УДК 005.521

О МЕТОДЕ АППРОКСИМАЦИИ ЭКОНОМИЧЕСКИХ ДАННЫХ, ОСНОВАННОМ НА ЗАДАЧЕ П. Л. ЧЕБЫШЁВА И ЕЕ ОБОБЩЕНИИ

И. Ю. Выгодчикова

Саратовский государственный университет
E-mail: VigodchikovaIY@info.sgu.ru

Статья посвящена равномерному методу аппроксимации динамики экономических показателей, основанному на исследованиях П. Л. Чебышева.

Ключевые слова: уравнение регрессии, аппроксимация, оценка, альтернанс, сжатие данных, прогнозирование.

About the Economic Indicator's Approximation by Method Based at the P. L. Chebyshev's Investigations and it's Generalizing

I. Y. Vigodchikova

This paper is devoted to the uniform method of economic indicator's approximation based at the P. L. Chebyshev's investigations.

Key words: regress equation, approximation, estimation, alternance, compression of the data, forecasting.

Эконометрическое моделирование и прогнозирование на основе построенных моделей вполне успешно применяется при анализе процессов с высоким уровнем риска, например при анализе динамики курсов валют, при оценке ожидаемых темпов инфляции и уровня безработицы, поскольку приемы эконометрики при современных информационных технологиях являются простыми в выполнении и наглядными для интерпретации.

Рассмотрим задачу наилучшей аппроксимации набора наблюдений $(\mathrm{x}_i, y_i), i=\overline{1,n}$ линейным уравнением $\hat{y}=\beta_0+\beta_1\mathrm{x}$. Точки $(\mathrm{x}_i, y_i), i=\overline{1,n}$ не лежат в точности на линии регрессии, а точки $(\mathrm{x}_i, \hat{y}_i), i=\overline{1,n}$, где $\hat{y}_i=\beta_0+\beta_1\mathrm{x}_i$, принадлежат этой линии, поэтому ошибки в *i*-ом наблюдении выражаются разностью между фактическим и расчетным значением зависимой переменной: $\varepsilon_i = y_i - \hat{y}_i$. Линейная регрессионная модель имеет вид $\mathrm{y_i}=\beta_0+\beta_1\mathrm{x_i}+\varepsilon_i$.

Поскольку измерить случайные ошибки ε_i невозможно, нужно *оценить* коэффициенты β_0, β_1 по имеющимся данным $(\mathrm{x}_i, y_i), i=\overline{1,n}$. Чаще всего для этого применяют метод наименьших квадратов (МНК)[1]. В качестве оценок неизвестных параметров β_0, β_1 берут такие значения $\hat{\beta}_0, \hat{\beta}_1$, которые минимизируют сумму квадратов отклонений значений y_i от $\hat{y}_i$:

$$Q(\beta_0,\beta_1):=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1\mathrm{x_i}\right)^2\to\min_{\beta_0,\beta_1}. \quad (1)$$

Иногда применяется прогнозирование экономических показателей на основании ортогональных многочленов и построения моделей динамики, например методом максимального правдоподобия. Однако при анализе численности населения эти способы иногда приводят к искажению результата – первый по причине возможности незапланированных сбоев в поведении многочлена с ростом его степени, а второй вследствие недостаточно надежных вероятностных характеристик рассматриваемого явления.

В данной работе предлагается метод аппроксимации данных, основанный на исследованиях П. Л. Чебышёва[2] о равномерном наилучшем приближении функции алгебраическим полиномом фиксированной степени и обобщении этого метода, который позволяет выявить дополнительные свойства динамического ряда. Исследование задачи П. Л. Чебышёва весьма интересно уже потому, что ее решение в линейном случае обладает симметричностью относительно входных данных. При рассмотрении обратной задачи (зависимая и независимая переменные меняются местами) решением будет прежняя линейная функция, естественно, если $\hat{y}=\beta_0+\beta_1\mathrm{x}$, а $\hat{x}=\omega_0+\omega_1 y$, то $\omega_0=-\beta_0/\beta_1$, $\omega_1=1/\beta_1$, что особенно ценно при анализе перекрестных данных. Также немаловажен факт наличия перспектив исследования. Разработаны методы решения обобщений этой задачи[3].

Указанный метод целесообразно применять для оценки сглаженных или достаточно стабильных данных, например динамики численности населения, числа работников на предприятии, объема инвестиций в экономику, индекса цен, процентных ставок крупных банков, объема ВВП, остаточной стоимости объектов жилого фонда и т. п.

Приведем формулировку задачи П. Л. Чебышёва для дискретного случая. Пусть $\mathrm{T}=\{\mathrm{t}_0<\mathrm{t}_1<\ldots<\mathrm{t}_N\}$ – дискретная сетка значений некоторой функции $\mathrm{y_k}=\mathrm{y}(\mathrm{t_k})$, $k=\overline{0,N}$, состоящая из «узлов» t_k, $k=\overline{0,N}$, $\mathrm{p_n}(\mathrm{A},\mathrm{t})=\mathrm{a}_0+\mathrm{a}_1\mathrm{t}+\ldots+\mathrm{a_n}\mathrm{t^n}$ – алгебраический полином степени не выше n с вектором коэффициентов $A=(a_0,a_1,\ldots,a_n)\in R^{n+1}$, N и n – целые неотрицательные числа. Требуется минимизировать максимальное по всем узлам сетки Т уклонение алгебраического полинома от значений дискретной функции в этих узлах:

$$\varphi(A):=\max_{k=0,N}\left|y_k-p_n(A,t_k)\right|\to\min_{A\in R^{n+1}}. \quad (2)$$

Решение такой задачи всегда существует вне зависимости от соотношений между N и n, а при $N\ge n$ оно единственно. Обозначим через $\rho:=\min_{A\in R^{n+1}}\varphi(A)$ минимальное значение целевой функции задачи (2).

Решение этой задачи сводится к поиску так называемого экстремального базиса, то есть такого (n + 2) – точечного подмножества сетки Т $\sigma^*=\left\{t_{j_0}{}^*<t_{j_1}{}^*<\ldots<t_{j_{n+1}}{}^*\right\}\subset T$, для которого выполняются соотношения

$$p_n\left(A^*,t_{j_k}\right)=y_{j_k}+(-1)^{k+1}\cdot h,\ \ k=\overline{0,n+1}, \quad (3)$$

причем $\varphi\left(A^*\right)=|h|$. В таком случае $A^*=\left(a_0{}^*,\ldots,a_n{}^*\right)\in R^{n+1}$ – решение задачи (2). Обычно для этих целей применяют алгоритмическую процедуру Вале – Пуссена, осуществляющую целенаправленный перебор базисов, начиная с произвольного.

Явление изменения знака уклонения фактических значений показателя от значений алгебраического полинома в (n + 2) различных узлах на одинаковую по модулю величину при переходе от одного узла к следующему по возрастанию, записанное математически в (3), часто называют *альтернансом.*

Соотношения (3) – это система с (n + 2) неизвестными – компонентами вектора $A^*=\left(a_0{}^*,\ldots,a_n{}^*\right)\in R^{n+1}$ и величиной h – и таким же числом уравнений. При $N>n$ эта система всегда имеет единственное решение.

В данной работе рассматриваем случай $n=1$. Система (3) запишется в виде (в целях упрощения обозначений берем базис $\sigma=\{t_0<t_1<t_2\}\subset T$, символы «*» и двойные индексы опускаем)

$$a_0+a_1t_k=y_k+(-1)^{k+1}\cdot h,\quad k=\overline{0,2}. \quad (4)$$

Несложно отыскать решение системы (4):

$$a_0=\frac{-h\cdot(t_1+t_0)+y_0\cdot t_1-y_1\cdot t_0}{t_1-t_0},$$

$$a_1=\frac{y_1-y_0+2\cdot h}{t_1-t_0},$$

$$h=-\frac{y_1}{2}+\frac{y_0}{2}\cdot\frac{t_2-t_1}{t_2-t_0}+\frac{y_2}{2}\cdot\frac{t_1-t_0}{t_2-t_0}.$$

Рассмотрим применение этого метода на примере аппроксимации динамики численности населения в России в 1991–2009 гг.[4] Исходные данные представлены в табл. 1.

Таблица1

Исходные данные численности населения, тыс. чел.

Дата	Городское	Сельское	Сумма
01.01.1991	109 405,1	38 868,6	148 274
01.01.1992	109 357,7	39 157	148 515
01.01.1993	108 668,4	39 893,3	148 562
01.01.1994	108 304,8	40 051,1	148 356
01.01.1995	108 321,7	40 138,2	148 460
01.01.1996	108 310,6	39 981	148 292
01.01.1997	108187,8	39 848	148036
01.01.1998	108 110,8	39 691,3	147 802
01.01.1999	108 053,2	39 486,2	147 539
01.01.2000	107 419,5	39 476	146 896
01.01.2001	107 071,3	39 261,9	146 333
01.01.2002	106 725,3	38 924	145 649
09.10.2002	106 420,9	38 737,7	145 159
01.01.2003	106 321,2	38 642,4	144 964
01.01.2004	105 818,4	38 349,8	144 168
01.01.2005	104 719,3	38 754,9	143 474
01.01.2006	104 104,8	38 648,7	142 754
01.01.2007	103 778,4	38442,6	142 221
01.01.2008	103 773	38 235,8	142 009
01.01.2009	103 690,4	38 213,6	141 904
01.01.2010	103 705	38 209	141 914

Сначала берем за условно нулевую дату 01.01.1991 и переводим значения дат в годы, прошедшие от нулевой даты. Далее рассматриваем задачу (2) для аппроксимации численности городского населения (y_k, $k=\overline{0,19}$) в зависимости от года (t_k, $k=\overline{0,19}$), при этом выбираем линейный полином, n = 1. В результате решения задачи (например с использованием алгоритма Вале – Пуссена) получена оценка зависимости численности населения от номера года от 01.01.1991 в линейной форме:

$$p_1(t)=\hat{y}=110\,104{,}1198-349{,}76\,t.$$

Экстремальным базисом оказались 3,8 и 15 годы. Если «убрать» из рассмотрения хотя бы одно из указанных наблюдений, решение задачи изменится, если же убирать любые другие наблюдения, решение останется прежним. Максимальная абсолютная ошибка аппроксимации 749,1 тыс. чел. Можно заметить, согласно полученной модели, что за рассматриваемый период городское население убывало приблизительно на 350 тыс. чел. в год (табл. 2).

Таблица 2

Вывод остатков по итогам применения алгоритма Вале – Пуссена

Т (годы)	Расчетное городское	Расчетное фактическое	Ошибка (%)	Абсолютная ошибка
0,0000	110 104,1198	699,019781	0,64%	699,02
1,0000	109 754,3593	396,6593429	0,36%	396,66
2,0027	109 403,6407	735,2406571	0,68%	735,24
3,0027	109 053,8802	749,080219	0,69%	749,08
4,0027	108 704,1198	382,419781	0,35%	382,42
5,0027	108 354,3593	43,75934292	0,04%	43,759
6,0055	108 003,6407	− 184,1593429	− 0,17%	184,16
7,0055	107 653,8802	− 456,919781	− 0,42%	456,92
8,0055	107 304,1198	− 749,080219	− 0,69%	749,08
9,0055	106 954,3593	− 465,1406571	− 0,43%	465,14
10,0082	106 603,6407	− 467,6593429	− 0,44%	467,66
11,0082	106 253,8802	− 471,419781	− 0,44%	471,42
11,7781	105 984,6126	− 436,2874059	− 0,41%	436,29
12,0082	105 904,1198	− 417,080219	− 0,39%	417,08
13,0082	105 554,3593	− 264,0406571	− 0,25%	264,04
14,0110	105 203,6407	484,3406571	0,46%	484,34
15,0110	104 853,8802	749,080219	0,72%	749,08
16,0110	104 504,1198	725,719781	0,70%	725,72
17,0110	104 154,3593	381,3593429	0,37%	381,36
18,0137	103 803,6407	113,2406571	0,11%	113,24
19,0137	103 453,8802	− 251,119781	− 0,24%	251,12
				749,08

При этом максимальная абсолютная ошибка аппроксимации составила 749,1 тыс. чел., или 0,72%.

Если использовать метод наименьших квадратов, получается следующая оценка:

$$p_1(t) = \hat{y} = 109\,822{,}63 - 333{,}15t.$$

Максимальная по периодам абсолютная ошибка аппроксимации – 837,58, все коэффициенты и уравнение регрессии значимы.

Исключая из рассмотрения данные за 1999 г., по методу наименьших квадратов получаем $\hat{y} = 109\,866{,}225895631 - 330{,}3255324t$, вычисляем прогноз по этой модели на пропущенный год, получаем (t = 18,0137) 103 915,841 тыс. чел., что на 225,441 тыс. чел. отличается от реальных данных. Решение по методу Чебышёва от исключения данных за 2009 г. не изменится, при этом прогноз на этот год 103 803,641 тыс. чел., что отличается от фактических данных всего на 113,241 тыс. чел. Так, конечно, бывает не во всех случаях, но даже этот пример говорит о праве метода равномерной аппроксимации данных на существование.

Используя метод наименьших модулей (рисунок), получаем следующую оценку:

$$\hat{y} = 109\,672{,}63 - 313{,}8t.$$

	Коэффициент	Стандартная ошибка	t-статистика	Р-значение
const	109 672	432,208	253,7	5,25e − 035 ***
t	−313,800	38,0855	− 8,239	1,08e − 07 ***

Медиана зависимой переменной 107 071,3
Стандартное отклонение зависимой переменной 1994,052
Сумма модулей ошибок 9184,583
Сумма квадратов остатков 5 333 762

Демонстрация результата метода наименьших модулей с применением программы Gretl

Также рассматривается задача построения экспоненциальной функции динамики городского населения:

$$\hat{y}_k = A\exp(bt).$$

После логарифмирования этой функции получена оценка параметров ln и b этой модели по алгоритму Вале – Пуссена (исходные данные численности населения также предварительно прологарифмированы):

$$\hat{y}_k = 110\,157{,}9709\exp(0{,}003293693t).$$

Однако максимальная абсолютная ошибка отклонения расчетных данных от исходных здесь получилась 769,07 тыс. чел.

Наконец, рассмотрим линейный аналог задачи (2) с учетом авторегрессии 1-го порядка:

$$\varphi 1(A) := \max_{k=1,N} \left| y_k - a_0 - a_1 t_k - a_2 t_{k-1} \right| \to \min_{a_0, a_1, a_2 \in R^3}. \quad (5)$$

Пользуясь для решения указанной задачи приближенным методом «Поиск решения» электронной таблицы MSExcel, получаем результат

$$\hat{y}_k = 7669{,}784 - 323{,}4522\, t_k + 0{,}925\, y_{k-1},\ k = 1,..,19.$$

При этом максимальная абсолютная ошибка аппроксимации составила 517,757 тыс.чел., или 0,48%.

Точное решение:

$$\hat{y}_k = 41\,775{,}62 - 146{,}39\, t_k + 0{,}619\, y_{k-1},\ k = 1,..,19.$$

При этом максимальная абсолютная ошибка аппроксимации составила 536,037 тыс.чел., или 0,52%. Полученную зависимость можно интерпретировать следующим образом: темпы сокращения населения со временем замедляются (ежегодно приблизительно на 38,1%).

Если, опять же, использовать для оценки параметров авторегрессионной модели 1-го порядка классический метод наименьших квадратов, получаем модель

$$\hat{y}_k = 19\,215{,}326 - 63{,}42 t_k + 0{,}823\, y_{k-1},\ k = 1,..,19.$$

При этом максимальная абсолютная ошибка аппроксимации составила 744,06 тыс.чел., или 0,71%.

В ходе исследования выявлены следующие полезные свойства оценок, полученных путем применения задачи П. Л. Чебышёва и ее обобщения:

– информативность цели. Минимальное значение целевой функции несет информацию о максимальной абсолютной ошибке аппроксимации исходных данных;

– возможность сильного сжатия данных. При устойчивой динамике обширный набор данных заменяется двумя коэффициентами полинома наилучшего приближения (как, впрочем, и в методе наименьших квадратов, однако можно сократить объем исходных данных до 3 значений показателей, соответствующих экстремальному базису, для задачи (2) и до 4 узлов для задачи (5));

– зависимость решения от значений исследуемого показателя. Лишь в нескольких (для линейного случая – в трех) рассматриваемых точках при исключении из анализа остальных данных решение задачи и прогнозные значения показателя, как, впрочем, и ошибки аппроксимации, останутся прежними;

– наглядность проверки результата, математическая четкость и «красота» его интерпретации, что, несомненно, повышает привлекательность этого метода при внедрении в учебный процесс.

Примечания

1. См.: *Себер Дж.* Линейный регрессионный анализ. М., 1980.
2. См.: *Демьянов В. Ф., Малоземов В. Н.* Введение в минимакс. М., 1972.
3. См.: *Выгодчикова И. Ю.* Применение алгебраических полиномов к моделированию экономических процессов // Математическое и информационное обеспечение экономической деятельности : сб. науч. ст. Вып.1. Саратов, 2006. С.16–21.
4. См.: Демографический ежегодник России 2009. М., 2010. С. 24. ; Распределение населения Российской Федерации по полу и возрастным группам (на 1 января 2010 г.) (тыс. чел.). URL: htpp://www.gks.ru (дата обращения: 25.01.2012).

УДК 338

РАЗВИТИЕ СОВРЕМЕННЫХ РЫНОЧНЫХ ИНСТРУМЕНТОВ КАК ОСНОВА ПОВЫШЕНИЯ КОНКУРЕНТОСПОСОБНОСТИ АВТОТРАНСПОРТНЫХ УСЛУГ

С. Н. Живайкин

Саратовский институт (филиал) Российского государственного технологического университета
E-mail: szhivajkin@yandex.ru

В статье рассмотрено современное положение транспортного комплекса Российской Федерации и обоснована классификация рыночных инструментов, обеспечивающая повышение конкурентоспособности автотранспортных услуг. Доказана необходимость оптимального использования различных рыночных инструментов в целях поддержания и эффективного функционирования как отдельного автотранспортного предприятия, так и автотранспортной системы в целом.

Ключевые слова: рыночные инструменты, сфера услуг, автотранспортные услуги, классификация услуг, конкурентоспособность, метод, эффективность.

The Development of Modern Market-based Instruments as a Basis for Improving the Competitiveness of Road Transport Services

S. N. Zhivaykin

In article modern position of a transport complex of Russian Federation is considered and the classification of market tools providing increase of competitiveness motor transportation mustache-meadow is proved. Necessity of optimum use of various market instrument tools with a view of maintenance and effective functioning, both the separate motor transportation enterprise, and motor transportation system as a whole is proved.

Key words: market tools, sphere of services, motor transportation services, classification of services, competitiveness, method, effectiveness.

Переход к рыночной экономике ознаменовался радикальными изменениями в структуре народного хозяйства. Наиболее важными из них следует считать сокращение доли отраслей материального производства и возрастание удельного веса отраслей непроизводственной сферы. По мере развития общества, роста производительных сил происходит увеличение занятости в этой сфере, возрастание технической оснащенности труда, внедрение все более совершенных технологий.

Сфера услуг, включающая воспроизводство разнообразных видов услуг, оказываемых предприятиями, организациями, а также физическими лицами, имеет ряд специфических особенностей по сравнению с материальным производством.

Во-первых, услуги производятся и потребляются в основном одновременно и не подлежат хранению, что порождает проблему регулирования их спроса и предложения.