ОЦЕНИВАНИЕ ПАРАМЕТРОВ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПРОПУСКАМИ НА ОСНОВЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

Р. И. Меркулов
Факультет прикладной математики и информатики, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: fpm.merkulovRI@bsu.by

В настоящее время временные ряды получили широкое распространение в биометрике, эконометрике, технике и других приложениях. Удобной математической моделью временных рядов являются модели в пространстве состояний, позволяющие с единых позиций описывать свойства таких моделей как $AR(p)$, $ARMA(p,q)$, $ARIMA(p,d,q)$, $VAR(p)$, $ARCH(q)$ и многих других. В данной работе рассматривается проблема оценивания параметров и прогнозирования значений временных рядов с пропусками, порожденных моделью $VAR(p)$, на основе сведения этой модели к форме модели в пространстве состояний.

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 такие актуальные, как оценивание параметров и прогнозирование [1].

Один из мощных инструментов, применимых к моделям в пространстве состояний, – фильтр Калмана – эффективный рекурсивный фильтр, который позволяет оценивать вектор состояния динамической системы, используя ряд неполных и зашумленных данных.

В [3] отмечается три подхода к оцениванию параметров векторной авторегрессии при наличии пропусков:

- оценивание по методу максимального правдоподобия;
- применение ЕМ-алгоритма;
- модификация известных оценок параметров для случая пропусков.

В данной работе используется иной подход, основанный на сведении моделей параметрических временных рядов к форме моделей в пространстве состояний и построении оценок параметров модели временного ряда на основе робастной модификации фильтра Калмана по неполным данным с заданным либо случайным шаблоном пропусков.

I. Модели в пространстве состояний

Рассмотрим модель, определенную уравнениями:

$$\beta_{t+1} = F_t\beta_t + \epsilon_t, \qquad (1)$$

$$z_t = H_t^T\beta_t + \eta_t, t > 1, \qquad (2)$$

где β_t – случайный вектор размерности k, называемый вектором состояния системы в момент времени t. Предполагается, что β_1 имеет нормальное распределение $N(m,P)$, а также, что k-мерные векторы ϵ_t независимы и одинаково распределены по закону $N(0,Q)$. Матрица F_t размерности $k\times k$, называемая матрицей переходов, может быть как частично или полностью неизвестной, так и известной. Уравнение (1) называется уравнением состояния.

Вектор z_t, именуемый вектором измерений, имеет размерность n. Векторы η_t взаимно независимы с векторами ϵ_t и одинаково распределены по закону $N(0,R)$. Уравнение (2) называется уравнением измерения.

Матрица измерений H_t размерности $n\times k$ является известной.

Важное различие между z_t и β_t заключается в том, что z_t наблюдаются, в то время как β_t, вообще говоря, частично наблюдаются либо не наблюдаются вовсе.

Уравнения (1) – (2) определяют модель в пространстве состояний [2]. Стоит отметить, что параметрические модели временных рядов могут быть сведены к форме моделей в пространстве состояний различными способами, то есть неоднозначно, что обусловлено произволом в задании вектора состояния, матриц переходов и измерений.

II. Модель $VAR(p)$

Рассмотрим модель $VAR(p)$ и одно из возможных ее представлений в форме модели в пространстве состояний.

Математическая модель $VAR(p)$:

$$z_t = A_1z_{t-1} + \cdots + A_pz_{t-p} + \epsilon_t, t = 0,1,\ldots,T, \qquad (3)$$

где $\epsilon_t \sim N(0,\Sigma)$, A_i – неизвестные матрицы параметров. Задача состоит в оценивании матриц A_i по налюдениям $\{z_0, z_1, \ldots, z_T\}$.

Модель (3) сначала представляется в виде модели $VAR(1)$, которая затем сводится к модели в пространстве состояний следующего вида:

$$\beta_{t+1} = \beta_t = \theta, \qquad (4)$$

$$z_t = H_t\beta_t + \epsilon_t, \qquad (5)$$

где $\theta = (a_{1,1}, a_{2,1}, \ldots, a_{n-1,n}, a_{n,n})^T$, $H_t = \begin{vmatrix} z_{t-1}^T & 0_n \cdots & 0_n \\ \vdots & \ddots & \vdots \\ 0_n & \cdots & z_{t-1}^T \end{vmatrix}$.

Для статистического анализа модели (4) – (5) применяется модифицированный алгоритм фильтра Калмана для оценивания $\hat{\beta}_t = \hat{\theta}$ по наблюдениям $\{z_0, z_1, \ldots, z_T\}$ с пропусками [4].

III. Компьютерные эксперименты

Для генерации пропусков в данных существуют два вида шаблонов: детерминированный и случайный. Ниже приведен пример детерминированного шаблона (см. рис. 1).

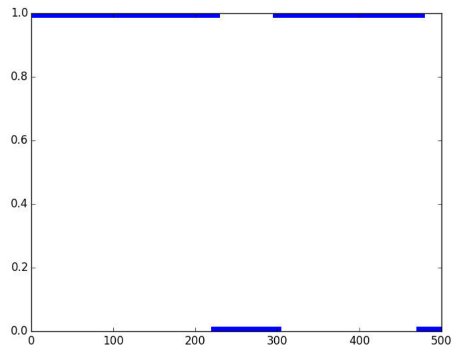

Рис. 1 – Пример детерминированного шаблона пропусков

Рассмотрим пример применения описанного подхода на реализации модельных данных длины 100, порожденных $VAR(1)$ моделью для двухкомпонетного временного ряда со случайными и детерминированными пропусками.

Матрица параметров, которую необходимо оценить по реализации временного ряда, имеет следующий вид:

$$A = \begin{vmatrix} 0.5 & 0.3 \\ 0.8 & -0.7 \end{vmatrix}$$

В таблице 1 приведены оценки матрицы параметров в зависимости от типа шаблона и пропусков и доли отсутствуюших наблюдений:

Таблица 1 – Оценки матрицы параметров

Пропуски	Случайный		Детерминированный	
20%	0.5246	0.2695	0.4516	0.2549
	0.8680	−0.6474	0.7624	−0.7071
50%	0.5927	0.2392	0.4260	0.1344
	0.8834	−0.6376	0.8912	−0.7450
70%	0.5328	0.2137	0.3114	0.2487
	0.8038	−0.6549	0.8141	−0.6573
90%	0.3378	0.1436	0.3688	0.3297
	0.6524	−0.5957	0.7421	−0.6617

Из таблицы 1 видно, что при одинаковой доле пропусков при случайном шаблоне оценки получаются более близкими к истинным, чем при детерминированном.

Далее рассматривается та же реализацию временного ряда, по которой построен прогноз для значений временного ряда. Отметим, что решение этой задачи аналогично присутствию детерминированных пропусков в конце временного ряда. Ниже приведен пример прогнозирования значений для одной из компонент векторного временного ряда (см. рис. 2).

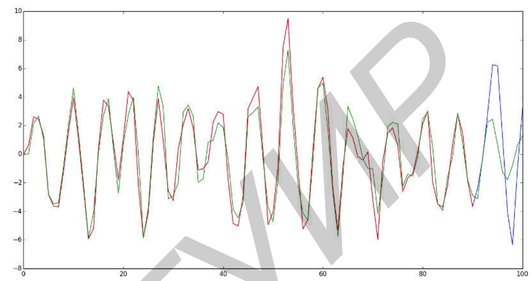

Рис. 2 – Прогнозирование значений временного ряда. Зеленый – оценки значений временного ряда, синий – отсутствующие наблюдения, красный – наблюдаемые значения

IV. Результаты

Приведены способы сведения моделей $VAR(p)$ к моделям в пространстве состояний, для которых описан подход построения оценок параметров модели и значений временного ряда при наличии пропусков, основанный на робастной модификации фильтра Калмана.

Проведена серия экспериментов для численного анализа точности оценивания параметров модели временных рядов как на модельных, так и на реальных данных. В качестве реальных данных был взят классический временной ряд популяции канадской рыси [5]. Результаты моделирования показывают достаточно высокую эффективность рассматриваемого подхода.

1. Канторович, Г. Г. Анализ временных рядов / Г. Г. Канторович // Экономический журнал ВШЭ. – 2002. – № 1. – С. 85–116.
2. Липцер, Р. Ш. Статистика случайных процессов (нелинейная фильтрация и смежные вопросы) / Р. Ш. Липцер, А. Н. Ширяев // Издательство: Главная редакция физико-математической литературы изд-ва «Наука», 1974. – С .696.
3. Литтл, Р. Дж. Статистический анализ данных с пропусками / Р. Дж. Литтл, Д. Б. Рубин // Издательство: «Финансы и статистика», 1991. – С .284.
4. Harvey, A. C. Forecasting, Structural Time Series Models and the Kalman Filter / A. C. Harvey // Cambridge: Cambridge University Press, 1989. – P .227 .
5. Encyclopedia of Mathematics [Electronic resource] / The European Mathematical Society, 2002. – Mode of access: http://www.encyclopediaofmath.org. – Date of access: 25.05.2015.