ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ ОТРАСЛИ С ПОМОЩЬЮ КОМПЛЕКСА СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

МАЙДУКОВ Г. Л., КАИРА З. С.

(Донецк)

Рассмотрим задачу прогнозирования показателей основных потребительских качеств добываемых углей (зольность и выход сортового топлива *) с помощью методов математической статистики.

Исходный временной ряд сформирован с помощью метода экспоненциального сглаживания по статистическим данным о среднединамической зольности горной массы A^c (в %), добытой шахтами Минуглепрома Украинской ССР за 1956–1976 гг. Тренд этого ряда удовлетворительно (коэффициент корреляции $r=0{,}798$, остаточное среднеквадратичное отклонение $\sigma=1{,}04\%$) описывается уравнением прямой **

$$A_1^c=17{,}862+0{,}391t. \qquad (1)$$

Оптимальное значение параметра сглаживания α из условия минимума среднеквадратичного остаточного отклонения принято равным 0,35. Результаты расчетов приведены в табл. 1, вариант 1.

Таблица 1

Прогнозируемая зольность добываемого угля и ошибки прогноза

Вариант	Показатели	Годы						
		1	2	3	4	5	6	7
1	A_1^c	30,16	31,17	32,18	33,19	34,20	35,21	36,22
	Δ_1	0,50	0,64	0,79	0,93	1,08	1,22	1,36
2	A_2^c	29,54	30,81	32,17	33,59	35,10	36,67	38,35
	Δ_2	0,32	0,37	0,43	0,49	0,57	0,63	0,70
3	A_3^c	29,71	30,4	31,09	31,78	32,47	33,30	34,50
	Δ_3	2,27	2,29	2,33	2,36	2,40	2,44	2,48
4	A_4^c	30,10	30,81	31,45	32,00	32,54	33,10	33,68
	Δ_4	1,93	2,01	2,04	2,07	2,10	2,13	2,17
5	A_5^c	30,91	31,72	32,43	33,22	34,01	34,72	35,43
	Δ_5	1,80	1,91	2,01	2,11	2,22	2,42	2,62

При прогнозе по варианту 2 использован ряд коэффициентов регрессии, вычисленных методом экспоненциального сглаживания. Первая пара значений ($\hat{a}_0$ и $\hat{a}_1$) получена по 10 начальным точкам временного ряда, каждая последующая – путем его удлинения на одну точку. Характер изменения коэффициентов регрессии в ряду полученных моделей хорошо ($r_{\hat{a}_0}=-0{,}994$, $r_{\hat{a}_1}=0{,}991$) аппроксимируется уравнениями

$$\hat{a}_0=20{,}95-0{,}22t, \qquad (2)$$

$$\hat{a}_1=-0{,}2214+0{,}0383t. \qquad (3)$$

Прогнозируемая зольность, вычисленная подстановкой в уравнение экстраполированных по (2) и (3) коэффициентов регрессии, и ее вероятные ($p=0{,}9$) ошибки приведены в табл. 1, вариант 2.

В качестве третьей прогнозной модели использовано многофакторное уравнение регрессии, в котором независимыми переменными являются факторы, оказывающие (по мнению специалистов или по результатам экспериментов) наибольшее влияние на рост зольности выдаваемой шахтами горной массы. Это уровни ***: узкозахватной

---

* Зольность характеризует массовое содержание в угле минеральных веществ; выход сортового топлива – массовое содержание в подвергнутом рассортировке угле кусков, оставшихся на сите с отверстиями ячеек 6×6 мм².

** Здесь и далее нижний индекс при A^c обозначает вариант прогноза.

*** Доля в процентах в общем объеме.

выемки (x_1), выемки механизированными комплексами (x_2), механизированной навалки (x_3), добычи угля из подготовительных забоев (x_4), конвейерной откатки по основным горизонтальным выработкам (x_5), добычи из забоев, где управление кровлей осуществляется полной или частичной закладкой породы (x_6), а также относительные (на 1000 *т* добычи) масса выдаваемой породы (x_7) и протяженность отремонтированных выработок (x_8); кроме того, введено время (x_9), вычисляемое как разность календарного и 1955 гг.

В результате получена линейная модель

$$A_3{}^c=0{,}217-0{,}030x_1+0{,}277x_2+0{,}128x_3+0{,}399x_4-0{,}706x_5+0{,}291x_6+ +0{,}002x_7-0{,}034x_8+0{,}709x_9. \quad (4)$$

Модель (4), для которой множественное корреляционное отношение $R=0{,}993$, а F-критерий превышает табличное значение при 0,05%-ном уровне значимости, характеризует возмущения функции отклика в исследованном диапазоне вследствие

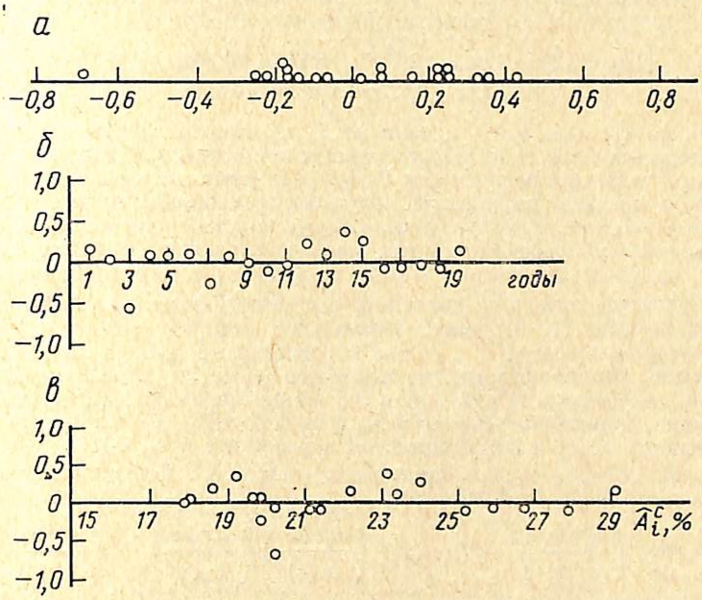

Рис. 1. Исследование остатков многофакторной модели (5): *а* – общий график; *б* – график временно́й последовательности; *в* – график зависимости остатков от предсказываемых значений A^c

варьирования или динамики независимых переменных. Для получения прогнозных значений $A_3{}^c$ в (4) подставлялись экстраполированные по методу наименьших квадратов значения x_i.

Для x_1, x_2, x_3 наиболее адекватными оказались логистические уравнения, уровень насыщения каждой из этих величин равен 90, 80 и 100% соответственно. Для других независимых переменных более подходящими (по величине остаточного среднеквадратичного отклонения и коэффициенту корреляции) являются уравнения экспоненты ($y=e^{a_0+a_1t}$).

Остатки (разности между фактическими и предсказанными значениями) исследовались с помощью графиков (рис. 1), где, несмотря на некоторую неупорядоченность точек, аномальность не обнаруживается. Из рис. 1, *б* следует, что эффект времени на данных не сказывается, т. е. остатки не обнаруживают явной зависимости. «Горизонтальная полоса» значений $\hat{A}_i{}^c$, соответствующих величинам остатков e_i на рис. 1, *в*, не указывает на какую-либо ненормальность, поэтому использование в рассматриваемом случае метода наименьших квадратов оправдано.

Учитывая, что развитие технико-экономических показателей редко подчиняется строгому линейному закону, была также рассчитана многофакторная показательная модель. После отсева статистически незначимых (по критерию Стьюдента) факторов она приобрела вид

$$A_4{}^c=93{,}45\,\frac{(x_1+0{,}01)^{0{,}0169}(x_2+0{,}01)^{0{,}013}}{x_4^{0{,}359}x_5^{0{,}126}x_6^{0{,}155}}. \quad (5)$$

Значения зольности (см. табл. 1, вариант 4) также предсказывались подстановкой экстраполируемых величин факторов-аргументов в (5).

Доверительные интервалы значений, определенных по (4) и (5), вычислялись по формуле

$$\Delta=\pm K\sqrt{\sum_{i=1}^{n} a_i^2\sigma_{x_i}^2}, \tag{6}$$

где K – табулированная функция, зависящая от количества наблюдений и периода упреждения [1]; a_i – коэффициент регрессии фактора i; σ_{x_i} – среднеквадратичное остаточное отклонение, вычисленное для тренда фактора i.

Для преобразования исходных переменных в новый линейный набор величин, обладающих свойством ортогональности и расположенных в порядке убывания дисперсий, был использован метод главных компонент. Рассчитанная с его помощью модель после перехода к натуральным переменным имеет вид

$$A_5{}^c=22{,}215-0{,}342x_1+2{,}745x_2+5{,}146x_3+0{,}394x_4-1{,}637x_5+ \\ +2{,}111x_6-0{,}413x_7-0{,}058x_8-0{,}637x_9. \tag{7}$$

Зольность по (7) (см. табл. 1, вариант 5) прогнозировалась так же, как по (4) и (5), т. е. подстановкой экстраполированных значений факторов в модель. Доверительные интервалы при вероятности $P=0{,}9$ рассчитывались по (6).

Полученные частные результаты могут быть обобщены путем построения комбинированной зависимости из условия минимума дисперсии ошибки прогноза [2], т. е. это может быть представлено как последовательное объединение результатов двух любых частных прогнозов. Тогда остатки e_i, т. е. алгебраическая разность фактических y_i и расчетных $\hat{y}_i$ значений, рассматриваются как реализация случайных величин, $i=1, 2, \ldots$, имеющих нормальное распределение с параметрами, не зависящими от y_i и времени t, а также независимыми для различных наблюдений.

Объединение частных прогнозов, например y_1 и y_2, обусловлено допущением, что получаемая в результате двумерная случайная величина $(\varepsilon_1, \varepsilon_2)$ также нормально распределена, характеризуется нулевым математическим ожиданием, определенными дисперсиями $\sigma_i^2(\varepsilon)$ и коэффициентом корреляции $\rho(\varepsilon_1, \varepsilon_2)$.

Нахождение объединяющего прогноза $\hat{y}_{12}$ сводится к вычислению значения y, минимизирующего квадратичную форму

$$\varphi(y)=\left(\frac{y-y_1}{\sigma_1(\varepsilon)}\right)^2-2\rho(\varepsilon_1,\varepsilon_2)\left(\frac{y-y_1}{\sigma_1(\varepsilon)}\right)\left(\frac{y-y_2}{\sigma_2(\varepsilon)}\right)+\left(\frac{y-y_2}{\sigma_2(\varepsilon)}\right)^2. \tag{8}$$

Применяя обычную технику нахождения безусловных экстремумов, получим

$$\hat{y}_{12}(y_1, y_2)=\frac{\sigma_2^2(\varepsilon)-\rho(\varepsilon_1,\varepsilon_2)\,\sigma_1(\varepsilon)\,\sigma_2(\varepsilon)}{\sigma_1^2(\varepsilon)-2\rho(\varepsilon_1,\varepsilon_2)\,\sigma_1(\varepsilon)\,\sigma_2(\varepsilon)+\sigma_2^2(\varepsilon)}\,y_1+ \\ +\frac{\sigma_1^2(\varepsilon)-\rho(\varepsilon_1,\varepsilon_2)\,\sigma_1(\varepsilon)\,\sigma_2(\varepsilon)}{\sigma_1^2(\varepsilon)-2\rho(\varepsilon_1,\varepsilon_2)\,\sigma_1(\varepsilon)\,\sigma_2(\varepsilon)+\sigma_2^2(\varepsilon)}\,y_2 \tag{9}$$

или

$$\hat{y}_{12}(y_1,\ y_2)=c_1y_1+c_2y_2, \tag{10}$$

где $c_1+c_2=1$. Веса c_i в (10) независимы от фиксируемых значений y_1 и y_2, что позволяет рассматривать данную функцию как объединяющую частные прогнозы и имеющую ряд простых свойств. При $\rho(\varepsilon_1, \varepsilon_2)=0$ имеем

$$c_1=\frac{\sigma_2^2(\varepsilon)}{\sigma_1^2(\varepsilon)+\sigma_2^2(\varepsilon)},\qquad c_2=\frac{\sigma_1^2(\varepsilon)}{\sigma_1^2(\varepsilon)+\sigma_2^2(\varepsilon)}.$$

Поэтому

$$c_1{:}c_2=\frac{1}{\sigma_1^2(\varepsilon)}:\frac{1}{\sigma_2^2(\varepsilon)},$$

т. е. прогнозы y_1 и y_2 взвешиваются пропорционально их так называемым точностям. При $\sigma_1(\varepsilon):\sigma_2(\varepsilon)\to 0$ имеем $c_1\to 1$, следовательно, существенно более точная частная зависимость y_1 будет иметь в (9) преобладающий вес. Для допустимых прогнозов, когда $\sigma_1^2(\varepsilon)$ и $\sigma_2^2(\varepsilon)$ достаточно близки, удельные веса c_i различаются мало.

Для получения оценок искомых параметров c_i, $i=1, \ldots, 5$, используем прием, применяемый в методе группового учета аргументов (МГУА) [3]. Блок-схема вычислительных операций приведена на рис. 2.

Значения обобщающих коэффициентов c_i частных прогнозов зольности, вычисленные по (9), для моделей: 1) –0,098; 2) –0,902; 3) –1,017; 4) –0,017; 12) –0,012; 34) –0,988; 1234) –0,994; 5) –0,006 *.

Окончательно после подстановки частных значений A_i^c имеем

$$A_0^c=0{,}012A_1^c+0{,}0107A_2^c+0{,}167A_3^c-0{,}9988A_4^c+0{,}0061A_5^c. \tag{11}$$

Обобщенный по пяти моделям прогноз зольности приведен в табл. 2. Те же значения коэффициентов приняты и при расчете доверительных интервалов.

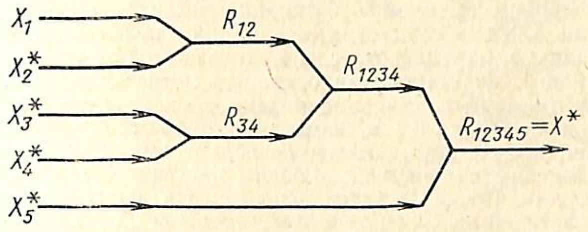

Рис. 2. Блок-схема процесса обобщения частных прогнозов

Наибольший весовой коэффициент соответствует варианту 4 прогноза, что является вполне обоснованным, так как в нем оказалась минимальная дисперсия остатков.

Столь большой вклад одной из частных моделей в обобщенный результат в нашем расчете случаен, что может быть показано на примере прогноза среднегодового суммарного выхода сортового топлива γ в угольной промышленности Украинской ССР.

Таблица 2

Прогнозируемая зольность горной массы, %

Показатели	Годы						
	1	2	3	4	5	6	7
A_0^c	30,08	30,79	31,44	31,99	32,53	33,09	33,68
ΔA_0^c	1,96	1,99	2,02	2,05	2,08	2,11	2,15

Варианты прогноза среднегодового выхода сортового топлива γ в отрасли найдены с использованием моделей: 1) экспоненциального сглаживания, 2) многофакторного регрессионного анализа, 3) главных компонент, 4) группового учета аргументов, 5) экстраполяции по нелинейным моделям вида $y_t=t^{a_x}e^{a_0+a_1t}$. Для получения обобщающего прогноза использовано (10).

Обобщающие коэффициенты c_i для вычисления прогнозируемых значений γ — для моделей: 1) —0,521; 2) —0,479; 3) —0,254; 4) —0,746; 12) —0,245; 34) —0,755; 1234) —0,892; 5) —0,108.

Сопоставляя полученную комплексную модель

$$\gamma_0=0{,}114\gamma_1+0{,}104\gamma_2+0{,}172\gamma_3+0{,}502\gamma_4+0{,}108\gamma_5 \tag{12}$$

с (11), нетрудно заметить, что в (12) вклад частичных прогнозов относительно равномернее: если в первой на долю варианта 4 приходилось 99,88%, то во второй — только 50,2%. Однако вариация прогнозируемых значений в обоих случаях, увеличиваясь к концу прогнозируемого ряда значений, остается небольшой (4,50 и 5,99% соответственно).

Результаты обобщенного прогноза, полученные по комплексу статистических моделей, приведены в табл. 3.

Показатели A_0^c и γ_0 характеризуются высокой близостью фактическим значениям, полученным в отрасли в 1976, 1977 и первых трех месяцах 1978 г., что

* Нумерация промежуточных моделей соответствует блок-схеме.

Таблица 3

Прогнозируемый среднегодовой суммарный выход сортового топлива

Показатели	Годы						
	1	2	3	4	5	6	7
γ_0	39,95	39,67	39,43	39,15	39,02	38,85	38,69
$\Delta\gamma_0$	2,97	3,03	3,08	3,15	3,21	3,27	3,33

обусловлено сохранением тенденций и инерционностью развития техники и технологии добычи угля. Однако следует иметь в виду незамкнутость использованных моделей, т. е. наличие параметров, с помощью которых возможно управляющее воздействие на процесс формирования основных потребительских качеств добываемой горной массы (например, создание механизированных комплексов для выемки пластов мощностью от 0,5 до 0,8 м, изменение принципа разрушения угля и др.).

Таким образом, предлагаемый метод позволяет получить достаточно обоснованную прогнозную оценку получения угольной продукции с определенными качественными параметрами. Предполагается использовать данный метод и при разработке прогнозов показателей на уровне производственных единиц и объединений.

ЛИТЕРАТУРА

1. *Четыркин Е. М.* Статистические методы прогнозирования. М.: Статистика, 1975.
2. *Ершов Э. Б.* Об одном методе объединения частных прогнозов.— В кн. Статистический анализ экономических временных рядов и прогнозирование. М.: Наука, 1973.
3. *Ивахненко А. Г., Лапа В. Г.* Предсказание случайных процессов. Киев: Наукова думка, 1971.

Поступила в редакцию
25 VII 1978

О ВЫЯВЛЕНИИ ПОТРЕБИТЕЛЬСКОГО СПРОСА

ШЕРМЕНЕВ А. М.

(Москва)

Будем предполагать, что отношение потребительского предпочтения на множестве товаров задано функцией полезности. Спрос и условия замещения одних товаров другими определяют множество безразличия, интерпретируемое как данный уровень благосостояния. Возникает вопрос: можно ли по этому уровню восстановить потребительское предпочтение?

Пусть S — гиперповерхность в R^n. Нас интересуют такие функции полезности, для которых S — гиперповерхность безразличия. Подобных функций, конечно, очень много. Выделим среди них класс F вида

$$u(x_1, \ldots, x_n) = u_1(x_1) + \ldots + u_n(x_n)$$

и будем искать в нем решение задачи. В заметке приводятся необходимые и достаточные условия для гиперповерхности $S \subset R^n$, при которых существует функция $u = u_1(x_1) + \ldots + u_n(x_n)$ такая, что S — множество уровня $u = c$. Единственность (если $u = u_1(x_1) + \ldots + u_n(x_n)$ существует, то она определена однозначно с точностью до линейного преобразования $u \to au + b$) установлена А. С. Тангяном [1]. Короткое доказательство этого свойства приводится здесь для полноты. Соответствующие функциям из F отношения предпочтения называются независимыми [2, 3], причем независимость не имеет строгого экономического обоснования, но является весьма правдоподобной и распространенной гипотезой (см., например, [3]).

Для композиции $z = g(f(x))$ двух функций $y = f(x)$ и $z = g(y)$ будем использовать