Компьютерные инструменты в образовании, 2018
№ 2: 14–24
УДК: 330.42 : 338.27
http://ipo.spb.ru/journal

ОБ ОДНОМ МЕТОДЕ МОДЕЛИРОВАНИЯ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

Логинова Н. В.[1]

[1]Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

Аннотация

В экономической науке давно изучаются возможности применения математического аппарата для проведения наиболее полного и точного анализа и построения надежного прогноза исследуемых экономических процессов. Существует достаточно много математических моделей, позволяющих сконструировать динамику изменения социально-экономических данных. В настоящей статье продемонстрировано применение двух видов дискретных вероятностных цепочек для прогнозирования динамики экономических показателей. Проведена проверка входных данных с помощью статистических критериев. Произведён анализ согласованности полученных результатов с эмпирической динамикой. Показано, что выполнение определенных критериев для входных данных является важным условием для получения правдоподобного прогноза.

Ключевые слова: *динамические системы, вероятностные цепочки, статистические критерии, экстраполяция, эконометрический анализ.*

Цитирование: Логинова Н. В. Об одном методе моделирования динамики социально-экономических процессов // Компьютерные инструменты в образовании. 2018. № 2. С. 14–24.

1. ВВЕДЕНИЕ

Современные исследователи для изучения экономических процессов используют различные научные подходы, среди которых особый интерес вызывает математическое моделирование. Модели многих естественных процессов описываются с помощью динамических систем. Мы рассматриваем класс дискретных динамических систем, которые могут быть использованы в моделировании различных процессов, связанных с распределением ресурсов. Подобного рода задачи возникают в общественных науках и экономике. Если распределение между участниками процесса выражается в относительных долях, то определение закона развития динамики такого процесса означает преобразование некоторого вероятностного вектора. Поэтому эти модели, предложенные и разработанные М. Сонисом [9–11] получили название вероятностных цепочек. Они являются обобщением хорошо известных цепей Маркова, где функции преобразования координат вероятностного вектора (так называемые порождающие функции) могут быть заданы различными способами. В зависимости от способа задания функций выделяют несколько классов цепочек. Наибольшее распространение получили модели с логистическим и линейно-логарифмическим ростом. Применение вышеуказанных методов для исследования финансовых показателей было представлено в работах Афанасьевой Е. В. [1, 2].

При построении разного рода математических моделей особое значение имеет оценка их качества. Необходимо понимать, насколько достоверны построенные модели экономических процессов, насколько надёжно проектировать на их основе прогноз дальнейшего изменения. Нужно отметить, что для таких моделей пока не предложен простой метод анализа устойчивости метода. Классический анализ бифуркаций неподвижных точек или периодических орбит, который возможен в дискретных динамических системах, требует применения математических пакетов, поскольку получаемые системы имеют довольно большой порядок. Наглядный пример исследования бифуркаций системы 3 порядка в терминах собственных значений матрицы Якоби был проведен М. Сонисом. Но уже для системы порядка 5 такой анализ трудно выполнить вручную.

Тем не менее, первый шаг в оценке надежности модели может быть проведен на основании проверки входных данных. Для этого в представленной работе проводится оценка входных данных с помощью статистических критериев. Применимость модели оценивается по достаточному числу проведенных экспериментов.

В настоящей работе описывается применение двух видов вероятностных цепочек к разным наборам данных. Моделирование реализовано с помощью разработанной автором программы NLPC Predictor. Для статистически достоверных данных обе модели показывают надежные результаты, в то время как невыполнение статистических критериев приводит к получению неправильного прогноза. Рассматриваемые показатели проверяются как в программе, так и с помощью пакета Gretl. Оценка достоверности выбранных методов моделирования основана на согласованности результатов прогноза с эмпирическими данными.

2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Вероятностным вектором назовем набор величин p_i, обладающих следующими свойствами:
$p_i \geq 0, i = 1, \ldots, N$, где N — число групп, участвующих в процессе

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

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

$$\bar{p}_t = \begin{pmatrix} p_{1t} \\ p_{2t} \\ \vdots \\ p_{nt} \end{pmatrix}, \quad t = 0, 1, 2, \ldots, \quad 0 \leq p_{kt} \leq 1, \quad \sum_{k=1}^{n} p_{kt} = 1.$$

Обозначение $(1, n)$ интерпретируется в контексте решаемой задачи, например поведение одной популяции на n территориях или же распределение ресурсов между n территориями или между n популяциями и т. д.

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

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

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

В зависимости от порождающих функций, которыми они задаются, выделяют различные виды вероятностных цепочек. Далее будут рассмотрены вероятностные цепочки с логистическим и линейно-логарифмическим ростом.

Вероятностные цепочки с логистическим ростом определяются набором порождающих функций

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

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

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

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

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

где $A_1, \ldots, A_n$ и элементы матрицы $[a_{ij}]_{i,j=1}^{n}$ — параметры, определяемые в контексте задачи. Логарифмически-линейные вероятностные цепочки могут быть представлены в следующем виде:

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

где $k = 1, 2, \ldots, n$, $t = 0, 1, 2, \ldots, T$ [10].

3. СТАТИСТИЧЕСКИЕ КРИТЕРИИ ДЛЯ ПРОВЕРКИ ДАННЫХ

Перед построением вероятностных цепочек необходимо провести проверку входных данных. Проверка необходима для того, чтобы построенные модели были реалистичны и соответствовали эмпирической динамике. Произвести оценку входных данных можно с помощью существующих пакетов: SPSS, Maple, Matlab, Mathcad, Mathematica и другие. Мы использовали программу Gretl, предназначенную для построения линейных и нелинейных эконометрических моделей. Программа содержит простой метод наименьших квадратов для линейных моделей и нелинейный метод наименьших квадратов, основанный на алгоритме Левенберга-Маквардта для нелинейных моделей. Она содержит встроенный в систему калькулятор тестовых характеристик, который позволяет вычислить различные статистические характеристики и вероятности случайных величин при проверке разных общих гипотез по одной или двум популяциям [13]. Программа позволяет проанализировать эмпирические данные с помощью одновыборочного и двухвыборочного критериев Стьюдента, проверить статистическую гипотезу о равенстве коэффициента вариации определенному значению, найти пропорцию. Также Gretl позволяет проверить равенство коэффициента вариации между двумя выборками и равенство пропорции между двумя выборками. Для проверки данных были взяты такие статистические характеристики, как критерий Фишера, критерий Стьюдента и критерий Шапиро-Уилка.

Критерий Стьюдента. Применяется для проверки нулевой гипотезы $H_o : E(X) = m$ о равенстве математического ожидания $E(X)$ некоторому известному значению. В качестве известного значения m берётся математическое ожидание по всему набору исследуемых совокупностей. С учётом предполагаемой независимости наблюдений $V(\bar{X}) = \sigma^2/n$. Используя несмещенную оценку дисперсии $s_X^2 = \sum_{t=1}^{n}(X_t - \bar{X})^2/(n-1)$, получаем следующую t-статистику:

$$t = \frac{\bar{X} - m}{s_X/n}, k = 1, 2, \dots, n, t = 0, 1, 2, \dots, T.$$

При нулевой гипотезе распределение этой статистики $t(n-1)$. При превышении значения статистики по абсолютной величине критического значения данного распределения (при заданном уровне значимости) нулевая гипотеза отвергается [3, 4].

Критерий Фишера для проверки равенства дисперсий у нескольких выборок. Пусть выборка объёмом N случайной величины X разделена на k групп с количеством наблюдений n_i в i-ой группе.

Межгрупповая дисперсия: $\hat{\sigma}_{BG}^2 = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2/(k-1)$.

Внутригрупповая дисперсия: $\hat{\sigma}_{WG}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} n_i(\bar{x_{ij}} - \bar{x}_i)^2/(N-k)$.

$$F = \frac{\hat{\sigma}_{BG}^2}{\hat{\sigma}_{WG}^2}.$$

Если статистика превышает критическое значение, то гипотеза о равенстве дисперсий в выборках отвергается, в противном случае дисперсии можно считать одинаковыми [5–7].

Критерий Шапиро-Уилка для проверки нормальности распределения. Пусть имеется выборка $(x_1, \dots, x_n)$. Вычисления статистики производятся по формулам [8]:

$$W = \frac{b^2}{S^2},$$

$$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2,$$

$$b = \sum_{i=1}^{k} a_{ni}(x_{n-i+1} - x_i)^2.$$

4. ОПИСАНИЕ ПРОГРАММЫ NLPC PREDICTOR

Для построения вероятностных цепочек с логистическим и линейно-логарифмическим ростом было разработано десктопное приложение NLPC Predictor. Программа предоставляет возможность пользователю загрузить массивы данных, произвести их анализ с помощью статистических критериев, построить графическое представление динамики эмпирических данных, а также построить два вида вероятностных цепочек. Кроме того, программный продукт позволяет произвести пользователю экстраполяцию на любой выбранный период, выгрузить полученные результаты в excel. Для разработки был использован язык C# 6, среда Visual Studio 2015.

Пример 1. Рассмотрим данные по потреблению основных продуктов питания по Российской Федерации за период 1990–2005 [12].

Таблица 1. Потребление основных продуктов питания по РФ
(килограммов на душу населения в год)

год	картофель	овощи	фрукты	мясо	молоко	яйца	хлеб
1990	106	89	35	75	387	297	120
1991	113	86	35	69	347	288	121
1992	118	77	32	60	282	263	125
1993	127	71	29	59	294	251	123
1994	123	68	28	57	281	238	124
1995	124	76	29	55	254	216	122
1996	123	74	31	51	233	208	118
1997	122	76	33	50	230	211	118
1998	114	74	30	48	220	217	117
1999	108	79	27	45	214	221	119
2000	109	79	32	45	215	229	117
2001	109	81	35	47	219	235	120
2002	106	80	39	50	227	244	121
2003	109	84	39	52	231	246	120
2004	108	85	43	54	232	243	119
2005	109	87	46	55	234	250	121

В результате проверки данных по одностороннему критерию Стьюдента были получены значения для каждого из видов продуктов питания: 0,3322, 1,42, 0,0192, 1,5331, 0,4334, 0,7514, 0,6187. Во всех случаях вычисленный критерий получился меньше критического значения равного 2,1314 для уровня значимости 0,05.

В результате проверки данных по одностороннему критерию Фишера дисперсии случайных величин признаются одинаковыми при критическом значении 2,17 для уровня значимости 0,05.

В результате проверки критерием Шапиро-Уилка все данные получились распределены согласно нормальному закону при критическом значении 0.844 для уровня значимости 0,05.

Во всех трёх критериях была принята нулевая гипотеза, что говорит о том, что на основе выбранных эмпирических данных можно построить достоверные модели с логистическим и линейно-логарифмическим ростом.

Затем проводим проверку гипотез в Gretl. Из имеющихся гипотез во встроенном калькуляторе тестовых характеристик была выбрана проверка о равенстве средних и равенстве дисперсий.

Проверяем гипотезу о равенстве средних по данным картофеля и хлеба. Получаем средние 114,25 и 120,313. Вычисленная тестовая статистика равна −3,14038. Нулевая гипотеза о равенстве средних принимается при уровне значимости 0,05. Проверяем гипотезу о равенстве средних по данным мяса и молока. Получаем средние 256,25 и 241,063. Вычисленная тестовая статистика равна 1,07551. Нулевая гипотеза о равенстве средних принимается при уровне значимости 0,05. Аналогично проводим этот тест для остальных данных.

Далее проверим гипотезу об отношении двух дисперсий (вариаций). Получаем дисперсии 29,7958 по данным овощей и 35,7167 по данным фруктов. Вычисленная тестовая статистика равна 1,1987. Принимается нулевая гипотеза о том, что дисперсии равны при уровне значимости 0,05. Получаем дисперсиии 54,067 по данным картофеля и 29,7958 по данным овощей. Вычисленная тестовая статистика равна 1,8146. Принимается нулевая гипотеза о том, что дисперсии равны при уровне значимости 0,05. Также выполняем тест для остальных данных.

Таким образом, в результате проведённой проверки экономических показателей статистическими критериями принималась нулевая гипотеза, что говорит о том, что исследуемые эмпирические данные могут быть использованы для построения вероятностных цепочек.

Рассмотрим результаты моделирования, представленные в графическом виде. Экстраполяция выполнялась на 5 лет (рис. 1, 2).

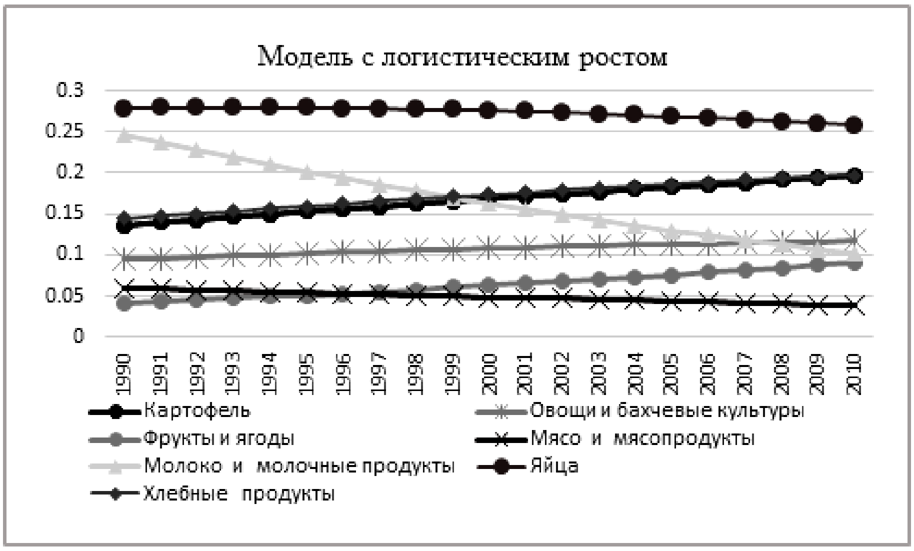

Рис. 1. Модель с логистическим ростом

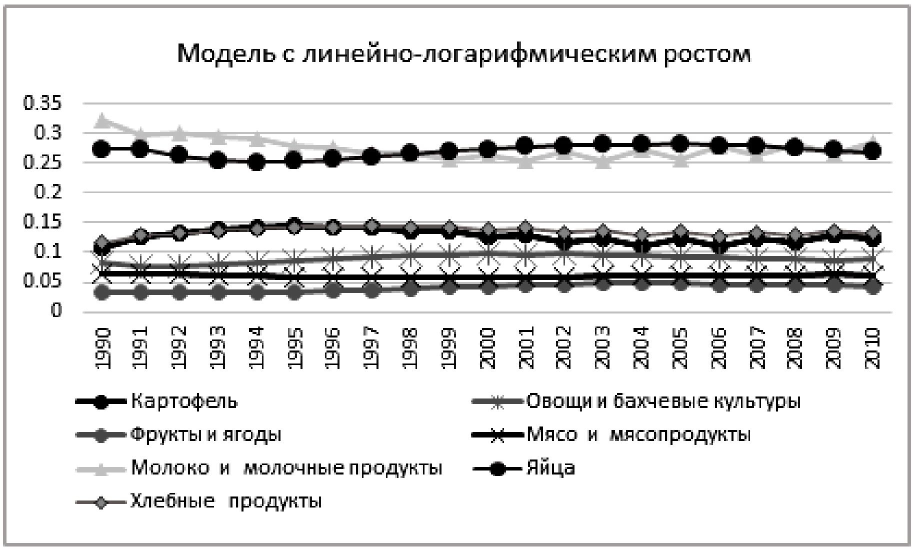

Рис. 2. Модель с линейно-логарифмическим ростом

Далее построим график на основе эмпирических данных (рис. 3).

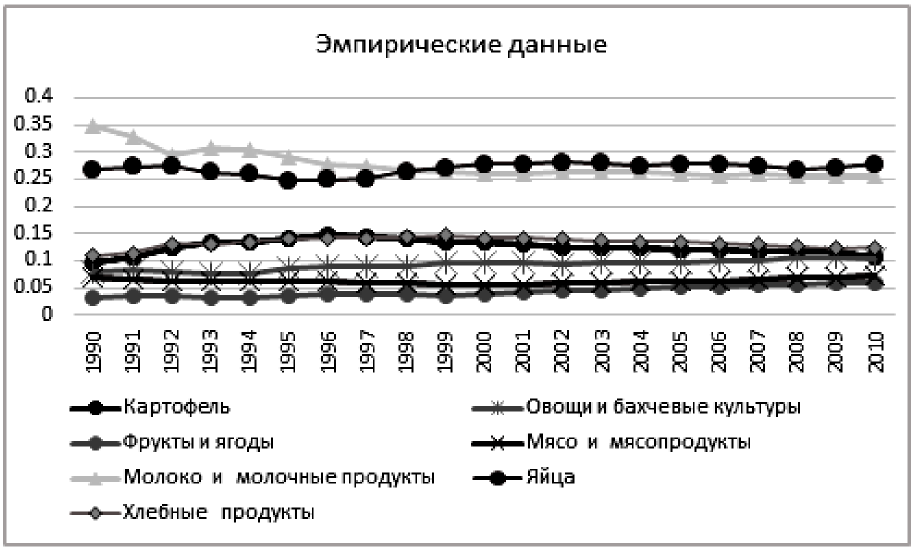

Рис. 3. Эмпирические данные

Из построенных графиков видно, что построенная экстраполяция для цепочек с логистическим ростом согласуется с эмпирическими данными за период с 2006 по 2010 года по данным потребления мяса. Кроме того, экстраполяция для цепочек с линейно-логарифмическим ростом согласуется с эмпирической динамикой данных потребления картофеля, мяса, фруктов.

Пример 2. Рассмотрим данные по урожайности сельскохозяйственных культур по Российской Федерации с 1990 по 2005 года [12].

Таблица 2. Урожайность сельскохозяйственных культур по РФ (центнеров с гектара)

год	пшеница	рожь	ячмень	овес	кукуруза	просо
1990	21,0	21,0	20,5	14,8	31,5	12,3
1991	17,3	16,5	15,1	12,1	29,1	6,7
1992	19,6	18,5	19,1	14,4	29,0	9,1
1993	18,2	15,8	18,1	15,2	32,0	8,9
1994	15,3	15,6	17,1	14,2	19,4	6,1
1995	13,9	13,2	12,7	12,2	28,7	9,5
1996	15,5	14,9	15,1	13,9	23,5	6,4
1997	18,4	19,2	17,6	16,1	31,3	12,7
1998	13,5	10,2	13,8	11,8	16,3	8,4
1999	15,7	14,7	14,3	11,3	19,7	9,3
2000	16,1	15,8	16,7	14,7	21,2	8,2
2001	20,6	18,8	20,1	17,1	18,0	7,9
2002	20,7	19,0	19,7	15,6	28,5	8,5
2003	17,1	18,6	19,6	16,8	32,0	13,9
2004	19,8	15,4	18,0	15,1	40,3	11,9
2005	19,3	15,7	18,1	14,4	38,5	11,2

Сначала проверим вышеуказанные показатели программой NLPC Predictor с помощью статистических критериев.

В результате проверки данных по одностороннему критерию Стьюдента были получены значения для каждой из сельскохозяйственных культур: 0,0362, 0,0398, 0,0095, 0,2659, 9,02476, 0,2307. Не во всех случаях вычисленный критерий получился меньше критического значения равного 2,2281 для уровня значимости 0,05.

В результате проверки данных сельскохозяйственных культур по одностороннему критерию Фишера дисперсии случайных величин признаются не одинаковыми при критическом значении 2,04 для уровня значимости 0,05.

В результате проверки критерием Шапиро-Уилка часть данных получилась не распределенной по нормальному закону при критическом значение равным 0,792 для уровня значимости 0,05.

В результате проверки с помощью статистических критериев не во всех случаях была принята нулевая гипотеза. В данном случае можно предположить, что построенные модели вероятностных цепочек будут не достоверны.

Затем проводим проверку гипотез в Gretl. Проверяем гипотезу о равенстве средних по данным пшеницы и овса. Получаем средние 17,625 и 14,3562. Вычисленная тестовая статистика равна 4,36141. Нулевая гипотеза о равенстве средних отвергается при уровне значимости 0,05. Далее проводим этот тест с остальными данными.

Проверим гипотезу об отношении двух дисперсий (вариаций). Сравним дисперсии выборок данных сельскохозяйственных культур. Проверяем нулевую гипотезу о том, что дисперсии равны по данным пшеницы и данным кукурузы. Получаем дисперсии 5,95267 и 50,7198. Вычисленная тестовая статистика равна 8,52052. Нулевая гипотеза о том, что дисперсии равны при уровне значимости 0,05, отвергается. То же самое делаем для остальных данных. В результате почти во всех случаях была принята альтернативная гипотеза.

Таким образом, в результате проверки данных статистическими характеристиками, имеющимися в стороннем пакете GRETL и NLPC Predictor, получаем вывод о том, что исследуемые эмпирические данные нельзя использовать для построения вероятностных цепочек. Можно предположить, что полученные вероятностные цепочки не будут согласовываться с эмпирической динамикой.

Рассмотрим результаты моделирования, представленные в графическом виде. Экстраполяция выполнялась на 5 лет (рис. 4, 5).

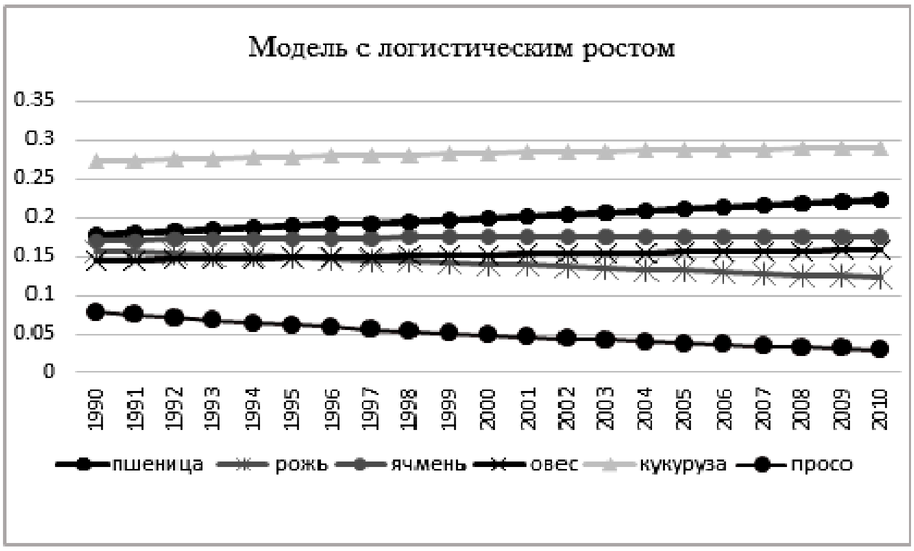

Рис. 4. Модель с логистическим ростом

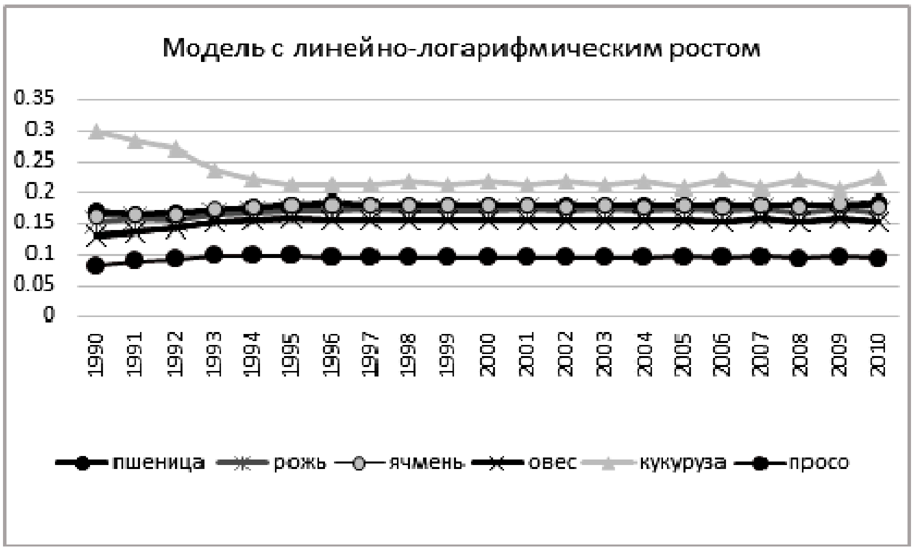

Рис. 5. Модель с линейно-логарифмическим ростом

Далее построим график на основе эмпирических данных (рис. 6).

Рис. 6. Эмпирические данные

Из построенных графиков видно, что экстраполяция для цепочек с логистическим ростом согласуется с эмпирическими данными за период с 2006 по 2010 годы для данных об урожайности пшеницы, ржи и ячменя. Для остальных данных построенные вероятностные цепочки не согласуются с эмпирической динамикой.

5. ЗАКЛЮЧЕНИЕ

Таким образом, в представленной работе на примере двух наборов данных были построены вероятностные цепочки с логистическим и линейно-логарифмическим ростом с помощью разработанной программы NLPC Predictor. Также была проведена проверка входных данных, выбранных для построения вероятностных цепочек с использованием статистических критериев. В результате было выяснено, что в случае если после проверки данных статистическими критериями принималась нулевая гипотеза, то построенные результаты прогноза лучше согласовывались с эмпирической динамикой. Если же после проверки данных принималась альтернативная гипотеза, то результаты не согласовывались или же согласовывались лишь частично с эмпирической динамикой.

Список литературы

1. *Афанасьева Е. В.* Моделирование процессов распределения ресурсов с помощью вероятностных цепочек // Дифференциальные уравнения и процессы управления. 2011. № 3. С. 84–137.
2. *Афанасьева Е. В.* Моделирование процессов потребления экономических ресурсов с помощью вероятностных цепочек (на примере стран Западной Европы) // Научно-технические ведомости СПбГПУ: Информатика. Телекоммуникации. Управление. СПб.: Политехн. ун-та. 2011. № 3. С. 93–97.
3. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика: начальный курс. Академия народного хозяйства (М.). 6-е изд., перераб. и доп. М. : Дело, 2009.
4. *Носко В. П.* Эконометрика: в 2 кн.: учебник для вузов. М.: Дело, 2011.
5. *Вербик М.* Путеводитель по современной эконометрике. М.: Научная книга, 2008.
6. *Айвазян С. А., Иванова С. С.* Эконометрика. М.: Маркет ДС, 2010.
7. *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 2008.
8. *Елисеева И. И.* Эконометрика. М.: Финансы и статистика, 2005.
9. *Sonis M.* Discrete Non-Linear Probabilistic Chains (M. Drachlin and E. Litsyn eds) // Functional-Differential Equations, Ariel, Israel, 2003. № 10. P. 445–487.
10. *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.
11. *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: SpringerVerlag, 2001. P. 139–157.
12. Федеральная служба государственной статистики [Online]. Available: http://www.gks.ru (дата обращения: 26.04.2018).
13. *Allin Cottrell.* Department of Economics, Wake Forest University. Riccardo "Jack"Lucchetti. Department of Economics, Marche Polytechnic University. Gretl User’s Guide. Gnu Regression, Econometrics and Time-series Library [Online]. Available: http://gretl.sourceforge.net/gretl-help/gretl-guide.pdf (дата обращения: 26.04.2018).

Поступила в редакцию 19.03.2018, окончательный вариант — 26.04.2018.

Computer tools in education, 2018
№ 2: 14–24
http://ipo.spb.ru/journal

ON A METHOD OF MODELING OF SOCIO-ECONOMIC PROCESSES DYNAMICS

Loginova N. V.[1]

[1]Saint-Petersburg State University, Saint Petersburg, Russia

Abstract

In economic science, researchers have been long studying the possibility of using mathematical apparatus to perform a complete and accurate analysis and build a reliable forecast of the studied economic processes. There is a large number of different ways to create mathematical models that allow the construction the dynamics of changes in socio-economic data. In this article, author applies two types of discrete probability chains, which are generalization of well known Markov processes, to the modeling of dynamics of two types of data. The initial data are checked by using several statistical criteria. The analysis of the consistency of results with empirical dynamics has been made. The results shows that the satisfaction of statistical criteria is important for successful application of discrete probabilistic chains.

Keywords: *dynamical systems, probability chains, statistical criteria, extrapolation, econometric analysis.*

Citation: N. V. Loginova, "On a Method of Modeling of Socio-Economic Processes Dynamics," *Computer tools in education*, no. 2, pp. 14–24, 2018 (in Russian).

Received 19.03.2018, the final version — 26.04.2018.

Natalia V. Loginova, graduate student, Computer Science Department SPbSU; 198504, Russia, St.Petersburg, Starii Petergof, Universitetski pr. 28, faculty of Mathematics and Mechanics SPbGU, computer science department., natalia.loginowa@gmail.com

Логинова Наталья Владиславовна, аспирантка кафедры информатики СПбГУ; 198504, Россия, Санкт-Петербург, Старый Петергоф, Университетский пр., д. 28, математико-механический факультет СПбГУ, кафедра информатики, natalia.loginowa@gmail.com