ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПРОПУСКАМИ НА ОСНОВЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

Меркулов Р. И.
Кафедра математического моделирования и анализа данных, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: merkylovecom@mail.ru

Прогнозирование временных рядов является одной из основных задач в современном анализе данных. На практике в данных часто содержатся искажения: пропуски и выбросы, что обусловливает необходимость разработки робастных методов построения оценок параметров моделей временных рядов либо модификации известных подходов. В данной статье описывается подход, позволяющий обобщить несколько классов параметрических моделей временных рядов, а также применить рекурсивную процедуру оценивания параметров и построения прогнозирующих статистик, устойчивую к выбросам и пропускам в наблюдениях.

Введение

В настоящее время временные ряды получили широкое распространение в эконометрике, теории управления и иных приложениях. Удобной математической моделью временных рядов являются модели в пространстве состояний, позволяющие с единых позиций описывать свойства таких параметрических моделей временных рядов как $AR(p)$, $ARIMA(p,d,q)$, $ARCH(q)$, $VAR(p)$ и многих других. В данной работе рассматривается проблема оценивания параметров и прогнозирования значений временных рядов с пропусками, порождаемых моделью $VAR(p)$, на основе сведения этой модели к форме модели в пространстве состояний. Одним из мощных инструментов, применимых к моделям в пространстве состояний, является фильтр Калмана: эффективный рекурсивный фильтр, который позволяет оценивать вектор состояния динамической системы, используя неполные и зашумленные данные [1].

В [2] описан 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 со случайным либо детерминированным шаблоном пропусков.

I. Модели в пространстве состояний

Рассмотрим модель, определённую уравнениями:

$$\beta_{t+1} = F_t \beta_t + \epsilon_t,$$
$$z_t = H_t^T \beta_t + \eta_t,$$

где β_t – случайный вектор размерности k, называемый вектором состояний в момент времени t. Предполагается, β_1 что имеют нормальное распределение $N(m,P)$, а также, что k-мерные вектора ϵ_t независимые и одинаково распределённые по закону $N(0,Q)$, $\eta_t \sim N(0,R)$. Матрица F_t размерности $k \times k$, называемая матрицей переходов, является неслучайной. Первое из приведённых уравнений называется уравнением состояния. Оно полностью определяет распределение β_t для любых $t \geq 1$.

Вектор z_t, именуемый вектором измерений, имеет размерность n. Второе из приведённых уравнений называется уравнением измерения. Оно позволяет определить распределение z_t. Оба эти уравнения определяют распределение Гауссовского процесса $(\beta_t, z_t), t \geq 1$. Матрица измерений H_t размерности $n \times k$ является неслучайной.

Важное различие между β_t и z_t заключается в том, что z_t наблюдаются, в то время как β_t, вообще говоря, частично наблюдаются либо не наблюдаются вовсе. В любом случае информацией, доступной на момент времени t, является $z_1, \cdots z_t$ [2].

II. Робастная модификация фильтра Калмана

Представление в форме модели в пространстве состояний позволяет применить рекурсивную процедуру фильтрации Калмана для оценивания параметров модели временного ряда и построения прогнозирующих статистик.

Робастная модификация фильтра Калмана на случай пропусков в наблюдениях состоит в том, что в момент времени $t+1$ отклонение прогнозного значения по t предыдущим наблюдениям полагается равным нулю, тогда вектор состояния переходит на следующую итерацию без обновления значений:

$$z_{\widetilde{t+1}} = 0,$$

$$\beta_{\widehat{t+1}|t+1} = \beta_{\widehat{t+1}|t}.$$

Эта модификация была предложена и обоснована математиками Гарвеем (Harvey) и Копманом (Coopman) [3].

III. Многомерный случай

Описанный подход может быть обобщен на многомерный случай. Ниже приведена процедура сведения $VAR(1)$ модели к форме пространства состояний.

Модель $VAR(1)$ описывается следующим уравнением:

$$z_t = Az_{t-1} + \epsilon_t,$$

где $z_t \in R^n$, $\epsilon_t \sim N_n(0, \Sigma)$.

Представим матрицу A в виде расширенного вектора параметров модели:

$$\beta = (a_{1,1}, a_{2,1}, \cdots, a_{n-1,n}, a_{n,n})^T,$$

тогда, положив матрицы состояний и переходов:

$$F_t = I_{n^2 \times n^2},$$

$$H_t = \begin{bmatrix} z_{t-1}^T & 0_n & 0_n \\ \vdots & \ddots & \vdots \\ 0_n & 0_n & z_{t-1}^T \end{bmatrix}.$$

IV. Компьютерное моделирование

Смоделируем двумерный временной ряд, порождённый моделью $VAR(1)$. Затем будем оценивать матрицу А по реализации временного ряда с различной долей пропусков как по детерминированному, так и по стохастическому шаблону. В качестве матрицы А была взята следующая матрица:

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

В таблице ниже приведены оценки параметров с различными шаблонами и долей пропусков:

Таблица 1 – Результаты оценивания параметров

Доля пропусков	Случайный	Детерминированный
20%	$\begin{bmatrix} 0.525 & 0.270 \\ 0.868 & -0.647 \end{bmatrix}$	$\begin{bmatrix} 0.452 & 0.255 \\ 0.762 & -0.707 \end{bmatrix}$
50%	$\begin{bmatrix} 0.593 & 0.239 \\ 0.883 & -0.638 \end{bmatrix}$	$\begin{bmatrix} 0.426 & 0.134 \\ 0.891 & -0.745 \end{bmatrix}$

Также решим задачу прогнозирования временных рядов с помощью процедуры фильтрации Калмана. Для валидации модели определим последние 10% наблюдений тестовыми значениями, а модель обучим на 90% значений временного ряда.

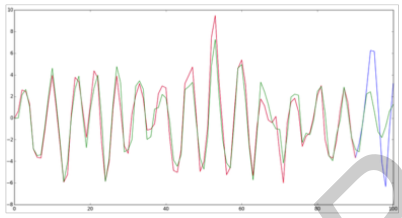

Рис. 1 – Прогнозирование временного ряда

V. Результаты

Приведены способы сведения различных параметрических моделей скалярных и многомерных временных рядов 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 как на модельных, так и на реальных данных для случайного и детерминированного шаблона пропусков с различной долей пропусков (от 10% до 90%).

1. Канторович, Г.Г. Анализ временных рядов / Экономический журнал ВШЭ. –2002. –№1. – С .85–116.
2. Меркулов, Р.И. Оценивание параметров и прогнозирование временных рядов с пропусками на основе моделей в пространстве состояний / Р.И. Меркулов // Информационные технологии и системы 2016 (ИТС-2016): материалы междунар. науч. конф. – Минск, 2016. – С. 302–304.
3. Harvey, A.C., Forecasting, Structural Time Series Models and the Kalman Filter, – Cambridge University Press, Cambridge, 1989. – 227 p.