

ПОСТРОЕНИЕ ТРЕНДОВЫХ МОДЕЛЕЙ ЭКОНОМИЧЕСКОЙ СИСТЕМЫ

Рассмотрен подход прогнозирования экономических показателей, основанный на построении трендовых моделей.

Задачами экономико-статистического прогнозирования являются выявление перспектив ближайшего или более отдаленного будущего в исследуемой области на основе реальных процессов деятельности; выработка оптимальных тенденций и перспективных планов с учетом составленного прогноза и оценки принятого решения с позиций его последствий в прогнозируемом периоде.

Временной ряд, динамический ряд (РД) - это последовательность упорядоченных по времени показателей, характеризующих уровень развития изучаемого явления. В составе динамического ряда можно выделить четыре компоненты: 1) общую тенденцию развития, или тренд; 2) регулярные колебания относительно тренда (типа циклов); 3) сезонные колебания; 4) остаток, или случайную компоненту, отражающую влияние разнообразных факторов стохастического характера.

Одной из важнейших задач исследования динамических рядов является установление общих закономерностей, или тенденций, развития. Для решения этой задачи используются разнообразные приемы уменьшения колеблемости динамического ряда (сглаживающие фильтры), среди которых можно выделить два основных метода: сглаживание ряда с помощью скользящей средней и аналитическое выравнивание. Аналитическое выравнивание динамического ряда - это метод выражения тенденций развития в виде функции изучаемого показателя от времени, называемой моделью тренда.

Рассмотрим задачу сглаживания ряда динамики, т.е. построение трендовой модели, на реальном примере с целью применения модели для решения задач анализа и прогнозирования социально-экономических показателей. Данные о продаже авиабилетов за 3 года ($N=36$ - длина РД) представлены графически на рис. 1.

Для этого РД характерны внутригодовые, повторяющиеся устойчиво из месяца в месяц изменения в уровнях. Это стационарный периодический РД, т.е. общей тенденции развития нет, но явно выделяется сезонная составляющая временного ряда и, естественно, случайная компонента. Поэтому для сглаживания данного РД будем использовать в качестве моделей прогноза мультипликативную модель и модель тригонометрического тренда - ряд Фурье [1]. Если в РД отсутствует тенденция, то уровень временного ряда рассматривается как функция сезонности и случайности:

$$y_t = f(S, \epsilon),$$

где y - фактические уровни РД, S - сезонная составляющая, ϵ - случайная компонента с математическим ожиданием $E\epsilon_t = 0$ и дисперсией

$$D\epsilon_t = \sigma^2 < \infty, \quad \cos(\epsilon_t, \epsilon_k) = 0, \quad \forall t \neq k.$$

При мультипликативной модели уровень РД можно представить как произведение его составляющих:

$$y_t = \bar{y} \cdot \frac{\bar{y}_t}{\bar{y}} \cdot \frac{y_t}{\bar{y}_t}, \quad t = \overline{1, N}, \quad (1)$$

где \bar{y}_t/\bar{y} - коэффициент сезонности (K_s); y_t/\bar{y}_t - отражает влияние случайного фактора, \bar{y}_t - средний уровень ряда соответствующего периода внутри года (месяца, квартала) за несколько лет.

Чем больше коэффициент сезонности, тем больше амплитуда колебаний уровней ряда относительно его среднего уровня, тем существеннее влияние сезонности. Чем меньше влияние случайной составляющей, тем в большей мере рассматриваемая модель адекватно описывает исходный РД. На рис. 2 изображен коэффициент сезонности K_s для каждого месяца.

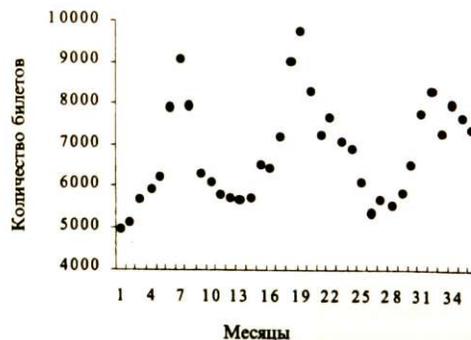


Рис. 1. Периодический ряд динамики продаж авиабилетов

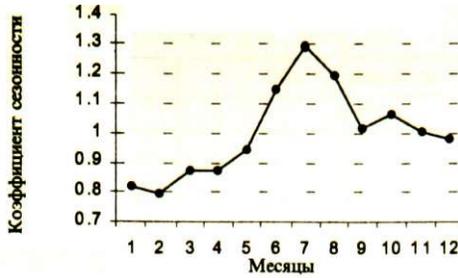


Рис. 2. Коэффициент сезонности

Оценочные значения для модели (1) представим в виде

$$\hat{y}_t = \bar{y} (K_s)_t, \quad t = \overline{1, N}.$$

Прогнозирование РД с помощью модели (1) сводится к прогнозированию среднего уровня (\bar{y}_p) с последующей корректировкой его на сезонную компоненту - умножение на K_s :

$$\hat{y}_t = \bar{y}_p (K_s)_t, \quad t > N. \quad (3)$$

Теоретически любой стационарный временной ряд может быть представлен как сумма среднего значения (\bar{y} заменяется часто параметром a_0) и ряда синусоид и косинусоид, что и называется рядом Фурье:

$$y_t = a_0 + \sum_{i=1}^n a_i \cos \frac{2\pi i}{N} t + \sum_{i=1}^n b_i \sin \frac{2\pi i}{N} t + \varepsilon_t, \quad (4)$$

где $a_0, a_1, \dots, b_1, \dots$ - неизвестные параметры, и n - число гармоник, $t = \overline{1, N}$.

Для решения задачи идентификации параметров данного уравнения применим классический метод наименьших квадратов (МНК) и дискретный фильтр Калмана. Параметры уравнения (4), оцениваемые МНК, определяются формулами [1]:

$$\hat{a}_0 = \sum_t y_t / N = \bar{y},$$

$$\hat{a}_i = \frac{2 \sum_t y_t \cos \frac{2\pi i}{N} t}{N} \quad \text{и} \quad \hat{b}_i = \frac{2 \sum_t y_t \sin \frac{2\pi i}{N} t}{N}. \quad (5)$$

Применим дискретный фильтр Калмана [2] для идентификации параметров ряда Фурье. Так как для (4) параметры $a_0, a_p, b_p, i = \overline{1, n}$ должны быть постоянными, то модель динамической системы имеет вид

$$\begin{aligned} x(k+1) &= x(k), \\ z(k) &= H(k)x(k) + \eta, \end{aligned} \quad (6)$$

где $z(k)$ - известный выход (исходные значения y_k), η - случайные погрешности измерения, шум на выходе с ковариацией R (моделируется белым гауссовским шумом), $x(k)$ - вектор состояния, $k = \overline{1, N}$ - момент времени. В качестве наблюдаемого процесса будем рассматривать

$$y(k) = a_0 + \sum_{i=1}^n a_i \cos \frac{2\pi i}{N} k + \sum_{i=1}^n b_i \sin \frac{2\pi i}{N} k,$$

что в применении к оцениванию по методу фильтрации Калмана-Бьюси можно записать для каждого момента времени к следующим образом:

$$y(k) = H(k)x(k),$$

где матрица измерений

$$\begin{aligned} H(k) &= \left[1 \cdot \cos \frac{2\pi}{N} k \cdot \sin \frac{2\pi}{N} k \cdot \cos \frac{2 \cdot 2\pi}{N} k \cdot \sin \frac{2 \cdot 2\pi}{N} k \cdot \dots \right. \\ &\quad \left. \dots \cos \frac{2n\pi}{N} k \cdot \sin \frac{2n\pi}{N} k \right], \\ x(k) &= [a_0 \quad a_1 \quad b_1 \quad a_2 \quad b_2 \quad \dots \quad a_n \quad b_n]^T. \end{aligned}$$

Оценка состояния системы, описываемой уравнениями (4), в момент времени $k+1$ по наблюдениям $z(1), z(2), \dots, z(k)$, которая минимизирует ковариацию ошибки оценки $\varepsilon(k) = \sigma(k) - x(k)$, удовлетворяет рекуррентному уравнению

$$\begin{aligned} (k+1) &= \hat{x}(k) + K_\Phi(k) [z(k) - \hat{z}(k)]; \\ \hat{z} &= H(k) \hat{x}(k) + \eta(k); \end{aligned}$$

$$\Gamma(k+1) = \Gamma(k) H'(k) [H(k) \Gamma(k) H'(k) + R]^{-1} H(k) \Gamma(k);$$

$$K_\Phi(k) = \Gamma(k) H'(k) [H(k) \Gamma(k) H'(k) + R]^{-1}; \quad (7)$$

$$k = \overline{1, N},$$

если $\Gamma(k)$ - ковариация ошибки, $K_\Phi(k)$ - коэффициент фильтрации. В качестве начальных значений возьмём $\sigma(1) = [5000 \ 0 \ 0 \ 0 \ 0 \ \dots \ 0 \ 0]^T$, $\Gamma(1) = \Gamma_0$.

Рассмотрим построение ряда Фурье (4) для исходных данных в обоих случаях с разным числом гармоник. Выбор ряда Фурье, который наилучшим образом отражает исходный временной ряд, основывался на расчёте коэффициентов детерминации R^2 (см. таблицу), являющихся критерием адекватности построенной модели, характеризующих так называемую долю «объяснённой» дисперсии, чем ближе R^2 к 1, тем лучше выбрана модель. Коэффициент детерминации определяется как

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / (N-p)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (N-1)}, \quad R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / (N-p)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (N-1)},$$

где p - число оцениваемых параметров.

**Коэффициент детерминации для уравнений
с разным числом гармоник**

Число гармоник		R^2 при применении фильтра Калмана
3	0,65893	0,65878
4	0,74412	0,74388
5	0,74493	0,74523
6	0,80935	0,81259
7	0,80775	0,81028

Таблица показывает, что уже уравнение с четырьмя гармониками хорошо описывает исходный РД, но в качестве модели прогноза возьмём ряд Фурье с шестью гармониками (что характерно для сезонных колебаний), который объясняет 81% вариации уровней. Для мультипликативной модели (2) коэффициент детерминации равен 0,732618.

Оценочные значения для модели тригонометрического тренда (4) имеют вид

$$\hat{y}(k) = \hat{a}_0 + \sum_{i=1}^6 \hat{a}_i \cos \frac{2\pi i}{N} k + \sum_{i=1}^6 \hat{b}_i \sin \frac{2\pi i}{N} k. \quad (8)$$

Полученные оценки коэффициентов ряда Фурье (4) методами МНК и дискретным фильтром Калмана мало отличаются друг от друга, поэтому для дальнейших исследований выбрана модель (8) с параметрами, значения которых были получены с помощью фильтра Калмана.

Графики выровненных динамических рядов по моделям (2) и (8) приведены на рис. 3: точками изображены наблюдаемые величины, сплошной и пунктирной линиями - значения, полученные моделированием.

Исследуем полученные остатки (разность между наблюдаемыми и модельными значениями) моделей (2) и (8). Выборочное среднее $\bar{\epsilon} \approx 0$; оценка среднеквадратического отклонения $\hat{\sigma}_\epsilon = 633,5933$ и $\hat{\sigma}_\epsilon = 428,5408$ соответственно для моделей (2) и (8).

Рассмотрим гипотезу о согласии распределения остатков с нормальным распределением. На рис. 4 изображены гистограммы (оценка плотности распределения) остатков моделей (2) и (8) с наложенной на них плотностью нормального распределения.

Значение статистики $c^2 = 1,76335$ и вероятности $p = 0,62294$ для остатков модели (2), $c^2 = 1,8707$ и вероятности $p = 0,599966$ для остатков модели (8); критическое значение распределения хи-квадрат χ_0^2 с параметрами $1-\alpha/2$ (α - заданный уровень значимости, $\alpha = 0,05$) и числом степеней свободы $f = 3$ равно $\chi_0^2 = 7,81$. Так как $\chi^2 < \chi_0^2$, то нет оснований отвергать гипотезу о нормальном распределении остатков модели на уровне значимости $\alpha = 0,05$. Аналогичный вывод можно сделать по значениям вероятностей. Так как вероятность p неправильного отвержения гипотезы, когда она верна, довольно большая, то гипотеза о нормальности остатков моделей (2) и (8) принимается на уровне значимости $\alpha = 0,05$.

Если предположить, что структура наблюдаемого процесса существенно не изменится в ближайшее будущее, то, используя одну из построенных моделей (либо мультипликативную, либо ряд Фурье), можно построить прогноз количества продаж авиабилетов на следующий год. На основе полученного уравнения тренда даётся точечная оценка прогноза. Однако более надёжный прогноз предполагает оценку его в интервале пред-

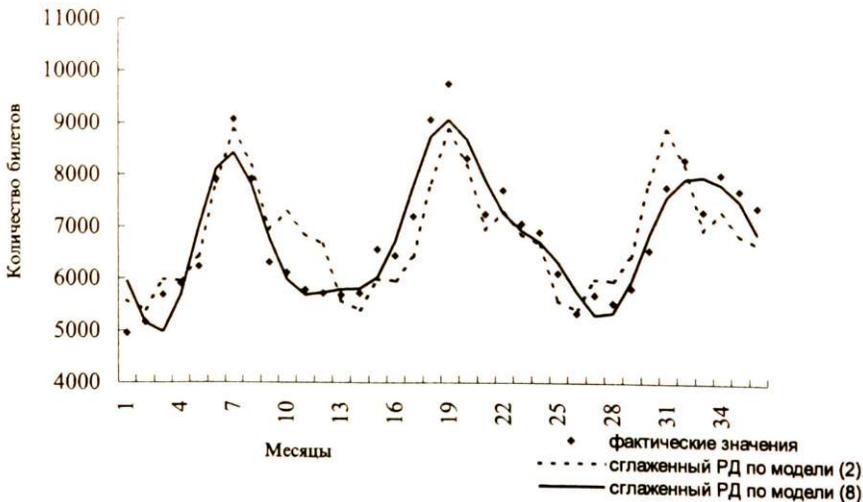


Рис. 3. График исходных и сглаженных значений по формулам (2) и (8)

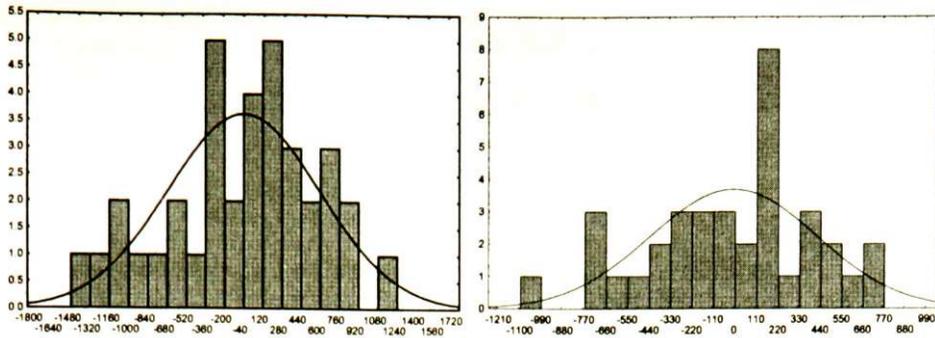


Рис. 4. Гистограммы остатков построенных моделей (2) и (8) соответственно

сказания. Интервал предсказания определим как доверительный интервал для предсказанного значения:

$$y_{1,2} = y_{t+l} \mp t_0 S, \quad (9)$$

где t_0 - табличное значение распределения Стьюдента с параметрами: $1-\alpha/2$ (α - заданный уровень значимости, обычно $\alpha=0,05$) и числом степеней свободы $f=N-k$; S - колеблемость уровней РД относительно полученного тренда, которая определяется формулой

$$S^2 = \sum_i (y_i - \hat{y}_i)^2 / (N - k), \quad (10)$$

где k - число параметров в уравнении тренда; l - глубина предсказания (период упреждения), $y_{1,2}$ границы интервала предсказания.

Построенные доверительные интервалы содержат исходный РД с заданной доверительной вероятностью $\alpha=1-\alpha=0,95$ (см. рис. 5). Результаты прогнозирования и построения интервала предсказания изображены на рис. 5.

Итак, в поставленной задаче для построения трендовой модели использованы мультипликативная модель и ряд Фурье, параметры которого оценивались с применением фильтра Калмана и по МНК. Сравнительный анализ результатов моделирования для конкретного числового примера - объемов продаж авиабилетов - подтверждает применимость предложенных подходов построения трендовых моделей и обеспечивает высокое качество аппроксимации данных и прогнозирования поведения экономической системы.

Использование трендовых моделей хозяйствующими субъектами позволит получить значительный экономический эффект, так как организация может заблаговременно просчитать возможное развитие событий, например при увеличении объемов продаж авиабилетов организовать дополнительные авиарейсы на маршрутах, по которым ожидается увеличение потока пассажиров.

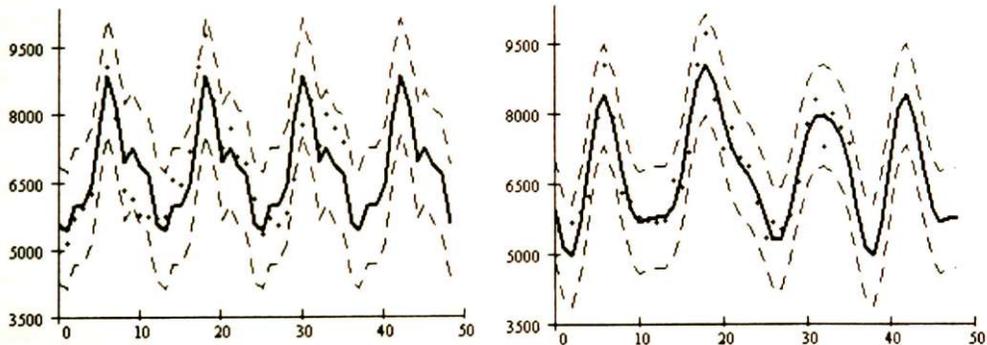


Рис. 5. Графики продаж авиабилетов с построенным прогнозом по моделям (3) и (8) соответственно

ЛИТЕРАТУРА

1. *Статистика*: Учебник / Под ред. И.И. Елисейевой. М.: ВИТРЭМ, 2002.
2. Брамер К., Зиффлинг Г. *Фильтр Калмана-Бьюси*. М.: Наука, 1982.

Статья представлена кафедрой прикладной математики факультета прикладной математики и кибернетики Томского государственного университета, поступила в научную редакцию «Экономические науки» 20 апреля 2004 г.