ЭКОНОМИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ 1965 том I, вып. 4

МНОГОШАГОВЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ ПРИ ПОСТРОЕНИИ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ ПРОИЗВОДИТЕЛЬНОСТИ ТРУДА

Г. Н. ВЕСЕЛАЯ, А. А. ФРЕНКЕЛЬ

(МОСКВА)

Целью данной работы является построение при помощи регрессионного анализа экономико-статистической модели, связывающей показатель производительности труда с основными факторами, влияющими на его уровень.

Математически задачу можно сформулировать следующим образом.

Пусть имеется N наблюдений. Приняты следующие обозначения: y — зависимая переменная (показатель производительности труда), $x_1, x_2, \ldots, x_n$ — независимые переменные (факторы), влияющие на уровень производительности труда.

Нужно, пользуясь результатами наблюдений, получить некоторое представление о математической модели изучаемого объекта. Естественно в таком случае аппроксимировать неизвестную функцию полиномами, определяя коэффициенты полиномов (коэффициенты регрессии) по результатам наблюдений при помощи метода наименьших квадратов.

Исходной информацией для построения модели служили годовые отчеты за 1960—1963 гг. заводов химических (искусственных и синтетических) волокон. Для расширения моделируемой совокупности использовался метод заводо-лет, т. е. показатели по каждому предприятию брались за ряд лет и рассматривались как данные самостоятельных предприятий.

При построении модели не учитывался ряд заводо-лет, поскольку в годовых отчетах не содержались данные по отобранным факторам. Кроме того, заводо-годы исключались в том случае, когда на предприятиях происходил резкий ассортиментный сдвиг в выпускаемой продукции (удельный вес стоимости новой, т. е. несравнимой, товарной продукции в стоимости всей товарной продукции составлял более $^2/_3$).

В итоге в моделируемую совокупность вошло 33 заводо-года ($N = 33$).

В качестве показателя производительности труда была взята выработка на одного работающего в рублях — y.

Были отобраны следующие четыре фактора, влияющие на уровень производительности труда:

x_1 — *стоимость производственных фондов* (тыс. руб.), равная сумме среднегодовых стоимостей основных и нормируемых оборотных фондов в производстве. Среднегодовая стоимость нормируемых оборотных фондов в производстве определяется как разность между всей среднегодовой стоимостью нормируемых оборотных фондов и среднегодовой стоимостью готовой продукции. Показатель стоимости производственных фондов характеризует размер производства;

x_2 — *коэффициент фондовооруженности*, определяется как отношение среднегодовой стоимости основных фондов (за вычетом среднегодовой стоимости зданий и сооружений) к числу промышленно-производственных рабочих, работающих в наибольшую смену. Число рабочих, работающих в наибольшую смену, находится как отношение среднегодовой численности промышленно-производственных рабочих к коэффициенту сменности;

x_3 — *коэффициент электровооруженности*, равный отношению всей потребленной электроэнергии к числу отработанных за год человеко-часов. Коэффициенты фондовооруженности и электровооруженности отражают уровень технического развития предприятий;

x_4 — *удельный вес несравнимой (новой) товарной продукции во всем объеме произведенной предприятием товарной продукции* (%), определяемый как отношение стоимости несравнимой товарной продукции к стоимости всей товарной продукции. Этот показатель характеризует ассортиментные сдвиги в номенклатуре выпускаемой продукции.

Конечно, на производительность труда в промышленности химических волокон влияет также целый ряд других факторов (специализация и кооперирование, автоматизация и механизация производства, организация управления и т. д.). Однако для некоторых из них до сих пор не найдено сколько-нибудь приемлемого измерителя, а другие не содержатся в отчетности предприятий. Поэтому в данном исследовании мы ограничимся четырьмя приведенными факторами, тем более что по экономическим представлениям они оказывают значительное влияние на уровень производительности труда и, кроме того, их можно получить из существующей отчетности предприятий.

Для удобства вычислений величины всех анализируемых показателей были представлены в нормированном виде, как относительные величины: $y' = y \cdot 10^{-4}$, $x_1' = x_1 \cdot 10^{-4}$, $x_2' = x_2 \cdot 10^{-2}$, $x_3' = x_3 \cdot 10^{-2}$, $x_1' = x_4 \cdot 10^{-2}$.

На первом этапе неизвестная функция аппроксимировалась линейным уравнением, которое имеет вид:

$$y' = 0{,}7181 + 0{,}0736x_1' - 0{,}1168x_2' + 1{,}3749x_3' - 1{,}1078x_4'. \qquad (1)$$

Для статистической оценки уравнения регрессии подсчитывалась остаточная дисперсия:

$$S^2_{\text{ост}} = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N - p} = 0{,}6211 \cdot 10^{-3}, \qquad (2)$$

где y_i — отчетное значение выработки; $\hat{y}_i$ — расчетное значение выработки, получаемое при подстановке в рассматриваемую формулу значений факторов, характеризующих деятельность i-го завода; p — число коэффициентов регрессии.

Кроме того, подсчитывалась дисперсия относительно среднего значения $\bar{y}$:

$$S^2_{\text{ср}} = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N - 1} = 0{,}7193 \cdot 10^{-3}. \qquad (3)$$

Для оценки адекватности проверялась гипотеза о равенстве двух дисперсий, в связи с чем определялось дисперсионное отношение

$$F = \frac{S^2_{\text{ср}}}{S^2_{\text{ост}}} = 1{,}16, \qquad (4)$$

которое оказалось ниже критического даже для 0,20 уровня значимости. Линия регрессии не имеет смысла, так как остаточная дисперсия не отличается от рассеяния относительно среднего.

Далее уравнение регрессии искалось в виде полинома 2-й степени (в нашем случае полный полином состоит из 15 членов):

$$y' = b_0 + \sum_{i=1}^{4} b_i x_i' + \sum_{\substack{i=1\\k=1}}^{4} b_{ik} x_i' \, x_k'. \tag{5}$$

Результаты этого этапа исследований приводятся в табл. 1, из которой видно, что остаточная дисперсия для полного квадратного уравнения уменьшилась почти вдвое по сравнению с линейным случаем.

Таблица 1

	I	II	III	IV	V	VI	VII
b_0	0,1581	0,1522	0,1416	0,1424	0,1469	0,1449	0,1437
t_0	3,14	3,60	8,10	5,88	7,78	7,78	8,50
b_1	—0,0757	—0,0736	—0,0720	—0,0654	—0,0681	—0,0586	—0,0566
t_1	1,20	1,17	1,17	1,17	1,17	1,7	1,20
b_2	—0,1501	—0,412	—0,1402	—0,1416	—0,1473	—0,1414	—0,1414
t_2	2,85	4,10	4,10	4,70	5,65	6,50	6,10
b_3	—0,0227	—0,0193	—	—	—	—	—
t_3	0,35	0,31					
b_4	0,0351	0,0360	0,0376	0,0293	0,0203	0,0256	—
t_4	0,56	0,60	0,66	0,57	0,51	0,16	
b_{11}	0,3897	0,3902	0,3967	0,3905	0,3707	0,3684	0,3646
t_{11}	1,98	2,05	2,34	2,23	2,19	2,20	2,30
b_{12}	1,2449	1,2056	1,2126	1,176	1,205	1,169	1,1681
t_{12}	3,05	3,35	3,56	2,50	4,12	4,00	4,10
b_{13}	—0,4786	—0,4837	—0,5228	—0,5346	—0,4898	—0,5435	—0,5486
t_{13}	1,58	1,60	2,00	2,04	2,34	3,00	3,10
b_{14}	—0,4141	—0,3983	—0,3887	—0,4018	—0,4419	—0,4836	—0,5232
t_{14}	1,07	1,07	1,10	1,17	1,37	1,60	2,70
b_{22}	—0,0245						
t_{22}	0,03	—	—	—	—	—	—
b_{23}	0,7361	0,7415	0,7251	0,7283	0,7690	0,7313	0,7315
t_{23}	3,27	3,40	3,52	3,52	5,13	5,40	5,90
b_{24}	—0,0607	—0,0793	—0,0735				
t_{24}	0,21	0,28	0,29	—	—	—	—
b_{33}	0,1192	0,1026	0,4221	0,4406			
t_{33}	0,45	0,40	0,30	0,30	—	—	—
b_{34}	—0,3022	—0,3091	—0,3472	—0,3436	—0,2406	—	—
t_{34}	0,58	0,50	0,60	0,66	0,65		
b_{44}	0,0392	0,0389	0,0414	0,0495	0,0511	0,0521	0,0520
t_{44}	0,91	0,90	1,05	1,80	2,00	2,08	2,10
$S^2_{ост}$	$0{,}3355 \cdot 10^{-3}$	$0{,}3187 \cdot 10^{-3}$	$0{,}3044 \cdot 10^{-3}$	$0{,}2910 \cdot 10^{-3}$	$0{,}2790 \cdot 10^{-3}$	$0{,}2720 \cdot 10^{-3}$	$0{,}2610 \cdot 10^{-3}$
F	2,14	2,22	2,36	2,47	2,58	2,64	2,75
$F_{кр}$	1,9	1,9	1,8	1,8	1,7	1,7	1,7
η	0,859	0,858	0,858	0,858	0,855	0,853	0,853
$S^2_{ср}$	$0{,}7193 \cdot 10^{-3}$						

Так как все коэффициенты регрессии взаимно связаны, то доверительные границы для некоторого k-го коэффициента регрессии b_k можно устанавливать только после того, как заранее выбраны все остальные коэффициенты регрессии. Отсюда следует, что нельзя проверять нулевую гипотезу для каждого коэффициента регрессии в отдельности.

Поэтому была предпринята попытка проранжировать факторы по величине их значимости. Для всех членов, входящих в уравнение регрессии, подсчитывалась величина

$$t_i = \frac{b_i}{S_{\text{ост}} \sqrt{c_{ii}}}, \tag{6}$$

где c_{ii} — i-й диагональный элемент обратной матрицы системы нормальных уравнений.

Член, для которого t_i было наименьшим, исключался, и уравнение регрессии строилось в виде неполного квадратного полинома. Исключение членов продолжалось до тех пор, пока оно вело к уменьшению остаточной дисперсии. Последний, седьмой, вариант был признан окончательным. В натуральном масштабе он имеет вид:

$$\begin{aligned} y = 14371{,}25 - 0{,}5659x_1 - 1413{,}705x_2 + 0{,}3646 \cdot 10^{-4}x_1^2 + 0{,}1168x_1x_2 - \\ - 0{,}054486x_1x_3 - 0{,}05232x_1x_4 + 73{,}1516x_2x_3 + 5{,}200x_4^2. \end{aligned} \tag{7}$$

При интерпретации необходимо учитывать, что коэффициенты регрессии — именованные величины, а произведение i-го коэффициента регрессии на соответствующий x_i имеет наименование (руб.).

Дальнейшее исключение членов приводило к росту остаточной дисперсии.

Во всех вариантах для определения тесноты связи находилось множественное корреляционное отношение

$$\eta = \sqrt{1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}}. \tag{8}$$

Кроме того, определялось дисперсионное отношение F, которое сравнивалось с критическим значением для 0,05 уровня значимости.

В последнем варианте $F = 2{,}75$, что превышает критическое значение для 0,01 уровня значимости со степенями свободы 32 и 24.

Таким образом, была получена интерполяционная формула, описывающая изучаемые показатели.

Для того, чтобы экономически интерпретировать полученную модель, будем поочередно закреплять все факторы, кроме одного, на постоянном (среднем) уровне. Средние значения факторов следующие: $\bar{x}_1 = 6098{,}6$, $\bar{x}_2 = 8{,}60$, $\bar{x}_3 = 10{,}78$, $\bar{x}_4 = 7{,}13$.

Подставим в уравнение (7) $\bar{x}_2$, $\bar{x}_3$, $\bar{x}_4$, тем самым определив влияние x_1:

$$y_{x_1} = -0{,}5258x_1 + 0{,}00003646x_1^2 + 9050{,}2, \tag{9}$$

$$\Delta\Delta_{x_1} y = -0{,}5275 + 0{,}00007292x_1. \tag{10}$$

С ростом производственных фондов на 1 тыс. руб. выработка растет на $(-0{,}5275 + 0{,}00007292x_1)$ руб.

Например, для Калининского комбината искусственного волокна № 513, где величина производственных фондов равна 8484 тыс. руб., с увеличением производственных фондов на 1 тыс. руб. выработка растет на 0,091 руб.

Чтобы определить влияние фондовооруженности x_2 на выработку, закрепим на среднем уровне x_1, x_3, x_4:

$$y_{x_2} = 87{,}23x_2 + 19852{,}38, \tag{11}$$

$$\Delta_{x_2}y = 87{,}23 \text{ руб.} \tag{12}$$

С увеличением фондовооруженности на 1 тыс. руб. на одного рабочего выработка растет на 87,23 руб.

Для того чтобы найти влияние электровооруженности x_3 на выработку, установим на среднем уровне x_1, x_2, x_4:

$$y_{x_3} = 294{,}5x_3 + 16427{,}54, \tag{13}$$

$$\Delta_{x_3}y = +294{,}5 \text{ руб.} \tag{14}$$

С увеличением электровооруженности на 1 *квт-час* выработка растет на 294,5 руб.

Для установления влияния ассортиментных сдвигов (удельного веса новой продукции x_4) на выработку зафиксируем на среднем уровне x_1, x_2, x_3:

$$y_{x_4} = -319{,}079x_4 + 5{,}2x_4^2 + 21623{,}3, \tag{15}$$

$$\Delta_{x_4}y = -313{,}879 + 10{,}4x_4. \tag{16}$$

Так, для того же Калининского комбината искусственного волокна № 513, где удельный вес новой продукции составляет 3,8%, с ростом этого удельного веса на 1% выработка падает на 276,439 руб. Рост удельного веса новой продукции будет вести к уменьшению выработки, так как всякий новый вид продукции требует значительных трудовых затрат на его освоение.

Модель (7) можно рекомендовать для анализа влияния факторов на уровень производительности труда для предприятий химических волокон. Ее также можно использовать для проверки текущих и составления перспективных планов по производительности труда. При этом необходимо отметить, что эксплуатацию по формуле (7) нужно проводить крайне осторожно, учитывая, с одной стороны, незначительный объем выборки, а с другой стороны — тот факт, что при моделировании рассматривались не все факторы, влияющие на производительность труда.

По мере накопления новой информации следует строить новые модели, тем самым производя корректировку на изменения, происшедшие в отрасли.

ЛИТЕРАТУРА

1. Статистические модели и методы в экономическом анализе и планировании. Сб. научн. тр. Новосибирск, 1963 (ротапринт).
2. В. П. Хайкин, В. С. Найденов, С. Г. Галуза. Корреляция и статистическое моделирование в экономических расчетах. М., «Экономика», 1964.
3. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.
4. А. Хальд. Математическая статистика с техническими приложениями. М., Изд-во иностр. лит., 1956.
5. M. Ezekiel, K. Fox. Methods of Correlation and Regression Analysis. New-York, 1959.
6. E. J. Williams, D. Sc. Regression Analysis. New-York, 1959.

Поступила в редакцию
10 III 1965