ISSN 2079-0031 Вестник НТУ "ХПИ", 2014, № 35 (1078)

УДК 658.012

В.А. ГОЛОВКО, асп., НТУ "ХПИ",
ЯМЕН ХАЗИМ, асп., НТУ "ХПИ"

ПРОГНОЗИРОВАНИЕ КОРРЕЛИРОВАННОГО ВРЕМЕННОГО РЯДА ПО МАЛОЙ ВЫБОРКЕ ИСХОДНЫХ ДАННЫХ

Рассмотрен метод оценки прогнозируемого значения случайного процесса, отсчеты которого коррелированны. Выделение детерминированных полиномиальной и гармонической составляющих наблюдаемого процесса с учетом малого объема выборки реальних статистических данных осуществляется методом наименьших квадратов. Расчет случайной составляющей прогнозируемого значения выполнен с использованием значения корреляционной функции, соответствующего интервалу между наблюдениями процесса. Библиогр.: 8 назв.

Ключевые слова: коррелированный временной ряд, малая выборка, прогнозирование.

Постановка проблемы и анализ литературы. Пусть совокупность значений случайного процесса, наблюдаемого через равные промежутки времени ΔT, образует последовательность $\{y_1, y_2, \mathbf{K}, y_n, \mathbf{K}\}$. Ограничимся n наблюдениями, имеющимися к текущему моменту времени. Будем считать, кроме того, что эти наблюдения коррелированны.

Прогнозирование случайных процессов – важнейший элемент сложной задачи описания изменения показателей динамических процессов, протекающих в системах различного назначения (в экономике, технике, медицине, социологии, военном деле и т.д.). К настоящему времени известны более ста различных методов построения прогнозирующих моделей, детальный анализ и классификация которых приведены в [1, 2]. В задачах, когда реальный объем статистических данных ограничен, наиболее применимыми оказываются экстраполяционные методы прогнозирования [3 – 5], к которым относятся метод наименьших квадратов и совокупность методов сглаживания. При этом, если выборка исходных данных мала, применение методов сглаживания, принципиальная особенность которых состоит в учете неравноважности "старых" и "новых" наблюдений, не дает заметных преимуществ перед методом наименьших квадратов [6 – 8]. Рассмотрим возможность применения метода наименьших квадратов для прогнозирования, например, случайного процесса спроса, наблюдения которого коррелированы.

Пусть временной ряд представлен совокупностью n коррелированных отчетов $\{y_1, y_2, \ldots, y_n, \ldots\}$. Поставим задачу краткосрочного прогнозирования ряда. Наблюдаемый процесс, предположительно, представляет собой суперпозицию детерминированных процессов, содержащих полиномиальную (трендовую) и гармоническую составляющие, измерения которого искажены случайным процессом, формируемым за счет ошибок измерений и, возможно, коррелированным за счет шумового влияния неконтролируемых факторов.

Цель статьи – решить задачу выделения и аналитического описания детерминированных и случайных составляющих процесса и его прогнозирование.

Основной результат. Построение искомой математической модели проведем поэтапно в следующей последовательности.

Этап 1. Выделение детерминированного тренда. Для аналитического описания поведения детерминированной составляющей наблюдаемого процесса введем полиномиальную модель

$$y(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_d t^d \; . \tag{1}$$

Пусть наблюдения процесса проводятся в моменты времени $\{\Delta T, 2\Delta T, \ldots, n\Delta T\}$. Надлежащим масштабированием приведем последовательность моментов наблюдения к виду $\{1, 2, \ldots, n\}$.

Параметры модели найдем методом наименьших квадратов [5, 6]. С этой целью используем матрицу H и векторы A и Y:

$$H = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 2^d & 3^d & \cdots & n^d \end{pmatrix}; \quad A = \begin{pmatrix} a_0 \\ \cdots \\ a_d \end{pmatrix}; \quad Y = \begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix}.$$

Тогда наилучшие в смысле наименьших квадратов оценки коэффициентов полинома (1) определяются по формуле

$$\hat{A} = \left(H^T H\right)^{-1} H^T Y \; .$$

Этап 2. Выделение периодической составляющей. После выделения из последовательности наблюдений процесса трендовой составляющей необходимо провести анализ наличия или отсутствия колебаний спроса. Отметим, что в задаче краткосрочного прогнозирования сезонными колебаниями спроса можно пренебречь. В связи с этим для описания суточных колебаний используем простейшую модель

$$\tilde{y}(t) = b\mathrm{Sin}\frac{2\pi}{T_0}t + c\mathrm{Cos}\frac{2\pi}{T_0}t \ , \quad T_0 = 24 \text{ часа.} \tag{2}$$

Выбор этой модели определяется следующими её достоинствами.

Во-первых, она позволяет оценить неизвестные амплитуду и фазу гармонической составляющей, так как $c\mathrm{Cos}wt + b\mathrm{Sin}wt = A\mathrm{Sin}(wt + \varphi)$, $A = \sqrt{c^2 + b^2}$, $\varphi = \mathrm{arctg(c/b)}$, $w = 2\pi / T_0$.

Во-вторых, модель (2) линейна относительно оцениваемых параметров. В модели (2) $\tilde{y}(t) = y(t) - \hat{y}(t)$ – результат, получающийся после удаления из наблюдаемого процесса трендовой полиномиальной составляющей. При этом для моментов наблюдений $\{\Delta T,\ 2\Delta T,\ \mathbf{K},\ n\Delta T\}$ получим соответствующую совокупность $\{\tilde{y}_1, \tilde{y}_2, \mathbf{K}, \tilde{y}_n\}$. Неизвестные параметры b и c вновь найдем методом наименьших квадратов. Введем

$$I_1 = \sum_{j=1}^{n}\left(b\mathrm{Sin}\frac{2\pi}{T_0}j\Delta T + c\mathrm{Cos}\frac{2\pi}{T_0}j\Delta T - \tilde{y}_j\right)^2 . \tag{3}$$

Далее

$$\frac{dI_1}{dc} = c\sum_{j=1}^{n}\mathrm{Cos}^2\frac{2\pi}{T_0}j\Delta T + b\sum_{j=1}^{n}\mathrm{Sin}\frac{2\pi}{T_0}j\Delta T \cdot \mathrm{Cos}\frac{2\pi}{T_0}j\Delta T -$$

$$-\sum_{j=1}^{n}\tilde{y}_j\mathrm{Cos}\frac{2\pi}{T_0}j\Delta T = 0 , \tag{4}$$

$$\frac{dI_1}{dc} = c\sum_{j=1}^{n}\mathrm{Sin}\frac{2\pi}{T_0}j\Delta T \cdot \mathrm{Cos}\frac{2\pi}{T_0}j\Delta T + b\sum_{j=1}^{n}\mathrm{Sin}^2\frac{2\pi}{T_0}j\Delta T -$$

$$-\sum_{j=1}^{n}\tilde{y}_j\mathrm{Sin}\frac{2\pi}{T_0}j\Delta T = 0 . \tag{5}$$

Решение системы уравнений (4), (5) определяет параметры b и c аналитического описания гармонической составляющей процесса.

Этап 3. Обработка случайной составляющей. После выделения из наблюдаемого процесса детерминированных полиномиальной (1) и периодической (2) составляющих получим случайную последовательность $T = \{\tau_1, \tau_2, \mathbf{K}, \tau_n\}$, $\tau_j = y_j - \hat{y}_j - \tilde{y}_j$, $j = 1, 2, \mathbf{K}, n$.

Найдем оценку дисперсии случайной величины, выборка значений которой образует последовательность $\tau_1, \tau_2, \mathbf{K}, \tau_n$. Эта оценка равна $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}\tau_j^2$. Рассчитаем теперь величину коэффициента корреляции между соседними членами в последовательности T, которая определяется по формуле $K_{n-1} = \frac{1}{(n-1)\hat{\sigma}_n^2}\sum_{j=1}^{n-1}\tau_j\tau_{j+1}$. Дополним последовательность T неизвестным прогнозируемым значением спроса τ_{np}. Запишем соотношение для коэффициента корреляции между соседними членами в дополненной последовательности $\tau_1, \tau_2, \mathbf{K}, \tau_n, \tau_{\text{пр}}$:

$$K_n = \frac{1}{n\sigma_{n+1}^2}\sum_{j=1}^{n}\tau_j\tau_{j+1} = \frac{1}{n\sigma_{n+1}^2}\left(\sum_{j=1}^{n-1}\tau_j\tau_{j+1} + \tau_n\tau_{\text{пр}}\right) =$$

$$= \frac{(n-1)\sigma_n^2}{n\sigma_{n+1}^2}K_{n-1} + \frac{1}{n\sigma_{n+1}^2}\tau_n\tau_{\text{пр}}. \qquad (6)$$

Заметим теперь, что даже если число наблюдаемых членов в последовательности T не слишком велико, можно приближенно считать, что $\sigma_n^2 \approx \sigma_{n+1}^2$, $K_n \approx K_{n-1}$.

Тогда соотношение (6) упростится к виду: $K_{n-1} \cong \frac{n-1}{n}K_{n-1} + \frac{1}{n\sigma_n^2}\tau_n\tau_{\text{пр}}$, откуда

$$\tau_{\text{пр}} \cong \frac{K_{n-1}\cdot\sigma_n^2}{\tau_n}. \qquad (7)$$

Теперь с использованием (1), (2), (7) получим оценку прогнозируемого значения процесса:

$$y(t_{\text{пр}}) = \hat{y}(t_{\text{пр}}) + \tilde{y}(t_{\text{пр}}) + \tau_{\text{пр}} =$$

$$= \sum_{i=0}^{d}\hat{a}_i t_{\text{пр}}^i + \hat{b}\,\mathrm{Sin}\frac{2\pi}{T_0}t_{\text{пр}} + \hat{c}\,\mathrm{Cos}\frac{2\pi}{T_0}t_{\text{пр}} + \frac{K_{n-1}\sigma_n^2}{\tau_n}.$$

Выводы. Таким образом, с использованием традиционных методов получены простые соотношения для оценки прогнозируемого значения коррелированного случайного процесса. Прогнозируемое значение является суммой соответствующих значений для детерминированных полиномиальной и гармонической составляющих и значения случайной составляющей, рассчитанной с использованием корреляционной функции процесса.

Список литературы: 1. *Вартанян В.М.* Моделирование динамических процессов по временным рядам / *В.М. Вартанян, Ю.А. Романенков, Д.С. Ревенко, В.Ю. Кащеева*. – Х.: НАУ им. Н. Е. Жуковского "ХАИ", 2012. – 266 с. **2.** *Янч Э.* Прогнозирование научно-технического процесса / *Э. Янч.* – М.: Прогресс, 1980. – 568 с. **3.** *Льюис К.* Методы прогнозирования экономических показателей / *К. Льюис*. – М.: Финансы и статистика, 1986. – 133 с. **4.** *Шелобаев С.И.* Математические методы и модели / *С.И. Шелобаев*. – М. : ЮНИТИ-ДАНА, 2009. – 496 с. **5.** *Линник Ю.В.* Метод наименьших квадратов и основы теории обработки наблюдений / *Ю.В. Линник*. – М.: Физматгиз, 1968. – 352 с. **6.** *Вучков И.* Прикладной линейный регрессионный анализ / *И. Вучков, И. Бояджиева, Е. Солаков*. – М.: Финансы и статистика, 1987. – 239 с. **7.** *Більчук В.М.* Теорія ймовірностей, випадкові процеси та математична статистика / *В.М. Більчук*. – Х.: ХУПС, 2009. – 436 с. **8.** *Красс М.С.* Математические методы и модели для магистрантов в экономике / *М.С. Красс, Б.П. Чупрынов*. – СПб.: ПИТЕР, 2010. – 496 с.

Bibliography (transliterated): **1.** *Vartanjan V.M.* Modelirovanie dinamicheskih processov po vremennym rjadam / *V.M. Vartanjan, Ju.A. Romanenkov, D.S. Revenko, V.Ju. Kashheeva* . – H.: NAU im. N. E. Zhukovskogo "HAI", 2012. – 266 s. **2.** *Janch Je.* Prognozirovanie nauchno-tehnicheskogo processa / *Je. Janch*. – M.: Progress, 1980. – 568 s. **3.** *L'juis K.* Metody prognozirovanija jekonomicheskih pokazatelej / *K. L'juis*. – M.: Finansy i statistika, 1986. – 133 s. **4.** *Shelobaev S.I.* Matematicheskie metody i modeli / *S.I. Shelobaev*. – M.: JuNITI-DANA, 2009. – 496 s. **5.** *Linnik Ju.V.* Metod naimen'shih kvadratov i osnovy teorii obrabotki nabljudenij / *Ju.V. Linnik*. – M.: Fizmatgiz, 1968. – 352 s. **6.** *Vuchkov I.* Prikladnoj linejnyj regressionnyj analiz / *I. Vuchkov, I. Bojadzhieva, E. Solakov*. – M.: Finansy i statistika, 1987. – 239 s. **7.** *Bil'chuk V.M.* Teorija jmovirnostej, vipadkovi procesi ta matematichna statistika / *V.M. Bil'chuk* . – H.: HUPS, 2009. – 436 s. **8.** *Krass M.S.* Matematicheskie metody i modeli dlja magistrantov v jekonomike / *M.S. Krass, B.P. Chuprynov*. – SPb.: PITER, 2010. – 496 s.

Поступила (*received*) *26.02.2014*

Работу представил д-р техн. наук, проф НТУ "ХПИ" Раскин Л.Г.

Golovko Vitaliy, graduate student
National Technical University "Kharkiv Polytechnic Institute"
Str. Frunze, 21, Kharkiv, Ukraine, 61002
tel./phone: (057) 707-66-28, e-mail: Sova_vika92@mail.ru

Yamen Hazim, graduate student
National Technical University "Kharkiv Polytechnic Institute"
Str. Frunze, 21, Kharkiv, Ukraine, 61002
tel./phone: (057) 707-66-28, e-mail: Sova_vika92@mail.ru