

# PREDICTION OF SELF-SIMILAR TIME SERIES CONTAINING POLYNOMIAL TRENDS

Omelchenko A.V., Rozdimakha E.A., Fedorov A.V.  
Kharkov National University of Radio Electronics  
14, Lenin Ave., Kharkov, 61166, Ukraine  
Ph.: (057) 7021113, e-mail: rozdy@mail.ru

**Abstract** — The analysis of prediction methods for time series that contain deterministic and stochastic trends is done. The method to extract a deterministic trend at presence of correlated noise is considered. Estimations for prediction accuracy depending on a prediction step for different algorithms are obtained by modeling.

## ПРОГНОЗИРОВАНИЕ САМОПОДОБНЫХ ВРЕМЕННЫХ РЯДОВ СОДЕРЖАЩИХ ПОЛИНОМИАЛЬНЫЕ ТРЕНДЫ

Омельченко А. В., Роздымаха Е. А., Федоров А. В.  
Харьковский национальный университет радиоэлектроники  
пр. Ленина, 14, Харьков, 61166, Украина  
тел.: (057) 7021113, e-mail: rozdy@mail.ru

**Аннотация** — Выполнен анализ методов прогнозирования временных рядов, содержащих детерминированный и стохастический тренды. Рассмотрен метод выделения детерминированного тренда на фоне коррелированной помехи. Методом моделирования получены оценки качества прогноза в зависимости от шага прогнозирования для различных алгоритмов.

### I. Введение

Задачи прогнозирования временных рядов возникают в самых различных областях человеческой деятельности. От точности прогнозов зависит эффективность принимаемых решений.

Развитие инфокоммуникационных сетей сопровождается их планированием. При этом требуется прогнозировать трафик для принятия управленческих решений по эффективному использованию ресурсов сетей и обслуживания абонентов.

Актуальность задач прогнозирования в современных условиях очень высока, поскольку от них зависит принятие того или иного решения в различных сферах человеческой деятельности.

Целью работы является анализ применимости различных подходов к прогнозированию временных рядов, содержащих как детерминированный тренд в виде полинома с неизвестными коэффициентами, так и стохастическую компоненту, описываемую моделью ARFIMA.

### II. Основная часть

Предположим, что наблюдается случайная последовательность (временной ряд)

$$z_t = \vec{x}_t \vec{a} + \xi_t, \quad (1)$$

где  $\vec{a}$  — неизвестный вектор коэффициентов полиномиального тренда;  $\vec{x}_t = (1, t, \dots, t^m)$  — вектор-строка полиномиальных регрессоров, где  $m$  — порядок модели;  $\xi_t$  — последовательность случайных величин в виде самоподобного процесса с известными параметрами.

Необходимо найти оптимальный в среднеквадратическом смысле прогноз последовательности  $\hat{z}_{t+h}$  для отсчета  $z_{t+h}$ ,  $h \geq 0$ , т.е. такое  $\hat{z}_{t+h}$ , при котором обеспечивается минимум показателя

$$M[(\hat{z}_{t+h} - z_{t+h})^2]. \quad (2)$$

В задаче прогнозирования последовательности вида (1) необходимо выделить детерминированный тренд на фоне коррелированной помехи.

Известно [1], что наилучшую линейную оценку параметров регрессии обеспечивает обобщенный метод наименьших квадратов (ОМНК)

$$\hat{\vec{a}} = (X^T R^{-1} X)^{-1} X^T R^{-1} \vec{z}, \quad (3)$$

где  $X$  — регрессионная матрица, определяемая выражением

$$X = \begin{bmatrix} x_{0,0} & x_{0,1} & \dots & x_{0,m-1} \\ x_{1,0} & x_{1,1} & \dots & x_{1,m-1} \\ \dots & \dots & \dots & \dots \\ x_{t-1,0} & x_{t-1,1} & \dots & x_{t-1,m-1} \end{bmatrix}, \quad (4)$$

$R = M[\vec{\xi} \vec{\xi}^T]$  — положительно определенная матрица размера  $t \times t$ .

Известно также [2], что минимум критерия (2) при известном тренде обеспечивает решение

$$\hat{z}_{t+h} = M[z_{t+h} | z_0, \dots, z_{t-1}], \quad (5)$$

где  $M[z_{t+h} | z_0, \dots, z_{t-1}]$  — условное математическое ожидание.

Для модели (1) с неизвестным вектором коэффициентов  $\vec{a}$  используем адаптивный алгоритм

$$\hat{z}_{t+h} = \vec{r}_c^T R^{-1} (\vec{z} - X \hat{\vec{a}}) + \vec{x}_{t+h} \hat{\vec{a}}, \quad (6)$$

где  $\vec{r}_c = M[z_{t+h} \vec{z}]$  — вектор взаимных корреляций величины  $z_{t+h}$  с отсчетами вектора  $\vec{z}$ , а  $\hat{\vec{a}}$  находится согласно (3). В данной работе показано, что прогноз (6) является несмещенным и асимптотически оптимальным (при  $t \rightarrow \infty$ ).

В случае самоподобной в широком смысле последовательностей вид матрицы  $R$  определяется коэффициентом корреляции последовательности [3]

$$r(k) = \frac{1}{2} [(k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}], \quad (7)$$

где параметр  $\beta \in (0,1)$  и связан с параметром Хёрста  $H$  соотношением  $\beta = 2 - 2H$ .

Для описания последовательностей, обладающих свойствами самоподобия, широко используется модель авторегрессионного дробно-интегрируемого процесса скользящего среднего (ARFIMA). Последовательность  $\xi_t$  удовлетворяет модели ARFIMA( $p, d, q$ ) с порядком интегрируемости  $|d| < \frac{1}{2}$  если она удовлетворяет уравнению [4]

$$A(L)(1-L)^d \xi_t = \Psi(L)\varepsilon_t, \quad (9)$$

где  $L$  — оператор сдвига:  $L^k \xi_t = \xi_{t-k}$ ;  $A(L)$  — многочлен степени  $p$  от  $L$ ,  $\Psi(L)$  — многочлен степени  $q$  от  $L$ ,  $\varepsilon_t$  — последовательность независимых одинаково распределенных случайных величин с нулевым средним и дисперсией  $\sigma^2$ . Величина  $d$  связана с параметром Хёрста выражением  $d = H - 0,5$ .

Для любого вещественного  $d$  и неотрицательно-го целого  $k$  справедливо [4]

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k, \quad (10)$$

где  $\Gamma(\cdot)$  — гамма-функция.

Последовательность  $\xi_t$  может быть представлена в виде

$$\Phi(L)X_t = \sum_{k=0}^{\infty} \varphi_k X_{t-k} = \varepsilon_t, \quad (11)$$

где  $\Phi(L)$  — обобщенный многочлен авторегрессионного представления модели ARFIMA. Такой вид ARFIMA-модели удобно использовать для генерирования и прогнозирования стохастических трендов. Для простейшей модели ARFIMA(0,  $d$ , 0) коэффициенты  $\varphi_k$  определяются из выражения (10) и равны

$$\varphi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}. \quad (12)$$

На практике для прогнозирования стохастического тренда временных рядов также часто используются авторегрессионные модели невысокого порядка. В таком случае коэффициенты  $\varphi_k, k = \overline{1, p}$  могут быть оценены по методу Юла-Уокера [5].

При использовании авторегрессионного представления последовательностей (11) асимптотически оптимальный прогноз (6) может быть представлен в виде

$$\hat{z}_{t+h} = \sum_{k=1}^p \varphi_k (z_{t+h-k} - \bar{x}_{t+h-k} \hat{a}) + \bar{x}_{t+h} \hat{a}, \quad (13)$$

где  $p$  — порядок модели, а вместо неизвестных значений последовательности используются их спрогнозированные значения.

Обычно качество прогноза оценивается показателем, обратным отношению дисперсии стохастической компоненты временного ряда к дисперсии ошибки прогноза,

$$SN^{-1} = \frac{M[(\hat{z}_{t+h} - z_{t+h})^2]}{M[\varepsilon_t^2]}. \quad (14)$$

Для исследования эффективности разработанных алгоритмов прогнозирования было проведено статистическое моделирование. При моделировании в качестве стохастического тренда  $\xi_t$  использовался

процесс ARFIMA, сгенерированный согласно выражениям (11, 12) с параметрами  $H = 0,9$ ;  $p = q = 0$ ,  $d = 0,4$ . В качестве порождающего процесса  $\varepsilon_t$  был использован белый гауссовский шум с нулевым средним и единичной дисперсией. В качестве модели тренда — линейный полином.

В программной среде Wolfram Mathematica были реализованы и исследованы алгоритмы прогнозирования на основе авторегрессионной модели (13) первого, пятидесятого порядка, разработанного алгоритма линейного прогноза (6).

На рис. 1 представлены зависимости показателя (14) от величины шага прогнозирования  $h$ . Как видно из рисунка при прогнозировании на один шаг все алгоритмы дают близкое качество прогноза, а при прогнозировании на большое число шагов наилучшим по показателю (14) является прогноз (6).

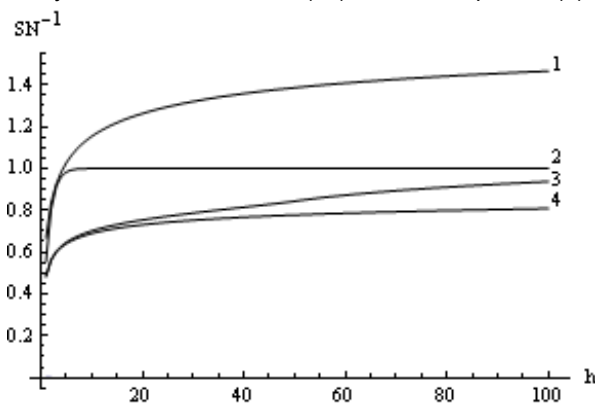


Рис. 1. Зависимость показателя  $SN^{-1}$  для алгоритмов прогнозирования: 1 — использование последнего отсчета; 2 — авторегрессии первого порядка; 3 — авторегрессии порядка 50; 4 — асимптотически оптимальный прогноз (6).

Fig. 1. Dependence of the factor  $SN^{-1}$  for different predictors: 1 for the last sample using; 2 for autoregression of the first order; 3 for autoregression of the order 50; 4 for asymptotically optimal predictor (6).

### III. Заключение

Для прогнозирования временных рядов со свойствами самоподобия и детерминированным трендом применим алгоритм (6), обладающий свойствами несмещенности и асимптотической оптимальности.

Методом статистического моделирования установлено, что для прогнозирования процессов ARFIMA на один шаг приемлем прогноз на основе процесса авторегрессии небольшого порядка. При прогнозировании на большое число шагов следует использовать асимптотически оптимальный линейный прогноз (6).

### IV. References

- [1] Seber G., Lee A. *Linear Regression Analysis, Second Edition*. Wiley, 2003. 582 p.
- [2] Cyplakov A. *Vvedenie v prognozirovaniye v klassicheskikh modelyakh vremennykh ryadov* [Introduction into predicting at classical models of time series]. *Kvantil'*, 2006, vol. 1, pp. 3-19.
- [3] Cybakov B.S. Model' teletrafika na osnove samopodobnogo sluchajnogo processa [Model of teletraffic based on self-similar random process]. *Radiotekhnika*, 1999, vol. 5, pp. 24-31.
- [4] Konjuhovskij P.V., Podkorytova O.A. "Dlinnaya pamjat" v obmennyykh kursah ["Long memory at the exchanging of course]. *Vestnik SPbGU*, 2007, vol. 5, No 3, pp. 102-109.
- [5] Box G., Jenkins G. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970, p. 406.