СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ И ФИЛЬТРА КАЛМАНА

С. В. Лобач, М. И. Малафеева

Белорусский Государственный университет, пр. Независимости, 4,
220030 г. Минск, Беларусь, lobachS@bsu.by

Исследуется применение моделей в пространстве состояний и фильтра Калмана для предсказания и восстановления пропущенных значений временных рядов. Для решения этих задач используется представление временных рядов в форме моделей в пространстве состояний.

Ключевые слова**:** модели в пространстве состояний, прогнозирование временных рядов, фильтр Калмана, моделирование, временные ряды с пропусками.

STATISTICAL ANALYSIS OF TIME SERIES BASED ON STATE SPACE MODELS AND KALMAN FILTER

S. V. Lobach, M. I. Malafeeva

Belarusian State University, Independence Ave., 4,
220030 Minsk, Belarus, lobachS@bsu.by

The application of state space models is explored and the Kalman filter for forecasting and recovering missing values in time series. To solve these problems, the representation of time series in the form of state space models is used.

Keywords: state space models, time series forecasting, Kalman filter, modeling, time series with missing data.

Введение

Многие математические модели временных рядов могут быть сведены к моделям в пространстве состояний. В свою очередь модели в пространстве состояний позволяют применить к исходной модели временного ряда широкий спектр процедур, включая алгоритм, определяющий фильтр Калмана [1].

Фильтр Калмана для статистического анализа ARMA-модели временного ряда

Рассмотрим ARMA(4, 1) скалярного временного ряда:

$$y_t = y_{t-4} + \varepsilon_t + \varepsilon_{t-1},\ t = 0,1,\ldots,T,$$

$$\varepsilon_t = N(0,1),\ y_0 = 1.$$

Данную модель временного ряда можно представить в форме модели в пространстве состояний:

$$\begin{cases} y_t = \begin{pmatrix} 1\,0\,0\,0 \end{pmatrix} x_t\ , \\ x_{t+1} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} x_t + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \varepsilon_t \end{cases},\quad t = 0,1,\ldots,T, \tag{1}$$

где $\varepsilon_t \sim N(0,1)$, $y_0 = 1$, $\{\varepsilon_t\}$ – последовательность независимых одинаково распределенных случайных величин с нулевым математическим ожиданием и единичной дисперсией.

Требуется найти оценки $\hat{y}_t$, используя фильтр Калмана.

Для нахождения оценок будем пользоваться следующими формулами фильтра Калмана [2]:

$$x_{t|t-1} = Fx_{t-1},$$

$$P_{t|t-1} = FP_{t-1}F' + RQR',$$

$$x_t = x_{t|t-1} + K_t v_t,$$

$$P_t = \left[I - K_t Z \right] P_{t|t-1},$$

$$v_t = y_t - Zx_{t|t-1},$$

$$K_t = P_{t|t-1} Z' \left(ZP_{t|t-1} Z' + H \right)^{-1},$$

где

$$\begin{cases} y_t = Z_t x_t \\ x_{t+1} = F_t x_t + R_t \varepsilon_t \end{cases},\quad t = 0,1,\ldots,T,\qquad T = 100, \tag{2}$$

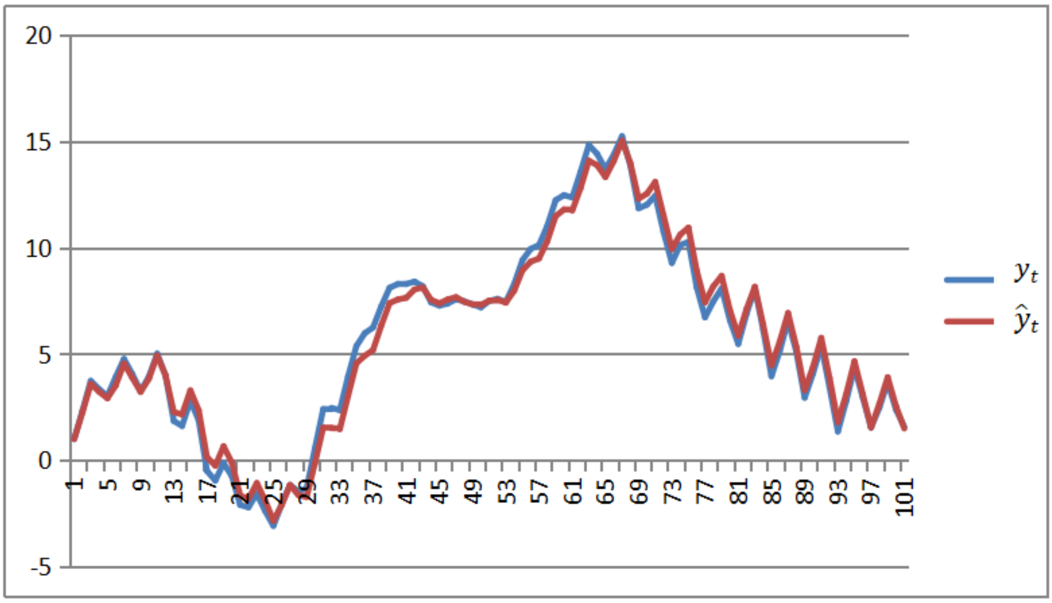
$$E\left(\varepsilon_{t-1}\varepsilon_{\tau-1}\right) = \begin{cases} Q, t = \tau, \\ 0, t \neq \tau, \end{cases}$$

$$E\left(\varepsilon_t \varepsilon_\tau\right) = \begin{cases} H, t = \tau, \\ 0, t \neq \tau. \end{cases}$$

В нашем случае: $Z = (1 0\ 0 0)$, $F = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, $R = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$.

Зададим: $H = 1$, $Q = 1$, $x_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $P_0 = \begin{pmatrix} 16 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix}$.

На рисунке представлены графики исходного временного ряда y_t и полученного в результате фильтрации ряда оценок $\hat{y}_t$, определенные уравнениями (1), (2).



Графики исходного временного ряда y_t и ряда оценок $\hat{y}_t$

Прогнозирование стационарной случайной последовательности на основе моделей в пространстве состояний

Рассмотрим стационарный в широком смысле процесс со спектральной плотностью:

$$\hat{f}(\lambda) = \frac{\left| e^{i\lambda} + 1 \right|^2}{\left| e^{2i\lambda} + \frac{1}{2} e^{i\lambda} + \frac{1}{2} \right|^2}. \tag{3}$$

Гауссовский случайный процесс ξ_t со спектральной плотностью (3) подчиняется следующему стохастическому разностному уравнению [3]:

$$\xi_{t+2}+\frac{1}{2}\left(\xi_{t+1}+\xi_t\right)=\varepsilon\left(t+2\right)+\varepsilon\left(t+1\right), \qquad (4)$$

где $\varepsilon_t,\ t=0,1,\ldots,T,$ последовательность независимых нормально распределенных случайных величин с нулевыми математическими ожиданиями и единичной дисперсией.

Введём вспомогательную случайную величину θ_t, тогда получим систему (4) эквивалентную (3):

$$\begin{cases}\theta_{t+1}=-\frac{1}{2}\theta_t-\frac{1}{2}\xi_t+\frac{1}{2}\varepsilon\left(t+1\right),\\ \xi_{t+1}=\theta_t+\varepsilon\left(t+1\right).\end{cases}$$

Для решения задачи прогнозирования строим дополнительные случайные последовательности $\eta_1\left(t,s\right)$ и $\eta_2\left(t,s\right)$ ($\eta_1\left(t,s\right)$ – для θ_t, $\eta_2\left(t,s\right)$ – для ξ_t) по формулам:

$$\begin{cases}\eta_1\left(t+1,s\right)=-\frac{1}{2}\eta_1\left(t,s\right)-\frac{1}{2}\eta_2\left(t,s\right),\\ \eta_2\left(t+1,s\right)=\eta_1\left(t,s\right),\\ \eta_1\left(s,s\right)=m_s,\eta_2\left(s,s\right)=\xi_s.\end{cases}$$

Для того, чтобы начать строить прогноз с момента времени s, нужно знать начальное значение $\eta_1\left(s,s\right)$. Для этого построим ещё одну последовательность m_s:

$$\begin{cases}m_{s+1}=-\frac{1}{2}m_s-\frac{1}{2}\xi_s+\frac{1-\gamma_s}{2\left(1+\gamma_s\right)}\left(\xi_{s+1}-m_s\right),\\ \gamma_{s+1=}\frac{\gamma_s}{1+\gamma_s},\\ m_0=0,\gamma_0=1.\end{cases}$$

Указанные выше формулы можно также применить для восстановления пропущенных значений временного ряда.

Результаты представлены в табл. 1, 2, 3.

Таблица 1

Ретроспективный прогноз с момента s = 90

t	θ_t	η_{1t}	ξ_t	η_{2t}
91	0.467	0.425	0.233	0.156
92	-0.247	-0.290	0.673	0.425
93	-0.331	-0.067	-0.483	-0.290

Таблица 2

Восстановление пропущенных значений (s = 70)

t	θ_t	η_{1t}	ξ_t	η_{2t}
71	0.243	-0.284	0.969	-0.088
72	-0.331	0.186	0.793	-0.284
73	-0.003	0.049	0.125	0.186

Таблица 3

Восстановление пропущенных значений (s = 50)

t	θ_t	η_{1t}	ξ_t	η_{2t}
51	-0.509	1.053	-3.069	0.123
52	0.382	-0.588	-3.322	1.053
53	0.205	-0.232	-2.148	-0.588

Исходя из полученных результатов, можно сделать вывод о том, что подход на основе применения моделей в пространстве состояний и фильтра Калмана является достаточно эффективным для решения задач прогнозирования временных рядов с пропусками.

Библиографические ссылки

1. *Chui Charles K., Guanrong Chen* Kalman Filtering with Real-Time Applications. Springer, 2009.
2. *Durbin J., Harvey A., Koopman S. J., Shephard N.* State Space and Unobserved Component Models: Theory and Applications. Cambridge University Press, 2004.
3. *Липцер Р. Ш., Ширяев А.Н.* Статистика случайных процессов. М : Наука, 1974.