

УДК 004.94
DOI: 10.15827/0236-235X.123.444-447

Дата подачи статьи: 23.04.18
2018. Т. 31. № 3. С. 444–447

Адекватные междисциплинарные модели в прогнозировании временных рядов статистических данных

*Б.М. Пранов*¹, д.т.н., профессор, *boris.pranov@gmail.com*

¹ *Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, г. Москва, 119571, Россия*

В статистических исследованиях для моделирования и прогнозирования временных рядов обычно используют многофакторные линейные модели. Область их применения весьма обширна.

Они достаточно эффективны в ситуации, когда множество точек, изображающих исследуемые объекты в многомерном пространстве параметров, располагается вблизи некоторого линейного подпространства (или его сдвига относительно начала координат). Этот эффект легко обнаруживается с помощью факторного анализа. Если же такое подпространство (линейное множество) отсутствует, для построения более точных моделей применяют нелинейные зависимости.

В экономике для описания зависимости прибыли предприятия от числа работников и стоимости основных средств используют функцию Кобба–Дугласа. Оказывается, если рассматривать пожары и другие явления социума как своеобразную его «продукцию», то функция Кобба–Дугласа с высокой степенью точности позволяет аппроксимировать соответствующие временные ряды. В результате получается целый ряд интересных моделей в новых предметных областях.

По итогам расчетов выяснилось, что временной ряд общего числа пожаров на территории РФ очень хорошо аппроксимируется функцией Кобба–Дугласа, а прогностические значения, рассчитанные по таким моделям, весьма близки к реальным. Аналогичная адекватная аппроксимация временных рядов общего числа пожаров пригодна для значительного количества европейских стран, а также США.

Такое моделирование применимо, в частности, и в туристической отрасли – в статье рассматриваются модели полного дохода гостиниц в зависимости от числа работников и величины основных фондов.

Ключевые слова: моделирование, прогнозирование, статистические данные, аппроксимация, временной ряд.

Начавшийся в середине прошлого века очередной экономический кризис обусловил появление научных разработок, направленных на оптимизацию деятельности пожарной охраны и полиции, а также других органов. Так как во многих городах мира затраты на содержание охраны и полиции сокращались, возникла необходимость повысить эффективность их деятельности. В результате проведенных исследований был построен ряд моделей для обоснования минимального уровня как пожарной, так и полицейской защиты. Аналогичные исследования продолжают до сих пор.

Остановимся на необходимом уровне пожарной защиты. Ясно, что обеспечение такого уровня лишь на настоящий момент не является полным решением задачи. Необходимо предвидеть уровень пожарной опасности на некоторое время вперед и обеспечивать соответствующую защиту с упреждением, то есть разработать методы прогнозирования пожарной опасности. Традиционно для прогнозирования временных рядов пожаров используют однофакторные линейные модели вида

$$y_i = a + bx_i + \varepsilon_i, \quad (1)$$

где y – фактические данные пожарной статистики; x – соответствующий момент времени; a и b – оцениваемые коэффициенты; b – так называемый «тренд» модели; ε_i – ошибка модели; i – номер момента времени (года) [1]. Рассматриваются также линейные многофакторные модели вида

$$y_i = a_0 + a_1x_{1i} + \dots + a_kx_{ki} + \varepsilon_i, \quad (2)$$

где i – номер момента времени; x_1, \dots, x_k – пара-

метры, с помощью которых оценивается временной ряд y ; a_0, a_1, \dots, a_k – оцениваемые коэффициенты. При этом для получения модели вида (2) необходимо иметь временные ряды параметров x_1, \dots, x_k . Значения этих параметров обычно берутся из официальной статистики (Росстат, МЧС).

К сожалению, для многих объектов и явлений модели (1) и (2) не дают приемлемого описания, так как графики наблюдаемых значений результирующей переменной имеют явно криволинейное строение и не укладываются в достаточно узкую полосу на плоскости для модели (1) или не расположены в достаточной близости к гиперплоскости в модели (2).

Используются также нелинейные математические модели, где в правой части зависимости (1) помещают, помимо первых степеней некоторых параметров x_i , их вторые степени и, возможно, парные произведения [2].

Однако при взгляде на график общего числа пожаров, произошедших в какой-либо конкретной стране (в России, США или Европе) на протяжении ряда лет, заметно, что его линия совершает непериодическое волнообразное движение и не может быть удовлетворительно описана линейной зависимостью.

В такой ситуации для более точной аппроксимации статистических данных следует использовать нелинейные зависимости. Одной из самых популярных моделей, используемых в экономике на протяжении уже более 100 лет, является функция

Кобба–Дугласа, моделирующая величину *валового внутреннего продукта* (ВВП) в зависимости от двух факторов – численности работников и оценки основных фондов объекта (страны, региона, предприятия) [3]. В экономике довольно успешно используются модели, построенные на основе так называемой производственной функции вида

$$Y = A \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \dots \cdot x_n^{a_n}, \quad (3)$$

где Y – моделируемая величина; x_1, \dots, x_n – наблюдаемые параметры объекта [4]. С помощью моделей вида (3) довольно успешно аппроксимируют количество продукции, выпускаемой объектом исследования за определенный период времени.

Оказалось, что та же самая производственная функция (3) может достаточно успешно оценивать количество пожаров в административно-территориальной единице за определенный период времени. В частности, в качестве такой единицы можно взять всю страну. В работах [5, 6] построен ряд моделей вида

$$Y = A \cdot X_1^\alpha \cdot X_2^\beta, \quad (4)$$

где Y – общее число пожаров (в Российской Федерации или в какой-либо другой стране); X_1 – численность населения; X_2 – величина ВВП.

Отметим любопытное свойство модели (3). Если Y – оцениваемое количество продукции, выпускаемое данным объектом исследования, то в качестве параметров x_1, \dots, x_n берутся такие показатели, которые способствуют выпуску продукции. Обычно это численность работников и оценка основных фондов предприятия. Применяя эту метафору как основу для выбора модели (3) при оценке общего числа пожаров, можно считать, что пожары являются своеобразной «продукцией» социума, и для их оценки также можно использовать аналогичные показатели – численность населения (аналог числа работников предприятия) и объем основных фондов (административно-территориальной единицы).

Таким образом, в зависимости (4) под величиной Y понимается продукция объекта – страны, региона, предприятия. Практика почти столетнего моделирования показывает, что это соотношение удовлетворительно описывает зависимость выпускаемой продукции как для макро-, так и для микроэкономического моделирования [6]. Можно сделать предположение, что общее число пожаров, гибель людей на них, получение травм, ущерб от пожаров можно рассматривать как своего рода «продукцию» человеческого сообщества при затратах как материальных, так и человеческих ресурсов. В работе [5] зависимость (2) была применена к аппроксимации общего числа пожаров в России и США. Оказалось, что самым удивительным образом зависимость оказалась неожиданно точной – для России коэффициент корреляции с фактическими данными составляет 0,994, а для США – 0,834. Довольно любопытным оказался тот

факт, что зависимость (4) очень точно отражает динамику общего числа пожаров для большинства стран Европы [5].

Попытаемся оценить предикторные возможности как линейной модели (2), так и производственной функции (4). Обычно производственная функция в качестве параметров использует X_1 – количество населения (млн чел.) и X_2 – основные фонды РФ (fixed assets – трлн руб.).

Используем статистические данные по населению и основным фондам РФ за 2003–2015 гг. (Росстат). Статистика пожаров приведена на сайте МЧС.

Расчет линейной модели приводит к следующей зависимости:

$$Y = -179,868 + 3,059 \cdot X_1 - 0,772 \cdot X_2. \quad (5)$$

Коэффициент корреляции статистических данных с оценкой по модели (5) составляет 0,991. Отметим, что коэффициент корреляции общего числа пожаров Y с оценкой основных фондов X_2 равен $-0,986$. Этот довольно высокий отрицательный коэффициент корреляции может навести на мысль, что при увеличении основных фондов РФ соответствующее число пожаров уменьшается, а это противоречит здравому смыслу. На рисунке 1 представлена графическая аппроксимация общего числа пожаров с помощью линейной модели (5).

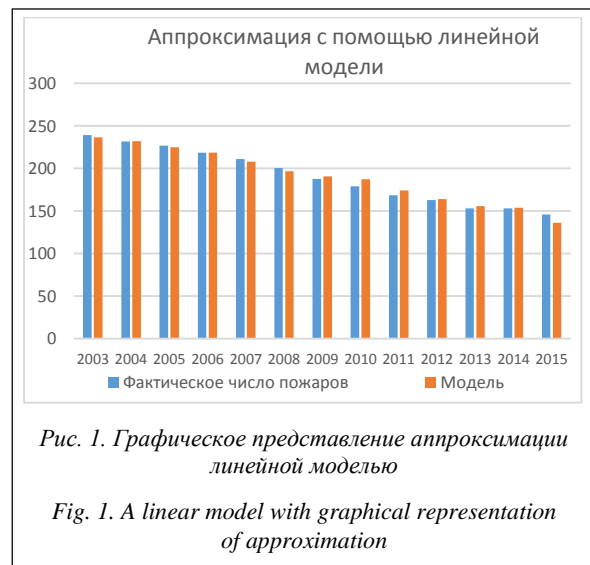


Рис. 1. Графическое представление аппроксимации линейной моделью

Fig. 1. A linear model with graphical representation of approximation

Подставляя в модель (5) сведения о численности населения и объеме основных фондов за 2016 год, получаем прогноз общего числа пожаров на 2016 год:

$$Y = -179,868 + 3,059 \cdot 146,54 - 0,772 \cdot 181,62 = 128,19.$$

Исходя из того, что в настоящий момент статистика пожаров за 2016 год известна и общее число пожаров составляет 139,70 тыс., получаем, что ошибка прогноза на 2016 год по модели (5) составляет $\eta_{\text{лин.}} = (128,19 - 139,70) / 139,70 \% = -8,2 \%$.

Из представленного графика видно, что прогнозная прямая (тренд) направлена вниз и уже в

2015 году дает заниженный результат, а для 2016 года – еще более заниженный.

Рассмотрим реализацию модели (4). Ее расчет можно проводить с помощью как Microsoft Excel (надстройка «Поиск решения») [7], так и статистических пакетов (в данном случае использована программа SPSS) [8]. В результате расчета получена зависимость вида

$$Y = 1317,368 \cdot X_1^{(-0,132)} \cdot X_2^{(-0,298)} = 144,74, \quad (6)$$

что приводит к ошибке

$$\eta_{\text{лин.}} = (144,74 - 139,70) / 139,70 \% = 3,6 \%$$

На рисунке 2 представлена аппроксимация фактических данных с помощью модели (6).

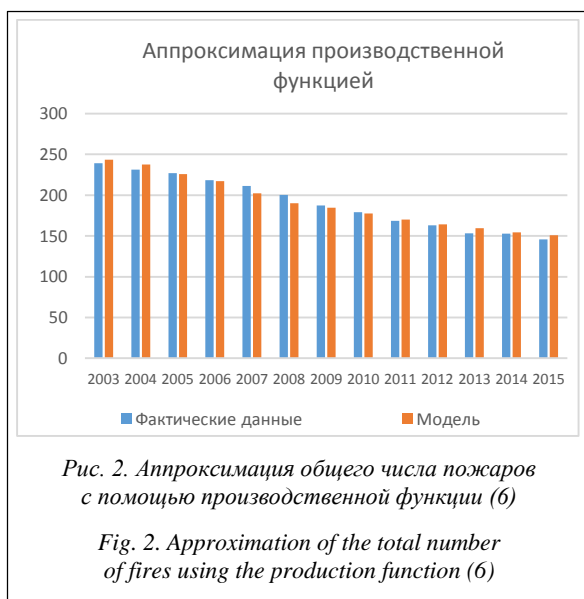


Рис. 2. Аппроксимация общего числа пожаров с помощью производственной функции (6)

Fig. 2. Approximation of the total number of fires using the production function (6)

Коэффициент корреляции двух временных рядов на графике рисунка 2 равен 0,988. Из рисунка 2 и расчета погрешности для нелинейной модели (6) видно, что ее возможности для прогнозирования более обоснованны.

Обратимся теперь к еще одному применению модели (4) производственной функции [9]. Здесь в качестве аппроксимируемой величины Y рассматривается полный доход гостиниц и ресторанов РФ за период с 2005 по 2014 гг. (есть официальная статистика Росстата), X_1 означает численность работников гостиниц и ресторанов, X_2 – основных фондов гостиниц и ресторанов. Расчеты показывают, что модель имеет вид

$$Y = 4,110 \cdot X_1^{0,836} \cdot X_2^{0,785}. \quad (7)$$

Сравнение рассчитанных по этой модели данных с фактическими представлено на рисунке 3.

Отметим хорошо наблюдаемый на графике факт – в 2010 году произошло падение выручки, что соответствует кризисному падению экономики в это время. Любопытно, что моделирующая кривая сглаживает этот скачок. Коэффициент корреляции расчетных данных с фактическими составляет 0,971, что отражает объективную динамику модели.

Поскольку модель (7) очень хорошо аппроксимирует данные в наблюдаемом диапазоне, ее можно использовать и как инструмент прогнозирования, то есть продолжить расчеты по данной формуле уже за пределами этого интервала.

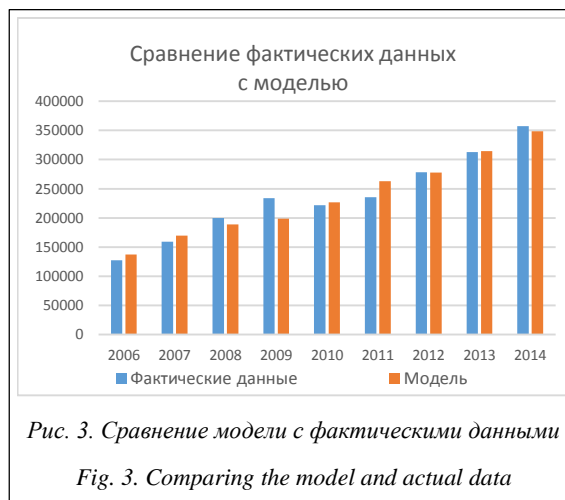


Рис. 3. Сравнение модели с фактическими данными

Fig. 3. Comparing the model and actual data

Описанная в настоящей статье модель основана на статистических данных по Российской Федерации в целом. Аналогичные модели можно строить также и для отдельных регионов, городов, предприятий.

Таким образом, можно сделать следующие выводы.

Аппроксимация отечественной статистики имеет высокое качество (у всех построенных моделей наблюдается высокая корреляция с исходными данными).

Можно строить модели для разных уровней административно-территориальной подчиненности (для городов, их кластеров, сельской местности); при этом в качестве X_1 и X_2 следует брать соответствующие показатели (население городское, сельское, ВВП города или сельской местности и т.д.).

Поскольку такое социальное явление, как преступность, также можно рассматривать в качестве своего рода «продукции» населения и ВВП (или соответствующих их разновидностей), аналогичные расчеты следовало бы провести и для аппроксимации, и, возможно, для прогнозирования преступности.

Кроме того, было бы интересно исследовать, имеют ли содержательный смысл коэффициенты модели оценки количества пожаров.

Литература

1. Смит Г., Дрейпер Н. Прикладной регрессионный анализ. М.: Диалектика, 2016. 912 с.
2. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. М.: Мир, 1985. 509 с.
3. Интрилигатор И. Математические методы оптимизации и экономическая теория. М.: Айрис Пресс, 2002. 553 с.
4. Клейнер Г.Б. Производственные функции: теория, методы, применение. М.: Финансы и статистика, 1986. 354 с.
5. Пранов Б.М. О некоторых подходах к моделированию

и прогнозированию временных рядов пожарной статистики // Технологии техносферной безопасности. 2014. № 5. URL: <http://agps-2006.narod.ru/ttb/2014-5/23-05-14.ttb.pdf> (дата обращения: 20.04.2018).

6. Leontief W. *Input-output economics*. Oxford Univ. Press, 1986, 436 p.

7. Мадера А.Г. *Математические модели в управлении*. М.: Изд-во РГГУ, 2007. 564 с.

8. Бюль А., Цефель П. *SPSS: искусство обработки информации*. СПб: ДиаСофтЮП, 2005. 608 с.

9. Пранов Б.М. Модели динамики и прогнозирования в сфере гостеприимства // Вестн. РМАТ. 2017. № 1. С. 24–27.

Software & Systems

DOI: 10.15827/0236-235X.123.444-447

Received 23.04.18

2018, vol. 31, no. 3, pp. 444–447

Adequate interdisciplinary models in forecasting time series of statistical data

B.M. Pranov¹, *Dr.Sc. (Engineering), Professor, boris.pranov@gmail.com*

¹ *Russian Presidential Academy of National Economy and Public Administration, 2. Москва, 119571, Russian Federation*

Abstract. Statistical studies commonly use multivariate linear models to model and predict time series. Their application area is quite extensive. They are quite effective in a situation when a set of points depicting the objects under investigation in a multidimensional parameter space is located near a certain linear subspace (or its shift relative to the origin). Factor analysis easily reveals this effect. If there is no such subspace (linear set), nonlinear dependencies are used to construct more accurate models.

In the economy, the Cobb-Douglas function is used to describe the dependence of enterprise profits on the number of employees and the value of fixed assets. It turns out that if we consider fires and other phenomena of society as a kind of its “production”, then the Cobb-Douglas function allows approximating a corresponding time series with a high degree of accuracy. As a result, we get a number of interesting models in the new subject areas.

The results of calculations showed that the Cobb-Douglas function is good at approximating the time series of the total number of fires in the territory of the Russian Federation. The prognostic values calculated by such models are very close to the real ones. A significant number of European countries, as well as the United States use a similarly adequate approximation of the time series of the total number of fires.

Such modeling is also appropriate for a tourism industry. The paper considers the models of total hotel income depending on the number of employees and the size of fixed assets.

Keywords: modeling, forecasting, statistical data, approximation, time series.

References

1. Smith G., Draper N. *Applied Regression Analysis*. Moscow, Dialectics Publ., 2016, 912 p.
2. Gill F., Murray W., Wright M. *Practical Optimization*. Moscow, Mir Publ., 1985, 509 p.
3. Intraligator I. *Mathematical Methods of Optimization and Economic Theory*. Moscow, Iris Press, 2002, 576 p.
4. Kleiner G.B. *Production Functions: Theory, Methods, Application*. Moscow, Finansy i statistika Publ., 1986, 354 p.
5. Pranov B.M. Some new approach to modeling and forecasting of time series of fire statistics. *Technology of Technosphere Safety*. 2014, iss. 5. Available at: <http://agps-2006.narod.ru/ttb/2014-5/23-05-14.ttb.pdf> (accessed April 20, 2018).
6. Leontief W. *Input-Output Economics*. Oxford, Univ. Press, 1986.
7. Madera A.G. *Mathematical Models in Management*. Moscow, RGGU Publ., 2007, 564 p.
8. Bühl A., Zöfel P. *SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows*. Pearson Studium Publ., 744 p. (Russ. ed.: St. Petersburg, DiSoftTU Publ., 2005, 608 p.).
9. Pranov B.M. Dynamics and Forecasting Models in Hospitality. *Vestn. RIAT*. 2017, no. 1, pp. 24–27 (in Russ.).