Адаптивный нестационарный регрессионный анализ

Красоткина О. В., Моттль В. В., Марков М. Р., Мучник И. Б.
krasotkina@uic.tula.ru
Тула, ТулГУ; Москва, ВЦ РАН;
США, Markov Processes International; США, университет Rutgers

Задача восстановления числовой регрессионной зависимости $y_t : T \to \mathbb{R}$ в некотором множестве наблюдений $t \in T$ является одной из ключевых в интеллектуальном анализе данных. В качестве наиболее проблематичного аспекта этой задачи обычно рассматривается выбор наиболее подходящего подмножества $\hat{I} \subseteq I$ в множестве доступных числовых признаков объектов $\{x_t^{(i)}, i \in I\}$, относительно которых искомая зависимость ищется, чаще всего, в классе линейных моделей.

Для многих приложений типичны ситуации, когда совокупность наблюдений $t \in T$ естественно рассматривать как упорядоченную последовательность $T = \{1, \ldots, N\}$, допуская, что коэффициенты регрессии могут принимать разные значения в разные моменты времени:

$$y_t = \sum\nolimits_{i \in \hat{I}} \beta_t^{(i)} x_t^{(i)} + e_t, \quad t = 1, \ldots, N, \tag{1}$$

где e_t — последовательность независимых ошибок наблюдения с нулевым средним. Принципиальная специфика модели нестационарной регрессии заключается в том, что число переменных $(\beta_t^{(i)},\ i \in \hat{I},\ t = 1, \ldots, N)$, подлежащих оцениванию, всегда во много раз превышает число наблюдений N. В результате оказывается, что скрытые коэффициенты регрессии невозможно оценить без принятия некоторых априорных предположений об их последовательности, т. е. без дополнительной регуляризации задачи.

Задача оценивания модели нестационарной регрессии (1) интенсивно изучалась в мировой литературе. Популярным инструментом ее решения является метод FLS — Flexible Least Squares [1]:

$$\begin{aligned} J\big(\beta_t^{(i)},\ t = 1 \ldots, N,\ i \in \hat{I} \mid \hat{I}, \rho\big) = \\ = \sum_{t=1}^{N} \Big(y_t - \sum_{i \in \hat{I}} \beta_t^{(i)} x_t^{(i)}\Big)^2 + \rho \sum_{t=2}^{N} \sum_{i \in \hat{I}} \Big(\beta_t^{(i)} - \beta_{t-1}^{(i)}\Big)^2 \to \min. \end{aligned} \tag{2}$$

Подмножество регрессоров $\hat{I} \subseteq I$ и коэффициент $\rho > 0$ являются параметрами критерия. Здесь первое слагаемое отвечает за аппроксимацию наблюдений, а второе слагаемое регулирует изменчивость искомых коэффициентов регрессии во времени. Чем больше ρ, тем более плавной будет последовательность оценок, что уменьшает фактическую «размерность» задачи, делая ее промежуточной между $|\hat{I}|$ и $N|\hat{I}|$. При $\rho \to \infty$ критерий сводится к обычному методу наименьших квадратов $\hat{\beta}_1^{(i)} = \ldots = \hat{\beta}_N^{(i)}$.

Если рассматривать априорную модель последовательности коэффициентов регрессии как совокупность независимых скрытых случайных процессов $\beta_t^{(i)} = \beta_{t-1}^{(i)} + \xi_t^{(i)}$, каждый из которых порождается нормальным белым шумом $\xi_t^{(i)}$, дисперсия которого в ρ раз меньше дисперсии шума e_t в модели наблюдения (1), то критерий FLS (2) максимизирует апостериорную плотность распределения вероятности на множестве реализаций скрытого случайного процесса.

Для выбранного множества регрессоров $\hat{I}$ и фиксированного значения коэффициента сглаживания ρ минимизация квадратичной функции (2) сводится к решению, вообще говоря, очень большой системы линейных уравнений относительно $N|\hat{I}|$ переменных, имеющей однако блочно-трехдиагональную матрицу с блоками $|\hat{I}|\times|\hat{I}|$. Эта особенность допускает применение метода прогонки, обеспечивающего линейную вычислительную сложность решения системы относительно длины временного ряда N. Методу прогонки эквиваленты квадратичное динамическое программирование [2, 3] и фильтр-интерполятор Калмана-Бьюси [4].

Как правило, эффективное подмножество регрессоров $\hat{I} \subset I$ выбрать априори невозможно, и уже хотя бы поэтому задачу нестационарного регрессионного анализа следует рассматривать как задачу интеллектуального анализа данных. Более того, нестационарность модели порождает проблему, совершенно новую для интеллектуального анализа данных, а именно, неизбежную необходимость выбирать степень изменчивости во времени коэффициентов регрессии. Часто по физической природе исследуемого явления нестационарными являются коэффициенты лишь при некоторых регрессорах, что порождает также проблему разделения множества регрессоров на стационарные и нестационарные.

В данной работе мы рассматриваем модификацию критерия FLS (2), позволяющую, во-первых, автоматически выбирать подмножество эффективных регрессоров, во-вторых, автоматически отбирать, далее, еще меньшее подмножество нестационарных регрессоров, и, в-третьих, определять для них индивидуальные коэффициенты сглаживания последовательности коэффициентов регрессии. В отличие от (2), адаптивный критерий применяется к полному множеству доступных регрессоров I:

$$J(\beta_t^{(i)},\ t=1\ldots,N,\ \delta^{(i)},\ \lambda^{(i)},\ i\in I \mid \rho)=$$
$$=\sum_{t=1}^{N}\Big(y_t-\sum_{i\in I}\beta_t^{(i)}x_t^{(i)}\Big)^2+\rho\sum_{t=2}^{N}\sum_{i\in I}\tfrac{\delta^{(i)}+\lambda^{(i)}}{\delta^{(i)}\lambda^{(i)}}\Big(\beta_t^{(i)}-\tfrac{\delta^{(i)}}{\delta^{(i)}+\lambda^{(i)}}\beta_{t-1}^{(i)}\Big)^2\to\min,$$
$$\prod_{i\in I}\tfrac{\delta^{(i)}\lambda^{(i)}}{\delta^{(i)}+\lambda^{(i)}}=1,\quad \text{т.е.}\quad \sum_{i\in I}\log\tfrac{\delta^{(i)}\lambda^{(i)}}{\delta^{(i)}+\lambda^{(i)}}=0. \tag{3}$$

Функцию адаптации выполняют вспомогательные переменные $\delta^{(i)}\geqslant 0$ и $\lambda^{(i)}\geqslant 0$, определяющие априорную модель случайных последовательностей коэффициентов регрессии, порождаемых независимыми реализациями нормального белого шума с нулевым средним значением:

$$\beta_t^{(i)}=\tfrac{\delta^{(i)}}{\delta^{(i)}+\lambda^{(i)}}\beta_{t-1}^{(i)}+\xi_t^{(i)},\quad \mathsf{E}(\xi_t^{(i)})^2=\tfrac{\delta^{(i)}\lambda^{(i)}}{\delta^{(i)}+\lambda^{(i)}}. \tag{4}$$

Если $\lambda^{(i)}\to 0$, то $\mathsf{E}\big(\xi_t^{(i)}\big)^2\to 0$, и последовательность коэффициентов при i-м регрессоре всегда будет оставаться постоянной $\beta_t^{(i)}=\beta_{t-1}^{(i)}$ с некоторым априори неизвестным значением. Если же $\delta^{(i)}\to 0$, то $\mathsf{E}\big(\xi_t^{(i)}\big)^2\to 0$ вместе с $\frac{\delta^{(i)}}{\delta^{(i)}+\lambda^{(i)}}\to 0$, и i-я последовательность коэффициентов превращается в нулевую константу. Совокупность ненулевых переменных $\delta^{(i)}$ образует множество активных регрессоров $\hat I=\{i:\delta^{(i)}>0\}\subseteq I$, а ненулевые переменные $\lambda^{(i)}$ выделяют среди них подмножество регрессоров с нестационарными коэффициентами $\hat I_{\text{var}}=\{i:\lambda^{(i)}>0\}\subseteq\hat I$.

Произведение дисперсий возмущающего шума $\mathsf{E}\big(\xi_t^{(i)}\big)^2$ по всем регрессорам определяет объем эллипсоида рассеяния случайного вектора $(\beta_t^{(i)}, i\in I)$ вокруг его условного математического ожидания (4) — чем меньше этот объем, тем интенсивнее подавляется общее отклонение всех коэффициентов регрессии как друг от друга во времени, так и от нуля. Роль переменных $\delta^{(i)}$ и $\lambda^{(i)}$ заключается в управлении соотношением между уровнями вариабельности коэффициентов при разных регрессорах, а не их общей вариабельностью в нестационарной модели, поэтому объем эллипсоида рассеяния зафиксирован в критерии (3) ограничением-равенством $\prod_{i\in I}\frac{\delta^{(i)}\lambda^{(i)}}{\delta^{(i)}+\lambda^{(i)}}=1$.

Эту роль берет на себя параметр ρ адаптивного критерия (3), не входящий в число варьируемых переменных и показывающий, во сколько раз дисперсия шума наблюдения $\mathsf{E}\big(e_t^{(i)}\big)^2$ в (1) предполагается большей, чем среднегеометрическое значение дисперсий шума в компонентах модели вариабельности коэффициентов регрессии (4).

При таких предположениях минимум критерия (2) соответствует максимуму апостериорной плотности распределения скрытой последовательности коэффициентов регрессии относительно заданного временного ряда $(y_t, x_t^{(i)}, i \in I)$. Точку минимума легко найти, применяя итерационный алгоритм Гаусса-Зайделя к двум группам переменных $(\delta^{(i)}, \lambda^{(i)}, i \in I)$ и $(\beta_t^{(i)}, i \in I, t = 1, \ldots, N)$, начиная с некоторых значений $\delta^{(i),0}$ и $\lambda^{(i),0}$, удовлетворяющих ограничению в (2).

Заметим, что при фиксированных значениях $\delta^{(i)}$ и $\lambda^{(i)}$, в частности, $\delta^{(i),k}$ и $\lambda^{(i),k}$ на k-й итерации, критерий является квадратичной функцией блочно-трехдиагональной структуры относительно последовательности векторных переменных $(\beta_t^{(i)}, i \in I)$, $t = 1, \ldots, N$, и его минимизация осуществляется эквивалентными методами прогонки, квадратичного динамического программирования, либо фильтрации-интерполяции Калмана-Бьюси [1, 2, 3, 4] за время, пропорциональное длине временного ряда N. Нетрудно доказать, что после того, как последовательность $(\beta_t^{(i),k}, i \in I, t = 1, \ldots, N)$ найдена, очередные значения $\delta^{(i),k+1}$ и $\lambda^{(i),k+1}$, минимизирующие (3), вычисляются по формулам, в которых $a^{(i)} = \frac{\delta^{(i)}}{\delta^{(i)} + \lambda^{(i)}}$ и $0 \lessapprox a_0 < a_1 \lessapprox 1$:

$$
\begin{aligned}
0 < a^{(i),k+1} &= \left(a_0, \frac{\sum_{t=2}^N \beta_{t-1}^{(i),k} \beta_t^{(i),k}}{\sum_{t=2}^N \left(\beta_{t-1}^{(i),k}\right)^2}, a_1 \right) < 1; \\
\delta^{(i),k+1} &= \frac{a^{(i),k+1}}{1 - a^{(i),k+1}} \lambda^{(i),k+1}; \\
\lambda^{(i),k+1} &= \frac{1}{a^{(i),k+1}} \frac{\sum_{t=1}^N \left(\beta_t^{(i),k} - a^{(i),k+1} \beta_{t-1}^{(i),k}\right)^2}{\left[\prod_{j \in I} \sum_{t=1}^N \left(\beta_t^{(j),k} - a^{(j),k+1} \beta_{t-1}^{(j),k}\right)^2\right]^{1/|I|}}, \quad i \in I.
\end{aligned}
$$

Итерационный процесс обычно сходится за 10–15 итераций, проявляя явную тенденцию к практическому обнулению вспомогательных переменных, отвечающих за подавление части регрессоров $\delta^{(i),k} \to \hat{\delta}^{(i)} \gtrapprox 0$, при этом переменные, ответственные за нестационарность коэффициентов регрессии, также почти обнуляются $\lambda^{(i),k} \to \hat{\lambda}^{(i)} \gtrapprox 0$. Предельные значения, оставшиеся существенно ненулевыми $\delta^{(i),k} \to \hat{\delta}^{(i)} > 0$, выделяют подмножество эффективных регрессоров $\hat{I} \subset I$. Однако переменные $\lambda^{(i),k}$ стремятся к нулю, как правило, для большего числа регрессоров, подавляя нестационарность коэффициентов регрессии для части регрессоров, выделенных как эффективные $i \in \hat{I}$. Остальные предельные значения $\hat{\lambda}^{(i)} > 0$ указывают подмножество эффективных регрессоров с нестационарными коэффициентами $\hat{I}_{\text{var}} \subseteq \hat{I}$.

Значение параметра ρ не может быть определено путем дополнительной оптимизации критерия (2) и подбирается с помощью процедуры

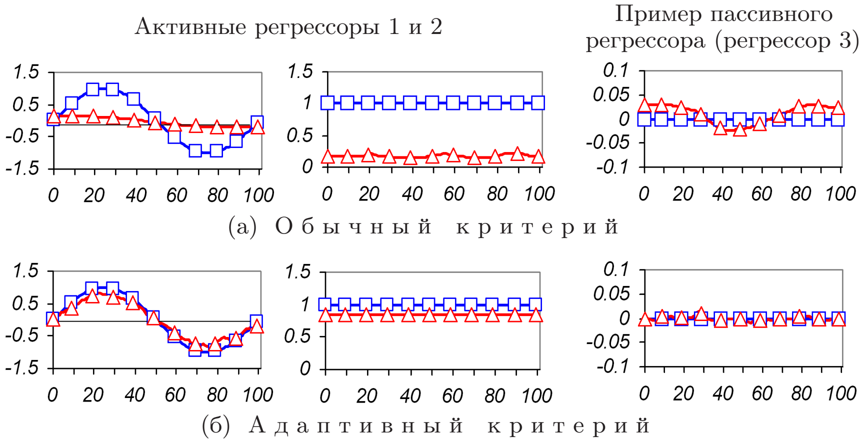

Рис. 1. Результаты экспериментов по оценивианию нестационарной регрессии: □ — модельные последовательности коэффициентов, △ — оцененные последовательности.

скользящего контроля, особенности использования которой в задаче оценивания нестационарной регрессии рассмотрены в работе [3].

Эффект выделения регрессоров, адекватных анализируемому временному ряду, и среди них регрессоров, входящих в модель с нестационарными коэффициентами, иллюстрирует следующий модельный эксперимент.

Модельный временной ряд $(y_t, x_t^{(i)}, i \in I)$ вида (1) построен как последовательность ста зашумленных линейных комбинаций, $t = 1, \ldots, 100$, ста регрессоров $I = \{1, \ldots, 100\}$, в качестве которых использовались независимые реализации нормального белого шума. Среди ста последовательностей коэффициентов регрессии две приняты отличными от нуля, одна из которых образована одним периодом синусоиды $x_t^{(1)} = \sin\left(\frac{2\pi t}{100}\right)$, а вторая является единичной константой $x_t^{(2)} \equiv 1$, остальные же коэффициенты регрессии тождественно равны нулю $x_t^{(3)} = \ldots = x_t^{(100)} \equiv 0$.

Таким образом, если неизвестно, какие именно регрессоры являются активными, то оцениванию подлежат десять тысяч коэффициентов регрессии при наличии всего лишь ста наблюдений. Естественно, что обычный критерий FLS (2), хотя в нем параметр сглаживания ρ подбирался процедурой скользящего контроля, оказался не способен в этом примере даже приближенно восстановить модель нестационарной регрессии, «размывая» вклад двух активных регрессоров на все сто регрессоров, что хорошо видно на Рис. 1(а). В то же время адаптивный критерий (3), как показывает Рис. 1(б), практически полностью подавляет пассивные регрессоры. Вся нестационарность модели оказывается сконцентрированной в изменении коэффициента только при первом регрессоре, а коэффициент при втором активном регрессоре идентифицирован как стационарный.

Работа выполнена при поддержке РФФИ, проект № 06-01-00412.

Литература

[1] *Kalaba R., Tesfatsion L.* Time-varying linear regression via flexible least squares. — International Journal on Computers and Mathematics with Applications. — Vol. 17. — Pp. 1215–1245.

[2] *Костин А. А., Красоткина О. В., Марков М. Р., Моттль В. В., Мучник И. Б.* Алгоритмы динамического программирования для анализа нестационарных сигналов. — ЖВМиМФ, 2004, — Т. 44, № 1. — С. 70–86.

[3] *Markov M., Krasotkina O., Mottl V., Muchnik I.* Time-varying regression model with unknown time-volatility for nonstationary signal analysis // 8th IASTED Int. Conf. on Signal and Image Processing, Honolulu, USA. — Pp. 14–16.

[4] *Wells C.* The Kalman Filter in Finance. — Kluwer Academic Publishers, 1996.