ИДЕНТИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ С ПРОПУСКАМИ НА ОСНОВЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

Р. И. Меркулов, В. И. Лобач

Белорусский государственный университет
Минск, Беларусь
e-mail: fpm.merkulov@bsu.by, lobach@bsu.by

Рассматриваются модели в пространстве состояний и применение к ним процедуры фильтрации Калмана для оценивания параметров моделей и построения прогнозирующих статистик. На основе сведения основных типов моделей параметрических временных рядов к моделям в пространстве состояний приводятся алгоритмы построения оценок параметров и прогнозирующих статистик с применением модифицированной процедуры фильтрации Калмана, позволяющей работать с неполными данными.

Ключевые слова: модель VAR; модель ARMA; модели в пространстве состояний; фильтр Калмана; данные с пропусками.

IDENTIFICATION OF TIME SERIES WITH MISSING OBSERVATIONS BASED ON STATE SPACE MODELS

R. I. Merkulov, V. I. Lobach

Belarusian State University
Minsk, Belarus

This article considers state space models and the use of Kalman filtration procedures for parameters estimation and construction of predictive statistics. Based on reduction of main types of parametric time series models to state space models algorithms are considered for constructing parameters estimates and predictive statistics using modified Kalman filtration procedure which allows to handle missing data.

Keywords: VAR model; ARMA model; state space model; Kalman filter; missing data.

ВВЕДЕНИЕ

Временные ряды имеют множество применений в теории управления, эконометрике, биометрике, технике и других приложениях [1].

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 столь актуальные, как оценивание параметров и прогнозирование.

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.

Модели в пространстве состояний и фильтр Калмана [2] в совокупности представляют важный инструмент статистического анализа временных рядов.

Пропуски в наблюдениях являются типичными искажениями модельных предположений в анализе данных.

Задачи оценивания параметров и прогнозирования значений временных рядов по неполным данным часто возникают в технических, эконометрических и других приложениях.

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

1) оценивание по методу максимального правдоподобия;
2) применение ЕМ-алгоритма;
3) 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 шаблоном пропусков.

МОДЕЛИ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

Рассмотрим модель, определенную уравнениями:

$$\beta_{t+1} = F_t \beta_t + e_t, \tag{1}$$

$$z_t = H_t' \beta_t + \eta_t,\ t \geq 1, \tag{2}$$

где β_t – случайный вектор размерности k, называемый вектором состояний в момент времени t. Предполагается, что β_1 имеет нормальное распределение $N(m, P)$, а также, что k-мерные вектора e_t независимы и одинаково распределены по закону $N(0, Q)$, $\eta_t \sim N(0, R)$. Матрица F_t размерности $n \times k$, называемая матрицей переходов, является известной. (1) называется уравнением состояния. Оно полностью определяет распределение β_t для любых $t \geq 1$.

Вектор z_t, именуемый вектором измерений, имеет размерность n. Вектора η_t взаимно независимы с векторами e_t и одинаково распределены по закону $N(0, R)$. (2) называется уравнением измерения. Оно позволяет определить распределение z_t. Оба эти уравнения определяют распределение гауссовского процесса (β_t, z_t), $t \geq 1$. Модель (1) – (2) называется моделью в пространстве состояний.

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

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. В любом случае информацией, доступной на момент времени t, являются наблюдения $z_1, \ldots, z_t$.

Проблемами, с которыми сталкиваются, используя модели в пространстве состояний, являются проблемы, связанные с численным вычислением условных математических ожиданий. Могут быть выделены три вида этой проблемы. Проблема фильтрации заключается в вычислении $E\{\beta_t|z_1, \ldots, z_t\}$, которое является оптимальной оценкой β_t по информации, доступной на момент времени t. Проблема сглаживания заключается в вычислении $E\{\beta_s|z_1, \ldots, z_t\}$, в случае когда $s < t$. Проблемы прогнозирования заключаются в численном вычислении $E\{\beta_s|z_1, \ldots, z_t\}$, $E\{z_s|z_1, \ldots, z_t\}$, когда $s > t$.

Математические модели временных рядов могут быть сведены к моделям в пространстве состояний различными способами, т. е. неоднозначно, что обусловлено возможностью различным образом задавать вектор состояний, а также матрицы переходов и измерений.

ФИЛЬТР КАЛМАНА. РОБАСТНАЯ МОДИФИКАЦИЯ ФИЛЬТРА КАЛМАНА

Ковариационный фильтр Калмана – это алгоритм, используемый для рекуррентного вычисления «отфильтрованных» векторов состояния. Ниже приведены формулы, задающие фильтр в классическом случае. Этот фильтр используется для фильтрации и прогнозирования зашумленных данных.

Введем следующие обозначения:

$$\widehat{\beta_{(t|t)}} = E\{\beta_t|z_1, \ldots, z_t\},$$
$$\widehat{\beta_{(t|t-1)}} = E\{\beta_t|z_1, \ldots, z_{t-1}\},$$
$$\widehat{z_{(t|t-1)}} = E\{z_t|z_1, \ldots, z_{t-1}\},$$
$$\Sigma_{(t|t)} = E\{(\beta_t - \widehat{\beta_{(t|t)}})(\beta_t - \widehat{\beta_{(t|t)}})'\},$$
$$\Sigma_{(t|t-1)} = E\{(\beta_t - \widehat{\beta_{(t|t-1)}})(\beta_t - \widehat{\beta_{(t|t-1)}})'\},$$
$$M_{(t|t-1)} = E\{(z_t - \widehat{z_{(t|t-1)}})(z_t - \widehat{z_{(t|t-1)}})'\},$$
$$\tilde{z}_t = z_t - \widehat{z_{(t|t-1)}} = z_t - H_t'\widehat{\beta_{(t|t-1)}},$$

Таким образом, $\widehat{\beta_{(t|t-1)}}$и $\widehat{z_{(t|t-1)}}$ прогнозные значения β_t, z_t, составленные в момент времени $t-1$. Матрицы $\Sigma_{(t|t)}, \Sigma_{(t|t-1)}, M_{(t|t-1)}$ – среднеквадратичные ошибки прогнозирования, а $\tilde{z}_t$ – остатки МНК в регрессии z_t на предыдущие значения.

Фильтр Калмана определяется следующими соотношениями [2]:

$$\widehat{\beta_{(t|t)}} = \widehat{\beta_{(t|t-1)}} + K_t\tilde{z}_t, \tag{3}$$

где коэффициенты фильтра вычисляется следующим образом:

$$K_t = \Sigma_{t|t-1}H_t\left(H_t'\Sigma_{t|t-1}H_t + R\right)^{-1}, \tag{4}$$
$$\Sigma_{t|t} = (I - K_tH_t')\Sigma_{t|t-1}, \tag{5}$$
$$\widehat{\beta_{(t+1|t)}} = F_t\widehat{\beta_{(t|t)}}, \tag{6}$$
$$\Sigma_{t+1|t} = F_t\Sigma_{t|t}F_t' + Q \quad , \tag{7}$$
$$\widehat{z_{(t+1|t)}} = H_{t+1}'\widehat{\beta_{(t+1|t)}}, \tag{8}$$
$$M_{t+1|t} = H_{t+1}'\Sigma_{t+1|t}H_{t+1} + R. \tag{9}$$

Рассмотрим не классическую реализацию фильтра Калмана для зашумленных данных, а его робастную модификацию, которую можно было бы применить к временному ряду с пропущенными значениями в определенные периоды времени.

Для того, чтобы модифицировать фильтр Калмана, применим следующую процедуру. Если в момент времени t наблюдение отсутствует, то вместо истинного значения наблюдения z_t будем использовать его оценку по предыдущим $t-1$ наблюдениям, таким образом, остаток регрессии $\tilde{z}_t$ на предыдущие значения становится тождественно равным нулю. Это влечет за собой то, что в формуле (3) пропадает второе слагаемое, т. е. для вектора состояния этап коррекции, по сути, пропускается, так как нет значения, по которому было бы возможно провести коррекцию.

Таким образом, если в момент времени t наблюдаемое значение отсутствует, то формулы (3) и (4) обращаются в следующие формулы:

$$\tilde{z}_t = 0,$$
$$\widehat{\beta_{(t|t)}} = \widehat{\beta_{(t|t-1)}}.$$

Эта модификация была предложена и обоснована Гарвеем (Harvey) и Копманом (Coopman) [4].

На основе вышеописанных формул может быть запущен процесс построения оценок параметров модели, заложенных в вектор состояния динамической системы, а также может быть построен прогноз. Оценки, полученные таким способом, являются оптимальными, то есть несмещенными, состоятельными и эффективными [4].

Однако, как видно из формул (3)–(9), для запуска рекуррентного вычислительного процесса необходимо задавать начальные приближения для вектора состояния $\hat{\beta}_{(1|0)}$, а также ковариационной матрицы $\Sigma_{(1|0)}$.

СВЕДЕНИЕ VAR(1) МОДЕЛИ К МОДЕЛИ В ПРОСТРАНСТВЕ СОСТОЯНИЙ. ПРИМЕНЕНИЕ ПРОЦЕДУРЫ РОБАСТНОЙ ФИЛЬТРАЦИИ КАЛМАНА

Модель VAR(1) имеет следующий вид:

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

или более подробно:

$$\begin{pmatrix} z_{t,1} \\ \vdots \\ z_{t,n} \end{pmatrix} = \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{1,n} \\ \vdots & \ddots & \vdots \\ \beta_{n,1} & \cdots & \beta_{n,n} \end{pmatrix} \begin{pmatrix} z_{t-1,1} \\ \vdots \\ z_{t-1,n} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \vdots \\ \epsilon_{t,n} \end{pmatrix},$$

где $z_t = (z_{t,1}, \dots, z_{t,n})^T \in \mathrm{R}^n, \epsilon_t = (\epsilon_{t,1}, \dots, \epsilon_{t,n})^T, \epsilon_t \sim N_n(0, Q)$.

Представим матрицу A в виде расширенного вектора параметров модели:

$$\tilde{\theta} = \begin{pmatrix} \beta_{1,1} \\ \beta_{2,1} \\ \vdots \\ \beta_{n-1,n} \\ \beta_{n,n} \end{pmatrix}, F_t = I_{n^2 \times n^2}, H_t = \begin{pmatrix} y'_{t-1} & 0_n & 0_n \\ \vdots & \ddots & \vdots \\ 0_n & 0_n & y'_{t-1} \end{pmatrix}.$$

Тогда исходная модель представляется в форме модели в пространстве состояний.

Проведена серия компьютерных экспериментов для численного анализа точности оценивания параметров VAR(p) модели временных рядов как на модельных, так и на реальных данных. В качестве реальных данных были взяты квартальные макроэкономические данные по экономике США в период с 1959 по 2009 г. Результаты моделирования показывают достаточно высокую эффективность рассматриваемого подхода.

Таким образом, VAR(1) представлена в форме модели в пространстве состояний, к которой возможно применение процедуры фильтрации Калмана. В данном случае вектор ненаблюдаемых переменных представляет собой расширенный вектор, составленный из столбцов матрицы A. Благодаря этому мы можем применить процедуру фильтрации Калмана (в том числе и робастную модификацию) для оценивания вектора ненаблюдаемых переменных, а следовательно, и элементов матрицы A. Учитывая то, что любая модель VAR(p) может быть сведена к VAR(1) модели, то эта процедура может быть применена для оценивания параметров любой векторной авторегрессионной модели.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Канторович Г. Г. Анализ временных рядов // Эконом. журн. ВШЭ. 2002. № 1. С. 85–116.
2. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов (нелинейная фильтрация и смежные вопросы. М. : Наука, 1974.
3. Литтл Р. Дж., Рубин Д. Б. А Статистический анализ данных с пропусками М. : Финансы и статистика, 1991.
4. Harvey A. C. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge : Cambridge University Press, 1989.