

ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ
И
ПРОЦЕССЫ УПРАВЛЕНИЯ
N 3, 2011
Электронный журнал,
рег. Эл. N ФС77-39410 от 15.04.2010
ISSN 1817-2172

<http://www.math.spbu.ru/diffjournal>
e-mail: jodiff@mail.ru

Моделирование динамических систем

УДК 519.65

МОДЕЛИРОВАНИЕ ПРОЦЕССОВ РАСПРЕДЕЛЕНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ РЕСУРСОВ С ПОМОЩЬЮ ВЕРОЯТНОСТНЫХ ЦЕПОЧЕК

Е.В. Афанасьева

Россия, Санкт-Петербург, e-mail: ekaterina.v.afanasyeva@gmail.com

Аннотация

В данной статье описаны результаты применения нелинейных вероятностных цепочек двух видов (с логистическим ростом и логарифмически-линейных) в моделировании процессов распределения социально-экономических ресурсов. Исследована область применимости указанных вероятностных цепочек, и на ряде примеров показан нетривиальный характер поведения асимптотики логарифмически-линейной динамики.

Введение

Одним из методов анализа социально-экономических процессов является математическое моделирование. Первые попытки построения математических моделей в данной области относятся к 17–18 вв. Уже в начале 20 в. применение математического аппарата закрепилось в экономике, а позднее мате-

математические модели стали использоваться и в других общественных науках, включая социологию.

В настоящее время исследователи выделяют следующие методы, которые применялись на протяжении нескольких последних десятков лет при моделировании социально-экономических процессов: балансовые, методы исследования операций, эконометрические [11].

Балансовый метод, получивший широкое распространение в СССР для прогнозирования и планирования экономических процессов, основывается на выявлении взаимосвязей между ресурсами и их потреблением. Методы исследования операций в общественных науках часто были связаны с решением экстремальных задач по планированию, когда необходимо сочетать принципы оптимальности, баланса и эффективного использования имеющихся ресурсов. Эконометрические методы, которые находят широкое применение и по сей день, основаны на построении статистических моделей.

Все перечисленные методы объединяет такое свойство, как линейность (или сводимость к линейности), которое, несмотря на простоту получаемых моделей, не позволяет адекватно отражать сложные нестационарные процессы, которые преобладают в современном обществе.

Среди новых методов моделирования социально-экономических процессов можно выделить методы, основанные на нечеткой логике, а также на теории нейронных сетей. Они дают возможность с относительно небольшими затратами строить гибкие модели, отражающие реальность с ее неопределенностями и неточностями.

Особняком стоят методы теории нелинейных динамических систем. Историю данной теории принято связывать с именами таких ученых, как А. Пуанкаре, А.М. Ляпунов, Л.И. Мандельштам, А.А. Андронов, А.Н. Колмогоров, Э. Лоренц, В.И. Арнольд и др. [6]. Первоначально зародившиеся в естественных и точных областях знания, методы нелинейной динамики за последние несколько десятилетий распространились и на сферу общественных и гуманитарных наук.

Предметом указанной теории являются динамические системы, для которых процесс перехода из одного состояния в другое описывается с помощью нелинейных моделей. Для систем с непрерывным временем такие модели строятся на основе дифференциальных уравнений, а для систем с дискретным временем — на основе разностных уравнений.

Согласно принципам нелинейной динамики, системы могут быть неустой-

чивыми по отношению к малым отклонениям. Это проявляется в колебательных и хаотических режимах поведения при движении в областях действия странных аттракторов [15].

В научном сообществе теория нелинейных динамических систем в применении к анализу социально-экономических процессов находит как сторонников, так и противников. С одной стороны, ученые все больше убеждаются в том, что большинство процессов в живой и неживой природе, включая социальные системы, являются нелинейными. Так, Ю.А. Данилов пишет: «Среди множества почетных титулов, которые принес нашему веку прогресс науки, «век нелинейности» — один из наименее звучных, но наиболее значимых и заслуженных...» [8]. С другой стороны, попытки перенести методы естественных наук, в том числе и методы нелинейной динамики, в сферу социального и гуманитарного знания часто подвергаются критике. Однако не подлежит сомнению тот факт, что внедрение указанных методов в область общественных наук уже произошло, поэтому речь должна идти скорее о повышении качества и корректности исследований в данной сфере [5].

Одним из методов теории нелинейных динамических систем для моделирования социально-экономических процессов являются нелинейные вероятностные цепочки. Этот математический аппарат, предложенный М. Сонисом, может использоваться для оценки динамики процессов распределения ресурсов в обществе. Среди нелинейных вероятностных цепочек особенно выделяют цепочки двух видов: с логистическим ростом и логарифмически-линейные — получившие наиболее широкое применение как методы нелинейной динамики в социально-экономических исследованиях [23]. Именно эти два вида нелинейных вероятностных цепочек и используются в данной работе.

Обзор программных продуктов для построения социально-экономических моделей

Прежде всего, стоит отметить, что в ходе исследования программных продуктов, позволяющих строить социально-экономические модели, не было найдено ни одного, который работал бы непосредственно с нелинейными вероятностными цепочками, описанными в предыдущем разделе. Однако существует немало пакетов, которые могут использоваться на отдельных этапах построения и оценивания моделей на основе вероятностных цепочек. Их можно разделить на две большие категории: статистические и эконометрические пакеты, а также пакеты для математического и статистического моделиро-

вания.

Среди программных продуктов и вычислительных сред первой категории можно выделить SPSS, Statistica, Eviews, STATA, Gauss, Microfit, MicroTSP, SYSTAT, MINITAB, STADIA, среды SAS и R и др. [9].

К программным продуктам второй категории можно отнести Maple, Matlab, Mathcad, Mathematica и др.

Хотя ни одна из описанных выше систем не позволяет осуществлять все необходимые операции по работе с нелинейными вероятностными цепочками, статистические пакеты могут использоваться для проверки статистических гипотез об адекватности моделей на основе логарифмически-линейной вероятностной цепочки, эконометрические пакеты — для оценки систем уравнений (с использованием, однако, предварительно обработанных данных), а пакеты для математического моделирования — для построения интерполяции и экстраполяции эмпирической динамики, а также для оценивания коэффициентов вероятностной цепочки с логистическим ростом. Без сомнения, все это делает актуальным вопрос о разработке специализированного приложения для работы с нелинейными вероятностными цепочками, которое объединяло бы в себе некоторые возможности перечисленных пакетов, но в то же время не было перегружено сложными статистическими и математическими инструментами, а содержало бы только необходимые модули для исследования моделей на основе вероятностных цепочек.

1 Нелинейные вероятностные цепочки

1.1 Определение вероятностной цепочки

Понятие нелинейной вероятностной цепочки естественным образом возникает при изучении различных социально-экономических процессов, связанных, например, с распределением доходов по группам населения или численности работающих по отраслям экономики. При этом приходится иметь дело со статистическими таблицами численных характеристик, представляющих долю некоторой группы от общего числа. Таким образом, мы работаем с набором величин p_i , обладающих следующими свойствами:

1. $p_i \geq 0$, $i = 1, \dots, N$, где N — число групп, составляющих таблицу;

$$2. \sum_{i=1}^N p_i = 1.$$

Подобная совокупность p_i называется вероятностным вектором. В каждый из моментов времени, в который производятся подсчеты (например, по годам) динамика изменений характеристик группы выражается своим вероятностным вектором, поэтому процесс, иллюстрируемый этими данными, можно рассматривать как некое преобразование вероятностных векторов. Задание этого преобразования в явном виде делает возможным построение некоторой модели изучаемого процесса.

Теория нелинейных дискретных вероятностных цепочек (то есть последовательностей, образуемых вероятностными векторами) позволяет получать статистические оценки относительной динамики многих социально-экономических процессов.

В данной работе представлены модели некоторых из этих процессов, построенные с применением вероятностных цепочек. Дадим определение нелинейной вероятностной цепочки, следуя [32], где приведены основные свойства вероятностных цепочек и социально-экономическая интерпретация этого понятия.

Определение 1 *Дискретная вероятностная $(1, n)$ -цепочка — это последовательность векторов вида*

$$\bar{p}_t = \begin{pmatrix} p_{1t} \\ p_{2t} \\ \vdots \\ p_{nt} \end{pmatrix}, t = 0, 1, 2, \dots \quad (1)$$

$$0 \leq p_{kt} \leq 1, \sum_{k=1}^n p_{kt} = 1.$$

Обозначение $(1, n)$ можно понимать следующим образом: изучается поведение одной популяции на n территориях, или же распределение ресурсов между n территориями или между n популяциями, и т.д.. В дальнейшем в этой главе будем рассматривать, не умаляя общности, вероятностные цепочки на примере интерпретации задачи о распределении популяции между n территориями.

Любую нелинейную вероятностную цепочку можно представить набором строго положительных порождающих функций — преобразований над вероятностными векторами. Так, i -ый элемент вероятностного вектора в момент времени t будет иметь вид

$$p_{k,t+1} = \frac{P_k(p_0, \dots, p_t)}{\sum_{i=1}^n P_i(p_0, \dots, p_t)}, \quad (2)$$

где $P_k(p_0, \dots, p_t)$ — строго положительная порождающая функция, и $k = 1, \dots, n$, $t = 0, 1, \dots$

В зависимости от порождающих функций, которыми они задаются, выделяют различные виды вероятностных цепочек. В данной работе используются два вида нелинейных вероятностных цепочек: с логистическим ростом и логарифмически-линейные.

Вероятностные цепочки с логистическим ростом определяются набором порождающих функций

$$P_k(p) = \gamma_k p_k, \quad (3)$$

где $\gamma_k > 0$, $k = 1, 2, \dots, n$ — скорость прироста или снижения доли популяции на территории k , и задаются формулой

$$p_{k,t+1} = \frac{\gamma_k p_{kt}}{\sum_{s=1}^n \gamma_s p_{st}}, \quad k = 1, 2, \dots, n, \quad t = 0, 1, 2, \dots, T, \quad (4)$$

а T определяет длину временного интервала, на протяжении которого рассматривается поведение популяции.

Логарифмически-линейные вероятностные цепочки задаются порождающими функциями Кобба-Дугласа¹:

$$P_k(p) = A_k p_1^{a_{k1}} p_2^{a_{k2}} \dots p_n^{a_{kn}}, \quad -\infty \leq a_{kj} \leq +\infty, \quad A_k > 0, \quad k = 1, \dots, n, \quad (5)$$

где A_1, \dots, A_n — суммарные преимущества территорий $1, \dots, n$, а элементы матрицы $[a_{ij}]_{i,j=1}^n$ отражают суммарные характеристики «гибкости» относительного прироста долей популяции на данных территориях. Так, например, положительное значение коэффициента a_{ij} говорит о том, что увеличение доли популяции на j -ой территории влечет за собой увеличения доли популяции

¹Функция Кобба-Дугласа — экономическая модель, отражающая зависимость объема производства от создающих его факторов производства: затрат труда и капитала [7].

на i -ой территории, и наоборот. Отрицательное значение свидетельствует о том, что, напротив, при увеличении доли популяции на j -ой территории доля популяции на i -ой территории снижается, и наоборот. В литературе, где описывается применение этой модели, такие отношения между группами i и j называются соответственно дополняющими и конкурирующими [32].

Таким образом, логарифмически-линейные вероятностные цепочки могут быть представлены в следующем виде:

$$p_{k,t+1} = \frac{A_k p_{1,t}^{a_{k1}} p_{2,t}^{a_{k2}} \dots p_{n,t}^{a_{kn}}}{\sum_{s=1}^n A_s p_{1,t}^{a_{s1}} p_{2,t}^{a_{s2}} \dots p_{n,t}^{a_{sn}}}, \quad (6)$$

где $k = 1, \dots, n$, $t = 0, 1, 2, \dots, T$, и T несет тот же смысл, что и в (4).

Опишем процесс интерполяции и экстраполяции динамики вероятностных цепочек с логистическим ростом и логарифмически-линейных вероятностных цепочек на примере распределения популяции между n территориями.

1.2 Построение моделей на основе вероятностных цепочек

1.2.1 Вероятностные цепочки с логистическим ростом: построение и оценивание модели

Система (4) имеет решение вида

$$p_{k,t} = \frac{p_{k0} \gamma_k^t}{\sum_{s=1}^n p_{s0} \gamma_s^t}, \quad k = 1, 2, \dots, n, \quad t = 0, 1, 2, \dots, T. \quad (7)$$

Кроме того, систему (4) можно переписать следующим образом:

$$p_{k,t+1} - p_{kt} = p_{k,t+1} \sum_{s=1}^n \left(1 - \frac{\gamma_s}{\gamma_k} \right) p_{st}, \quad k = 1, 2, \dots, n, \quad t = 0, 1, 2, \dots, T. \quad (8)$$

Заметим, что выражение $\frac{p_{k,t+1} - p_{kt}}{p_{k,t+1}}$, которое обозначает относительный прирост доли популяции p_{kt} , представляет собой линейную комбинацию всех долей популяции. Следовательно, коэффициенты $a_{ks} = 1 - \frac{\gamma_s}{\gamma_k}$ можно интерпретировать как косвенное влияние доли популяции на территории s на

относительное приращение доли популяции на территории k . Такие коэффициенты образуют матрицу взаимного влияния для вероятностной цепочки с логистическим ростом и описывают образцы территориального соревнования за статистическую популяцию.

При оценке динамики поведения популяций удобно рассматривать все доли популяции по отношению к одной, для которой скорость прироста фиксируется и, не умаляя общности, считается равной единице. Такая доля популяции (или, в зависимости от поставленной задачи, доля доходов, потребляемых ресурсов и пр.) в литературе называется *numeraire* [20, 26] (стандарт).

Примем за стандарт первую долю популяции и зафиксируем для нее $\gamma_1 = 1$. Тогда формулу (4) можно переписать в виде

$$z_{kt} = \frac{p_{kt}}{p_{1t}} = z_{k0} \gamma_k^t, \quad k = 2, \dots, n, t = 0, 1, 2, \dots, T \quad (9)$$

или

$$z_{k,t+1} = \gamma_k z_{kt}, \quad k = 2, \dots, n, t = 0, 1, 2, \dots, T. \quad (10)$$

Пусть имеются эмпирические данные о распределении популяции по n территориям за $T + 1$ лет:

$$\begin{aligned} & \tilde{p}_{1t}, \tilde{p}_{2t}, \dots, \tilde{p}_{nt}, \quad t = 0, 1, 2, \dots, T, \\ & 0 \leq \tilde{p}_{kt} \leq 1, \quad k = 1, 2, \dots, n, \quad \sum_{k=1}^n \tilde{p}_{kt} = 1. \end{aligned} \quad (11)$$

Эти эмпирические данные образуют следующую динамику:

$$\tilde{z}_{kt} = \frac{\tilde{p}_{kt}}{\tilde{p}_{1t}}, \quad k = 2, \dots, n, t = 0, 1, 2, \dots, T. \quad (12)$$

Чтобы представить эту эмпирическую динамику с помощью вероятностных цепочек с логистическим ростом, т.е. приблизить динамику (9) к данным (12), необходимо минимизировать среднеквадратическую ошибку. Для получения оценки относительного прироста долей популяции нужно свести к минимуму корень среднеквадратической ошибки

$$s_1 = \left(\frac{1}{T(n-1)} \sum_{t=0}^{T-1} \sum_{k=2}^n (\tilde{z}_{k,t+1} - \hat{\gamma}_k \tilde{z}_{k,t})^2 \right)^{1/2}, \quad (13)$$

а для вычисления приближенного начального значения z_{k0} , $k = 2, 3, \dots, n$ — корень среднеквадратической ошибки

$$s_2 = \left(\frac{1}{T(n-1)} \sum_{t=0}^{T-1} \sum_{k=2}^n (\tilde{z}_{k,t+1} - \gamma_k^t \hat{z}_{k0})^2 \right)^{1/2}, \quad (14)$$

В результате могут быть получены оценки скорости изменения долей популяции

$$\hat{\gamma}_k = \frac{\sum_{t=0}^{T-1} \tilde{z}_{k,t+1} \tilde{z}_{kt}}{\sum_{t=0}^{T-1} \tilde{z}_{kt}^2}, \quad k = 2, 3, \dots, n, \quad \gamma_1 = 1 \quad (15)$$

и начальных значений долей популяции на территориях $2, 3, \dots, n$

$$\hat{z}_{k0} = \frac{1 - \hat{\gamma}_k^2}{1 - \hat{\gamma}_k^{2T}} \sum_{t=0}^T \hat{\gamma}_k^t \tilde{z}_{kt}, \quad k = 2, 3, \dots, n. \quad (16)$$

Зная z_{k0} , $k = 2, 3, \dots, n$, можно выразить начальное состояние системы в терминах долей популяции $p_{10}, p_{20}, \dots, p_{n0}$:

$$\begin{cases} \hat{p}_{10} = \frac{1}{n + \sum_{k=2} \hat{z}_{k0}}, \\ \hat{p}_{k0} = \hat{p}_{10} \hat{z}_{k0}, \quad k = 2, 3, \dots, n. \end{cases} \quad (17)$$

Интерполяция динамики распределения популяции по n территориям, в свою очередь, может быть представлена с помощью полученной оценки начального состояния:

$$\begin{cases} \hat{p}_{1t} = \frac{1}{n + \sum_{k=2} \hat{z}_{k0} \hat{\gamma}_k^t}, \\ \hat{p}_{kt} = \hat{z}_{k0} \hat{\gamma}_k^t \hat{p}_{1t}, \quad k = 2, 3, \dots, n, \quad t = 0, 1, 2, \dots, T. \end{cases} \quad (18)$$

Можно также провести экстраполяцию этой динамики, если рассматривать $t = T + 2, T + 3, \dots$

Таким образом, имея статистические данные (11) о распределении популяции по n территориям за $T + 1$ лет, можно построить модель (4) этого процесса, а также предсказать его поведение на ближайшие несколько лет.

Чтобы оценить точность построенной модели, по формуле

$$R = \frac{\sum_{t=0}^T \sum_{k=1}^n \hat{p}_{kt} \tilde{p}_{kt}}{\left(\sum_{t=0}^T \sum_{k=1}^n \hat{p}_{kt}^2 \right)^{1/2} \left(\sum_{t=0}^T \sum_{k=1}^n \tilde{p}_{kt}^2 \right)^{1/2}}. \quad (19)$$

вычисляется коэффициент корреляции между эмпирической динамикой (12) и ее приближением (18).

1.2.2 Логарифмически-линейные вероятностные цепочки: построение модели

Приняв за стандарт первую популяцию (см. 1.2.1), можно представить любую вероятностную цепочку в виде системы уравнений:

$$\begin{cases} p_{1,t+1} = \frac{1}{1 + \sum_{j=2}^n P_j(p_t)}, \\ p_{k,t+1} = p_{1,t+1} \times P_k(p_t), \end{cases} \quad (20)$$

где $k = 2, \dots, n$, $t = 0, 1, 2, \dots, T$, $P_k(p_t)$ — строго положительная порождающая функция, и

$$p_t = \begin{pmatrix} p_{1,t} \\ \vdots \\ p_{n,t} \end{pmatrix}. \quad (21)$$

Перепишем эту систему более подробно для случая логарифмически-линейной цепочки:

$$\begin{cases} p_{1,t+1} = \frac{1}{1 + \sum_{j=2}^n A_j p_1^{a_{j1}} p_2^{a_{j2}} \dots p_n^{a_{jn}}}, \\ \frac{p_{k,t+1}}{p_{1,t+1}} = A_k p_1^{a_{k1}} p_2^{a_{k2}} \dots p_n^{a_{kn}}. \end{cases} \quad (22)$$

где k , t и p_t принимают те же значения.

Приведенную систему (22) удобно представить следующим образом:

$$\begin{aligned} \ln p_{k,t+1} - \ln p_{1,t+1} &= \ln A_k + \sum_{j=1}^n a_{kj} \ln p_{j,t}, \\ k &= 2, \dots, n, \\ t &= 0, 1, 2, \dots, T. \end{aligned} \quad (23)$$

Таким образом, получаем систему из $(n - 1)$ уравнения:

$$\begin{aligned} \ln p_{2,t+1} - \ln p_{1,t+1} &= \ln A_2 + a_{21} \ln p_{1,t} + a_{22} \ln p_{2,t} + \dots + a_{2n} \ln p_{n,t}, \\ &\vdots \\ \ln p_{n,t+1} - \ln p_{1,t+1} &= \ln A_n + a_{n1} \ln p_{1,t} + a_{n2} \ln p_{2,t} + \dots + a_{nn} \ln p_{n,t}, \end{aligned} \quad (24)$$

где $t = 0, 1, 2, \dots, T$.

Система (24) представляет собой систему внешне не связанных уравнений (Seemingly unrelated regressions, SUR), т.е. такую систему, которая характеризуется наличием корреляций между ошибками в ее уравнениях. Для оценивания подобных систем используется трехшаговый метод наименьших квадратов (в англоязычной литературе именуемый FGLS — Feasible Generalized Least Squares [28] и SURE — Seemingly Unrelated Regressions Estimator [20], [28]), который состоит в последовательном применении двухшагового метода наименьших квадратов — для оценки коэффициентов и погрешностей каждого уравнения по отдельности — и обобщенного метода наименьших квадратов — для оценки коэффициентов всей системы.

Однако выделяют два условия, при которых вместо трехшагового метода наименьших квадратов для оценки системы внешне не связанных уравнений можно использовать обычный метод наименьших квадратов, не проигрывая при этом в эффективности оценки [22, 28, 33]:

1. Ошибки не коррелируют между уравнениями системы;
2. Во всех уравнениях системы присутствуют одни и те же объясняющие (независимые) переменные.

В нашем случае выполняется второе условие, так как в каждом из уравнений системы (24) в правой части участвуют переменные $\ln p_{1,t}, \ln p_{2,t}, \dots, \ln p_{n,t}$, следовательно, для оценки неизвестных параметров можно использовать обычный метод наименьших квадратов. В литературе имеются доказательства достаточности выполнения второго условия для применения данного метода [19].

Перепишем теперь систему (24) в матричном виде. Пусть

$$Y = \begin{pmatrix} y_{2,T} & y_{3,T} & \dots & y_{n,T} \\ y_{2,T-1} & y_{3,T-1} & \dots & y_{n,T-1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{2,1} & y_{3,1} & \dots & y_{n,1} \end{pmatrix}, \quad (25)$$

$$X = \begin{pmatrix} 1 & \ln p_{1,T-1} & \ln p_{2,T-1} & \dots & \ln p_{n,T-1} \\ 1 & \ln p_{1,T-2} & \ln p_{2,T-2} & \dots & \ln p_{n,T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \ln p_{1,0} & \ln p_{2,0} & \dots & \ln p_{n,0} \end{pmatrix}, \quad (26)$$

$$A = \begin{pmatrix} \ln A_2 & \ln A_3 & \dots & \ln A_n \\ a_{21} & a_{31} & \dots & a_{n1} \\ a_{22} & a_{32} & \dots & a_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{2n} & a_{3n} & \dots & a_{nn} \end{pmatrix}, \quad (27)$$

где $y_{k,t} = \ln p_{k,t} - \ln p_{1,t}$, а T — количество лет, за которые исследуется распределение популяции по территориям.

Тогда система (24) примет вид

$$Y = XA, \quad (28)$$

или, с учетом ошибок

$$Y = XA + \mathcal{E}, \quad (29)$$

где $\mathcal{E} = [\epsilon_{ij}]$, $i = 1, \dots, n - 1$, $j = 1, \dots, T$.

Оценка \hat{A} матрицы A неизвестных коэффициентов системы (29) методом наименьших квадратов выглядит следующим образом:

$$\hat{A} = (X^T X)^{-1} X^T Y. \quad (30)$$

Интерполяция динамики логарифмически-линейной цепочки может проводиться следующим образом: в систему (22) подставляются эмпирические данные за нулевой год из исследуемого промежутка времени (т.е. при $t = 0$), а также найденные оценки неизвестных коэффициентов. Таким образом, вычисляются доли популяции за первый год из рассматриваемого интервала, и т.д., т.е. при известных долях p_{kt} популяции за год t рассчитываются значения за следующий $t + 1$ год. Так, получаем смоделированные данные за $1, 2, \dots, T$ год.

Остатки уравнений системы, т.е. матрицу \mathcal{E} , можно найти из уравнения (29), если подставить в Y смоделированные данные, в X — исходные эмпирические данные, а в A — найденные оценки неизвестных коэффициентов. Тогда, с учетом того, что смоделированные данные, в отличие от эмпирических, имеются за T лет, то и матрица \mathcal{E} на самом деле будет содержать не $T + 1$, а T строк. Всюду в дальнейшем под количеством остатков для уравнения системы будем понимать именно T .

1.2.3 Логарифмически-линейные вероятностные цепочки: проверка предпосылок регрессионного анализа

Поскольку оценка \hat{A} матрицы A неизвестных коэффициентов логарифмически-линейной цепочки вычисляется методом наименьших квадратов, чтобы гарантировать ее несмещенность, состоятельность и эффективность, необходимо проверить, удовлетворяет ли она основным предпосылкам регрессионного анализа (или условиям Гаусса-Маркова)[13]. Приведем здесь эти предпосылки [14]:

1. Математическое ожидание случайного отклонения равно нулю, т.е. $M(\epsilon_i) = 0$, $i = 1, \dots, k$, где k — количество остатков². Это означает, что случайный член не имеет систематического смещения в положительную или отрицательную сторону;
2. Дисперсия случайных отклонений постоянна, т.е. $D(\epsilon_i) = D(\epsilon_j) = \tau^2$, $\forall i, j = 1, \dots, k$. Данное условие говорит об отсутствии какой-либо величины, вызывающей большую ошибку, и называется гомоскедастичностью;
3. Случайные отклонения независимы друг от друга, т.е. $M(\epsilon_i, \epsilon_j) = 0$, $\forall i, j = 1, \dots, k$. Если это условие не выполняется, имеет место автокорреляция остатков;
4. Случайные отклонения не зависят от объясняющих переменных, т.е. $M(\epsilon_i, x_i) = 0$, $\forall i, j = 1, \dots, k$;
5. Модель линейна по параметрам.

Этих пяти предпосылок достаточно для получения уравнения регрессии, однако для оценки точности этого уравнения и его параметров необходимо также, чтобы случайные отклонения были распределены согласно нормальному закону.

Перечисленные условия относятся к парной линейной регрессии, однако они распространяются и на случай множественной регрессии, а следовательно, и на случай системы внешне не связанных уравнений. Однако, поскольку при оценке матрицы A неизвестных коэффициентов методом наименьших квадратов приходится решать матричное уравнение (см. (30)), к указанным предпосылкам добавляются также условия о том, что матрица $X^T X$ должна

²Как правило, оно совпадает с количеством наблюдений, однако в нашем случае это не так

быть неособенной (т.е. ее определитель должен быть отличным от нуля), и о том, что количество наблюдений должно превышать число неизвестных параметров (иначе получение надежных выводов является невозможным) [13].

Таким образом, чтобы гарантировать построение адекватной модели на основе логарифмически-линейной вероятностной цепочки, необходимо для каждого из уравнений системы (24) проверить следующие предпосылки:

- Равенство нулю математического ожидания остатков;
- Гомоскедастичность;
- Отсутствие автокорреляции остатков;
- Независимость остатков от объясняющих переменных;
- Нормальность распределения остатков.

Как можно заметить, были опущены условия о линейности модели по параметрам (ведь модель на основе логарифмически-линейной вероятностной цепочки является таковой по определению), а также о том, что матрица $X^T X$ должна быть неособенной — проверка данного свойства естественным образом происходит при нахождении решения матричного уравнения (30). Что касается условия о количестве наблюдений, проверка этой предпосылки не требует сколько-нибудь сложных математических выкладок и производится непосредственно в разработанном программном комплексе до начала работы с данными.

Проверка вышеприведенных условий сводится к проверке соответствующих статистических гипотез. Приведем теперь описание того, как эти процедуры были реализованы в данной работе.

Для проверки нормальности распределения остатков каждого из уравнений регрессии (24) в данной работе используется критерий Шапиро-Уилка, который считается одним из наиболее эффективных критериев проверки нормальности [16]. Этот критерий применяется для выборок размером от 5 до 50 наблюдений. Статистика критерия W выглядит следующим образом:

$$W = \frac{1}{s^2} \left(\sum_{i=1}^k a_{n-i+1} (x_{n-i+1} - x_i) \right)^2, \quad (31)$$

где $s^2 = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2$ — умноженная на n выборочная дисперсия, n — количество наблюдений, $k = [n/2]$, x_i — значения проверяемой перемен-

ной (в данном случае — остатков), упорядоченные по возрастанию, a_{n-i+1} — коэффициенты из специальной таблицы. В нашем случае проверяется нормальность распределения остатков, поэтому здесь и далее, где тестируются остатки, в качестве размера выборки n берется T .

Если статистика $W < W(\alpha)$, то нулевая гипотеза о нормальности распределения отклоняется на уровне значимости α . Если $W > W(\alpha)$, то гипотеза о нормальности распределения принимается. $W(\alpha)$ — критические значения статистики W также находятся из специальных таблиц. Все используемые в работе таблицы критических значений приведены в разработанной программе.

Поскольку нахождение коэффициентов a_{n-i+1} из таблиц не всегда является удобным, особенно если размер выборки велик, существует следующая аппроксимация критерия Шапиро-Уилка [10]:

$$\begin{aligned} W^* &= \left(1 - \frac{0.6695}{n^{0.6518}}\right) \frac{s^2}{B}, \\ B &= \left(\sum_{j=1}^k a_j (x_{n-j} - x_j)\right)^2, \\ a_0 &= \frac{0.899}{(n - 2.4)^{0.4162}} - 0.02, \\ a_j &= a_0 \left(z_j + \frac{1483}{(3 - z_j)^{10.845}} + \frac{71.6 \times 10^{-10}}{(1.1 - z_j)^{8.26}} \right), \\ z_j &= \frac{n - 2j + 1}{n - 0.5}, \end{aligned} \tag{32}$$

где s^2 по-прежнему представляет собой умноженную на n выборочную дисперсию, n — количество наблюдений, $k = [n/2]$, x_i — упорядоченные по возрастанию значения проверяемой переменной.

В этом случае нулевая гипотеза о нормальности распределения отклоняется на уровне значимости 0.05, если $W^* < 1$, и принимается в противном случае. В данной работе указанная аппроксимация применяется для выборок, в которых число наблюдений превышает 50.

Убедившись в том, что остатки в уравнениях (24) распределены согласно нормальному закону, можно осуществить проверку того, что их математическое ожидание равно нулю, с помощью t -критерия Стьюдента [13]. Тестовая статистика этого критерия в случае с неизвестной дисперсией имеет вид

$$t = \frac{\bar{x} - a_0}{s/\sqrt{n-1}}, \tag{33}$$

где s^2 — выборочная дисперсия, x — среднее арифметическое выборки, n — размер выборки, a_0 . В данном случае в качестве выборки рассматривается последовательность ошибок уравнения регрессии за каждый год из исследуемого промежутка времени, а $a_0 = 0$. Такая тестовая статистика распределена по закону Стьюдента со степенью свободы $df = n - 1$, и для проверки статистических гипотез по этому критерию используются таблицы критических значений Стьюдента. Так, в случае двустороннего критерия (когда альтернативная гипотеза имеет вид $a \neq a_0$), если $|t| > t_{\alpha, n-1}$ (где $t_{\alpha, n-1}$ находится из таблицы критических значений), то нулевая гипотеза о том, что математическое ожидание равно a_0 , отклоняется на уровне значимости α . В противном случае нулевая гипотеза принимается.

Для проверки отсутствия автокорреляции остатков в уравнениях, для этой цели используются тесты Дарбина-Уотсона, Бреуша-Годфри, Льюинга-Бокса, Кохрана-Оркатта, Q -тест Льюинга-Бокса и др. [4, 12]. В данной работе для проверки отсутствия автокорреляции был выбран тест Дарбина-Уотсона, который является наиболее удобным для реализации в программе и, в отличие от многих других тестов, не требует дополнительного построения регрессий остатков. Данный метод довольно прост, но имеет и свои недостатки: он обладает зонами неопределенности и выявляет корреляцию только между соседними наблюдениями. Впрочем, ситуации, когда на значения наблюдений оказывают влияние не соседние, а более ранние наблюдения, является достаточно редкой. Что касается проблемы с зоной неопределенности, можно прибегнуть к помощи еще одного теста — метода рядов, который будет описан чуть ниже.

Итак, тестовая статистика Дарбина-Уотсона имеет вид

$$d = \frac{\sum_{t=2}^n (\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=2}^n \epsilon_{t-1}^2}, \quad (34)$$

где ϵ_i - остатки уравнения регрессии, n — размер выборки.

Тест Дарбина-Уотсона, строго говоря, не является статистическим критерием, поскольку содержит зоны неопределенности, однако существуют пороговые значения d_L и d_U , на основе которых можно сделать определенные выводы о наличии или отсутствии автокорреляций:

1. Если $0 \leq d < d_L$, то принимается гипотеза о наличии положительной

- автокорреляции;
2. Если $d_L \leq d < d_U$, то вопрос о наличии или отсутствии автокорреляции остается открытым;
 3. Если $d_U \leq d < 4 - d_U$, то гипотеза о наличии автокорреляции отклоняется;
 4. Если $4 - d_U \leq d < 4 - d_L$, то вопрос о наличии или отсутствии автокорреляции остается открытым;
 5. Если $4 - d_L \leq d \leq 4$, то принимается гипотеза о наличии отрицательной автокорреляции.

Значения d_L и d_U находятся из специальных таблиц.

В данной работе также используется метод рядов, который основан на исследовании того, как распределяются знаки остатков уравнения регрессии. Итак, пусть n — объем выборки, n_1 — количество положительных остатков при n наблюдениях, n_2 — количество отрицательных остатков при n наблюдениях, k — количество рядов, т.е. идущих подряд остатков одинакового знака.

Для небольшого количества наблюдений ($n_1 < 20$, $n_2 < 20$) существуют специальные таблицы (для уровней значимости 0.01 и 0.05), который позволяют получить значения k_1 и k_2 по заданным n_1 и n_2 . Тогда, если $k_1 < k < k_2$, то можно говорить об отсутствии автокорреляции остатков. Если же $k \leq k_1$ или $k \geq k_2$, тогда можно говорить о наличии положительной или отрицательной автокорреляции остатков соответственно.

Для проверки гомоскедастичности используют тесты Парка, Глейзера, Уайта, Голдфелда-Квандта, ранговой корреляции Спирмена и др. [4, 12]. В данной работе применяется тест ранговой корреляции Спирмена, и его выбор объясняется теми же причинами, что и в случае с тестом Дарбина-Уотсона. Стоит также заметить, что этот тест одновременно служит и для проверки предпосылки о независимости остатков уравнений регрессии и объясняющих переменных.

Коэффициент ранговой корреляции Спирмена задается формулой

$$\rho = \frac{1 - \sum_{i=1}^n (r_i - s_i)^2}{\frac{1}{6}(n^3 - n)}, \quad (35)$$

где r_i и s_i — ранги наблюдений объясняющей переменной x_i и остатка ϵ_i , а n — объем выборки. Разумеется, в случае с системой внешне не связанных уравнений, такой коэффициент вычисляется для каждой объясняющей переменной и остатка каждого уравнения, т.е. если имеется m объясняющих переменных и k уравнений, то будет получено $m \times k$ коэффициентов корреляции.

Возможна и такая ситуация, когда нескольким значениям объясняющей переменной или нескольким значениям остатка приписываются одинаковые ранги (их называют связанными). В этом случае коэффициент ранговой корреляции Спирмена вычисляется по формуле [13]:

$$\rho = \frac{1 - \sum_{i=1}^n (r_i - s_i)^2}{\frac{1}{6}(n^3 - n) - (T_r + T_s)}, \quad (36)$$

$$T_r = \frac{1}{12} \sum_{i=1}^{m_r} (t_r^3 - t_r), \quad T_s = \frac{1}{12} \sum_{i=1}^{m_s} (t_s^3 - t_s),$$

где m_r и m_s — число групп одинаковых рангов у объясняющей переменной и у остатка, а t_r t_s — число рангов, которые входят в соответствующую группу неразличимых рангов у объясняющей переменной и у остатка.

Для проверки нулевой гипотезы об отсутствии корреляционной связи при $n > 10$ используется тестовая статистика вида

$$t = \frac{\rho \sqrt{(n-2)}}{\sqrt{(1-\rho^2)}}, \quad (37)$$

которая имеет t -распределение Стьюдента с $df = n - 2$ степенями свободы. Таким образом, нулевая гипотеза об отсутствии корреляционной связи (а следовательно, и гетероскедастичности) отклоняется на уровне значимости α , если $|t| > t_{\alpha, n-2}$, где $t_{\alpha, n-2}$ находится из таблицы критических значений Стьюдента.

1.2.4 Логарифмически-линейные вероятностные цепочки: оценивание модели

Предположим, что система (24) удовлетворяет всем условиям Гаусса-Маркова, и оценки ее неизвестных коэффициентов являются несмещенными, состоятельными и эффективными. Тем не менее, этого недостаточно для того,

чтобы делать выводы о качестве построенной модели. Необходимо также проверить значимость найденных оценок коэффициентов системы, значимость уравнений системы и значимость системы в целом.

Для оценки значимости коэффициентов $A[i, j]$ системы внешне не связанных уравнений (29) используется тестовая статистика [13] вида

$$t = \frac{A[i, j]}{s_i^2[(X^T \times X)^{-1}]_{ii}}, \quad i = 1, \dots, n - 1, \quad j = 1, \dots, n + 1, \quad (38)$$

где s_i^2 — несмещенная выборочная дисперсия остатков для i -ого уравнения, а A и X — матрицы из уравнения (29).

Согласно t -критерию Стьюдента, нулевая гипотеза о незначимости коэффициента $A[i, j]$ отвергается на уровне значимости α , если $|t| > t_{\alpha, k-p-1}$, где $t_{\alpha, k-p-1}$ находится из таблицы критических значений Стьюдента, и $df = k - p - 1$ — число степеней свободы при количестве объясняющих переменных p и объеме выборки k (в данном случае — $T - 1$).

Перейдем теперь к оценке значимости самих уравнений системы (29). Здесь возникает понятие коэффициента детерминации, который отражает долю вариации зависимой переменной, обусловленную изменчивостью независимых (объясняющих) переменных. Данный коэффициент для i -ого уравнения системы выражается формулой

$$R_i^2 = 1 - \frac{SSE_i}{SST[i, i]}, \quad (39)$$

где $SSE_i = \sum_{j=1}^{T-1} e_{ij}^2$, $i = 1, \dots, n - 1$, $SST = Q^T \times Q$, $Q = Y - M$, $M = [m_{ij}]$, $m_{ij} = \frac{1}{T} \sum_{j=1}^T y_{ij}$, а Y — матрица из уравнения (29).

Коэффициент R^2 обладает определенным недостатком: он увеличивается при добавлении новых объясняющих переменных. Поэтому используют также скорректированный коэффициент детерминации R_{adj}^2 , который задается формулой

$$R_{adj}^2 = 1 - \frac{k - 1}{k - p - 1}(1 - R^2), \quad (40)$$

где k — объем выборки (в нашем случае $k = T - 1$), а p — число объясняющих переменных.

Значимость i -ого уравнения системы проверяется с помощью критерия Фишера, где тестовая статистика имеет вид

$$F_i = \frac{R_i^2(k-p-1)}{(1-R_i^2)p}, \quad (41)$$

где k и p принимают те же значения.

Нулевая гипотеза о незначимости i -ого уравнения системы отвергается на уровне значимости α при $F > F_{\alpha, k_1, k_2}$, где $k_1 = p$, $k_2 = k - p - 1$, и F_{α, k_1, k_2} определяется из таблицы критических значений Фишера-Снедекора.

Для проверки значимости системы в целом также используется указанный критерий, однако в качестве R^2 выступает общий коэффициент детерминации системы (29). он определяется формулой

$$R_{total}^2 = 1 - \frac{n-1}{Tr(Cov^{-1} \times SST)}, \quad (42)$$

где SST — матрица, уже описанная выше, а Cov — матрица ковариаций остатков, которая имеет вид $Cov = \frac{1}{T-1} E^T \times E$ (E - матрица остатков для наблюдений $2, \dots, T$ в уравнениях $2, \dots, n$).

2 Моделирование распределения доходов по слоям населения

В этом и последующих разделах работы приведены примеры исследования различных социально-экономических процессов с помощью нелинейных вероятностных цепочек.

Первый из этих примеров касается процесса распределения доходов по слоям населения РФ.

Для построения моделей будем использовать данные Федеральной службы государственной статистики РФ за 1992–2009 гг. (см. таблицу 6 в приложении 4.3) [18]. Здесь все население РФ было условно разделено на пять 20-типроцентных групп в зависимости от дохода, и эти группы были упорядочены по возрастанию так, чтобы в первой оказались 20% населения с наименьшим доходом, а в пятой — 20% с наибольшим доходом.

Таблица 1: Оценки относительных приращений доходов в зависимости от выбранного стандарта

| Прирост / Стандарт | I-ст. | II-ст. | III-ст. | IV-ст. | V-ст. |
|--------------------|---------|---------|---------|---------|---------|
| I | 1 | 1.00008 | 0.99979 | 0.99821 | 0.96909 |
| II | 0.99936 | 1 | 0.99990 | 0.99841 | 0.96465 |
| III | 0.99858 | 0.99935 | 1 | 0.99912 | 0.96385 |
| IV | 0.99817 | 0.99884 | 1.00022 | 1 | 0.96449 |
| V | 1.01599 | 1.01404 | 1.01479 | 1.01440 | 1 |

2.1 Постановка задачи

Используя эмпирические данные из таблицы 6, построим с помощью разработанной программной системы модели, основанные на вероятностных цепочках с логистическим ростом и логарифмически-линейных вероятностных цепочках, а также проведем экстраполяцию на следующие 10 лет (2010–2019 гг.).

2.2 Вероятностная цепочка с логистическим ростом

2.2.1 Относительные приращения доходов

Как было указано в разделе 1.2.1, доля доходов одной из групп населения может быть взята за стандарт. Поскольку имеется пять таких групп, возникает пять возможных решений. В таблице 1 приведены пять различных вариантов расчета относительных приращений доли дохода в зависимости от выбранного стандарта.

Как можно заметить, во всех пяти случаях, т.е. независимо от выбранного стандарта, наблюдается следующая тенденция: относительный прирост дохода наиболее состоятельной группы населения является наибольшим из всех пяти.

2.2.2 Матрица взаимного влияния

На основе полученных значений относительных приращений можно вычислить матрицу взаимного влияния для логистической вероятностной цепочки. В общем случае такая матрица выглядит следующим образом:

$$[a_{ij}]_{i,j=1}^n = \begin{pmatrix} 0 & 1 - \frac{\gamma_2}{\gamma_1} & 1 - \frac{\gamma_3}{\gamma_1} & 1 - \frac{\gamma_4}{\gamma_1} & 1 - \frac{\gamma_5}{\gamma_1} \\ 1 - \frac{\gamma_1}{\gamma_2} & 0 & 1 - \frac{\gamma_3}{\gamma_2} & 1 - \frac{\gamma_4}{\gamma_2} & 1 - \frac{\gamma_5}{\gamma_2} \\ 1 - \frac{\gamma_1}{\gamma_3} & 1 - \frac{\gamma_2}{\gamma_3} & 0 & 1 - \frac{\gamma_4}{\gamma_3} & 1 - \frac{\gamma_5}{\gamma_3} \\ 1 - \frac{\gamma_1}{\gamma_4} & 1 - \frac{\gamma_2}{\gamma_4} & 1 - \frac{\gamma_3}{\gamma_4} & 0 & 1 - \frac{\gamma_5}{\gamma_4} \\ 1 - \frac{\gamma_1}{\gamma_5} & 1 - \frac{\gamma_2}{\gamma_5} & 1 - \frac{\gamma_3}{\gamma_5} & 1 - \frac{\gamma_4}{\gamma_5} & 0 \end{pmatrix} \quad (43)$$

Приведем здесь матрицу взаимного влияния в случае, когда за стандарт взята доля доходов третьей группы населения $R\Phi^3$, подставив в матрицу $[a_{ij}]_{i,j=1}^n$ соответствующие относительные скорости прироста:

$$[a_{ij}]_{i,j=1}^n = \begin{pmatrix} 0 & -0.0001 & -0.0002 & -0.0004 & -0.0150 \\ 0.0001 & 0 & -0.0001 & -0.0003 & -0.0149 \\ 0.0002 & 0.0001 & 0 & -0.0002 & -0.0148 \\ 0.0004 & 0.0003 & 0.0002 & 0 & -0.0146 \\ 0.0148 & 0.0147 & 0.0146 & 0.0144 & 0 \end{pmatrix} \quad (44)$$

Можно заметить, что все элементы приведенной матрицы, располагающиеся ниже главной диагонали, положительны, а все элементы выше главной диагонали отрицательны. Это можно проинтерпретировать следующим образом. Относительный прирост доли доходов первой группы населения получает отрицательное входное воздействие со стороны всех остальных долей доходов. Относительный прирост доли доходов второй группы получает отрицательное входное воздействие со стороны всех долей доходов, кроме первой, и т.д., а относительный прирост доли доходов пятой группы населения, напротив, получает положительное входное воздействие со стороны всех остальных долей доходов. Такие результаты можно объяснить тем, что, если зафиксировать относительное приращение доходов третьей (средней по

³Матрицы взаимного влияния для остальных четырех случаев можно найти в приложении к данной работе.

Таблица 2: Коэффициенты корреляции между эмпирической динамикой и динамикой с логистическим ростом для распределения доходов по слоям населения РФ

| I-ст. | II-ст. | III-ст. | IV-ст. | V-ст. |
|----------|----------|----------|----------|----------|
| 0.998890 | 0.999038 | 0.999048 | 0.999036 | 0.996337 |

уровню благосостояния) группы населения, то скорости изменения доходов для остальных групп оказываются упорядоченными по номеру группы.

Построив такие матрицы взаимного влияния для случаев, когда за стандарт были взяты доли доходов других групп, можно отметить следующие закономерности. В случае с долей доходов четвертой группы, взятой за стандарт, имеет место иерархия взаимных воздействий, аналогичная описанной выше. Для долей доходов первой, второй и пятой групп, взятых за стандарт соответственно, такие иерархии различаются, однако все пять случаев объединяет одно: доля доходов пятой (наиболее состоятельной) группы населения оказывает отрицательное воздействие на относительный прирост долей доходов других групп, но в то же время ее прирост испытывает положительное воздействие со стороны долей других групп.

2.2.3 Оценка построенной модели

Приведем теперь в таблице 2 коэффициенты корреляции между эмпирической динамикой и построенной динамикой с логистическим ростом для распределения доходов по слоям населения РФ. Как и ранее, рассматриваются пять случаев, соответствующих выбранной за стандарт доли доходов группы населения.

Указанные коэффициенты корреляции говорят о высокой точности построенной модели. Однако при доле дохода третьей группы, взятой за стандарт, аппроксимация динамики с логистическим ростом к эмпирической является наиболее точной. Поэтому при построении экстраполяции будем рассматривать случай, когда за стандарт берется доля дохода именно этой группы.

2.2.4 Экстраполяция динамики распределения доходов

В таблице 7 в приложении к работе представлена экстраполяция распределения доходов среди групп населения на 2010–2019 гг. График экстраполяции,

а также интерполяции этого распределения (вместе с исходными эмпирическими данными), приведен на рисунке 2 в приложении.

2.2.5 Выводы

Опираясь на модель, построенную для распределения доходов по пяти слоям населения РФ на основе вероятностной цепочки с логистическим ростом, мы можем дать следующий прогноз: доходы наиболее состоятельной группы населения растут за счет сокращения доходов остальных групп; доходы групп с наименьшими доходами (с первой по четвертую) сокращаются тем быстрее, чем меньше они на данный момент в иерархии доходов всех групп.

2.3 Логарифмически-линейная вероятностная цепочка

Попробуем решить ту же задачу с использованием другой модели — основанной на логарифмически-линейной вероятностной цепочке. Эта модель, впервые описанная в работе Д. Дендриноса (D. Dendrinos) и М. Сониса (M. Sonis) [21] и называемая также DS-моделью, неоднократно применялась в исследованиях распределения валового регионального продукта по регионам таких стран, как США [24], Бразилия [27], Греция [26], Индонезия [29], Испания и Португалия [20].

2.3.1 Проверка предпосылок регрессионного анализа

Прежде, чем исследовать динамику взаимных влияний долей доходов, необходимо удостовериться в том, что выполняются предпосылки регрессионного анализа. Проведем эту процедуру для каждой из пяти групп, выбранной в качестве стандарта, и будем далее рассматривать только те случаи, которые удовлетворяют предпосылкам.

1. Стандарт — первая группа.

По критерию Шапиро-Уилка ни для одного из уравнений системы остатки не являются нормально распределенными, следовательно, дальнейшее исследование матрицы взаимных влияний в этом случае является нецелесообразным, поскольку даст заведомо некорректные результаты.

2. Стандарт — вторая группа.

Получаем результаты, аналогичные первому случаю.

3. Стандарт — третья группа.

По критерию Шапиро-Уилка остатки нормально распределены только для последнего уравнения системы (где тестовая статистика $W = 0.924$), следовательно, имеет смысл рассматривать только результаты, касающиеся оценок для этого уравнения.

По t -критерию Стьюдента принимается гипотеза о равенстве нулю математического ожидания остатков последнего уравнения.

Тест Дарбина-Уотсона для остатков данного уравнения не позволяет однозначно сказать о наличии или отсутствии автокорреляции, поэтому воспользуемся методом рядов. Согласно этому тесту, статистика $k = 8$ лежит в интервале между $k_1 = 5$ и $k_2 = 14$, следовательно, принимается гипотеза об отсутствии автокорреляции.

Коэффициенты ранговой корреляции Спирмена между остатками последнего уравнения системы и объясняющими переменными являются незначимыми, что говорит об отсутствии гетероскедастичности.

Таким образом, в случае с долей доходов третьей группы, выбранной в качестве стандарта, можно проводить дальнейшие исследования только для доли доходов пятой группы населения.

4. Стандарт — четвертая группа.

В данном случае остатки также распределены по нормальному закону только для последнего уравнения системы, и гипотеза о равенстве их математического ожидания нулю принимается. Результаты теста рядов позволяют сказать об отсутствии автокорреляции остатков последнего уравнения, а результаты теста ранговой корреляции Спирмена — об их гомоскедастичности.

Следовательно, в данном случае также имеется возможность работать лишь с долей доходов пятой группы населения.

5. Стандарт — пятая группа.

По критерию Шапиро-Уилка, остатки нормально распределены только для последних двух уравнений системы, следовательно, далее будем рассматривать только эти уравнения. По t -критерию Стьюдента гипотеза о равенстве нулю математического ожидания остатков последних двух уравнений принимается. По тесту рядов можно сделать вывод об отсутствии автокорреляции, а по тесту ранговой корреляции Спирмена — о гомоскедастичности.

2.3.2 Оценка построенной модели

Найдем оценку матрицы неизвестных коэффициентов для логарифмически-линейной вероятностной цепочки. Фактически мы получим три оценки этой матрицы в соответствии с выбранными в качестве стандарта долями доходов третьей, четвертой и пятой групп населения. Приведем здесь оценку матрицы для случая, когда стандартом является доля доходов третьей группы⁴

$$[\hat{a}_{i,j}] = \begin{pmatrix} 1.438 & 0.707 & -0.773 & -1.236 \\ 3.509 & 1.809 & -1.686 & 0.239 \\ -4.158 & -1.243 & 1.025 & -3.391 \\ 5.147 & 2.779 & -2.739 & -2.045 \\ 4.335 & 3.194 & -3.295 & -4.73 \end{pmatrix}, \quad (45)$$

а также коэффициенты A_1, A_2, A_4, A_5 — суммарные преимущества всех групп, кроме третьей:

$$\begin{aligned} A_1 &= 1629845.26082146, \\ A_2 &= 22495.9848716216, \\ A_4 &= 3.17305452063293 \times 10^{-5}, \\ A_5 &= 3.16182422628969 \times 10^{-7} \end{aligned} \quad (46)$$

Для оценки значимости коэффициентов последнего уравнения (заметим, что исследуется только данное уравнения, поскольку остальные в данном случае не удовлетворяют предпосылкам регрессионного анализа) проведем t -тест Стьюдента. Согласно его результатам, коэффициент $a[5, 1] = -1.236$, отражающий положительное входное воздействие доли дохода первой группы на относительное приращение доли дохода пятой группы населения, является статистически значимым на уровне 0.01. Однако результаты F -теста Фишера говорят о статистической незначимости соответствующего уравнения системы, что не позволяет строить относительно его коэффициентов какие-либо выводы.

Перейдем теперь к случаям, когда в качестве стандартов были выбраны доли доходов четвертой и пятой групп населения.

Согласно результатам t -теста Стьюдента для первого из этих двух случаев, ни один из коэффициентов последнего уравнения не является статистически значимым, поэтому и построение выводов о взаимном влиянии при

⁴Оценки матриц для прочих случаев можно найти в приложении к работе.

доле доходов четвертой группы, взятой за стандарт, является недопустимым в рамках данной модели.

Если провести аналогичную проверку для случая с долей дохода пятой группы в качестве стандарта, выяснится, что коэффициент $a[3, 1] = 1.236$ является статистически значимым на уровне 0.05. Этот коэффициент отражает положительное входное воздействие доли дохода первой группы на относительное приращение доли дохода третьей группы населения. Однако проверка значимости соответствующего уравнения системы говорит о том, коэффициент детерминации этого уравнения является небольшим (0.58) и, кроме того, статистически незначимым. Это обстоятельство ставит под вопрос адекватность результатов, получаемых в случае, когда в качестве стандарта рассматривается доля доходов пятой группы населения.

2.3.3 Результаты

Таким образом, ввиду невыполнения основных предпосылок регрессионного анализа, не представляется возможным выявить какие-либо закономерности в логарифмически-линейной динамике распределения доходов по слоям населения РФ, поскольку полученные методом наименьших квадратов оценки неизвестных параметров в данном случае не являются эффективными.

Как демонстрирует приведенный пример, логарифмически-линейные вероятностные цепочки, вследствие высокой степени сложности и наличия дополнительных условий при построении моделей на их основе, не всегда дают наилучшее в смысле метода наименьших квадратов приближение к исходным данным.

3 Моделирование распределения исследователей по областям науки в РФ

3.1 Постановка задачи

Рассмотрим теперь еще один пример задачи о распределении ресурсов.

Пусть необходимо смоделировать распределение исследователей в РФ по областям науки. Под исследователями здесь подразумеваются работники, профессионально занимающиеся исследованиями и разработками и непосредственно осуществляющие создание новых знаний, продуктов, методов и систем, а также управление указанными видами деятельности.

Таблица 3: Оценки относительных приращений для шести случаев выбора стандарта

| Прирост / Ст. | Естеств. | Техн. | Мед. | С/х | Обществ. | Гум. |
|---------------|----------|---------|---------|---------|----------|---------|
| Естеств. | 1 | 1.01015 | 0.98714 | 1.00322 | 1.00137 | 0.97493 |
| Техн. | 0.98872 | 1 | 0.97568 | 0.99369 | 0.99239 | 0.96323 |
| Мед. | 1.00134 | 1.00960 | 1 | 1.00125 | 0.99939 | 0.97478 |
| С/х | 0.99484 | 1.00677 | 0.98034 | 1 | 0.99770 | 0.96814 |
| Обществ. | 0.99571 | 1.00837 | 0.98167 | 1.00055 | 1 | 0.96773 |
| Гум. | 1.02276 | 1.03273 | 1.00747 | 1.02463 | 1.02133 | 1 |

Для построения логарифмически-линейной модели будут использоваться данные Федеральной службы государственной статистики РФ за 1994–2009 по числу исследователей в естественных, технических, медицинских, сельскохозяйственных, общественных и гуманитарных науках (всего — шесть областей) [17]. Указанные данные приведены в таблице 8 в приложении к работе.

3.2 Вероятностная цепочка с логистическим ростом

3.2.1 Построение и оценка модели

Как и в разделе 2.2, начнем исследование с нахождения оценок относительных приращений долей исследователей в рассматриваемых областях науки. Поскольку таких областей шесть, приведем шесть вариантов расчета этих оценок в зависимости от того, доля исследователей какой из областей считается стандартом. В таблице 3 каждому из вариантов соответствует определенный столбец.

Основываясь на полученных оценках, можно отметить, что, независимо от фиксируемой области научной деятельности, во всех шести случаях наблюдается положительный прирост (причем наибольший из всех) доли исследователей в гуманитарной сфере и снижение доли исследователей — для технических дисциплин.

Следующий этап моделирования заключается в оценке матрицы взаимного влияния. Приведем здесь матрицу взаимного влияния при фиксированной технической области, поскольку именно этот вариант отличается наибольшим коэффициентом корреляции между эмпирической динамикой и динамикой с логистическим ростом⁵:

⁵Для этого и последующих примеров описание моделирования различных задач будет менее подроб-

$$[a_{ij}]_{i,j=1}^n = \begin{pmatrix} 0 & 0.0101 & 0.0005 & 0.0034 & 0.0018 & -0.0223 \\ -0.0102 & 0 & -0.0096 & -0.0068 & -0.0084 & -0.0327 \\ -0.0005 & 0.0095 & 0 & 0.0028 & 0.0012 & -0.0229 \\ -0.0034 & 0.0067 & -0.0028 & 0 & -0.0016 & -0.0258 \\ -0.0018 & 0.0083 & -0.0012 & 0.0016 & 0 & -0.0242 \\ 0.0219 & 0.0317 & 0.0224 & 0.0251 & 0.0236 & 0 \end{pmatrix} \quad (47)$$

Данная матрица взаимного влияния, как и остальные, приведенные для прочих вариантов стандарта в приложении , позволяет сделать вывод о том, относительное приращение доли исследователей гуманитарной сферы получает положительное входное воздействие со стороны долей исследователей остальных областей, но вместе с тем доля исследователей-гуманитариев отрицательно влияет на относительный прирост долей исследователей в других науках. Абсолютно противоположная закономерность характеризует долю исследователей в технической сфере.

3.2.2 Экстраполяция динамики распределения исследователей по областям науки

Как и в разделе 2.2, проведем экстраполяцию исследуемого распределения на ближайшие 10 лет, т.е. на интервал с 2010 по 2019 г. (для случая с зафиксированной в качестве стандарта второй долей исследователей). Спрогнозированные данные за этот промежуток времени приведены в таблице 9 в приложении. График спрогнозированных данных для случая с первой долей исследователей, взятой в качестве стандарта, представлен на рисунке 3 в приложении к данной работе. Они также иллюстрируют описанные выше закономерности.

3.2.3 Выводы

Результаты моделирования распределения исследователей по областям науки в РФ на основе вероятностной цепочки с логистическим ростом дают основания сделать вывод о том, что доля исследователей гуманитарных дисциплин увеличивается за счет снижения долей исследователей остальных сфер, и в

ном, включающим лишь наиболее значимые результаты. Что касается динамики распределения исследователей по областям науки, стоит отметить, что все шесть коэффициентов корреляции говорят о высокой точности построенной модели с логистическим ростом.

первую очередь — сферы технических наук (однако при этом, по данным за 2009 год, доля исследователей в гуманитарных науках сама по себе весьма незначительна (всего 2,4%), в то время как доля исследователей в технической области составляет 61,6%).

3.3 Логарифмически-линейная вероятностная цепочка

3.3.1 Построение и оценка модели

Как было показано в разделе 2.3 на примере задачи о распределении доходов по слоям населения РФ, модели на основе логарифмически-линейных цепочек достаточно сложны как в построении, так и в обработке и интерпретации вычислений, и отнюдь не всегда гарантируют значимые результаты. Поэтому в данном случае приведем только наиболее важные и статистически значимые из найденных оценок, не останавливаясь на прочих результатах.

Так, если рассматривать динамику изучаемого распределения относительно доли исследователей в естественных науках, можно выделить следующие значимые коэффициенты взаимного влияния:

$$\begin{aligned} a[6, 1] = 7.375, \quad a[6, 2] = 23.811, \quad a[6, 3] = 1.875, \\ a[6, 5] = 1.568, \quad a[6, 6] = 1.361. \end{aligned} \quad (48)$$

Эти коэффициенты говорят о том, что относительное приращение доли исследователей в гуманитарной сфере испытывает положительное входное воздействие со стороны долей исследователей других научных областей (кроме сельскохозяйственного направления, для которого соответствующий коэффициент оказался статистически незначимым). Кроме того, данное положительное воздействие отмечается для доли исследователей гуманитарной сферы и со стороны самой себя, причем соответствующий коэффициент $a[6, 6] = 1.361$ обладает наибольшим уровнем значимости (0.005) из всех приведенных выше.

В случае, если в качестве стандарта берется доля исследователей в сельскохозяйственной или общественной сферах, значимыми оказываются коэффициенты $a[6, 6] = 1.591$ и $a[6, 6] = 1.403$ соответственно, причем на достаточно высоком уровне значимости (0.005), что также свидетельствует в пользу описанной в предыдущем абзаце закономерности.

Если же динамика распределения исследователей по областям науки рассматривается по отношению к доле исследователей гуманитарной области,

выделяются следующие значимые коэффициенты взаимного влияния:

$$\begin{aligned} a[1, 1] &= -7.375, & a[1, 2] &= -23.811, & a[1, 3] &= -1.875, \\ a[1, 5] &= -1.568, & a[1, 6] &= -1.361, \\ a[4, 6] &= -1.591, & a[5, 6] &= -1.403, \end{aligned} \quad (49)$$

т.е. в данном случае относительное приращение доли исследователей в естественных науках испытывает отрицательное входное воздействие со стороны долей исследователей во всех остальных областях, кроме сельскохозяйственных. Доля исследователей гуманитарного направления, в свою очередь, оказывает отрицательное воздействие на относительный прирост доли исследователей в естественной, сельскохозяйственной и общественной областях (и соответствующие коэффициенты значимы на уровне 0.005).

3.3.2 Экстраполяция динамики распределения исследователей по областям науки

Как и ранее, проведем аппроксимацию эмпирической динамики распределения исследователей на основе логарифмически-линейной динамики на 2010–2019 гг. Спрогнозированные данные представлены в таблице 10 в приложении к работе.

3.3.3 Выводы

Таким образом, с какой бы точки зрения ни рассматривалась динамика распределения исследователей по областям науки в РФ, четко прослеживается следующая тенденция: доля исследователей гуманитарной сферы, получая достаточно мощное положительное входное влияние как со стороны долей исследователей других областей, так и со стороны самой себя, оказывает при этом отрицательное влияние на прирост долей исследователей в других научных сферах.

3.4 Сравнение моделей, основанных на цепочке с логистическим ростом и логарифмически-линейной цепочке

Как видно, при решении одной и той же задачи о моделировании распределения исследователей по областям науки в РФ двумя способами были получены не совпадающие, хотя и не противоречащие друг другу результаты. В связи с этим встает вопрос о том, какие из них являются наиболее достоверными,

и естественным образом возникает предположение о том, что это результаты второй модели, которая предполагает наличие сложных статистических процедур, отсеивающих «лишние» закономерности. Чтобы проверить истинность этого предложения, приведем таблицу 4 коэффициентов корреляции между эмпирической динамикой и динамикой с логистическим ростом, а также между эмпирической и логарифмически-линейной динамикой, для шести вариантов переменной-стандарта.

Таблица 4: Коэффициенты корреляции для сравнения точности логарифмически-линейной динамики и динамики с логистическим ростом

| Модель | Естеств. | Техн. | Мед. | С/х | Обществ. | Гум. |
|-----------|----------|----------|----------|----------|----------|----------|
| Логист. | 0.999878 | 0.999884 | 0.999877 | 0.999884 | 0.999884 | 0.999867 |
| Лог.-лин. | 0.999829 | 0.999829 | 0.999829 | 0.999829 | 0.999829 | 0.999829 |

Как видно, в данном случае динамика с логистическим ростом ничуть не уступает логарифмически-линейной динамике (а точнее, превосходит ее примерно на 0.00005, но это преимущество является пренебрежимо малым), т.е. обе вероятностные цепочки могут служить инструментом для моделирования. Это можно проиллюстрировать и графиками интерполяции, а также графиками энтропии распределения исследователей для эмпирической, логарифмически-линейной динамики и динамики с логистическим ростом (см. рисунки 3, 4 в приложении). Графики приведены для случая, когда в качестве стандарта рассматривается доля исследователей естественных наук.

4 Моделирование распределения организаций по формам собственности

Данный пример, приведенный в работе в качестве заключительного, демонстрирует некоторые свойства логарифмически-линейных вероятностных цепочек, о которых еще не было упомянуто в предыдущих разделах.

4.1 Постановка задачи

Используя данные Федеральной службы государственной статистики РФ о численности организаций по форме собственности за 1994–2009 гг.[17], по-

строим на основе вероятностных цепочек модель процесса распределения организаций между формами собственности⁶.

Приведем список форм собственности, по которым распределяются организации:

- государственная,
- муниципальная,
- частная,
- собственность общественных и религиозных организаций,
- прочие формы собственности, включая смешанную российскую, иностранную, совместную российскую и иностранную.

Будем рассматривать в данном случае только модель, основанную на ологарифмически-линейной вероятностной цепочке.

4.2 Логарифмически-линейная вероятностная цепочка: построение и оценка модели

Как и ранее, с учетом того, что имеется статистика о распределении организаций по пяти формам собственности, возникает пять возможных вариантов построения модели, когда в качестве стандарта берется доля организаций с одной из пяти форм собственности. Произведя проверку предпосылок регрессионного анализа для каждого из этих случаев и убедившись, что все необходимые условия выполнены, перейдем к описанию выявленных закономерностей взаимного влияния.

Рассмотрим случай, когда стандарт — доля государственных организаций.

Чтобы проиллюстрировать тенденции взаимного влияния долей организаций с различными формами собственности, приведем ниже таблицу взаимного влияния (см. таблицу 5). Ее столбцы соответствуют формам собственности, приращения долей организаций в которых испытывают влияние долей организаций других форм собственности (которые, в свою очередь, соответствуют строкам таблицы). Знак «+» означает положительное влияние, знак «++» — положительное влияние на высоком уровне значимости, знак «-»

⁶Таблица с эмпирическими данными приведена в таблице 11 в приложении к работе.

— отрицательное влияние и знак «—» — отрицательное влияние на высоком уровне значимости.

Таблица 5: Таблица взаимного влияния долей организаций с различными формами собственности (стандарт — доля государственных организаций)

| Форма собств. | Муниц. | Частн. | Обществ. и рел. | Смешан. |
|-----------------|--------|--------|-----------------|---------|
| Гос. | + | + | + | ++ |
| Муниц. | | | | + |
| Частн. | + | ++ | + | ++ |
| Обществ. и рел. | + | + | + | ++ |
| Смешан. | | | + | ++ |

Как видно из таблицы, если логарифмически-линейная динамика рассматривается относительно доли государственных организаций, имеют место следующие тенденции: приращения долей муниципальных, частных и общественно-религиозных организаций испытывают положительное входное воздействие со стороны доли государственных, частных и общественно-религиозных организаций. Кроме того, относительное приращение доли общественно-религиозных организаций получает положительное входное воздействие доли смешанных организаций, а приращение тех, в свою очередь, испытывает положительное влияние со стороны долей организаций всех остальных форм собственности.

Для случаев, когда в качестве стандарта фиксируются доли организаций других форм собственности, наблюдаются иные иерархии взаимных влияний, и они могут быть проиллюстрированы таблицами 12, 13, 14 и 15 из приложения 4.3 к данной работе. Однако во всех пять случаях отмечается отрицательное входное влияние, которое получает относительный прирост доли государственных организаций со стороны долей других организаций. Достаточно мощное положительное воздействие со стороны самих себя испытывают приращения долей частных и общественно-религиозных организаций. В целом отношения между долями организаций в представленных пяти формах собственности можно называть скорее конкурирующими, нежели дополняющими.

4.3 Экстраполяция динамики распределения организаций по формам собственности

Как можно заметить, в предыдущих примерах, приведенных в данной работе, прогнозирование социально-экономических процессов проводилось лишь на ближайшие десять лет. Это объясняется тем, что в условиях современной динамично развивающейся экономики построение прогнозов на более длительные сроки представляет собой рискованное предприятие. Попробуем, однако, в чисто экспериментальных целях указать в качестве опций для экстраполяции логарифмически-линейной динамики достаточно высокую верхнюю границу интервала — например, 2200 год. На рисунке 1 приведен фрагмент такой экстраполяции для случая с долей государственных организаций в качестве стандарта.

Если проследить всю логарифмически-линейную динамику с 2010 по 2200 год, можно отметить, что начиная примерно со второго десятилетия XXI века наблюдаются колебания долей организаций всех форм собственности, причем их амплитуда достаточно велика. Однако спустя сто лет, согласно прогнозу, такие колебательные движения прекращаются, и начинается изменение долей организаций, обратное тому, которое наблюдалось в интервале 1994–2009 гг. И в 2157 году, как это показано фрагменте таблицы, совершенно неожиданным образом доля государственных организаций принимает значение, равное единице, в то время как доли организаций остальных форм собственности становятся равными нулю. Данное поведение модели также иллюстрируется графиком с 1995 по 2157 г., который приведен на рисунке 7 в приложении к работе. Подобный характер логарифмически-линейной динамики, который наблюдается также и для долей организаций других форм собственности, взятых за стандарт, как бы отсылает пользователя к теории нелинейных динамических систем. Действительно, нелинейные вероятностные цепочки являются одним из методов данной теории, согласно которой нелинейные динамические системы могут быть неустойчивыми к малым отклонениям и проявлять разнообразные колебательные и хаотические режимы поведения.

Именно это свойство моделей, основанных на логарифмически-линейных цепочках, демонстрирует данный пример. Стоит отметить, что при тестировании разработанной программы использовалось более двадцати примеров задач о распределении некоторого социально-экономического ресурса, и в подавляющем большинстве случаев модели проявляли поведение, похожее на описанное выше. Список таких примеров приведен в приложении 4.3 к рабо-

| Год | государственная | муниципальная | частная | собственность общественных и религиозных организаций | прочие формы собственности, включая смешанную российскую, иностранную, совместную российскую и иностранную |
|------|--------------------|----------------------|----------------------|--|--|
| 2131 | 0,0692593349176473 | 0,0520413844279528 | 0,877428201776303 | 0,000672459983243464 | 0,00059861889485361 |
| 2132 | 0,0703603976681362 | 0,0521832169967789 | 0,876218444279152 | 0,000657056063544289 | 0,000580884992388972 |
| 2133 | 0,0715096491438719 | 0,052331296213575 | 0,874953595794913 | 0,000641955264460808 | 0,000563503583178985 |
| 2134 | 0,072711866245401 | 0,0524862475741221 | 0,873628277715674 | 0,000627149170241237 | 0,000546459294561412 |
| 2135 | 0,0739725994198931 | 0,0526487985261827 | 0,872236236075034 | 0,00061262962277403 | 0,000529736356115923 |
| 2136 | 0,0752983514597451 | 0,0528198019712501 | 0,870770139409796 | 0,000598388732431299 | 0,000513318426777899 |
| 2137 | 0,07669681125783 | 0,05300026693083 | 0,869221314538978 | 0,000584418897704873 | 0,000497188374657527 |
| 2138 | 0,0781771640580183 | 0,0531913991612631 | 0,867579395951625 | 0,000570712837685287 | 0,00048132799140784 |
| 2139 | 0,0797505101581305 | 0,0533946558178032 | 0,865831852765941 | 0,000557263643472938 | 0,000465717614652283 |
| 2140 | 0,0814304405992722 | 0,0536118203602027 | 0,863963338563734 | 0,00054406485803925 | 0,000450335618752054 |
| 2141 | 0,0832338454246959 | 0,0538451072752056 | 0,861954778987469 | 0,000531110599690169 | 0,000435157712939128 |
| 2142 | 0,0851820755765505 | 0,0540973118238525 | 0,859782060895109 | 0,000518395753803869 | 0,000420155950684187 |
| 2143 | 0,0873026586048833 | 0,0543720297003199 | 0,857414098126606 | 0,000505916274044222 | 0,000405297294146209 |
| 2144 | 0,0896319112795452 | 0,0546739887219543 | 0,85480988863306 | 0,000493669663837185 | 0,000390541471357138 |
| 2145 | 0,09221906184215 | 0,0550095666069466 | 0,85191387811907 | 0,000481655763657822 | 0,000375837668175701 |
| 2146 | 0,095133029418686 | 0,0553876308376587 | 0,8486483242449111 | 0,000469878074845322 | 0,000361119219698763 |
| 2147 | 0,0984741320971956 | 0,055820963162535 | 0,844900263981107 | 0,000458346060919301 | 0,000346294698243227 |
| 2148 | 0,102395524681325 | 0,0563288058493279 | 0,840497358029365 | 0,000447079305872433 | 0,000331232134109614 |
| 2149 | 0,107145345575117 | 0,0569417054457832 | 0,835161104377845 | 0,000436115361496927 | 0,000315729239758288 |
| 2150 | 0,113157233109935 | 0,0577114411619757 | 0,828406347794214 | 0,000425525242647166 | 0,000299452691228652 |
| 2151 | 0,121267945920122 | 0,0587332930948245 | 0,819301514059767 | 0,000415445132005918 | 0,000281801793280083 |
| 2152 | 0,13332505319104 | 0,0602016647887567 | 0,80580557996854 | 0,000406139599026061 | 0,000261562452637108 |
| 2153 | 0,154279287410017 | 0,0625663077185111 | 0,782520452759946 | 0,000398075761141399 | 0,000235876350385088 |
| 2154 | 0,202061902580181 | 0,0669811430739746 | 0,730369411755978 | 0,000391095836544288 | 0,000196446753321735 |
| 2155 | 0,383263939109466 | 0,0739676771837996 | 0,542295737705419 | 0,000359948436454168 | 0,000112697564861042 |
| 2156 | 0,983010826159506 | 0,00988502482618846 | 0,007083505658552... | 2,02443860293675E-05 | 3,98969723773659E-07 |
| 2157 | 1 | 3,56474878270867E-26 | 1,54381393132212E... | 1,51116967804525E-35 | 3,38178829686965E-60 |
| 2158 | NaN | NaN | 0 | NaN | 0 |

Рис. 1: Экстраполяция логарифмически-линейной динамики распределения организаций по форме собственности на длительный срок

те.

Однако при подборе задач для программы был выявлен и другой, гораздо реже встречающийся, тип поведения логарифмически-линейной динамики. Он заключается в следующем: спрогнозированная на достаточно длительный срок система входит в положение равновесия, и все доли распределяемого ресурса остаются практически неизменными. Такое поведение демонстрирует, в частности, логарифмически-линейная динамика распределения исследователей по областям науки в РФ (см. график на рисунке 5 в приложении; за стандарт принимается доля исследователей естественных наук).

Заключение

В рамках работы был разработан и реализован программный комплекс для работы с двумя видами дискретных нелинейных вероятностных цепочек: цепочки с логистическим ростом и логарифмически-линейной цепочки. Реализованы алгоритмы вычисления оценки неизвестных коэффициентов для обоих типов моделей, алгоритмы оценивания точности и значимости построенных моделей, алгоритмы проверки условий Гаусса-Маркова для логарифмически-линейной вероятностной цепочки, а также алгоритмы прогнозирования динамики с логистическим ростом и логарифмически-линейной динамики.

С помощью разработанной программной системы исследована асимптотика логарифмически-линейной динамики. На ряде примеров показано, что она может демонстрировать нетривиальное поведение при экстраполяции на длительный срок.

Приложения

Графики и таблицы

Таблица 6: Статистические данные о распределении доходов между слоями населения РФ за 1992–2009 гг.

| Год | I | II | III | IV | V |
|------|-------|-------|-------|-------|-------|
| 1992 | 0.060 | 0.116 | 0.176 | 0.265 | 0.383 |
| 1993 | 0.051 | 0.094 | 0.149 | 0.238 | 0.468 |
| 1994 | 0.054 | 0.101 | 0.151 | 0.227 | 0.467 |
| 1995 | 0.061 | 0.108 | 0.152 | 0.216 | 0.463 |
| 1996 | 0.061 | 0.107 | 0.153 | 0.218 | 0.461 |
| 1997 | 0.058 | 0.105 | 0.152 | 0.223 | 0.462 |
| 1998 | 0.060 | 0.105 | 0.150 | 0.215 | 0.470 |
| 1999 | 0.059 | 0.104 | 0.148 | 0.211 | 0.467 |
| 2000 | 0.059 | 0.104 | 0.151 | 0.219 | 0.467 |
| 2001 | 0.057 | 0.104 | 0.154 | 0.228 | 0.457 |
| 2002 | 0.057 | 0.104 | 0.154 | 0.227 | 0.458 |
| 2003 | 0.055 | 0.103 | 0.153 | 0.227 | 0.462 |
| 2004 | 0.054 | 0.101 | 0.151 | 0.227 | 0.467 |
| 2005 | 0.054 | 0.101 | 0.151 | 0.227 | 0.467 |
| 2006 | 0.052 | 0.099 | 0.150 | 0.226 | 0.473 |
| 2007 | 0.051 | 0.098 | 0.148 | 0.225 | 0.478 |
| 2008 | 0.051 | 0.097 | 0.148 | 0.225 | 0.479 |
| 2009 | 0.051 | 0.098 | 0.148 | 0.225 | 0.478 |

Таблица 7: Экстраполяция распределения доходов среди групп населения на 2010–2019 гг.

| Год | I | II | III | IV | V |
|------|-------|-------|-------|-------|-------|
| 2010 | 0.052 | 0.096 | 0.142 | 0.212 | 0.497 |
| 2011 | 0.052 | 0.095 | 0.141 | 0.210 | 0.501 |
| 2012 | 0.051 | 0.095 | 0.140 | 0.209 | 0.504 |
| 2013 | 0.051 | 0.094 | 0.139 | 0.207 | 0.508 |
| 2014 | 0.051 | 0.093 | 0.138 | 0.206 | 0.512 |
| 2015 | 0.050 | 0.093 | 0.137 | 0.204 | 0.515 |
| 2016 | 0.049 | 0.092 | 0.136 | 0.203 | 0.519 |
| 2017 | 0.049 | 0.091 | 0.135 | 0.201 | 0.523 |
| 2018 | 0.049 | 0.090 | 0.134 | 0.199 | 0.526 |
| 2019 | 0.048 | 0.089 | 0.133 | 0.198 | 0.530 |

Таблица 8: Статистические данные о распределении исследователей по областям науки в РФ

| Год | Естеств. | Техн. | Мед. | С/х | Обществ. | Гум. |
|------|----------|--------|-------|-------|----------|------|
| 1994 | 116391 | 345921 | 18866 | 18228 | 17917 | 7996 |
| 1995 | 114335 | 342906 | 16781 | 18077 | 18049 | 8542 |
| 1996 | 110295 | 317789 | 16256 | 16326 | 16033 | 8097 |
| 1997 | 107800 | 292675 | 15456 | 15317 | 15527 | 8333 |
| 1998 | 100269 | 265175 | 15944 | 13889 | 13777 | 7904 |
| 1999 | 100458 | 268307 | 16230 | 13799 | 13534 | 7884 |
| 2000 | 99834 | 274955 | 15539 | 14390 | 13259 | 7977 |
| 2001 | 97615 | 273597 | 15669 | 14259 | 13159 | 7877 |
| 2002 | 96887 | 266549 | 16849 | 14061 | 12571 | 7759 |
| 2003 | 96328 | 256921 | 22039 | 13735 | 12565 | 8187 |
| 2004 | 91698 | 258850 | 15907 | 14280 | 12467 | 8223 |
| 2005 | 91570 | 249358 | 15672 | 13724 | 12497 | 8300 |
| 2006 | 89304 | 248201 | 15896 | 13447 | 13721 | 8370 |
| 2007 | 94668 | 244475 | 16734 | 13743 | 13740 | 9489 |
| 2008 | 91117 | 232463 | 16713 | 13622 | 13012 | 8877 |
| 2009 | 89856 | 227403 | 16652 | 13217 | 13272 | 8837 |

Таблица 9: Экстраполяция динамики распределения исследователей по областям науки на 2010–2019 гг. на основе цепочки с логистическим ростом

| Год | Естеств. | Техн. | Мед. | С/х | Обществ. | Гум. |
|------|----------|--------|--------|--------|----------|--------|
| 2010 | 0.2480 | 0.6157 | 0.0425 | 0.0349 | 0.0338 | 0.0252 |
| 2011 | 0.2496 | 0.6128 | 0.0428 | 0.0350 | 0.0339 | 0.0259 |
| 2012 | 0.2513 | 0.6099 | 0.0431 | 0.0350 | 0.0340 | 0.0267 |
| 2013 | 0.2529 | 0.6070 | 0.0435 | 0.0351 | 0.0340 | 0.0275 |
| 2014 | 0.2546 | 0.6041 | 0.0438 | 0.0351 | 0.0341 | 0.0283 |
| 2015 | 0.2562 | 0.6011 | 0.0442 | 0.0352 | 0.0342 | 0.0291 |
| 2016 | 0.2579 | 0.5982 | 0.0445 | 0.0352 | 0.0343 | 0.0299 |
| 2017 | 0.2595 | 0.5952 | 0.0449 | 0.0353 | 0.0343 | 0.0308 |
| 2018 | 0.2612 | 0.5922 | 0.0452 | 0.0353 | 0.0344 | 0.0317 |
| 2019 | 0.2628 | 0.5892 | 0.0455 | 0.0353 | 0.0345 | 0.0326 |

Таблица 10: Экстраполяция динамики распределения исследователей по областям науки на 2010–2019 гг. на основе логарифмически-линейной цепочки

| Год | Естеств. | Техн. | Мед. | С/х | Обществ. | Гум. |
|------|----------|--------|--------|--------|----------|--------|
| 2010 | 0.2461 | 0.6153 | 0.0412 | 0.0359 | 0.0382 | 0.0235 |
| 2011 | 0.2464 | 0.6156 | 0.0408 | 0.0357 | 0.0383 | 0.0232 |
| 2012 | 0.2465 | 0.6161 | 0.0405 | 0.0356 | 0.0383 | 0.0230 |
| 2013 | 0.2465 | 0.6167 | 0.0403 | 0.0355 | 0.0382 | 0.0228 |
| 2014 | 0.2464 | 0.6173 | 0.0403 | 0.0354 | 0.0380 | 0.0226 |
| 2015 | 0.2462 | 0.6178 | 0.0403 | 0.0353 | 0.0379 | 0.0225 |
| 2016 | 0.2460 | 0.6182 | 0.0404 | 0.0353 | 0.0377 | 0.0225 |
| 2017 | 0.2458 | 0.6185 | 0.0405 | 0.0352 | 0.0375 | 0.0225 |
| 2018 | 0.2456 | 0.6187 | 0.0406 | 0.0352 | 0.0374 | 0.0225 |
| 2019 | 0.2454 | 0.6187 | 0.0408 | 0.0352 | 0.0373 | 0.0225 |

Таблица 11: Статистические данные о распределении организаций по формам собственности

| Год | Гос. | Муницип. | Частн. | Обществ. и рел. | Смешан. |
|------|------|----------|--------|-----------------|---------|
| 1994 | 325 | 171 | 1216 | 53 | 181 |
| 1995 | 322 | 198 | 1426 | 95 | 181 |
| 1996 | 233 | 184 | 1731 | 130 | 227 |
| 1997 | 143 | 178 | 2014 | 158 | 235 |
| 1998 | 148 | 183 | 2147 | 183 | 240 |
| 1999 | 150 | 198 | 2312 | 213 | 234 |
| 2000 | 151 | 217 | 2510 | 223 | 209 |
| 2001 | 155 | 231 | 2726 | 237 | 240 |
| 2002 | 157 | 239 | 2957 | 244 | 234 |
| 2003 | 161 | 246 | 3238 | 252 | 247 |
| 2004 | 159 | 248 | 3499 | 253 | 245 |
| 2005 | 160 | 252 | 3838 | 253 | 265 |
| 2006 | 149 | 264 | 3639 | 239 | 215 |
| 2007 | 142 | 263 | 3855 | 212 | 203 |
| 2008 | 135 | 258 | 3976 | 197 | 206 |
| 2009 | 130 | 253 | 4128 | 188 | 209 |

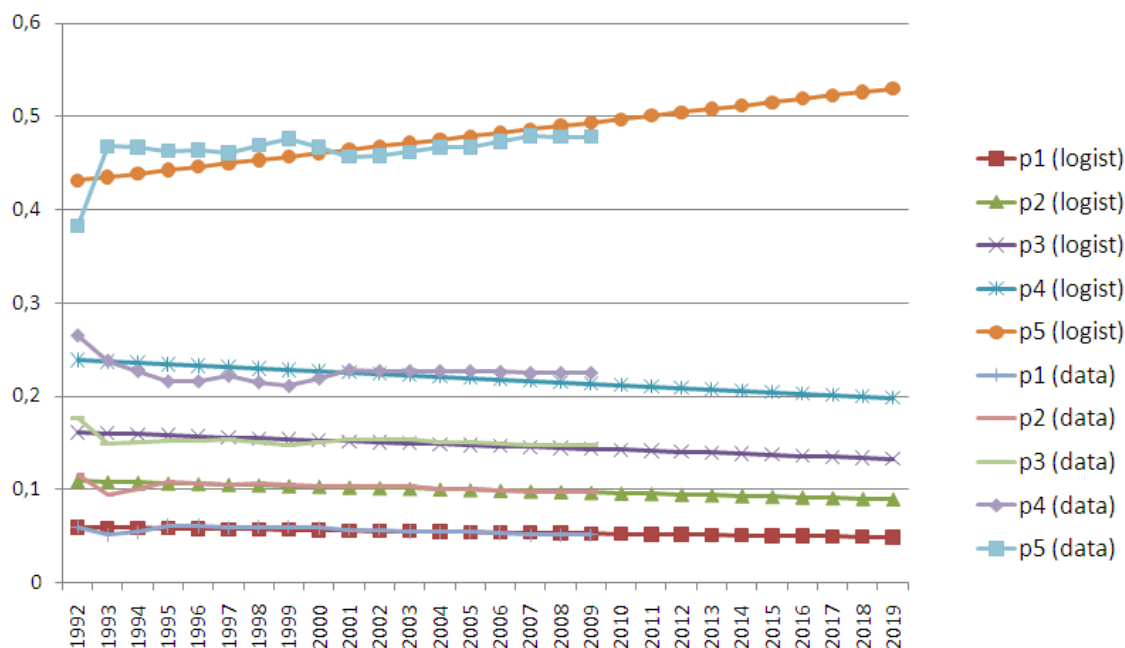


Рис. 2: Интерполяция на 1992–2009 гг. и экстраполяция на 2010–2019 гг. распределения доходов по слоям населения на основе цепочки с логистическим ростом (а также эмпирические данные за 1992–2009 гг.)

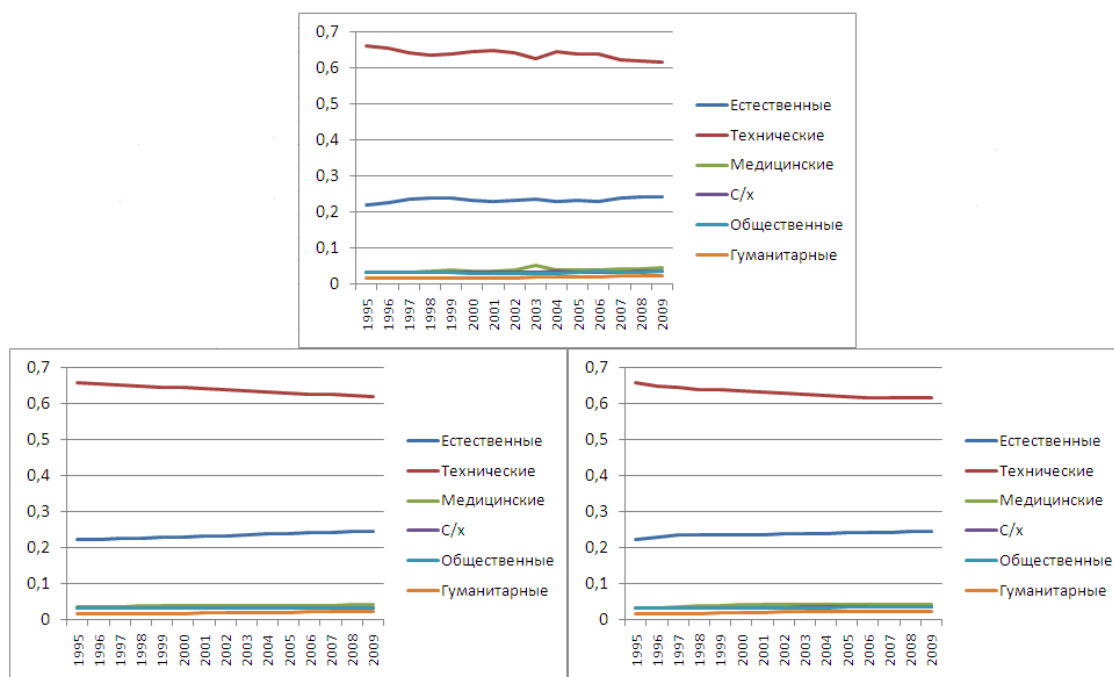


Рис. 3: Интерполяция динамики распределения исследователей по областям науки с логистическим ростом (слева) и логарифмически-линейная (справа), а также эмпирическая динамика (наверху) за 1995–2009 гг.

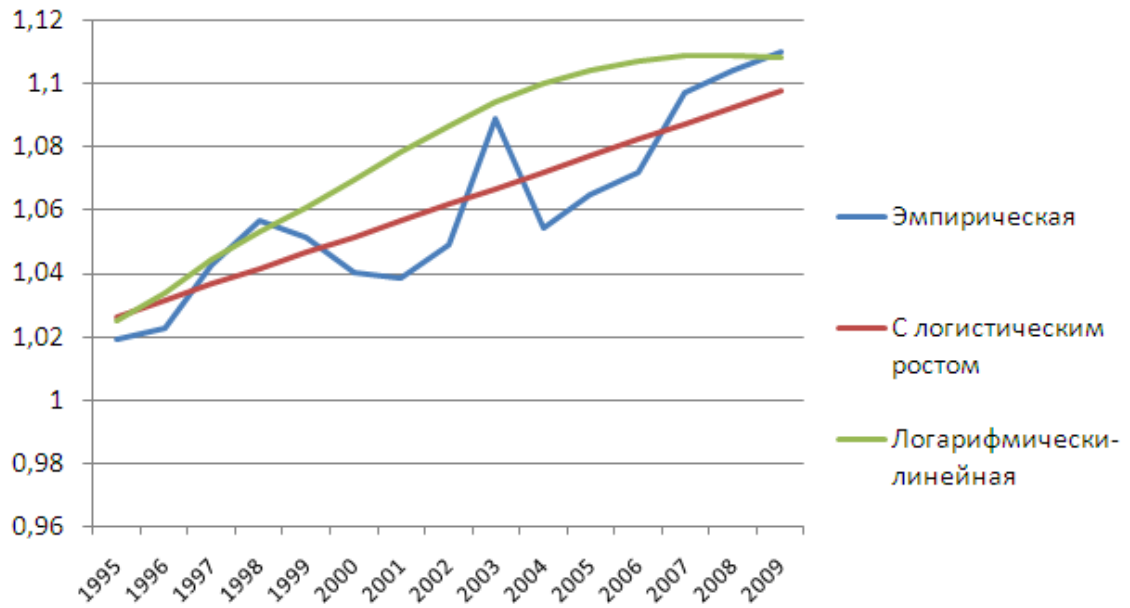


Рис. 4: Энтропия распределения исследователей за 1995–2009 гг.

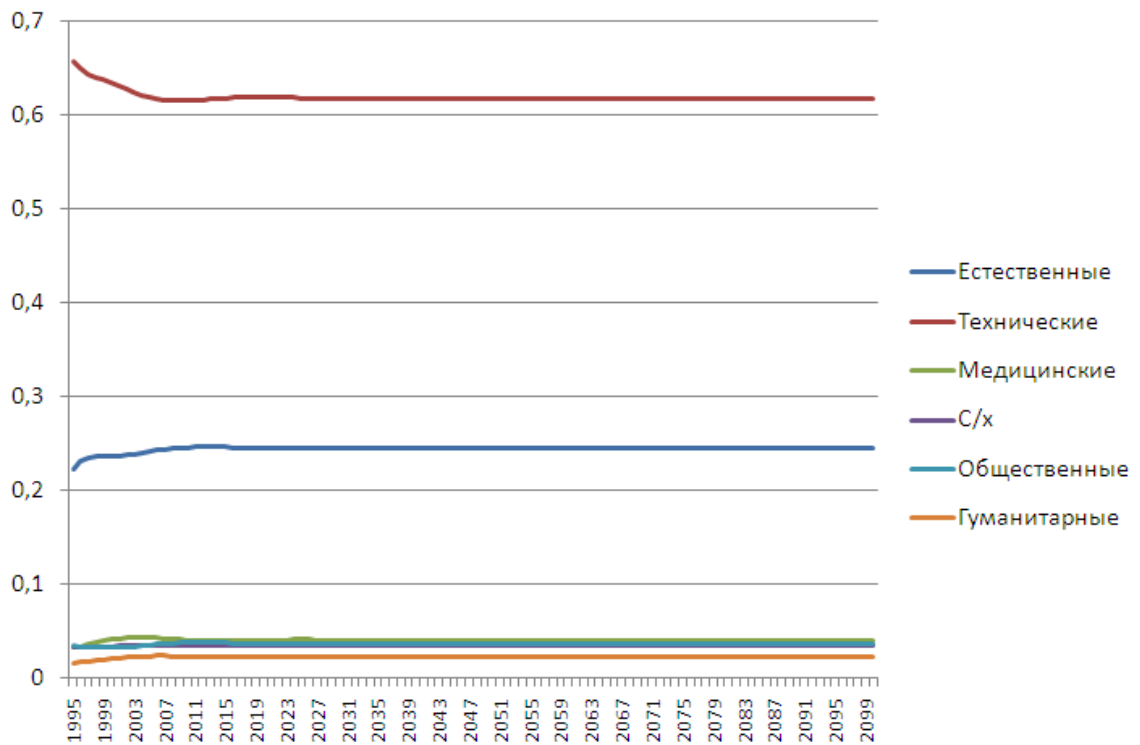


Рис. 5: Интерполяция и экстраполяция логарифмически-линейной динамики распределения исследователей по областям науки на 1995–2100 гг.

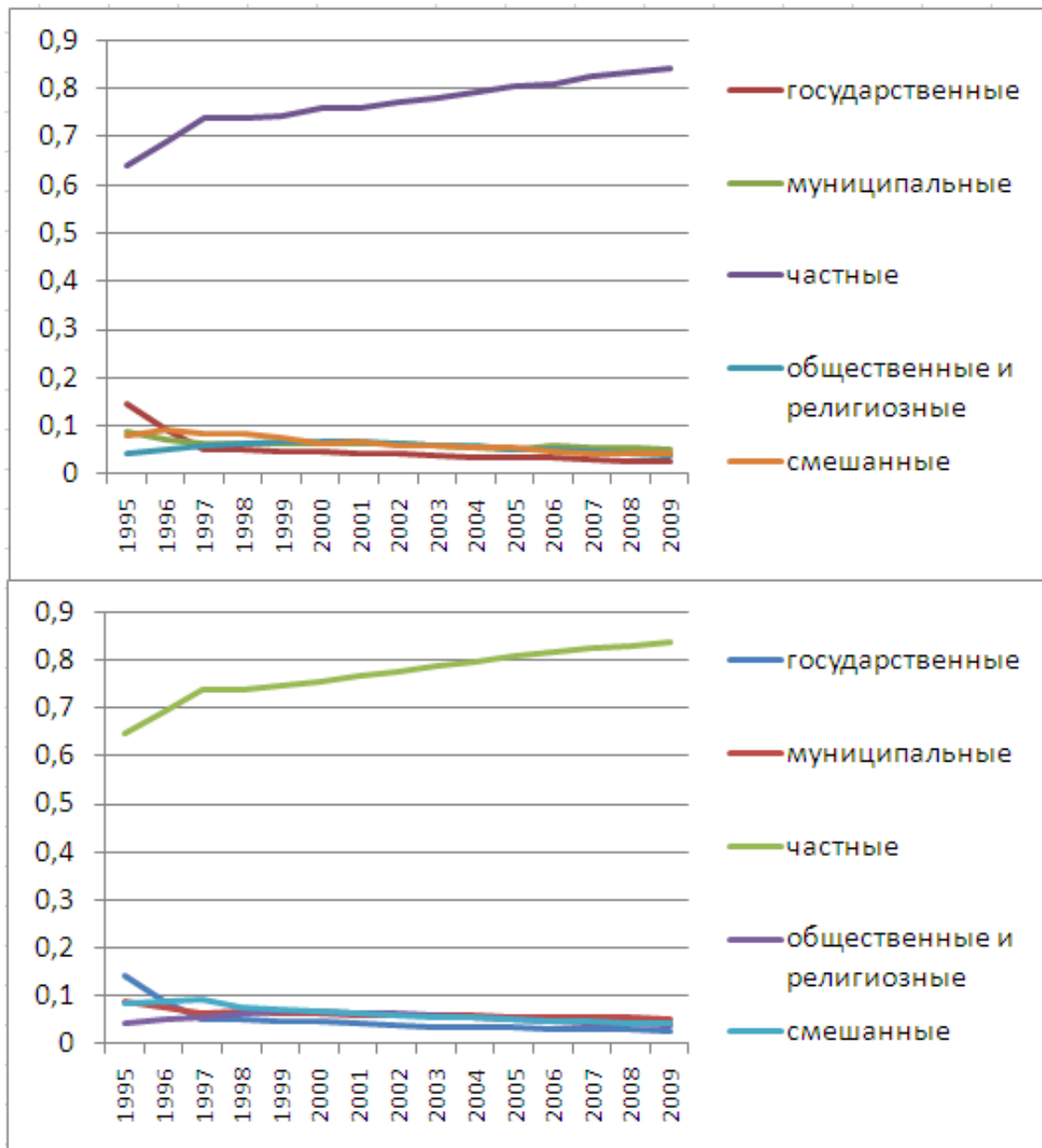


Рис. 6: Эмпирическая и логарифмически-линейная динамика распределения организаций по формам собственности за 1995–2009 гг.

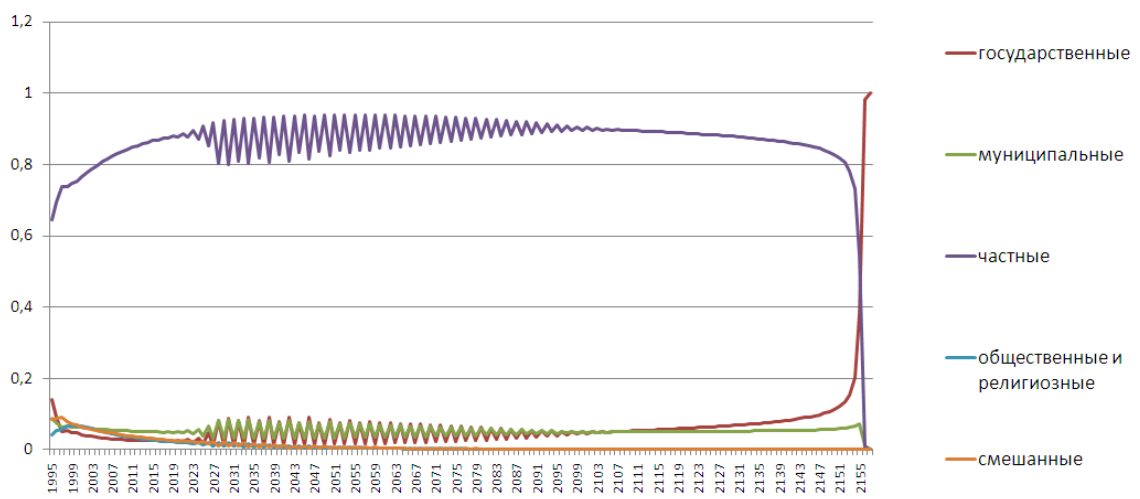


Рис. 7: Экстраполяция логарифмически-линейной динамики распределения организаций по формам собственности на 2010–2157 гг.

Матрицы и таблицы взаимного влияния

Ниже перечислены матрицы взаимного влияния для динамики распределения доходов по слоям населения РФ с логистическим ростом:

1. стандарт — первая группа;

$$\begin{pmatrix} 0 & 0.0006 & 0.0014 & 0.0018 & -0.0160 \\ -0.0006 & 0 & 0.0008 & 0.0012 & -0.0166 \\ -0.0014 & -0.0008 & 0 & 0.0004 & -0.0174 \\ -0.0018 & -0.0012 & -0.0004 & 0 & -0.0179 \\ 0.0157 & 0.0164 & 0.0171 & 0.0175 & 0 \end{pmatrix}$$

2. стандарт — вторая группа;

$$\begin{pmatrix} 0 & 0.0001 & 0.0007 & 0.0012 & -0.0140 \\ -0.0001 & 0 & 0.0007 & 0.0012 & -0.0140 \\ -0.0007 & -0.0007 & 0 & 0.0005 & -0.0147 \\ -0.0012 & -0.0012 & -0.0005 & 0 & -0.0152 \\ 0.0138 & 0.0138 & 0.0145 & 0.0150 & 0 \end{pmatrix}$$

3. стандарт — четвертая группа;

$$\begin{pmatrix} 0 & -0.0002 & -0.0009 & -0.0018 & -0.0162 \\ 0.0002 & 0 & -0.0007 & -0.0016 & -0.0160 \\ 0.0009 & 0.0007 & 0 & -0.0009 & -0.0153 \\ 0.0018 & 0.0016 & 0.0009 & 0 & -0.0144 \\ 0.0160 & 0.0158 & 0.0151 & 0.0142 & 0 \end{pmatrix}$$

4. стандарт — пятая группа.

$$\begin{pmatrix} 0 & 0.0046 & 0.0054 & 0.0048 & -0.0319 \\ -0.0046 & 0 & 0.0008 & 0.0002 & -0.0366 \\ -0.0054 & -0.0008 & 0 & -0.0007 & -0.0375 \\ -0.0048 & -0.0002 & 0.0007 & 0 & -0.0368 \\ 0.0309 & 0.0353 & 0.0361 & 0.0355 & 0 \end{pmatrix}$$

Перечислим теперь матрицы взаимного влияния для логарифмически-линейной динамики распределения доходов по слоям населения РФ:

1. стандарт — четвертая группа;

$$\begin{pmatrix} 2.211 & 1.48 & 0.773 & -0.463 \\ 5.195 & 3.495 & 1.686 & 1.925 \\ -5.183 & -2.269 & -1.025 & -4.417 \\ 7.886 & 5.517 & 2.739 & 0.694 \\ 7.63 & 6.489 & 3.295 & -1.435 \end{pmatrix}$$

2. стандарт — пятая группа.

$$\begin{pmatrix} 2.674 & 1.943 & 1.236 & 0.463 \\ 3.27 & 1.57 & -0.239 & -1.925 \\ -0.766 & 2.148 & 3.391 & 4.417 \\ 7.192 & 4.823 & 2.045 & -0.694 \\ 9.065 & 7.924 & 4.73 & 1.435 \end{pmatrix}$$

Приведем матрицы взаимного влияния для динамики распределения исследователей по областям науки с логистическим ростом:

1. стандарт — первая область (естественные науки);

$$\begin{pmatrix} 0 & 0.0113 & -0.0013 & 0.0052 & 0.0043 & -0.0228 \\ -0.0114 & 0 & -0.0128 & -0.0062 & -0.0071 & -0.0344 \\ 0.0013 & 0.0126 & 0 & 0.0065 & 0.0056 & -0.0214 \\ -0.0052 & 0.0062 & -0.0065 & 0 & -0.0009 & -0.0281 \\ -0.0043 & 0.007 & -0.0057 & 0.0009 & 0 & -0.0272 \\ 0.0223 & 0.0333 & 0.0209 & 0.0273 & 0.0264 & 0 \end{pmatrix}$$

2. стандарт — третья область (медицинские науки);

$$\begin{pmatrix} 0 & 0.0116 & -0.013 & 0.0069 & 0.0055 & -0.0206 \\ -0.0117 & 0 & -0.0249 & -0.0048 & -0.0061 & -0.0326 \\ 0.0129 & 0.0243 & 0 & 0.0197 & 0.0183 & -0.0075 \\ -0.0069 & 0.0048 & -0.0201 & 0 & -0.0014 & -0.0277 \\ -0.0056 & 0.0061 & -0.0187 & 0.0014 & 0 & -0.0263 \\ 0.0202 & 0.0316 & 0.0074 & 0.0269 & 0.0256 & 0 \end{pmatrix}$$

3. стандарт — четвертая область (сельскохозяйственные науки);

$$\begin{pmatrix} 0 & 0.0095 & -0.002 & 0.0032 & -0.0027 & -0.0213 \\ -0.0096 & 0 & -0.0076 & -0.0063 & -0.0069 & -0.0311 \\ 0.002 & 0.0075 & 0 & 0.0012 & 0.0007 & -0.0234 \\ -0.0032 & 0.0063 & -0.0012 & 0 & -0.0006 & -0.0246 \\ -0.0027 & 0.0069 & -0.0007 & 0.0006 & 0 & -0.0241 \\ 0.0209 & 0.0302 & 0.0228 & 0.024 & 0.0235 & 0 \end{pmatrix}$$

4. стандарт — пятая область (общественные науки);

$$\begin{pmatrix} 0 & 0.009 & 0.002 & 0.0037 & 0.0014 & -0.0199 \\ -0.0091 & 0 & -0.0071 & -0.0054 & -0.0077 & -0.0292 \\ -0.002 & 0.007 & 0 & 0.0017 & -0.0006 & -0.0219 \\ -0.0037 & 0.0053 & -0.0017 & 0 & -0.0023 & -0.0237 \\ -0.0014 & 0.0076 & 0.0006 & 0.0023 & 0 & -0.0213 \\ 0.0195 & 0.0283 & 0.0215 & 0.0231 & 0.0209 & 0 \end{pmatrix}$$

5. стандарт — шестая область (гуманитарные науки).

$$\begin{pmatrix} 0 & 0.012 & 0.0002 & 0.007 & 0.0074 & -0.0257 \\ -0.0121 & 0 & -0.012 & -0.0051 & -0.0047 & -0.0382 \\ -0.0002 & 0.0118 & 0 & 0.0068 & 0.0072 & -0.0259 \\ -0.007 & 0.0051 & -0.0069 & 0 & 0.0004 & -0.0329 \\ -0.0074 & 0.0047 & -0.0073 & -0.0004 & 0 & -0.0333 \\ 0.0251 & 0.0368 & 0.0252 & 0.0319 & 0.0323 & 0 \end{pmatrix}$$

Ниже приведены таблицы взаимного влияния для логарифмически-линейной динамики распределения организаций по формам собственности.

Таблица 12: Стандарт — муниципальные организации

| Форма собств. | Муниц. | Частн. | Обществ. и рел. | Смешан. |
|-----------------|--------|--------|-----------------|---------|
| Гос. | — | + | | + |
| Муниц. | | | | |
| Частн. | -- | ++ | | + |
| Обществ. и рел. | — | | ++ | |
| Смешан. | | | | + |

Таблица 13: Стандарт — частные организации

| Форма собств. | Муниц. | Частн. | Обществ. и рел. | Смешан. |
|-----------------|--------|--------|-----------------|---------|
| Гос. | — | — | — | |
| Муниц. | | | | |
| Частн. | -- | -- | -- | |
| Обществ. и рел. | -- | | ++ | |
| Смешан. | | | | + |

Таблица 14: Стандарт — общественные и религиозные организации

| Форма собств. | Муниц. | Частн. | Обществ. и рел. | Смешан. |
|-----------------|--------|--------|-----------------|---------|
| Гос. | -- | | + | + |
| Муниц. | | | | + |
| Частн. | -- | | + | + |
| Обществ. и рел. | -- | -- | -- | -- |
| Смешан. | — | | + | + |

Таблица 15: Стандарт — смешанные организации

| Форма собств. | Муниц. | Частн. | Обществ. и рел. | Смешан. |
|-----------------|--------|--------|-----------------|---------|
| Гос. | -- | — | | — |
| Муниц. | — | | | — |
| Частн. | -- | — | — | — |
| Обществ. и рел. | -- | | | ++ |
| Смешан. | -- | — | | — |

Примеры задач, моделируемых с помощью разработанной программы

Приведем список примеров задач, в которых процесс распределения ресурсов был смоделирован средствами разработанной программы:

1. Распределение эмигрантов, выбывших из РФ, по направлению передвижения;
2. Распределение эмигрантов, выбывших из РФ в страны СНГ, по направлению передвижения;
3. Распределение эмигрантов, выбывших из РФ в страны дальнего зарубежья, по направлению передвижения;
4. Распределение эмигрантов, выбывших из СПб, по направлению передвижения;
5. Распределение численности занятых в экономике по форме собственности организации;
6. Распределение видов затрат на ИКТ в организациях в РФ;
7. Распределение видов затрат домашних хозяйств в РФ;
8. Распределение по федеральным округам РФ внутренних иммигрантов в Центральный федеральный округ;
9. Распределение пассажирооборота по видам транспорта в РФ;
10. Распределение потребительских расходов домашних хозяйств в РФ;
11. Распределение иммигрантов, прибывших в РФ из стран СНГ;
12. Распределение иммигрантов, прибывших в РФ из стран дальнего зарубежья;
13. Распределение валового регионального продукта по федеральным округам РФ;
14. Распределение эмигрантов, выбывших из РФ, по федеральным округам;
15. Распределение иммигрантов, прибывших в РФ, по федеральным округам;

16. Распределение расходов бюджета РФ на социокультурные мероприятия;
17. Распределение расходов бюджета РФ;
18. Распределение доходов по источникам в СПб;
19. Распределение потребляемого электричества по странам Северной Европы;
20. Распределение потребляемой энергии по странам Восточной Европы;
21. Распределение Интернет-пользователей по странам Северной Европы;
22. Распределение абонентов сотовой связи по странам Восточной Европы.

В каждой из перечисленных задач логарифмически-линейная динамика исследуемого распределения демонстрирует нетривиальное поведение.

Список литературы

- [1] Афанасьева Е.В. Вероятностные цепочки: моделирование распределения ресурсов // Процессы управления и устойчивость: Труды 41-й международной научной конференции аспирантов и студентов / Под ред. Н.В. Смирнова, Г.Ш. Тамасяна. — СПб.: Издат. Дом С.-Петербур. гос. ун-та, 2010. — С. 547–552.
- [2] Афанасьева Е.В., Ампилова Н.Б. Моделирование процессов потребления ресурсов странами Западной Европы с помощью вероятностных цепочек // Технологии Microsoft в теории и практике программирования: материалы межвузовского конкурса-конференции студентов, аспирантов и молодых ученых Северо-Запада. — СПб.: Изд-во Политехн. ун-та, 2011. — С. 126–127.
- [3] Афанасьева Е.В., Ампилова Н.Б. Применение вероятностных цепочек для моделирования процессов распределения ресурсов // Технологии Microsoft в теории и практике программирования: материалы межвузовского конкурса-конференции студентов, аспирантов и молодых ученых Северо-Запада. — СПб.: Изд-во Политехн. ун-та, 2010. — С. 132–133.
- [4] Бородич С.А. Эконометрика: Учеб. пособие. — Минск: Новое знание, 2001. — 408 с.
- [5] Бородкин Л.И. Методология анализа неустойчивых состояний в политико-исторических процессах // Международные процессы. — 2005. — Т. 3. — № 7. — С. 4–16.
- [6] Буданов В.А. Синергетика: мировоззрение, методология, наука // Экономические стратегии. — 2010. — № 5. — С. 48–57.
- [7] Булатов А.С. Экономика. Учебное пособие. — М.: Юристъ, 2002. — 896 с.
- [8] Данилов Ю.А. Нелинейность // Знание — сила. — 1982. — № 11. — С. 34.
- [9] Магнус Я.Р., Катышев П.К., Пересецкий, А.А. Эконометрика. Начальный курс. — М.: Дело, 2004. — 576 с.
- [10] Кобзарь А.И. Прикладная математическая статистика. — М.: Физматлит, 2006. — 238 с.

- [11] Козлов Д.А. Методы нелинейной динамики в моделировании макроэкономических процессов // Научные труды: Институт народнохозяйственного прогнозирования РАН. — 2003. — Т. 12. — С. 157–173.
- [12] Кремер Н.Ш., Путко Б.А. Эконометрика. — М.: ЮНИТИ-ДАНА, 2002. — 311 с.
- [13] Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ-ДАНА, 2009. — 551 с.
- [14] Кузьмин П.И. Эконометрические модели. Учебно-методическое пособие. — Барнаул: Изд-во Алт. ун-та, 2004. — 199 с.
- [15] Кузнецов С.П. Динамический хаос (курс лекций). — М.: Издательство Физико-математической литературы, 2001. — 296 с.
- [16] Лемешко Б.Ю., Лемешко С.Б. // Метрология. — 2005. — № 2. — С. 3–23.
- [17] Российский статистический ежегодник 2010: Стат. сб. — М.: Росстат, 2010. — 813 с.
- [18] Социальное положение и уровень жизни населения России. Стат. сб. — М.: Росстат, 2010. — 507 с.
- [19] Эконометрика-3: курс лекций / Суслов В.И., Лапо В.Ф., Талышева Л.П., Ибрагимов Н.М. — Красноярск: Сиб. федерал. ун-т, 2008. — 194 с.
- [20] Dall'erba S. Competition, Complementarity and Increasing Disparities Among the Regions of Spain and Portugal // *Revue d'Economie Regionale et Urbaine*. — 2004. — № 2. — P. 311–330.
- [21] Dendrinos D.S., Sonis M. Nonlinear Discrete Relative Population Dynamics of the U.S. Regions // *Journal of Applied Mathematics and Computation*. — 1988. — № 22. — P. 25–44.
- [22] Dufour J.-M. Seemingly Unrelated Regressions. — [S.I.], 2008.
- [23] Healy R., Jackson R.W. Competition and Complementarity in Local Economic Development: A Nonlinear Dynamic Approach // *Studies in Regional and Urban Planning*. — 2001. — № 9. — P. 21–36.
- [24] Hewings G.J.D., Sonis M., Cuello F.A., Mansouri F. The Role of Regional Interaction in Regional Growth: Competition and Complementarity in the US Regional System // *Australasian Journal of Regional Science*. — 1996. — № 2. — P. 133–149.

- [25] Henningsen A., Hamann J.D. Systemfit: a Package to Estimate Simultaneous Equation Systems in R. — [S.I.], 2006.
- [26] Kamarianakis Y., Kaslis V. Geographical Competition-Complementarity Relationships between Greek Regional Economies // European Regional Science Association Conference Papers. — 2005.
- [27] Magalhaes A., Sonis M., Hewings G.J.D. Regional Competition and Complementarity Reflected in Relative Dynamics and Growth of GSP: A Comparative Analysis of the Northeast of Brazil and the Midwest States of the U.S. // Structure and Structural Change in the Brazilian Economy / J.J.M. Guilhoto, G.J.D. Hewings. — [S.I.], 1999.
- [28] Moon H.R., Perron B. Seemingly Unrelated Regressions // The New Palgrave Dictionary of Economics / S.N. Durlauf, L.E. Blume. — N.Y.: Palgrave Macmillan, 2008.
- [29] Nazara S., Sonis M., Hewings G.J.D. Interregional Competition and Complementarity in Indonesia // The Regional Economics Applications Laboratory Discussion Paper. — 2001.
- [30] Sonis M. Discrete non-linear probabilistic chains // Functional differential equations. — 2003. — Vol. 10. — № 3–4. — P. 593–639.
- [31] Sonis M., Azzoni C. R., Hewings G. J. D. The Three-sector Growth Hypothesis and the Euler-Malthus Economic growth model: Application to the analysis of GDP dynamics of Brazil, 1985–2004–2020 // The Fifth International Conference on Mathematical Modeling and Computer Simulation of Materials Technologies. — 2008. — P. 153–163.
- [32] Sonis M., Hewings G. Regional Competition and Complementarity: Comparative Advantages/Disadvantages and Increasing/Diminishing Returns in Discrete Relative Spatial Dynamics // Regional Competition Advances in Spatial Science / P. Batey, P. Friedrich. — Berlin: Springer-Verlag, 2001. — P. 139–157
- [33] Yoder J. Seemingly Unrelated Regression // WSU Econometrics II. — 2007. — P. 38–43.