula, the average generalized dependence, economic and mathematical modeling.

При исследовании различных технических проблем экономических и социальных явлений невозможно обойтись без использования математических методов создания моделей. Моделирование позволяет создать надежную основу для управления различными процессами с помощью компьютерных технологий.

При изучении экономических и технических предметов в высшей школе в той или

иной мере применяются знания, полученные при изучении элементарной и высшей математики. В частности, при создании динамических моделей социальных и экономических задач зачастую приходится находить аналитическую связь между переменными по заданным статистическим данным, другими словами – аппроксимировать некоторую функцию, для которой известны лишь отдельные значения при определенных значениях аргумента.

Решение указанной задачи состоит из двух важных этапов. Сначала устанавливается вид эмпирической формулы из некоторого заранее определенного класса функций, а затем с помощью известных методов математической статистики находят ее параметры. Удачно подобранный вид формулы зависимости между исследуемыми признаками является залогом создания модели, адекватной исследуемому явлению. При этом анализ механизма изучаемой зависимости, основанный на познании сущности ее природы, является главным основанием для выбора вида модели. Поскольку реальные социально-экономические процессы не отличаются простотой, а экзогенные переменные, от которых они зависят, зачастую взаимосвязаны, то на теоретическом уровне вид зависимости эндогенной переменной от отдельных факторов не всегда представляется возможным установить.

Цель исследования. В методическом обеспечении преподавания разделов эконометрики одной из недостаточно разработанных проблем подачи материала является вопрос о способах выбора эмпирической формулы. В большинстве источников с этой целью зачастую предлагаются ничем не обоснованные наборы функций, которые, по мнению их составителей, наиболее востребованы в практике обработки результатов наблюдений. Целью настоящего исследования является разработка методически обоснованного подхода к выбору вида уравнения регрессии из определенного класса монотонных гладких функций. В качестве теоретической основы в данной работе использованы принципы выбора общего вида функции регрессии, предложенные С.А. Айвазяном [1, с. 66]:

Качественное проведенное выявление класса функции, которые будут образовывать множество допустимых решений задачи, среди которых будет осуществляться поиск регрессионных зависимостей, закладывает предпосылки к дальнейшему получению качественной модели процесса. Однако эта часть регрессионного анализа недостаточно теоретически разработана и обоснована. При определении эмпирической формулы, используемой в качестве функции регрессии, в первую очередь необходимо изучить глубинную сущность моделируемой зависимости с целью использования известных априорных данных о механизме функционирования исследуемого явления. Неоценимую помощь в этом может оказать геометрический анализ имеющихся статистических данных. Для этого надо рассечь многомерную гиперповерхность, которую предполагается построить по заданному набору статистических данных, гиперплоскостями уровня. При этом будут

получаться парные корреляционные поля, по виду которых можно делать выводы о монотонности и порядке «гладкости» частных функций отклика, наличии асимптот, возможности линеаризации функций с помощью замены переменных и некоторых других их свойствах.

Для проверки гипотез о виде предполагаемой эмпирической формулы, полученной с помощью содержательного и геометрического анализа процесса, можно использовать методы, разработанные в математической статистике для проверки гипотез.

Материал и методы исследования. К сожалению, в учебной литературе по эконометрике зачастую присутствует формальный подход. Сначала предлагается научить студентов построению моделей, а лишь затем провести проверку полученных результатов на соответствие имеющейся априорной информации об исследуемом в процессе. В частности, в ряде учебных пособий при построении нелинейной парной регрессии по заданному набору экспериментальных точек предлагается не задумываясь применять встроенную в MS Excel функцию построения линии тренда, хотя заведомо известно, что по экономическим соображениям некоторые из них применять не следует.

При такой методической реализации преподавания регрессионного анализа у студентов зачастую складывается мнение о том, что лучшую модель, аппроксимирующую заданный набор точек, можно найти, если использовать многочлен высокой степени. Но если экономический или физический смысл решаемой задачи предполагает монотонность процесса, то очевидно, это приведет к получению неверных результатов, хотя по критерию минимизации погрешностей полиномиальная модель дает лучший результат. Так, например, не имеет смысла на большом интервале изменения переменных использовать квадратичную функцию (параболу) для моделирования функции спроса на изделие от его цены, поскольку эта зависимость строго монотонная и к тому же теоретически обладает асимптотой.

Сущностный анализ рассматриваемого явления должен являться основой при выборе вида эмпирической формулы, используемой при построении модели [2, 3]. Для этого надо учитывать не только основные теоретические соображения о характере процесса и свойствах его факторов, но и воздействие других экономических, технологических и социальных процессов, которые могут повлиять на функционирование моделируемого процесса.

При восстановлении эконометрических зависимостей часто приходится учитывать не только то, что они должны быть монотонными и достаточно «гладкими», но и некоторые другие их качественные характеристики. В частности, экономические соображения могут диктовать необходимость того, чтобы график искомой функции не имел точек перегиба. В настоящей работе представлен подход, с помощью которого можно определять наиболее подходящий вид эмпирической формулы для установления регрессионной зависимости

между двумя переменными, учитывающей указанную особенность [4, 5].

Свойство обобщенного среднего для монотонных двухпараметрических функций

Будем использовать в качестве эмпирических формул функции вида $y = f(x, a, b)$, зависящие от двух параметров:

$$y = (ax^p + b)^{1/q}, \quad \text{если } p \neq 0, q \neq 0; \quad (1)$$

$$y = (a \ln x + b)^{1/q}, \quad \text{если } q \neq 0; \quad (2)$$

$$y = e^{ax^p + b}, \quad \text{если } p \neq 0; \quad (3)$$

$$y = bx^a. \quad (4)$$

Эти функции (1–4) обладают общим свойством, связанным с так называемым обобщенным средним положительных чисел z_1, z_2, \dots, z_n

$$M_z(t) = \left(\frac{1}{n} \sum_{k=1}^n z_k^t \right)^{1/t}. \quad (5)$$

Средние гармоническое H_z (6), геометрическое G_z (7), арифметическое A_z (8) и квадратическое S_z (9) представляют собой частные случаи обобщенного среднего (5):

$$H_z = \left(\frac{1}{n} \left(\frac{1}{z_1} + \frac{1}{z_2} + \dots + \frac{1}{z_n} \right) \right)^{-1} = M_z(-1), \quad (6)$$

$$G_z = (z_1 \cdot z_2 \cdot \dots \cdot z_n)^{1/n} = \lim_{t \rightarrow 0} M_z(t) = M_z(0), \quad (7)$$

$$A_z = \frac{z_1 + z_2 + \dots + z_n}{n} = M_z(1), \quad (8)$$

$$S_z = \sqrt{\frac{z_1^2 + z_2^2 + \dots + z_n^2}{n}} = M_z(2). \quad (9)$$

При этом $\min z \leq H_z \leq G_z \leq A_z \leq S_z \leq \max z$.

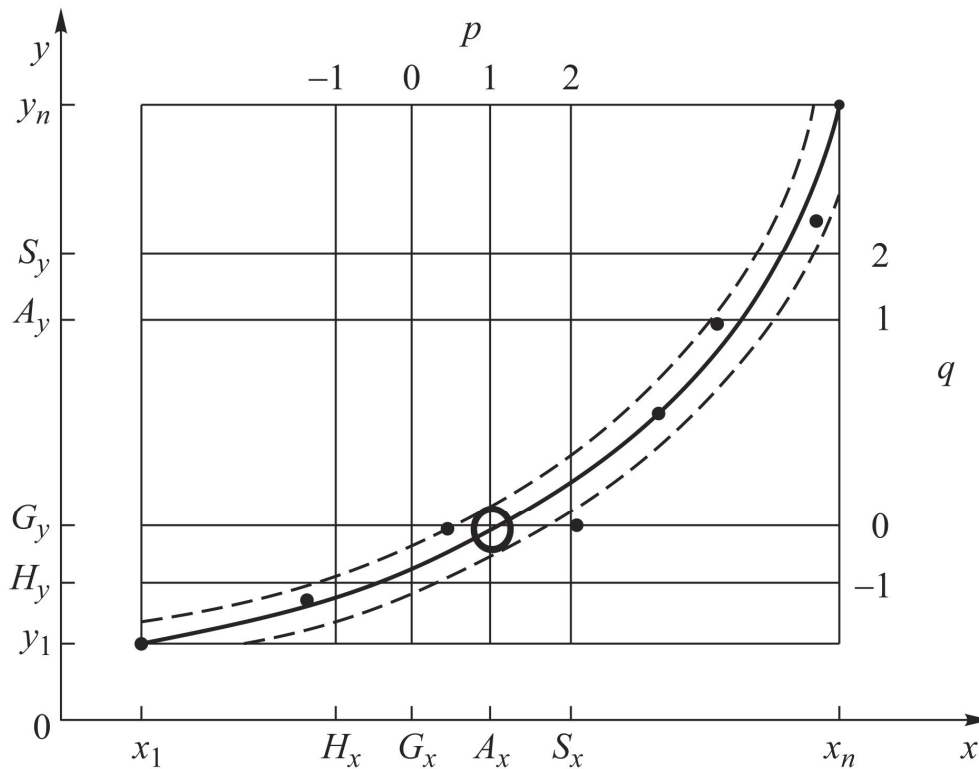
Авторами выявлено общее свойство функций (1–4):

$$f(M_x(p)) = M_y(q). \quad (10)$$

К примеру, если p и q взять равными нулю, то получим функцию $y = bx^a$, для которой имеет место равенство $f(M_x(0)) = M_y(0)$ или по-другому $f(G_x) = G_y$.

Беря p и q , равные $-1; 0; 1; 2$, получим 16 монотонных функций, среди которых, в частности, есть: линейная, квадратичная, гиперболическая, показательная, логарифмическая, степенная и другие функции, с которыми студентам приходится иметь дело при изучении эконометрики. Пусть на координатную плоскость (рис. 1) нанесены результаты измерений $K_i(x_i, y_i)$ ($i = 1, 2, \dots, n$). Под графиком искомой функциональной зависимости понимается кривая, проходящая через середину полосы разброса экспериментальных данных (рис. 1).

Предполагается, что построенная кривая отображает тенденцию, присущую ряду данных, и при этом является свободной от незначительных колебаний, обусловленных погрешностями измерений. Возьмем две крайние точки $K_1(x_1, y_1)$ и $K_n(x_n, y_n)$ кривой, полагая их достаточно надежными. Вычислим средние гармоническое $H_x = 2x_1x_n / (x_1 + x_n)$, геометрическое $G_x = \sqrt{x_1x_n}$, арифметическое $A_x = (x_1 + x_n) / 2$, квадратическое $S_x = \sqrt{(x_1^2 + x_n^2) / 2}$. Подсчитаем средние H_y, G_y, A_y, S_y .



Исследование моделируемой зависимости

Найденные значения отметим на соответствующих координатных осях и проведем через них прямые, которые пересекутся в шестнадцать точках, соответствующих формулам таблицы.

Двупараметрические эмпирические функции, соответствующие им средние и
линеаризующие преобразования

№ П/П	$y = f(x, a, b)$	p	q	$M_x(p)$	$M_y(q)$	x'	y'	a'	b'
1	$ax + b$	1	1	A_x	A_y	x	y	a	b
2	ba^x	1	0	A_x	G_y	x	$\ln y$	$\ln a$	$\ln b$
3	$1/(ax + b)$	1	-1	A_x	H_y	x	$1/y$	a	b
4	$a \ln x + b$	0	1	G_x	A_y	$\ln x$	y	a	b
5	$b + a/x$	-1	1	H_x	A_y	$1/x$	y	a	b
6	bx^a	0	0	G_x	G_y	$\ln x$	$\ln y$	a	$\ln b$
7	$x/(bx + a)$	-1	-1	H_x	H_y	$1/x$	$1/y$	a	b
8	$ba^{1/x}$	-1	0	H_x	G_y	$1/x$	$\ln y$	$\ln a$	$\ln b$
9	$1/(a \ln x + b)$	0	-1	G_x	H_y	$\ln x$	$1/y$	a	b
10	$ax^2 + b$	2	1	S_x	A_y	x^2	y	a	b
11	ba^{x^2}	2	0	S_x	G_y	x^2	$\ln y$	$\ln a$	$\ln b$
12	$1/(ax^2 + b)$	2	-1	S_x	H_y	x^2	$1/y$	a	b
13	$\sqrt{ax^2 + b}$	2	2	S_x	S_y	x^2	y^2	a	b
14	$\sqrt{ax + b}$	1	2	A_x	S_y	x	y^2	a	b
15	$\sqrt{a \ln x + b}$	0	2	G_x	S_y	$\ln x$	y^2	a	b
16	$\sqrt{a/x + b}$	-1	2	H_x	S_y	$1/x$	y^2	a	b

В качестве эмпирической формулы искомой аналитической зависимости следует выбирать ту, которая соответствует ближайшей к графику точке пересечения прямых. Так, на рисунке 1 кривая проходит через точку пересечения прямых, соответствующую показательной функции $y = ba^x$ (№ 2 в табл. 1).

Если таких точек пересечения прямых оказалось несколько, то надо дополнительно проверить, насколько удовлетворяют соответствующие им формулы априорным сведениям, а также теоретическим предпосылкам об искомой зависимости. Легко доказывается, что все представленные функции являются линеаризуемыми, поэтому для них методом наименьших квадратов можно найти значения параметров a , b и общую погрешность измерений. Тогда в качестве лучшей следует выбирать функцию, отвечающую минимальной сумме мер отклонений.

Аппроксимирующие аналитические выражения должны как можно более точно описывать реальные показатели. Однако повышение точности аппроксимации приводит, как правило, к усложнению аппроксимирующих функций, что затрудняет как определение значений входящих в эти формулы параметров, так и использование этих выражений для анализа явлений. Если вычисления по построенной модели значительно отличаются от имеющихся данных, то применяемые двухпараметрические эмпирические формулы можно дополнить еще одним параметром. Его можно включить в формулу, используя дополнительные дифференциально-геометрические условия, предъявляемые к восстанавливаемой зависимости. В частности, в качестве третьего параметра можно использовать p или q .

Представленный набор двухпараметрических формул может быть неограниченно увеличен путем придания p и q любых числовых значений. Решение вопроса о том, при каких p и q эмпирическая зависимость наилучшим образом будет аппроксимировать имеющиеся экспериментальные данные, зависит от спецификации исследуемой модели. Если же p или q известно, то другой параметр можно определить приближенно графически или более точно, решая неявное уравнение.

Результаты исследования и их обсуждение

Для успешного освоения эконометрики обучающиеся должны иметь запас знаний по основным понятиям и методам математической статистики: цели и методы сбора статистических данных, первичная обработка статистических данных, основные характеристики случайных величин, основные статистические распределения, таблицы распределений и их использование, получение доверительных интервалов, проверка статистических гипотез. Все темы должны быть закреплены решением практических задач, желательно на реальных статистических данных.

Вместе с разработкой соответствующего методического обеспечения [6–8] необходимо решать проблему внедрения информационных технологий в процесс обучения эконометрике, поскольку решение научно-практических задач обусловлено многочисленными расчетами, статистической обработкой и анализом данных. Поэтому овладение методологическими понятиями эконометрики на качественно новом уровне возможно только с применением информационных технологий [9–11].

Современные компьютерные программы обработки данных позволяют строить уравнения регрессий для отдельных функций рассмотренного класса и выбирать из них ту, для которой остаточная дисперсия и ошибка аппроксимации минимальны, а коэффициент детерминации максимален. Однако для получения обоснованно лучшего результата исследователю следует добавить отсутствующие двухпараметрические формулы из

указанного выше перечня в предлагаемый программой имеющийся комплект регрессионных моделей.

При этом важно донести до студентов тот факт, что каждая модель есть конгломерат объективной информации, субъективных сведений, практикуемых подходов, а также математических действий. В конце концов, любая математическая модель конкретного экономического объекта, даже построенная на базе общепризнанных положений экономической теории, содержит существенную долю субъективных решений. Поэтому и рекомендации, основанные на расчетах с использованием таких моделей, не могут быть вполне объективными.

Следует отметить, что в эконометрической литературе, а также в известных пакетах прикладных программ до сих пор отсутствует логически обоснованный набор эмпирических формул. Так, в справочнике [12] представлены 7 из 16 рассмотренных функций с алгоритмом выбора лучшей из них для аппроксимации данных. Но при этом не объясняется, каким образом был определен этот набор. Предложенная методика дает ответ на поставленный вопрос и позволяет более системно подходить к вопросам изучения разделов эконометрики и математического моделирования, связанных с использованием нелинейных моделей парной регрессии.

Заключение

Одной из основных задач регрессионного анализа является нахождение вида аналитической зависимости между переменными, когда известны отдельные значения эндогенной переменной при заданных значениях экзогенных признаков. Во многих методических пособиях по эконометрике для нахождения лучшего уравнения парной регрессии предлагается с учетом возросших возможностей компьютерной техники использовать некоторый набор аппроксимирующих функций, выбирая из них ту, для которой отклонение от имеющихся данных будет минимальным. Такой подход не всегда является оптимальным по причине того, что выбранная функция может и не удовлетворять экономическим или дифференциально-геометрическим требованиям к модели процесса (или просто отсутствовать в используемом наборе).

В работе представлен методический подход, позволяющий выбрать лучшую аппроксимирующую функцию из класса монотонных двухпараметрических функций, значение каждой из которых в точке, соответствующей обобщенному среднему по аргументу x , равно обобщенному среднему по переменной y . В отличие от других наборов эмпирических формул, традиционно предлагаемых в литературе по экономико-математическому моделированию для обработки результатов наблюдений, представленные функции объединены общим свойством, характеризующим логику их объединения для

решения поставленной задачи. К тому же эта логика позволяет при необходимости неограниченно расширить предлагаемый перечень формул и показывает направление исследований по их модификации с целью внедрения третьего параметра. При этом учитывается не только статистическая, но и аналитическая информация об исследуемом явлении.

Реализация предложенной методики была успешно осуществлена авторами при подготовке магистров по дисциплинам «Математическое моделирование» в МГТУ им. Н.Э. Баумана и «Эконометрика (продвинутый уровень)» в РЭУ им. Г.В. Плеханова. С уверенностью можно заявить о необходимости внедрения представленной методики в учебный процесс и обновления соответствующих разделов многочисленных учебных пособий, в которых представлены материалы о выборе вида уравнения регрессии из класса гладких монотонных функций.

Список литературы

1. Айвазян С.А. Методы эконометрики: Учебник / С.А. Айвазян; Московская школа экономики МГУ им. М.В. Ломоносова (МШЭ). – М.: Магистр: ИНФРА-М, 2010. – 512 с.
2. Клейнер Г. Б. Экономика. Моделирование. Математика. Избранные труды / Г.Б. Клейнер. Российская академия наук, Центральный экономико-математич. ин-т. – М. : ЦЭМИ РАН, 2016. – 856 с.
3. Берндт Э.Р. Практика эконометрики : классика и современность / Эрнст Р. Берндт ; науч. ред. и предисл. С. А. Айвазяна / Пер. с англ. Е. Н. Лукаша. – М. : ЮНИТИ, 2005. – 847 с.
4. Полежаев В.Д. Методы и модели в экономике: конспект лекций / В.Д. Полежаев, Л.Н. Полежаева, Е.Н. Казанцева. – Омск: Изд-во ОмГТУ, 2008. – 65 с.
5. Полежаев В.Д. Обоснование выбора вида функций, применяемых для аппроксимации данных в экономико-математическом моделировании / В.Д. Полежаев, Л.Н. Полежаева // VI Международная научно-практическая конференция имени А.И. Китова «Информационные технологии и математические методы в экономике и управлении (ИТиММ-2016). 26–27 мая 2016 г. : сборник научных статей. – М.: ФГБОУ ВО «РЭУ им. Г.В. Плеханова», 2016. – С. 18–25
6. Тихомиров Н.П. Научная школа «Повышение качества разработки и использования математического инструментария в решении проблем анализа прогнозирования и управления социально-экономическими процессами» // Вестник Российского экономического университета им. Г.В. Плеханова. 2007. – № 1. – С. 47–53.

7. Елисеева И.И. Статистическое образование: цели и условия // Формирование основных направлений развития современной статистики и эконометрики. Материалы I Международной научной конференции. Научный редактор В.Н. Афанасьев. 2013. – С. 10–13.
8. Мисюра В.В. Методологические проблемы преподавания эконометрики // Строительство – 2015: Современные проблемы строительства: материалы международной научно-практической конференции. – Ростов-на-Дону: Ростовский государственный строительный университет, 2015. – С. 412–414.
9. Бывшев В.А., Богомолов А.И., Костюнин В.И. Математико-статистическое обоснование оценки знаний студента при тестировании // Бизнес информ. 2012. – № 7. – С. 186–188.
10. Игнаткина Л.А., Перстенёва Н.П., Репина Е.Г. Особенности методики преподавания и оценки знаний студентов по дисциплине «Эконометрика» в Самарском государственном экономическом университете // Современные проблемы науки и образования. – 2012. – № 4; URL: <http://science-education.ru/ru/article/view?id=6551> (дата обращения: 23.04.2018).
11. Use of information and communication technologies for teaching physics at the technical university / Polezhaev V.D., Polezhaeva L.N., Kamenev V.V. // AIP Conference Proceedings "Information Technologies in Education of the XXI Century, ITE-XXI 2015: Proceedings of the International Scientific-Practical Conference "Information Technologies in Education of the XXI Century" 2017. С. 030013. DOI: 10.1063/1.4972452
12. Справочник по математике для экономистов / В.Е. Барбаумов, В.И. Ермаков, Н.Н. Кривенцова и др. / Под ред. В. И. Ермакова. 3-е изд., перераб. и доп. – М.: Инфра-М, 2009. – 464 с.