СТАТИСТИЧЕСКИЙ АНАЛИЗ ПАНЕЛЬНЫХ ДАННЫХ НА ОСНОВЕ МОДЕЛЕЙ В ПРОСТРАНСТВЕ СОСТОЯНИЙ

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

Предлагается подход к статистическому оцениванию параметров моделей панельных данных регрессионного типа, основанный на применении моделей в пространстве состояний и фильтра Калмана. Проведены компьютерные эксперименты, подтверждающие эффективность данного подхода.

Введение

Панельные данные представляют собой информацию, собранную для разных объектов в последовательные моменты времени. Применение таких наблюдений позволяет специфицировать и оценивать более сложные модели, чем применение одной пространственной выборки или одного временного ряда. Математически панельные данные описываются случайным полем. Примером панельных данных могут служить показатели хозяйственной деятельности совокупности предприятий, которые собираются каждый год. В этом случае получается массив данных, в котором содержатся и данные об однородных объектах за один и тот же период времени, и последовательные значения одной экономической переменной в различные периоды времени. Но если совокупность предприятий из года в год будет различной, то такие данные уже не будут являться панельными.

I. Панельные данные регрессионного типа

Панельные данные регрессионного типа в отличие от пространственных данных и временных рядов описываются переменными, содержащими два индекса:

$$y_{it} = \alpha + x_{it}\beta + u_{it}, i = 1, \dots, n, t = 1, \dots, T, \quad (1)$$

где i обозначает номер объекта, t – означает момент времени, α – неизвестный скаляр (общее среднее значение), β – $k \times 1$ вектор коэффициентов, x_{it} является i-ым наблюдением k объясняющих переменных. Во многих приложениях панельных данных предполагается, что ошибки наблюдений представляются в виде:

$$u_{it} = \mu_i + \nu_{it},$$

где μ_i обозначает индивидуальный специфичный эффект, ν_{it} – случайные величины. Предположим, что μ_i – последовательность независимых случайных величин с нулевым математическим ожиданием и дисперсией σ_μ^2, тогда ν_{it} описывается $AR(1)$ временным рядом:

$$\nu_{it} = \rho\nu_{it-1} + \epsilon_{it},$$

где $|\rho| < 1$, ϵ_{it} – последовательность независимых случайных величин с нулевым математическим ожиданием и дисперсией σ_ϵ^2, $\nu_{i0} \sim N(0, \frac{\sigma_\epsilon^2}{1-\rho^2})$. Проблема состоит в оценивании неизвестных параметров $\alpha, \beta, \sigma_\mu, \sigma_\epsilon, \rho$. Если ввести матрицу $N \times N$

$$C = \begin{bmatrix} (1-\rho^2)^{\frac{1}{2}} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 & 0 \\ & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{bmatrix},$$

то с помощью преобразования

$$w^* = (I_N \otimes C)u,$$

где $\otimes$ – кронекерово произведение матриц, получим модель (1) с независимыми ошибками наблюдений, и оценку параметров можно провести методом наименьших квадратов, однако этот процесс требует громоздких вычислений [1].

II. Панельные данные в форме моделей в пространстве состояний

Рассмотрим простейшую модель панельных данных:

$$y_i = (1, \dots, 1)^T\beta + \epsilon_i, i = 1, \dots, K, \quad (2)$$

$$y_i \in R^K, \epsilon_i \sim N_K(0, \Sigma_i).$$

Модель панельных данных (2) можно представить в виде следующей модели в пространстве состояний для каждого отдельного объекта (при фиксированном i):

$$\begin{cases} y_t = \beta + \epsilon_t, \\ \epsilon_t = \phi\epsilon_{t-1} + w_t, \\ w_t \sim N(0, \sigma^2), \epsilon_0 \sim N(0, \frac{\sigma^2}{1-\phi^2}). \end{cases} \quad (3)$$

Традиционный подход к проблеме оценивания параметра β сводится к проблеме нелинейной оптимизации. Чтобы избежать этой трудности, можно представить модель (3) в виде:

$$\begin{bmatrix}\epsilon_t\\ \beta_t\end{bmatrix} = \begin{bmatrix}\phi & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}\epsilon_{t-1}\\ \beta_{t-1}\end{bmatrix} + \begin{bmatrix}w_t\\ 0\end{bmatrix},$$

$$\begin{bmatrix}w_t\\ 0\end{bmatrix} \sim N(0,W), W = \begin{bmatrix}\sigma^2 & 0\\ 0 & 0\end{bmatrix},$$

$$\beta_t = \beta, (\epsilon_0, \beta_0)^T \sim N(0,G), G = \begin{bmatrix}\frac{\sigma^2}{1-\phi^2} & 0\\ 0 & 0\end{bmatrix},$$

$$y_t = \epsilon_t, t = 0, \ldots, T.$$

Обозначив

$$\begin{bmatrix}\epsilon_t\\ \beta_t\end{bmatrix} = x_t,$$

получим представление в форме моделей в пространстве состояний:

$$\begin{cases} x_t = \begin{bmatrix}\phi & 1\\ 0 & 1\end{bmatrix} x_{t-1} + \begin{bmatrix}w_t\\ 0\end{bmatrix} \\ y_t = [1,0]x_t, \end{cases} \qquad (4)$$

где y_t – наблюдения, x_t – ненаблюдаемые компоненты.

III. Компьютерные эксперименты

Для оценивания параметра β будем использовать процедуру фильтрации Калмана. Применительно к модели в пространстве состояний (4) рекурсивные уравнения фильтра выглядят следующим образом:

$$\hat{\beta_{t|t}} = \hat{\beta_{t|t-1}} + K_t\tilde{y_t},$$

$$\tilde{y_t} = y_t - \hat{y_{t|t-1}} = y_t - H_t^T\hat{\beta_{t|t-1}},$$

$$K_t = \Sigma_{t|t-1}H_t(H_t^T\Sigma_{t|t-1}H_t + W)^{-1},$$

$$\Sigma_{t|t} = (I - K_tH_t^T)\Sigma_{t|t-1},$$

$$\hat{\beta_{t+1|t}} = F_t\hat{\beta_{t|t}},$$

$$\Sigma_{t+1|t} = F_t\Sigma_{t|t}F_t^T,$$

$$\hat{y_{t+1|t}} = H_{t+1|t}^T\hat{\beta_{t+1|t}},$$

где

$$H_t = H = (1,0), F_t = F = \begin{bmatrix}\phi & 1\\ 0 & 1\end{bmatrix}.$$

В нашем случае матрица переходов F_t известна с точностью до параметра, то есть частично неизвестна. Для оценивания параметра ϕ необходимо либо переформулировать модель в пространстве состояний таким образом, чтобы этот параметр стал частью вектора состояния системы x_t, который можно оценить с помощью фильтра Калмана, либо оценивать эту матрицу с использованием функции правдоподобия или EM-алгоритма. В нашем случае задача состоит именно в оценивании параметра β, поэтому для простоты положим, что в данных компьютерных экспериментах матрица переходов процесса известна, то есть параметр ϕ задан явно [2].

Ниже приведены результаты экспериментов при различных значениях параметров ϕ – параметр матрицы переходов F_t, β – свободный член в исходной модели, T – число наблюдений. Эксперименты проведены на модельных данных, порождённых моделью (4) с заданным параметром ϕ.

Таблица 1 – Результаты компьютерных экспериментов

Эксперимент №	ϕ	T	β	$\hat{\beta}$
1	0.1	500	-0.65	-0.599
2	0.5	200	0.3	0.380
3	0.4	200	0.0	-0.050
4	-0.5	100	-0.2	-0.213

Из проведенных экспериментов видно, что оценки параметра β близки к истинным значениям параметра, породившим реализацию случайного процесса в заданных экспериментах.

IV. Список литературы

1. Harvey, A. C. Forecasting, Structural Time Series Models and the Kalman Filter / A. C. Harvey // Cambridge University Press, Cambridge – 1989. – 227 p.
2. Меркулов, Р.И. Оценивание параметров и прогнозирование экономических временных рядов с пропусками на основе моделей в ранстве состояний / Р. И. Меркулов // Проблемы прогнозирования и государственного регулирования социально-экономического развития: материалы XVII Междунар. науч. конф. – Минск, 2016 г. Т.3. – С. 253–254.