

ПОСТРОЕНИЕ ТРЕНДОВЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ СПРОСА НА УГОЛЬ

По данным о продажах фирмы, занимающейся поставками каменного угля в районы Томской области, проводится анализ спроса на уголь и строится прогноз с использованием трендовых моделей.

Прогнозирование является одной из наиболее важных задач при принятии управленческих решений. Для предсказания поведения спроса в будущем и вычисления ошибок прогнозирования требуется некоторая информация, касающаяся предыстории спроса. В связи с этим для анализа спроса приходится использовать данные о продажах. Если неудовлетворенный спрос откладывается до очередной поставки угля на склад, то данные о продажах можно считать условно эквивалентными данным о спросе. Под условной эквивалентностью подразумевается возможное искажение картины спроса вследствие того, что момент продажи связан с моментом выполнения заказа на поставку, поэтому временной характер продаж может отличаться от картины возникновения спроса во времени.

Данные о продажах угля представляют собой динамический временной ряд. На рис. 1 они изображены графически. На графике продаж угля имеются отчетливые годовые периоды, присутствуют резко выраженные пики, амплитуда колебаний постоянна. Пики объясняются наличием сезонности. Действительно, весной спрос на уголь резко возрастает в связи с началом навигации. Потребители пытаются как можно больше угля вывезти по “большой” воде – это гораздо дешевле. Однако су-

ществует ряд клиентов, которым выгоднее возить уголь автотранспортом или железной дорогой, что объясняет наличие спроса в другое время года.

Так как спрос на уголь имеет явно выраженный сезонный характер, для составления прогноза спроса исследуем следующие две модели представления временных рядов: аддитивную модель [1] и модель тригонометрического тренда [2], представленные ниже соответствующими уравнениями (1) и (2):

$$y_t = m_t + s_t + \varepsilon_t, \quad \forall t = \overline{1, T}, \quad (1)$$

где m_t – гладкая составляющая ряда, а s_t – сезонная компонента;

$$y_t = \alpha_0 + \sum_{j=1}^{\frac{m}{2}-1} \left[\alpha_{2j-1} \cos\left(\frac{2\pi j}{m}t\right) + \alpha_{2j} \sin\left(\frac{2\pi j}{m}t\right) \right] + \alpha_{m-1}(-1)^t + \varepsilon_t, \quad \forall t = \overline{1, T}. \quad (2)$$

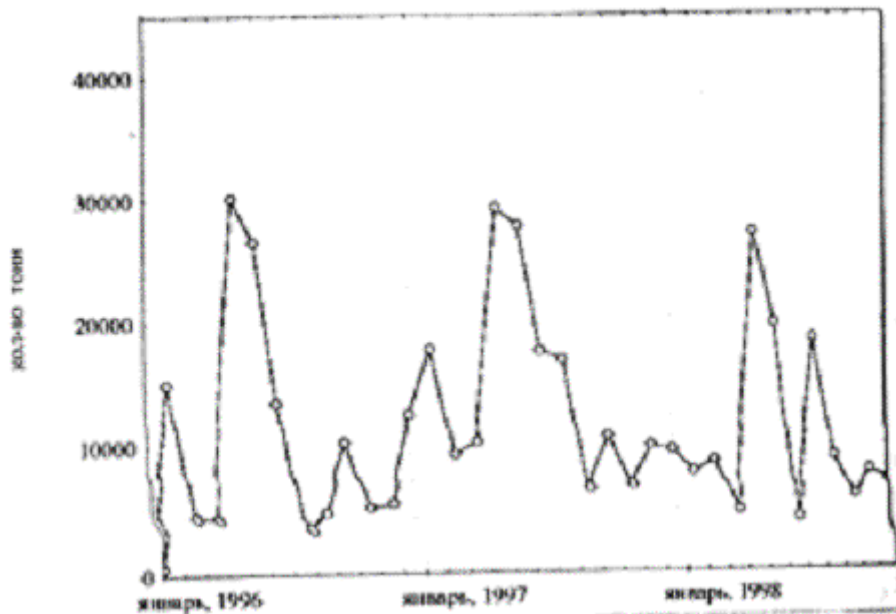


Рис. 1. График продаж угля

Эти модели состоят из детерминированной составляющей и ошибок наблюдения ε_t , где ошибки можно считать случайными с математическим ожиданием $E\varepsilon_t = 0$ и дисперсией

$$D\varepsilon_t = \sigma^2 < \infty, \text{cov}(\varepsilon_t, \varepsilon_k) = 0, \forall t \neq k.$$

Оценочные значения для моделей (1), (2) можно представить в виде (3), (4):

$$\hat{x}_t = \hat{\mu}_t + \hat{\varepsilon}_t \quad \forall t = \overline{1, T}. \quad (3)$$

$$\hat{x}_t = \hat{\alpha}_0 + \sum_{j=1}^{\frac{m-1}{2}} \left[\hat{\alpha}_{2j-1} \cos\left(\frac{2\pi j}{m}t\right) + \hat{\alpha}_{2j} \sin\left(\frac{2\pi j}{m}t\right) \right] + \hat{\alpha}_{m-1}(-1)^t, \quad \forall t = \overline{1, T}. \quad (4)$$

Воспользовавшись методом наименьших квадратов оценки параметров α_i , $i = \overline{0, m-1}$ модели (2), будем иметь следующий вид оцениваемых коэффициентов [2]:

$$\begin{cases} \hat{\alpha}_0 = \bar{y}; \\ \hat{\alpha}_{2j-1} = \frac{2}{T} \sum_{t=1}^T y_t \cos\left(\frac{2\pi j}{m}t\right), \quad j = 1, 2, \dots, \frac{m}{2} - 1; \\ \hat{\alpha}_{2j} = \frac{2}{T} \sum_{t=1}^T y_t \sin\left(\frac{2\pi j}{m}t\right), \quad j = 1, 2, \dots, \frac{m}{2} - 1; \\ \hat{\alpha}_{m-1} = \frac{1}{T} \sum_{t=1}^T y_t (-1)^t. \end{cases}$$

Так как анализ спроса проводится по данным о продажах за три года, а спрос является сезонным с лагом в один год, то $T = 36$, $m = 12$.

Для получения оценок аддитивной модели (1) был использован метод сезонной декомпозиции [1].

На рис. 2 изображены значения спроса на уголь, полученные по моделям (3) и (4): сплошной линией изображены наблюдаемый (фактический) спрос, пунктирными – значения, полученные моделированием.

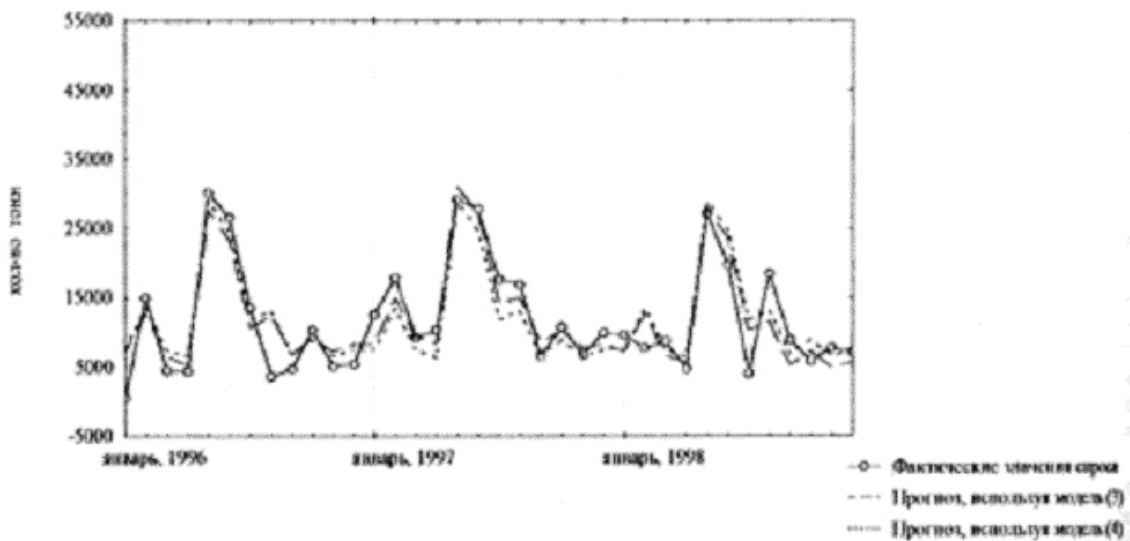


Рис. 2. График фактических объемов продаж угля с прогнозными значениями, полученные по формулам (3), (4)

Исследуя остатки модели (разность между наблюдаемыми и модельными значениями), мы можем оценить степень адекватности модели. Если модель подобрана правильно, то остатки будут вести себя достаточно хаотично, в известном смысле они будут напоминать белый шум. В остатках правильно построенной модели не должно быть систематической составляющей, резких выбросов, в чередовании их знаков не должно быть никаких закономерностей, и, наконец, остатки должны быть независимы друг от друга.

Проверим свойства остатков модели тригонометрического тренда (4). Выборочное среднее остатков $\bar{\mathcal{E}} \approx 0$; несмещенная оценка среднеквадратичного отклонения $\mathcal{S}_e = 4460$. Рассмотрим гипотезу о согласии распределения остатков с нормальным распределением. На рис. 3 изображена гистограмма (оценка плотности распределения) остатков модели с наложенной на нее плотностью нормального распределения. Значение статистики $\chi^2 = 2,689$, вероятности $p=0,44$. Так как вероятность p неправильного отвержения гипотезы, когда она верна, довольно большая, то гипотеза о нормальности остатков модели принимается на уровне значимости $\alpha=0,05$.

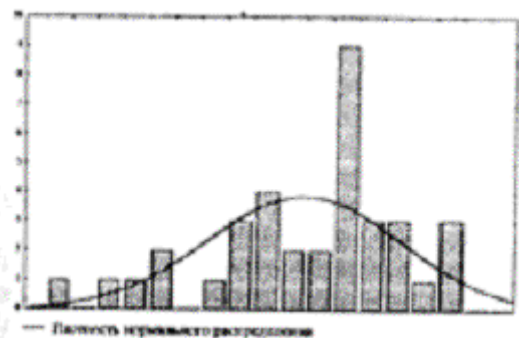


Рис. 3. Гистограмма остатков построенной модели (4).

Исследуем зависимость между остатками. Так как \mathcal{E}_t – нормальные случайные величины, понятия некоррелируемости и независимости эквивалентны. Гипотеза о независимости остатков имеет вид: $H_0: \rho_k=0, k=1,2,\dots$, тогда $H_1: \exists k: \rho_k \neq 0$ – альтернативная гипотеза.

Если статистика

$$\sqrt{T-k-2} \frac{\mathcal{F}_k}{\sqrt{1-\mathcal{F}_k^2}} > St_{T-k-2} \left(1 - \frac{\alpha}{2} \right),$$

то гипотеза H_0 отвергается на уровне значимости α (следовательно, остатки зависимы), иначе – нет оснований отвергать гипотезу H_0 .

Обычно гипотезу $H_0: \rho_k = 0$ проверяют для $k = \overline{1,5 \div 8}$. Если для всех k гипотеза принимается, то считается, что модель адекватно отображает экспериментальные данные. Проведем соответствующие вычисления для остатков модели (4).

Проверим гипотезу $H_0: \rho_k = 0$ для $k=1$: статистика критерия равна 0,765, что меньше значения $St_{T-k-2}(1-0,5\alpha) = 2,03$, следовательно, H_0 верна на уровне значимости $\alpha = 0,05$. Для остальных значений k результаты вычислений приведены в табл. 1, из которой видно, что гипотеза $H_0: \rho_k = 0$ принимается для всех k на уровне значимости $\alpha = 0,05$.

Т а б л и ц а 1

k	Статистика критерия	Табличное значение $St_{T-k-2}(1-0,5\alpha)$
2	0,209	2,03
3	1,255	2,03
4	0,814	2,04
5	0,856	2,05
6	0,984	2,05
7	0,088	2,05

График автокорреляционной функции остатков на рис. 4 показывает, что ни один из коэффициентов автокорреляции значимо не отличается от нуля при $\alpha = 0,05$.

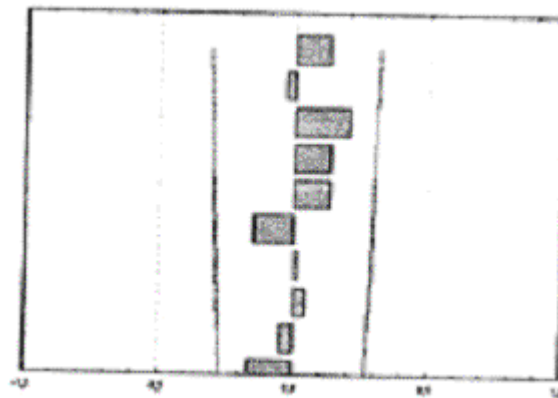


Рис. 4. Автокорреляционная функция остатков модели (4).

Так как остатки нормально распределены с нулевым средним, конечной дисперсией и независимы, конечной дисперсией и независимы, построенная прогнозная модель адекватно описывает реальные данные. Аналогично исследуются остатки аддитивной модели (3).

В качестве меры того, насколько хорошо модель описывает систему наблюдений, рассмотрим коэффициент детерминации R^2 , который дает долю объясненной дисперсии на базе выбранной модели, и, чем ближе R^2 к 1, тем лучше выбрана модель. Для модели тригонометрического тренда $R^2 = 0,792$, для аддитивной модели – 0,767, поэтому прогноз спроса будем строить на базе модели тригонометрического тренда (4). В связи с небольшим количеством данных есть основания полагать, что эта модель достаточно хорошо описывает их.

Предполагая, что структура спроса на уголь в 1999 г. существенно не изменится, используя предлагаемую нами модель тригонометрического тренда (4), построим прогноз продаж угля на 1999 г. Прогноз строится на основании уравнения [2]:

$$\begin{aligned} \hat{y}_{T+\tau} = & \hat{\sigma}_0 + \sum_{j=1}^{\frac{m-1}{2}} \left[\hat{\sigma}_{2j-1} \cos\left(\frac{2\pi j}{m}(T+\tau)\right) + \right. \\ & \left. + \hat{\sigma}_{2j} \sin\left(\frac{2\pi j}{m}(T+\tau)\right) \right] + \hat{\sigma}_{m-1} (-1)^{T+\tau}, \\ & \forall \tau = 1, 12, \quad T = 36, \quad m = 12. \end{aligned}$$

Доверительный интервал для прогнозных значений [2] имеет вид:

$$\hat{y}_{T+\tau} - t_{1-0,5\alpha}(T-m) \sqrt{\frac{m\hat{\sigma}_\varepsilon^2}{T}} \leq y_{T+\tau} \leq$$

$$\hat{y}_{T+\tau} + t_{1-0,5\alpha}(T-m) \sqrt{\frac{m\hat{\sigma}_\varepsilon^2}{T}},$$

где $\hat{\sigma}_\varepsilon^2$ – оценка дисперсии случайной составляющей, $t_{1-0,5\alpha}(T-m)$ – квантиль уровня $1-0,5\alpha$ распределения Стьюдента с числом степеней свободы $T-m$, $\alpha = 0,05$. Результаты прогнозирования и построения доверительного интервала изображены на рис. 5.

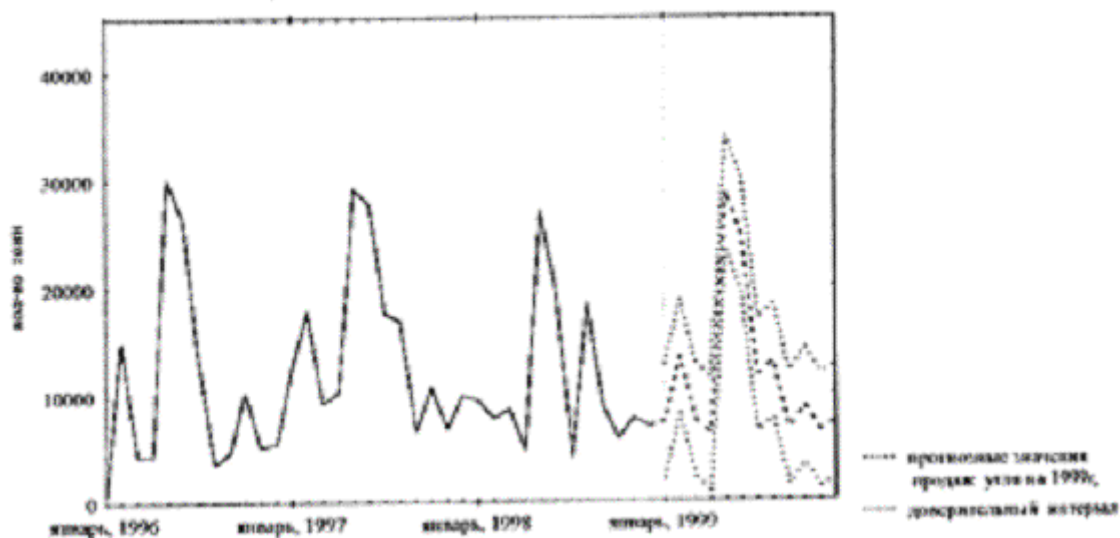


Рис. 5. График продаж с построенным прогнозом

В результате анализа временного ряда продаж угля удалось построить модель, отображающую исходные данные. Сезонная составляющая модели была представлена в виде тригонометрического тренда, что привело к функции оценки спроса вида (4). С помощью этой модели на основе наблюдаемых

значений продаж за 1996–1998 гг. был построен прогноз спроса на уголь на 1999 г.

Важно отметить, что прогноз является основным ориентиром, позволяющим определить потребности потребителей, и неверный прогноз может повлечь за собой значительное уменьшение прибыли.

ЛИТЕРАТУРА

1. Кендалл М. Временные ряды. М.: Финансы и статистика, 1981.
2. Глинский В.В., Ионин В.Г. Статистический анализ. М.: Филинь, 1998.

Статья представлена кафедрой прикладной математики факультета прикладной математики и кибернетики Томского государственного университета, поступила в научную редакцию 19 февраля 2000 г.