

О ВЛИЯНИИ ОШИБОК МОДЕЛИРОВАНИЯ НА ОЦЕНКУ СОСТОЯНИЯ СИСТЕМЫ

Первадчук В.П.¹, Севодина В.М.¹

¹ *Пермский национальный исследовательский политехнический университет*

Изучены ситуации, в которых экономические индикаторы сложных систем моделируются линейной однофакторной регрессией. Установлено, что в случае использования таких моделей возникает вероятность принятия ошибочного решения (риск), что связано с неправильной информацией, выдаваемой моделью, о нахождении параметров системы в допустимой области или вне ее. Предложены принципы определения границ множеств, соответствующих допустимым значениям факторов системы. Получены формулы для определения вероятности принятия ошибочного решения в таких системах. Описан метод управления риском, который во многих ситуациях оказывается более действенным по сравнению с известными методами. Результаты статьи проиллюстрированы на примере влияния площади квартиры на ее рыночную стоимость. Предложенный метод управления риском в данном примере дает значительное падение вероятности принятия ошибочного решения.

Ключевые слова: линейные регрессии, риски, допустимые состояния, качество модели, опасные ситуации, ложные оценки.

ABOUT THE IMPACT OF THE MODELING ERROR IN THE ASSESSMENT OF THE STATUS OF THE SYSTEM

Pervadchuk V.P.¹, Sevodina V.M.¹

¹ *Perm National Research Polytechnic University*

In the paper studied a situation in which economic indicators of complex systems are modeled by linear one-factorial regression. It is established, that in case of use of such models there is a probability of making wrong decisions (risk), which is connected with incorrect information given by the model, the determination of the parameters of the system in the permitted area or outside it. Proposed principles for determining the boundaries of the sets, corresponding to the admissible values of factors of the system. The formulas for determining the probability of making wrong decisions in such systems. The method of risk management, which in many cases appears to be more effective in comparison with the known methods. The results of this paper are illustrated by the example of the impact of the flat area on its market value. The proposed method of risk management in this example gives a significant fall in the probability of making wrong decisions.

Key words: linear regression, the risks, the available status, the quality of the model, dangerous situations, false estimates.

1. **Введение.** В современной экономике широко используются экономические и финансовые индикаторы, отражающие состояние сложной системы в данный момент времени и служащие для понимания процессов, происходящих в ней, для корректировки управления, а также дающих возможность для анализа и изменения поведения исследуемой системы в целом. В связи с таким широким использованием индикаторов системы важной является проблема обеспечения высокого качества результатов экономического анализа. Среди подобных исследований особо выделим проблему соотношения показателей, полученных моделированием какого-нибудь типа, к их фактическим величинам. Случайные ошибки, возникающие при измерении характеристик экономических явлений, в сочетании со случайными ошибками модели могут оказывать существенное влияние на конечный результат исследований. Поэтому актуальной является задача определения границ множеств, соответствующих допустимым значениям факторов системы, а, значит, дающих

приемлемую в рамках данной задачи вероятность ошибки (риск) и не допускающих ошибочных выводов о состоянии системы.

Проблема появления неправильных решений при измерении различных индикаторов экономических систем систематически начала изучаться в работах [1-3]. Дело в том, что ошибки измерения могут принципиально исказить представление о состоянии экономической системы в целом и, следовательно, повлиять на правильность принимаемого решения. Оказывается, что ошибки модели играют аналогичную роль, и более того, совместное появление названных ошибок может существенно усилить этот отрицательный эффект. Именно изучение влияния ошибок моделирования на оценку состояния сложной системы и представляет цель настоящей работы. Отметим, что некоторые частные случаи данной задачи изучены в [4]: определение границ допустимого множества параметров производилось через область изменения независимого фактора. В данной статье рассмотрен наиболее общий подход к определению допустимых множеств. Наконец, в работе указаны способы управления риском: наряду с уже известными [4], предложен новый метод снижения последствий ошибок, который базируется на корректировке параметров эконометрической модели.

2. **Постановка задачи и вычисление рисков.** Пусть на выходе системы или ее отдельных подсистем имеет место некоторый процесс y , зависящий от фактора x , то есть $y = y(x)$. Рассмотрим случай, когда зависимость между процессом y и фактором x устанавливается с помощью линейной регрессии, т.е. имеет форму $y = ax + b + \delta y$, где a, b – коэффициенты модели, а δy – ошибка модели, обладающая «классическими» свойствами, ошибка и факторы независимы. Будем предполагать, что значения переменной y измерены с ошибкой:

$$y = \tilde{y} + \Delta y,$$

где волна означает измеренное значение переменной.

Тогда рассматриваемая модель может быть представлена в следующем виде:

$$y = ax + b + \delta y + \Delta y. \quad (1)$$

Видно, что при использовании модели вида (1) совокупная ошибка складывается не только из случайной ошибки построения модели, но и из ошибки измерения переменной y .

Особо определим оценочную величину y_o :

$$y_o = ax + b = \tilde{y} - \delta y. \quad (2)$$

Предположим, что состояния рассматриваемой системы можно разбить на два класса: 1) множество состояний, когда система не способна выполнять свои целевые назначения; 2) множество состояний, при которых функционирование системы обеспечивает

выполнение каких-либо целевых установок. Первое множество состояний назовем (ср. с [2]) множеством опасных состояний, а второе - множеством допустимых фактических состояний и обозначим его через $\Omega_{\text{дон}}^{\phi}$. Рассмотрим случай, когда $\Omega_{\text{дон}}^{\phi} = (\underline{y}_{\phi,\text{дон}}, \overline{y}_{\phi,\text{дон}})$. Заметим, что при измерении y возникает ошибка Δy , и, следовательно, в общем случае $y \neq \tilde{y}$. Это неравенство может привести к ситуации, в которой фактическое значение $y \notin \Omega_{\text{дон}}^{\phi}$, а измеренное значение $\tilde{y} \in \Omega_{\text{дон}}^{\phi}$, то есть система находится в области опасного состояния, а измеренные величины об этом не сигнализируют. Таким образом, необходимо учитывать ошибку измерения Δy , а значит необходимо вводить новую область $\Omega_{\text{дон}}^u = (\underline{y}_{u,\text{дон}}, \overline{y}_{u,\text{дон}})$, лежащую внутри $\Omega_{\text{дон}}^{\phi}$ и позволяющую создать некоторый запас надежности. Методы построения подобных областей можно найти, например, в [2; 5].

В случае, когда для нахождения y используется модель вида (1), возникают дополнительные сложности с решением вопроса о нахождении \tilde{y} в $\Omega_{\text{дон}}^u$, так как \tilde{y} заменяется своими оценочными значениями из (2). Следуя [2], рассмотрим ситуации, соответствующие различным положениям в пространстве переменных системы относительно описанных множеств. Введем событие A_1 , означающее что $y \in \Omega_{\text{дон}}^{\phi}$. Будем говорить, что произошло событие B_1 , если $y_o \in \Omega_{\text{дон}}^u$. Через A_2, B_2 обозначим события, противоположные событиям A_1, B_1 соответственно.

Рассмотрим теперь возможные комбинации этих событий.

$A_1 \cap B_1$ – состояние системы, когда фактическое состояние системы допустимо, с помощью средств оценки поступает информация также о допустимом состоянии;

$A_2 \cap B_2$ – фактическое состояние находится вне области допустимых границ, оценочная информация подтверждает это состояние;

$A_1 \cap B_2$ – по оценочной информации фиксируется недопустимое состояние системы, хотя фактическое значение находится в допустимой области;

$A_2 \cap B_1$ – оценочная модель дает информацию о допустимом состоянии системы, фактическое значение находится вне допустимой области.

В первых двух случаях анализ состояния системы показывает корректность построенной эконометрической зависимости. Последние два события характеризуются потерями ввиду неправильной оценки состояния системы. Область $A_1 \cap B_2$ назовем (ср. с [2]) областью ложной оценки, так как она характеризуется появлением ложной информации о

выходе системы в область опасных состояний, хотя система находится в допустимых режимах. Область $A_2 \cap B_1$ назовем областью опасной ситуации, так как в этом случае модель, по сути, допускает использование данных, на которые она не рассчитана, что приводит к крайне недостоверным результатам.

Таким образом, при использовании модели (1) часть данных будет располагаться относительно допустимых областей таким образом, что вероятным станет наступление событий $A_1 \cap B_2$ и $A_2 \cap B_1$, которые приведут к ошибкам. Это означает, что совокупный риск использования модели можно оценить как вероятность P нахождения данных в области ложной оценки и области опасной ситуации.

Для численного определения таких вероятностей можно использовать подход, изложенный в [2].

Очевидно

$$P = P_{\text{опас сит}} + P_{\text{ложн оц}} \cdot \quad (3)$$

С учетом (1), (2) имеем

$$\begin{aligned} P_{\text{ложн оц}} &= P(A_1 \cap B_2) = P(y_{\text{оц}} > \overline{y_{\text{оц.доп.}}}; \underline{y_{\text{ф.доп.}}} < y_{\text{ф}} < \overline{y_{\text{ф.доп.}}}) + \\ &+ P(y_{\text{оц}} < \underline{y_{\text{оц.доп.}}}; \underline{y_{\text{ф.доп.}}} < y_{\text{ф}} < \overline{y_{\text{ф.доп.}}}) = \\ &= P(\delta y > \overline{y_{\text{оц.доп.}}} - ax - b; \underline{y_{\text{ф.доп.}}} - ax - b - \delta y < \Delta y < \overline{y_{\text{ф.доп.}}} - ax - b - \delta y) + \\ &+ P(\delta y < \underline{y_{\text{оц.доп.}}} - ax - b; \underline{y_{\text{ф.доп.}}} - ax - b - \delta y < \Delta y < \overline{y_{\text{ф.доп.}}} - ax - b - \delta y) = \\ &= \int_{\underline{y_{\text{оц.доп.}}} - ax - b}^{\overline{y_{\text{ф.доп.}}} - ax - b - \delta y} d\delta y \int_{\underline{y_{\text{ф.доп.}}} - ax - b - \delta y}^{\overline{y_{\text{ф.доп.}}} - ax - b - \delta y} W(\Delta y, \delta y) d\Delta y + \int_{-\infty}^{\underline{y_{\text{оц.доп.}}} - ax - b} d\delta y \int_{\underline{y_{\text{ф.доп.}}} - ax - b - \delta y}^{\overline{y_{\text{ф.доп.}}} - ax - b - \delta y} W(\Delta y, \delta y) d\Delta y. \end{aligned}$$

Аналогично

$$\begin{aligned} P_{\text{опас сит}} &= P(A_2 \cap B_1) = \\ &= \int_{\underline{y_{\text{оц.доп.}}} - ax - b}^{\overline{y_{\text{оц.доп.}}} - ax - b} \delta y \int_{-\infty}^{\underline{y_{\text{ф.доп.}}} - ax - b - \delta y} W(\Delta y, \delta y) \Delta y + \int_{\underline{y_{\text{оц.доп.}}} - ax - b}^{\overline{y_{\text{оц.доп.}}} - ax - b} \delta y \int_{\underline{y_{\text{ф.доп.}}} - ax - b - \delta y}^{\overline{y_{\text{ф.доп.}}} - ax - b - \delta y} W(\Delta y, \delta y) \Delta y. \end{aligned}$$

Здесь – совместная плотность распределения отклонений. Вид подынтегральной функции и основные факторы, подлежащие учету при ее формировании, определяются объектами или подсистемами процесса и их режимом работы, а также множеством других факторов.

Рассмотрим теперь вопрос управления рисками. Ранее, в работе [4], были предложены следующие способы: 1) увеличением объема исходных данных или сужением

области изменения независимой переменной добиться снижения риска; 2) снижением качества эконометрической модели обеспечить необходимый уровень риска. При дальнейшем рассмотрении оказалось, что оба этих метода имеют существенные недостатки. Так, действия в первом направлении на практике фактически могут состоять только в сужении области измерений, что ограничивает поле применения модели, при этом все же недостаточно снижая показатели риска. Действия во втором направлении позволяли получить приемлемые показатели вероятности ошибочных действий только при существенном снижении качества регрессии, что почти нивелировало полученное преимущество.

Имея в виду вышесказанное, предложим еще один способ управления риском, состоящий в перерасчете коэффициентов эконометрической модели с целью обеспечения нахождения данных исключительно в области $\Omega_{\text{дон}}^u$, что должно исключить высокую вероятность ошибки, возникающую при выходе полученных при помощи регрессионной зависимости значений из области $\Omega_{\text{дон}}^u$, и оставить только риск, связанный с ошибками δy и Δy .

На практике действия в этом направлении будут означать использование специальных коэффициентов регрессии, найденных путем решения задачи квадратичного программирования:

$$\begin{cases} \sum_{i=1}^n (y_i - (ax_i + b))^2 \rightarrow \min \\ \underline{y_{u.\text{дон}}} \leq ax + b \leq \overline{y_{u.\text{дон}}} \\ \underline{x} \leq x \leq \overline{x} \end{cases} \quad (4)$$

где \underline{x} и \overline{x} - минимальное и максимальное значение переменной x из выборки. Несмотря на то что предлагаемый способ также может привести к потере качества модели, на практике часто оказывается, что такой подход наиболее рационален. В следующем пункте данной статьи обсудим введенные понятия на конкретном примере.

3. Определение стоимости квартиры. Воспользуемся примером, рассмотренным в работе [4]. Построим линейную регрессионную модель, описывающую зависимость стоимости квартиры от ее площади. Полученная в [4] модель имеет вид:

$$y = -2,061 + 0,039x + \delta y, \quad (5)$$

где y – стоимость квартиры (млн руб.), x – общая площадь (кв. м) квартиры, δy – ошибка модели. Оценки параметров эконометрической модели здесь являются эффективными,

точность построенной модели характеризуется коэффициентом детерминации (R^2), равным 0,83. Таким образом, качество построенной модели (5) является достаточно высоким.

Специфика полученной модели состоит в возможности ошибки при оценке квартиры, ввиду влияния на цену отличных от площади, менее значимых факторов (наличие балкона, ремонт, планировка и т.п.). Исследуемую модель (5) будем считать эконометрической моделью с ошибкой в зависимой переменной [3]. Исходные данные отличаются от фактических значений на отклонение Δy ($y = \tilde{y} + \Delta y$).

Оценка рынка недвижимости показывает, что допустимыми фактическими границами являются $\Omega_{оон}^{\phi} = (1,57; 7)$. Определим теперь границы оценочного допустимого множества исходя из его назначения: создания некоторого запаса надежности. Определим используемую модель как модель, надежность которой оценивается по критерию возврата брака, т.е. по отказам потребителя принимать к рассмотрению квартиры, оцененные с помощью модели. В этом случае, согласно принятому ГОСТу [5], допустимым является установление коэффициента запаса экспертными методами. Для данного класса недвижимости в агентствах недвижимости принято давать запас надежности в 5% для нижней границы стоимости и в 30% для верхней границы стоимости (имеется в виду, что близость к предельным ценам «отпугивает» покупателей, и поэтому нижние цены завышаются либо квартиры переводят в другой ранг, а квартиры с верхними ценами занижают или переводят в другой ранг). Таким образом, определим $\Omega_{оон}^u = (1,75; 5,09)$.

Отклонение Δy считаем имеющим нормальное распределение с $m = 0$ и $\sigma = 0,333$. С помощью критерия согласия Пирсона была проверена гипотеза о нормальности распределения ошибки модели $\delta y (m = 0, \sigma = 0,138)$. Таким образом, плотность $W(\Delta y, \delta y)$ из (3) соответствует двумерному нормальному распределению, и, следовательно, вероятность из (3) может быть найдена. На рисунке 1 представлена зависимость риска от конкретной величины независимой переменной (площади квартиры).

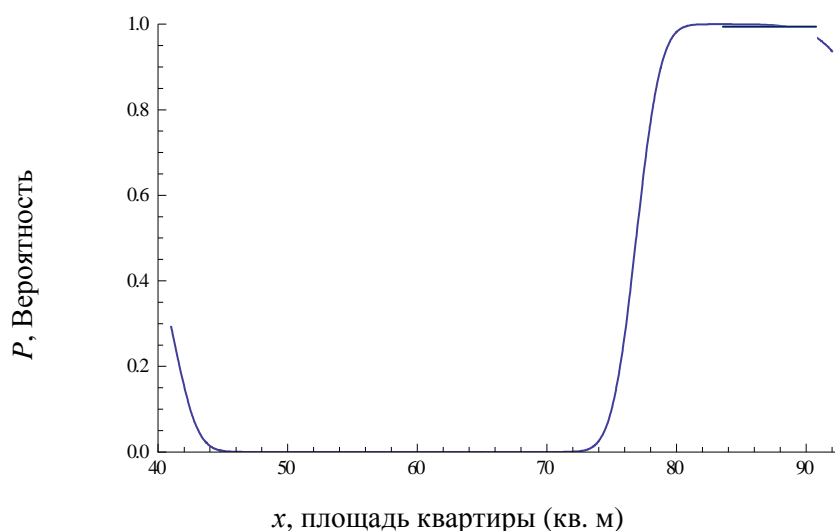


Рисунок 1

Как можно видеть из графика, риск довольно высок на краях рассматриваемой области изменения площади квартиры x . Особенно выделяется промежуток (86,95), на котором риск крайне высок – равен 1. Это связано с тем, что полученные при помощи модели значения y на этом интервале выходят за границы допустимой области $\Omega_{доп}^u$.

Применим для данной задачи описанный выше способ управления риском: построим новую регрессионную зависимость, используя специальные коэффициенты.

Решив задачу квадратичного программирования

$$\begin{cases} \sum_{i=1}^n (y_i - (ax_i + b))^2 \rightarrow \min \\ 1,75 \leq ax + b \leq 5,09 \\ 41 \leq x \leq 92 \end{cases},$$

получим коэффициенты $a = 0,056$, $b = -0,093$, коэффициент детерминации R^2 при этом равен 0,67. На рисунке 2 хорошо видно, что новая линия регрессии (отмеченная зеленым) целиком принадлежит области $\Omega_{доп}^u$ (границы которой отмечены синим пунктиром), в отличие от первоначальной линии регрессии (отмеченной желтым), которая выходит из допустимой оценочной области на множестве значений x (86;95).

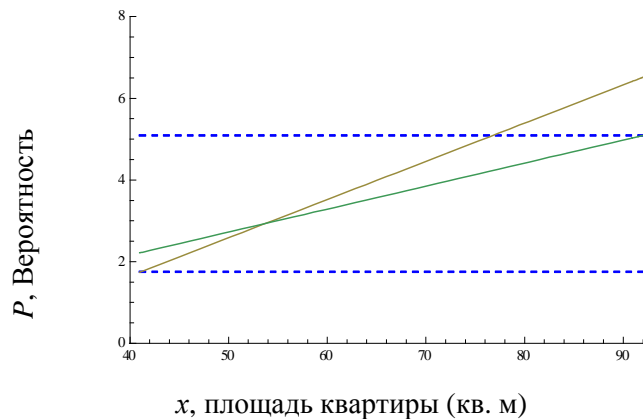


Рисунок 2

На рисунке 3 представлена кривая, описывающая риск, возникающий при использовании модели с коэффициентами, полученными специальным образом.

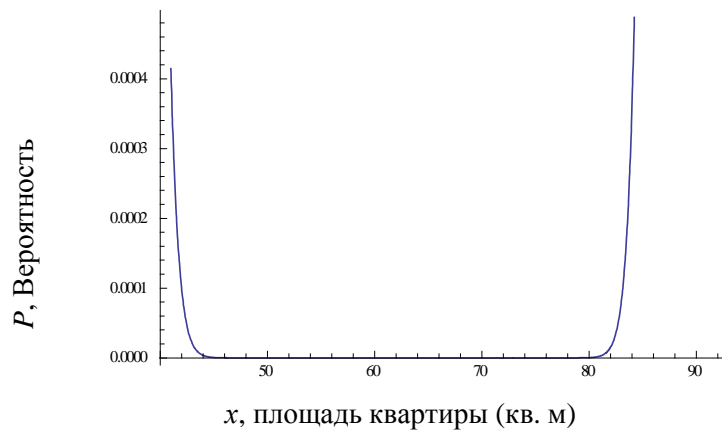


Рисунок 3

Видно, риск упал практически до 0 на всей области изменения x . Такой хороший результат получен потому, что в рамках данной модели удалось изменить регрессию таким образом, что даже суммарное максимальное значение всех ошибок, возникающих при использовании данной модели, не выводит переменную y из допустимой оценочной области. Кроме того, коэффициент детерминации для специальных параметров регрессии остается достаточно высоким, что позволяет считать новую модель высококачественной. Однако в общем случае эти условия могут не выполняться, и тогда следует ожидать значительного, но не полного падения риска. Отметим, что в [4] рассмотрена задача определения регрессии с $P < P_{кр} = 0,75$, результатом решения которой стала регрессия с $R^2 = 0,46$, т.е. предлагаемый здесь метод оказывается более рациональным по сравнению с методом из [4].

Выводы. В данной работе представлен случай, когда экономические индикаторы моделируются при помощи линейной однофакторной регрессии. Показано, что в случае использования такой модели может возникать вероятность принятия ошибочного решения (риск). Риск возникает в связи с неправильной информацией о нахождении параметров системы в допустимой области или вне ее. Предложен способ интерпретации и определения границ области, возникающей при введении запаса надежности. Получены формулы для определения вероятности принятия ошибочного решения в таких системах. Описан метод управления риском, который оказывается более действенным во многих ситуациях по сравнению с методами из [4]. Результаты статьи проиллюстрированы на примере влияния площади квартиры на ее рыночную стоимость. Предложенный метод управления риском в данном примере дает абсолютное падение вероятности ошибочного решения.

Список литературы

1. Живетин В.Б. Аэромеханический контроль. – Казань : Изд-во Казанского математического общества, 2000. – 195 с.
2. Живетин В.Б. Риски и безопасность экономических систем (математическое моделирование). - 2-е изд. – М. : Изд-во Института проблем риска, 2005. – 345 с.
3. Живетин В.Б. Научный риск. – Казань : Изд-во Казанского математического общества, 2003. – 355 с.
4. Севодин М.А., Севодина В.М. О некоторых рисках, возникающих при использовании линейных регрессионных зависимостей // Управление экономическими системами : электронный науч. журн. - 2013. - № 1. – URL: <http://uecs.ru/instrumentalnii-metody-ekonomiki/item/1947-2013-01-25-06-15-40>.
5. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика. – М. : Экзамен, 2003. – 512 с.

Рецензенты:

Елохова Ирина Владимировна, д.э.н., профессор, заведующий кафедрой «Управление финансами» Пермского национального исследовательского университета, г. Пермь.

Долгова Елена Владимировна, д.э.н., профессор кафедры «Информационные технологии и автоматизированные системы» Пермского национального исследовательского университета, г. Пермь.