УДК 332.1 JEL C53 DOI 10.26425/1816-4277-2019-1-94-99

**Гагарин Юрий Евгеньевич**
канд. техн. наук, Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени Н. Э. Баумана», г. Калуга
***e-mail:*** *g_ug@mail.ru*

**Гагарина Светлана Николаевна**
канд. экон. наук, ФГБОУ ВО «Калужский государственный университет имени К. Э. Циолковского» , г. Калуга
***e-mail:*** *g_ug@mail.ru*

# ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЙ С УЧЕТОМ НЕОПРЕДЕЛЕННОСТИ ИСХОДНЫХ ДАННЫХ

***Аннотация.*** *Рассмотрен математический метод составления прогнозов при неопределенности исходных данных. Для получения точечных и интервальных оценок параметров математических моделей, учитывающих погрешности в значениях функции и аргумента, предложено использовать методы конфлюэнтного анализа. Для линейных моделей найдены точечные оценки и доверительные интервалы прогнозов. Разработанный метод используется при решении задачи прогнозирования объема предоставляемых услуг организациями жилищно-коммунального комплекса. Проведено моделирование объема продаж при различных значениях интервалов времени и проанализировано влияние погрешности исходной информации на значения параметров модели и на прогнозируемые величины.*

***Ключевые слова:*** *прогнозирование, неопределенность, оценки параметров, исходные данные, доверительные интервалы, конфлюэнтный анализ.*

**Gagarin Yurii**
Candidate of Engineering Sciences, Bauman Moscow State Technical University, Kaluga Branch, Kaluga
***e-mail:*** *g_ug@mail.ru*

**Gagarina Svetlana**
Candidate of Economic Sciences, Kaluga State University named afler K. E. Tsiolkovsky, Kaluga
***e-mail:*** *g_ug@mail.ru*

# THE PREDICTION OF ENTERPRISES ACTIVITY INDICATORS TAKING INTO ACCOUNT THE INITIAL DATA UNCERTAINTY

***Abstract.*** *The mathematical method of making forecasts when the initial data uncertainty have been considered. To obtain point and interval estimates of the parameters of mathematical models that take into account errors in the values of the function and argument, the use methods of confluent analysis has been proposed. For linear models, point estimates and confidence intervals of predictions have been found. The developed method is used to solve the problem of forecasting the volume of services provided by organizations of the housing and utilities complex. The sales volume simulating for different values of time intervals has been conducted and the influence of the initial information error on the model parameter values and predicted values has been analyzed.*

***Keywords:*** *prediction, uncertainty, parameter estimates, input data, confidence intervals, confluent analysis.*

Прогнозирование занимает особое место в задачах анализа и синтеза экономических систем. С одной стороны, анализ систем можно проводить с целью прогнозирования поведения системы или изменения различных ее характеристик, с другой стороны – в процессе прогнозирования выявляют основные закономерности системы и отдельных ее элементов, что служит в дальнейшем дополнительными условиями в задачах синтеза.

Как правило, исходные данные являются случайными величинами, и результаты статистической обработки таких данных могут существенно отличаться по степени их обоснованности. Исходя из статистически обоснованных данных могут быть получены прогнозы. При отсутствии обоснованности прогнозы не заслуживают доверия. Степень обоснованности результатов зависит от характера и количества данных, согласованности между данными и методов их обработки.

При разработке планов развития предприятий основные показатели деятельности могут носить случайный характер, и поэтому в прогнозной модели необходимо учитывать неопределенности всех связываемых величин.

В силу различных причин реальные задачи содержат не одну, а несколько статистических величин с соизмеримыми дисперсиями, которые необходимо учитывать в процессе обработки исходной информации.



Исследования проведены при финансовой поддержке Российского фонда фундаментальных исследований и Правительства Калужской области (научный проект № 17-12-40010-ОГН).

При этом во всех задачах необходимо определять интервальные оценки решения, чтобы можно было оценить качество полученного решения и сделать выводы о том, достаточно ли исходной информации для получения надежного решения, удовлетворяет ли поставленным требованиям математический алгоритм решения или требуется повысить точность исходных данных, улучшить статистику либо искать другие способы решения задачи.

Особую актуальность получают исследования методов и алгоритмов прогнозирования с учетом неопределенности исходных данных. Методы «измерения с ошибками» позволяют учитывать погрешности как в значениях функции, так и в значениях ее аргумента при оценивании параметров математических моделей [5]. Если не учитывать погрешности всех исходных данных, то в задачах аппроксимации это приведет к смещенным и несостоятельным оценкам параметров [3]. В методах конфлюэнтного анализа (далее − МКА) учитывается случайный характер всех исходных величин, при этом в регрессионных методах аргумент должен быть детерминирован, поскольку невязка берется по одной величине.

В процессе прогнозирования необходимо построить математическую модель, которая основывается на идентификации, оценивании и диагностической проверке.

Рассмотрим задачу нахождения точечных и интервальных оценок прогноза линейной модели вида: $y = a + bx$. В качестве исходных данных рассмотрим изменение объема предоставляемых услуг организациями жилищно-коммунального комплекса в течение 60 месяцев.

При нахождении оценок прогноза будем учитывать случайный характер исходных данных и сравнивать два способа. Эти способы отличаются тем, что в первом значения параметров линейной модели находят методом наименьших квадратов (далее − МНК), во втором значения параметров будем находить с помощью МКА [1].

При оценивании параметров МНК модель имеет вид:

$$y_i = a + bx_i + \varepsilon_i, \quad i = \overline{1, n}.$$

Метод наименьших квадратов учитывает только погрешность значения функции $\varepsilon_i$, а значение аргумента $x$ считается детерминированной величиной. При использовании МНК уравнение регрессии имеет следующий вид:

$$\hat{y} = 5140{,}1 + 275x. \tag{1}$$

Методы конфлюэнтного анализа позволяют учитывать погрешности аргумента линейной функции x и самой функции y. Модель оценивания параметров имеет вид:

$$\begin{cases} y_i = a + b\xi_i + \varepsilon_i; \\ x_i = \xi_i + \delta_i; \quad i = \overline{1, n}, \end{cases} \tag{2}$$

где $\varepsilon_i$, $\delta_i$ – ошибки значений $y_i$, $x_i$ соответственно; $\xi_i$ – истинные значения $x_i$.

Наиболее часто в природе и технике случайные величины подчиняются нормальному распределению. Будем считать, что случайные величины $\varepsilon_i$, $\delta_i$ соответствуют нормальному распределению со средними значениями равными нулю, дисперсиями $\sigma^2(y_i)$, $\sigma^2(x_i)$, и коэффициентом корреляции $\rho_i=0$.

Минимизируемый функционал позволяет определить оценки параметров $a$, $b$ как:

$$F = \frac{1}{2}\sum_{i=1}^{n}\left[\left(x_i - \xi_i\right)^2 / \sigma^2\left(x_i\right) + \left(y_i - a - b\xi_i\right)^2 / \sigma^2\left(y_i\right)\right]. \tag{3}$$

В функционале (3) значения $\xi_i$ неизвестны. Оценки значений $\xi_i$ будут определяться из условия:

$$\left.\frac{\partial F}{\partial \xi_i}\right|_{\xi_i = \hat{\xi}_i} = 0; \quad i = \overline{1, n}. \tag{4}$$

При решении системы уравнений (4) необходимо, чтобы значения $\xi_i$ принадлежали области неопределенности измеренных величин $x_i$. Если значения $x_i$ распределены по нормальному закону, то область неопределенности будет соответствовать неравенству:

$$|x_i - \xi_i| \leq k\sigma(x_i), \tag{5}$$

где $k$ – коэффициент, который определяется исходя из условия доверия.

С учетом (4) неизвестные значения $\xi_i$ оценивают по формуле:

$$\xi_i = \frac{\sigma^2(y_i)x_i + b\sigma^2(x_i)(y_i - a)}{\sigma^2(y_i) + b^2\sigma^2(x_i)}\ . \tag{6}$$

Процесс нахождения оценок параметров $a$, $b$ состоит из двух основных этапов итерационного цикла:

1) получают первое приближение параметров $a$, $b$ при значениях $\xi_i = x_i$, $i = \overline{1, n}$;

2) по (6) определяют значения $\hat{\xi}_i$ и проверяют принадлежность значений $\hat{\xi}_i$ области возможных значений $x_i$.

Выходом из итерационного цикла является выполнение одного из следующих условий:

а) на очередном шаге итерационного цикла значение функционала (3) меньше заданного числа $\gamma$;

б) на соседних итерациях значение функционала $F$ отличаются несущественно, т. е.:

$$\left|\frac{F_\nu - F_{\nu+1}}{F_\nu}\right| \leq \gamma_1, \tag{7}$$

где $\gamma_1$ – заданное число;

в) исчерпан лимит итераций.

В модели (2) и в формулах (3), (6) значения $\sigma^2(y_i)$ принимали одинаковыми для всех точек наблюдений $\sigma^2(y_i)=\sigma^2$ и оценивали по формуле:

$$\hat{\sigma}^2 = S/(n-2), \tag{8}$$

где

$$S = \sum_{i=1}^{n}\left[y_i - \left(\hat{a} + \hat{b}x\right)\right]^2.$$

Средние квадратические отклонения $\sigma(x_i)$ предполагались известными и равными $\sigma(x_i)=0{,}3x$. В ограничении (5) значение $k=3$ и итерационный процесс нахождения оценок параметров заканчивался при выполнении условия, в котором сравнивались значения функционала (3) на соседних итерациях при $\gamma_1=0{,}003$.

Для выполнения условия (7) потребовалось 8 итераций. Количество итераций при решении задачи с учетом погрешностей аргумента $x$ зависит от значения $\gamma_1$ в условии (7). Значение $F$ на первой итерации соответствует решению задачи МНК.

Уравнение прямой с учетом погрешностей в значениях функции и аргумента имеет вид:

$$\hat{y} = 3936{,}2 + 324{,}6x\ . \tag{9}$$

Дисперсии оценок параметров определяют с помощью матрицы вторых производных от функционала (3) по $a$, $b$, и имеют следующий вид:

$$\left.\begin{aligned} D(\hat{a}) &= \sigma^2\sum_{i=1}^{n}x_i^2 \Bigg/ \left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]; \\ D(\hat{b}) &= \sigma^2 n \Bigg/ \left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]. \end{aligned}\right] \tag{10}$$

В таблице 1 приведены значения параметров *a*, *b*, средних квадратических отклонений σ(*a*), σ(*b*), σ(*y*), полученных двумя способами. Значение σ(*y*) для двух методов будет различно, поскольку зависит от значений параметров линейной модели и оценивается по формуле (8).

*Таблица 1*

**Значения параметров и средних квадратических отклонений при решении задачи двумя методами**

| Параметр | Решение МНК | Решение МКА |
|---|---|---|
| $a$ | 5 140,10 | 3 936,20 |
| $b$ | 275,00 | 324,60 |
| $\sigma(a)$ | 689,98 | 731,57 |
| $\sigma(b)$ | 19,67 | 20,85 |
| $\sigma(y)$ | 2 638,9 | 2 798,00 |

*Источник: [1]*

Определим доверительные интервалы для линейной функции [2]. В общем случае статистика $|\hat{y}-y|/\sqrt{D(\hat{y})}$ подчиняется распределению Стьюдента с *n–p* степенями свободы и интервальная оценка при заданной доверительной вероятности γ имеет вид:

$$P\left\{\hat{y}-t_\gamma\sqrt{D(\hat{y})}<y<\hat{y}+t_\gamma\sqrt{D(\hat{y})}\right\}=\gamma, \quad (11)$$

где $t_\gamma$ – квантиль распределения Стьюдента; $D(\hat{y})$ определяется как:

$$D(\hat{y})=\frac{\sigma^2}{n}+(x-\overline{x})^2\sigma^2\bigg/\sum_{i=1}^{n}(x_i-\overline{x})^2=\sigma^2\left(\frac{1}{n}+(x-\overline{x})^2\bigg/\sum_{i=1}^{n}(x_i-\overline{x})^2\right). \quad (12)$$

При определении интервальных оценок для доверительной вероятности $\gamma=0{,}95$ и $n-p=60-2=58$ степеней свободы квантиль распределения Стьюдента принимается равным $t_\gamma=2{,}0$.

Из формул (11), (12) видно, что ширина доверительного интервала зависит от значения переменной *x*: при $x=\overline{x}$ ширина минимальна, а по мере удаления *x* от $\overline{x}$ ширина доверительного интервала увеличивается.

Предполагая, что число экспериментальных точек достаточно велико можно считать, что оценки параметров *a*, *b* распределены относительно их математических ожиданий по нормальному закону. Для получения доверительных интервалов применяют безразмерную *t*-статистику Стьюдента, которая подчиняется *t*-распределению с *n*–2 степенями свободы [4]. При уровне значимости α доверительные интервалы для параметров *a*, *b* определяют как:

$$\hat{a}-t_{1-\alpha/2}\sqrt{D(\hat{a})}\le a\le\hat{a}+t_{1-\alpha/2}\sqrt{D(\hat{a})};$$

$$\hat{b}-t_{1-\alpha/2}\sqrt{D(\hat{b})}\le b\le\hat{b}+t_{1-\alpha/2}\sqrt{D(\hat{b})},$$

где $D(\hat{a})$, $D(\hat{b})$ вычисляют по формулам (10).

Для уровня значимости α=0,05 и $n-p=60-2=58$ степенях свободы квантиль распределения Стьюдента равен $t_{1-\alpha}=2{,}0$. В таблице 2 приведены 95 %-ные границы для параметров *a*, *b* при решении задачи двумя методами.

Доверительные интервалы прогнозов индивидуальных значений $\hat{y}_0$ определяют по (11). В таблице 2 приведены точечные и интервальные оценки при 95 %-ном уровне доверия прогноза $\hat{y}_0$ для $x_0=61$. Точечные значения прогноза рассчитаны по (1), (9). Учет одновременно погрешностей в значениях функции и аргумента приводит к уточнению величины прогноза.

*Таблица 2*

**Доверительные границы интервальных оценок параметров, точечные и интервальные оценки прогноза при решении задачи двумя методами**

| Решение МНК | Решение МКА |
|---|---|
| $3\ 760{,}0 \le a \le 6\ 520{,}1$ | $2\ 473 \le a \le 5\ 399{,}3$ |
| $235{,}6 \le b \le 314{,}3$ | $282{,}9 \le b \le 366{,}3$ |
| $\hat{y}_0 = 21\ 913$ | $\hat{y}_0 = 23\ 739$ |
| $16\ 458 \le \hat{y}_0 \le 27\ 369$ | $17\ 945 \le \hat{y}_0 \le 29\ 532$ |

*Источник: [2]*

Рассмотрим зависимость погрешности параметров $a$, $b$ и прогнозируемых значений $\hat{y}_0$ объема продаж на следующий год от числа измерений. Проведено моделирование объема продаж при различных значениях интервалов времени. За единицу интервала времени принимали месяц (количество измерений – 60), квартал (количество измерений – 20), полугодие (количество измерений – 10) и год (количество измерений – 5). С увеличением числа измерений погрешность параметров уменьшается. Учет погрешности аргумента $x$ приводит к увеличению значений $\sigma(a)$, $\sigma(b)$, $\sigma(y)$, а увеличение количества измерений приводит к уточнению прогнозируемого значения $\hat{y}_0$.

Для учета неопределенности информации при построении прогнозов авторы предлагают использовать методы конфлюентного анализа. Минимизируемый функционал, получаемый в МКА, учитывает суммы квадратов наикратчайших расстояний от наблюдаемых точек до функциональной зависимости, связывающей эти точки. В традиционном подходе МНК минимизируется сумма квадратов отклонений при фиксированных значениях аргумента. Учет в одном функционале погрешностей значений функций и аргументов позволяет получать несмещенные точечные и интервальные оценки как параметров, так и самих функциональных зависимостей.

Результаты математического моделирования показывают, что учет неопределенности информации приводит к увеличению погрешностей параметров математической модели (см. табл. 1). При этом значение минимизируемого функционала, в процессе выполнения итерационного цикла нахождения оценок параметров, уменьшается, т. е. уравнение прямой (9) с большей точностью описывает динамику изменения объема предоставляемых услуг. Изменение значений интервалов времени существенным образом влияет на точечные и интервальные значения параметров линейной зависимости и на значения прогнозируемых показателей, что отражается на достоверности прогноза.

*Библиографический список*


1. Гагарин, Ю. Е. Возможность интервального оценивания параметров модели при случайном характере множества факторов / Ю. Е. Гагарин, С. Н. Гагарина // Научные труды КГУ им. К. Э. Циолковского. Серия: Естественные науки. – Калуга: Издательство КГУ имени К. Э. Циолковского. – 2017. – С. 90–92.
2. Гагарина, С. Н. Интервальное прогнозирование объемов спроса на услуги субъектов естественных монополий с учетом неопределенности информации / С. Н. Гагарина, Ю. Е. Гагарин // Вестник университета. – 2013. –№ 22. – С. 101–110.
3. Грешилов, А. А. Математические методы принятия решений: Учеб. пособие для вузов / А. А. Грешилов. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2006. – 584 с.
4. Кендалл М. Статистические выводы и связи / М. Кендалл, А. Стьюарт. – М.: Наука, 1973. –900 с.
5. Fuller, W. A. Measurement error models. – New york ect.: Wiley, 1987. – 440 p.

*References*

1. Gagarin Yu. E., Gagarina S. N. Vozmozhnost' interval'nogo otsenivaniya parametrov modeli pri sluchaynom kharaktere mnozhestva faktorov [*The possibility of interval estimation of model parameters for the random nature of many factors*]. Nauchnye trudy KGU im. K. E. Tsiolkovskogo. Seriya: Estestvennye nauki. Kaluga: Izdatel'stvo KGU imeni K. E. Tsiolkovskogo, 2017, pp. 90–92.
2. Gagarina, S. N., Gagarin, Yu. E. Interval'noe prognozirovanie ob"emov sprosa na uslugi sub"ektov estestvennykh monopolii s uchetom neopredelennosti informatsii [*Interval forecasting of demand for services of natural monopolies subject to the uncertainty of information*]. Vestnik universiteta, 2013, I. 22, pp. 101–110.
3. Greshilov A. A. Matematicheskie metody prinyatiya reshenii: Ucheb. posobie dlya vuzov [*Mathematical decision making methods*]. M.: Izd-vo MGTU im. N. E. Baumana, 2006, 584 p.
4. Kendall, M., Stuart, A. Statisticheskie vyvody i svyazi [*The advanced theory of statistics*]. M.: Nauka, 1973, 900 p.
5. Fuller, W. A. Measurement error models. New york ect.: Wiley, 1987. 440 p.