ОЦЕНИВАНИЕ ПАРАМЕТРОВ ВРЕМЕННЫХ РЯДОВ С «ПРОПУСКАМИ» В ПРОСТРАНСТВЕ СОСТОЯНИЙ

С. В. Лобач

Белорусский государственный университет
Минск, Беларусь
E-mail: lobachS@bsu.by

Исследуется проблема оценивания параметров временных рядов, представленных в форме моделей в пространстве состояний при наличии пропусков. Предлагается использовать модифицированный ЕМ-алгоритм.

Ключевые слова: модели в пространстве состояний, фильтр Калмана, ЕМ-алгоритм.

ВВЕДЕНИЕ

Временные ряды широко используются для анализа и прогнозирования экономических показателей, однако на практике модельные предположения относительно свойств этих рядов далеко не всегда выполняются. В частности, возможны ситуации, когда в какие-то моменты времени значения временного ряда по каким-то причинам не определены или сильно искажены. Многие исследователи пытаются избавиться от пропусков с тем, чтобы потом проводить обработку данных классическими статистическими методами. Однако эти методы в общем случае имеют малую эффективность [1].

В данной работе рассматривается подход к анализу временных рядов, основанный на представлении их в форме моделей в пространстве состояний и использовании робастного фильтра Калмана.

ПРЕДСТАВЛЕНИЕ ВРЕМЕННЫХ РЯДОВ В ФОРМЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

Практически все известные традиционные модели временных рядов могут быть представлены в форме моделей в пространстве состояний. Рассмотрим, например, *ARMA*(*p*, *q*) модель временного ряда:

$$x_t = \alpha_1 x_{t-1} + \ldots + \alpha_p x_{t-p} + u_t + b_1 u_{t-1} + \ldots + b_q u_{t-q}, \tag{1}$$

записанную в операторной форме $\phi(B)x_t = \theta(B)u_t$, где $u_t \sim N(0, \sigma^2)$. Не исключено, что $\phi(B)$ имеет корни внутри или на единичном круге, и, следовательно, модель может быть нестационарного типа (*ARIMA*). Обобщенная модель в пространстве состояний имеет вид:

$$\begin{cases} y_t = Z_t \alpha_t + G_t \varepsilon_t, \\ \alpha_{t+1} = T_t \alpha_t + H_t \varepsilon_t, \; t = 1, \ldots, n, \end{cases} \tag{2}$$

где $\varepsilon_t \sim N(0, \sigma^2 I)$, $a_1 \sim N(\alpha_1, \sigma^2 P_1)$, ε_t и a_t взаимно некоррелированы. Матрицы системы Z_t, G_t, T_t, H_t неслучайны и обычно зависят от некоторых глобальных параметров. В одномерном случае с $s \times 1$ вектором состояний и $m \times 1$ вектором ошибок матрицы Z_t, G_t, T_t, H_t имеют размерности $1 \times s$, $s \times s$, $1 \times m$, $s \times m$ соответственно.

Временной ряд (1) может быть представлен в форме (2), если $Z_t = Z = (1, 0, \ldots, 0)$, $G_t = G = 1$,

$$T_t = T = \begin{pmatrix} \phi_1 & 1 & 0 & \ldots & 0 \\ \phi_2 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_{m-1} & 0 & 0 & \ldots & 1 \\ \phi_m & 0 & 0 & \ldots & 0 \end{pmatrix}, \quad H_t = H = \begin{pmatrix} \theta_1 + \phi_1 \\ \theta_2 + \phi_2 \\ \vdots \\ \vdots \\ \theta_m + \phi_m \end{pmatrix}. \tag{3}$$

Параметр $m = \max(p, q)$. Заметим, что представление (3) является неоднозначным, возможны другие формы.

МОДЕЛИ ПРОПУСКОВ

Рассмотрим основные модели пропусков, используемые в теоретических исследованиях и на практике.

Только некоторые определенные компоненты вектора y_t с номерами $i_1(t)$, $i_2(t)$, …, $i_{m_t}(t)$ $(1 \le i_1(t) \le \ldots \le i_{m_t}(t))$ доступны к наблюдению в момент времени t. Тогда наблюдаемым в момент времени t можно считать вектор $y_t^* = S_t \cdot y_t$, где S_t – $(m_t \times m)$-матрица, в которой в позициях с номерами $(1,\ i_1(t))$, $(2,\ i_2(t))$, …, $(m_t, i_{m_t}(t))$ находятся единицы, а остальные элементы матрицы – нулевые.

Только линейная комбинация $y_{1t} + \ldots + y_{mt}$ всех компонент вектора y_t известна в момент времени t. Действительно наблюдаемым становится скаляр $y_t^* = S_t \cdot y_t$, где S_t – m-строка, составленная из единиц.

Эти два случая можно учесть в исследуемой модели, представленной как модель в пространстве состояний

$$x_t = F_t \cdot x_{t-1} + w_t, \quad y_t^* = H_t^* \cdot x_t + v_t,$$

где $y_t^* = S_t \cdot y_t, H_t^* = S_t \cdot H_t, \upsilon_t^* = S_t \cdot \upsilon_t, \ R_t^* = E(\upsilon_t^* \cdot \upsilon_t^{*T}), \quad S_t$ – $(m_t \times m)$-матрица.

По сравнению с исходной моделью в полученной линейной модели размерность вектора y_t^* меняется с течением времени.

Не имеется никакой информации в момент t, то есть вектор наблюдений y_t полностью отсутствует в момент времени t.

Проблема калмановской фильтрации временных рядов при наличии пропусков может быть решена при использовании следующей теоремы относительно фильтра Калмана с наблюдениями, у которых переменная размерность. Эта теорема допускает и тот случай, когда размерность вектора x_t также является переменной.

Теорема. Рассмотрим базовую (2) модель в пространстве состояний в предположении, что размерность n_t вектора x_t и размерность m_t вектора y_t изменяются со временем. Если вектор наблюдений y_t пропущен в момент времени t, тогда рекуррентные формулы, определяющие фильтр Калмана [2], заменяются на

$$\hat{x}_t^t = \hat{x}_t^{t-1}, \quad P_t^t = P_t^{t-1}.$$

Доказательство полностью повторяет вывод уравнений, определяющих фильтр Калмана с очевидными модификациями, касающимися размерностей векторов и матриц.

РОБАСТНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ AR-МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ С «ПРОПУСКАМИ»

Пусть задан авторегрессионный временной ряд $\overline{y}_t = (y_1, \dots, y_t)$, часть наблюдений $y^*_{mis} = (y_{k_1}, \dots, y_{k_m})$ которого пропущена. Требуется оценить параметры временного ряда

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t,$$

где порядок p известен, $\overline{a} = (a_1, \dots, a_p)$ – вектор параметров,

ε_k – стационарный случайный процесс; $p << t$.

Итерационная процедура робастного построения авторегрессионных моделей временных рядов с пропусками выглядит следующим образом:

Шаг 1:

а) Определение оценок $\overline{a}^{(1)}$ коэффициентов авторегрессии в виде:

$$a^{(1)} = \arg\min_{a \in R^p} \sum_{k \in J_{obs}} \rho(| y_k - \sum_{i=1}^{p} a_i y_{k-i} |),$$

где J_{obs} – множество индексов, для которых целевая функция определена, ρ – некоторая монотонно возрастающая, дважды непрерывно дифференцируемая на положительной полуоси вогнутая вверх функция, $\rho(0) = 0$, $\rho''(x) < 0$.

б) заполнение пропусков их оценками

$$y^{*(1)}_{mis} = \arg\min_{y^*_{mis}} \sum_{k \in J_{mis}} \rho(| y_k - \sum_{i=1}^{p} a^{(1)}_i y_{k-i} |),$$

где J_{mis} – множество индексов, для которых целевая функция не определена.

Шаг *j* (j > 1)**:**

а) Определение оценок $\overline{a}^{(j)}$ коэффициентов авторегрессии в виде

$$a^{(j)} = \arg\min_{a \in R^p} \sum_{k \in J_{obs}} \rho(| y_k - \sum_{i=1}^{p} a_i y_{k-i} |),$$

где вместо пропусков наблюдений используются оценки $y^{*(j-1)}_{mis}$.

б) Заполнение пропусков их оценками

$$y^{*(j)}_{mis} = \arg\min_{y^*_{mis}} \sum_{k \in J_{mis}} \rho(| y_k - \sum_{i=1}^{p} a^{(j)}_i y_{k-i} |).$$

Итерационный процесс является сходящимся. В качестве критерия остановки можно использовать выполнение одного или одновременно двух требований по точности:

$$\| a^{(j)} - a^{(j-1)} \| < \delta_a, \quad \| y^{*(j)}_{mis} - y^{*(j-1)}_{mis} \| < \delta_y,$$

где δ_a, δ_y – заданные погрешности.

Исследования методом Монте-Карло показывают, что данный итерационный метод сохраняет устойчивость и при наличии в наблюдениях аддитивных выбросов.

ЛИТЕРАТУРА

1. *Литтл, Р. Дж. А.* Статистический анализ данных с пропусками / Р. Дж. А. Литтл, Д. Б. Рубин. М. : Финансы и статистика, 1991.

2. *Брайсон, А.* Прикладная теория оптимального управления / А. Брайсон, Хо Ю-Ши. М. : Мир, 1972.