УДК 519.233.33

Робастные GM-оценки в авторегрессии и тесты типа хи-квадрат Пирсона

М. В. Болдин*, М. Н. Петриев*

* *Кафедра теории вероятностей,*
Московский государственный университет им. М.В.Ломоносова,
Ленинские Горы 1, Москва, Россия, 119992

Аннотация. В работе рассматривается ситуация, когда наблюдения за авторегрессией первого порядка содержат грубые ошибки (выбросы или засорения), причём распределение засорений неизвестно и произвольно. В качестве альтернативы оценке наименьших квадратов неизвестного параметра, которая не применима в данной ситуации, предлагается обобщённая М-оценка. Установлена её асимптотическая качественная робастность в терминах равностепенной непрерывности семейства предельных функций распределения. На основании этой оценки строится тест типа хи-квадрат Пирсона для проверки гипотезы о виде распределения инноваций и показывается, что асимптотический уровень значимости такого теста будет устойчив к выбросам.

Ключевые слова: авторегрессия, проверка гипотез, GM-оценки, асимптотическая качественная робастность, тесты хи-квадрат Пирсона.

1. Введение

Рассмотрим авторегрессию первого порядка

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},$$

где $\{\varepsilon_t\}$ независимые одинаково распределенные случайные величины (н.о.р сл. в.) с неизвестной функцией распределения (ф.р.) $G(x)$ и лебеговой плотностью g(x), $\mathsf{E}\varepsilon_1 = 0$, $\mathsf{E}\varepsilon_1^2 < \infty$; $|\beta| < 1$, β - неизвестный параметр. Мы предполагаем, что наблюдения авторегрессии могут содержать грубые ошибки (выбросы или засорения), так что вместо $\{u_t\}$ наблюдаются величины:

$$y_t = u_t + z_t^{\gamma_n}\xi_t,\ t = 0, 1, \ldots, n. \tag{1}$$

В (1) $\{z_t^{\gamma_n}\}$ – н.о.р.сл.в., имеющие распределение Бернулли $Br(\gamma_n)$, $\gamma_n = \min(1, \frac{\gamma}{\sqrt{n}})$, $\gamma \geq 0$, γ — неизвестный уровень засорения; $\{\xi_i\}$ последовательность н.о.р. сл.в. с неизвестным и произвольным распределением μ. Последовательности $\{u_t\}, \{z_t^{\gamma}\}, \{\xi_t\}$ предполагаются независимыми между собой.

Схема засорений (1) является локальным вариантом известной схемы засорения данных во временных рядах из [4].

Целью данной работы является решение двух взаимосвязанных задач. Первая из них - построение робастной обобщенной M-оценки (GM-оценки) параметра β по наблюдениям $\{y_t\}$. Робастность оценки будет характеризоваться равностепенной непрерывностью семейства предельных распределений оценки в точке $\gamma = 0$. Подобные оценки уже исследовались ранее, см., например, [3], но лишь при не очень естественных моментных ограничениях на засорения. Подчеркнем еще раз – у нас нет никаких ограничений для μ.

С помощью построенной GM-оценки мы решаем вторую задачу – построения теста типа хи-квадрат для проверки гипотезы о виде неизвестной ф.р. $G(x)$. Подобная задача в авторегрессии без засорений решалась в [1] с помощью теста Колмогорова. Мы покажем, что наш хи-квадрат тест теперь уже в схеме с засорениями является асимптотически качественно робастным при гипотезе.

2. Построение робастной GM-оценки параметра авторегрессии

Пусть $\varphi(x)$ и $\psi(x)$ - некоторые *à priori* выбранные вещественнозначные функции. Определим процесс

$$l_n(\theta) := n^{-1/2} \sum_{t=1}^{n} \varphi(y_{t-1})\psi(y_t - \theta y_{t-1}).$$

Обощённая M-оценка определяется как $n^{1/2}$-состоятельное решение уравнения

$$l_n(\theta) = 0. \qquad (2)$$

Сформулируем необходимые в дальнейшем условия:
($\boldsymbol{i}$) $\mathsf{E}\psi(\varepsilon_1) = 0$.
($\boldsymbol{ii}$) Функция $\psi(x)$ дважды дифференцируема, причём сама ψ и обе ее производные ограничены.
($\boldsymbol{iii}$) $\sup|\varphi(x)| < \infty$.
($\boldsymbol{iv}$) $\mathsf{E}u_0\varphi(u_0)\int\limits_{\mathbb{R}} g(x)d\psi \neq 0$.

Для изучения поведения GM-оценки нам потребуется асимпмтотическое разложение процесса $l_n(\theta)$ в $n^{-1/2}$-окрестности истинного значения параметра β. Это разложение нам даёт следующая

Лемма 1. *Пусть выполнены условия (**i**)—(**iv**). Тогда для любых* $0 \le \Theta < \infty$, $0 \le \Gamma < \infty$, $\delta > 0$

$$\sup_{\gamma\le\Gamma} \mathsf{P}\big(\sup_{|\theta|\le\Theta} |l_n(\beta + n^{-1/2}\theta) - \tilde{l}_n(\beta) + \theta\Delta_1 - \gamma\Delta_2| > \delta\big) \to 0,\ n \to \infty,$$

где $\Delta_1 = \mathsf{E}u_0\varphi(u_0)\int\limits_{\mathbb{R}} g(x)d\psi$, $\Delta_2 = \mathsf{E}\varphi(u_0)\mathsf{E}\psi(\varepsilon_1 + \xi_1) + \mathsf{E}\varphi(u_0 + \xi_0)\psi(\varepsilon_1 - \beta\xi_0)$, *а* $\tilde{l}_n(\beta) := n^{-1/2}\sum_{t=1}^{n}\varphi(u_{t-1})\psi(\varepsilon_t)$.

Лемма 1 и стандартная техника (см., например, [1], гл. 6) совместно с асимптотической гауссовостью $\tilde{l}_n(\beta)$ влекут следующую теорему.

Теорема 1. *Пусть выполнены условия* ***(i)—(iv)****. Тогда с вероятностью, стремящейся к единице при $n \to \infty$, существует последовательность $\hat{\beta}_{n,GM}$ равномерно по $\gamma \le \Gamma\ n^{1/2}$ - состоятельных корней уравнения* (2)*. При этом*

$$\sup_{x,\gamma\le\Gamma}\Big|P(n^{1/2}(\hat{\beta}_{n,GM}-\beta)\le x)-\Phi\Big(\frac{x-\gamma\Delta_1^{-1}\Delta_2}{\sigma}\Delta_1\Big)\Big|\to 0,\ n\to\infty,$$

где $\Phi(x)$ - функция Лапласа, а $\sigma^2 := \mathsf{E}\psi^2(\varepsilon_1)\mathsf{E}\varphi^2(u_0)$.

Очевидно, предельное распределеие GM-оценки в схеме без засорений с $\gamma = 0$ — нормальное $N(0, \sigma^2/\Delta_1^2)$ с ф.р. $F^Y(x, 0, \mu) := F(x) = \Phi\big(\frac{x\Delta_1}{\sigma}\big)$. Введём обозначение:

$$F^Y(x,\gamma,\mu) := \lim_{n\to\infty}\mathsf{P}(n^{1/2}(\hat{\beta}^Y_{n,GM}-\beta)\le x) = \Phi\Big(\frac{x-\gamma\Delta_1^{-1}\Delta_2}{\sigma}\Delta_1\Big).$$

Из *Теоремы 1* прямо следует асимптотическая качественная робастность обощённых М-оценок.

Теорема 2. *Пусть выполнены условия* ***(i)—(iv)****. Тогда $\hat{\beta}_{n,GM}$ - асимптотически качественно робастна, то есть*

$$\sup_{x,\mu}\big|F^Y(x,\gamma,\mu)-F(x)\big|\to 0, \gamma\to 0.$$

3. Тесты типа хи-квадрат Пирсона

В этом разделе мы построим тест типа хи-квадрат Пирсона для проверки гипотезы $H_0 : G(x) = G_0(x)$, $G_0(x)$ - известная функция распределения. Тест будет основан на наблюдениях $y_0, \ldots, y_n$ из (1). Пусть $\hat{\beta}_n$ - любая оценка, являющаяся равномерно по $\gamma \le \Gamma\ n^{1/2}$ – состоятельной оценкой β. Например, годится оценка $\hat{\beta}_{n,GM}$ из *Теоремы 1*. Пусть $\hat{\varepsilon}_t = y_t - \hat{\beta}y_{t-1}$, $t = 1, \ldots, n$ - остатки, а

$$\hat{G}_n(x) := n^{-1}\sum_{t=1}^{n} I(\hat{\varepsilon}_t \le x), x \in \mathbb{R}.$$

$\hat{G}_n(x)$ - подобие эмпирической функции распределения

$$G_n(x) = n^{-1}\sum_{t=1}^{n} I(\varepsilon_t \le x),$$

которая может быть построена лишь гипотетически. Поскольку остатки $\{\hat{\varepsilon}_t\}$ зависимы между собой, свойства $\hat{G}_n(x)$ нуждаются в изучении.

Теорема 3. *Пусть* $\sup|g'(x)| < \infty$. *Пусть* $\Delta(x,\mu) = \mathsf{E}G(x+\beta\xi_1) + \mathsf{E}G(x-\xi_1) - 2G(x)$. *Тогда при любых* $\varepsilon > 0$ *и* $0 \le \Gamma < \infty$

$$\sup_{\gamma\le\Gamma} \mathsf{P}(|\hat{G}_n(x) - G_n(x) - \Delta(x,\mu)\gamma| > \varepsilon) \to 0,\ n\to\infty.$$

Доказательство *Теоремы 3* весьма кропотливо и основано на идеях [2].

Разобьём числовую ось на m непересекающихся полуинтервалов $\Delta_1,\ldots,\Delta_m$, $\Delta_j = (x_{j-1}, x_j]$, $x_0 = -\infty$, $x_m = \infty$, и пусть при H_0

$$p_j^0 := \mathsf{P}(\varepsilon_1 \in \Delta_j) = G_0(x_j) - G_0(x_{j-1}) > 0.$$

Пусть $\hat{\nu}_j$ есть число отстатков $\{\hat{\varepsilon}_t\}$, попавших в Δ_j. Очевидно,

$$\hat{\nu}_j = n[\hat{G}_n(x_j) - \hat{G}_n(x_{j-1})],\ j = 1,\ldots,m. \qquad (3)$$

Интересующая нас статистика хи-квадрат имеет вид

$$\hat{\chi}_n^2 = \sum_{j=1}^m \frac{(\hat{\nu}_j - np_j^0)^2}{np_j^0}.$$

Предельное распределение χ_n^2 при H_0 можно найти с помощью *Теоремы 3*, соотношения (3) и стандарного доказательства теоремы Пирсона. А именно, пусть $F_k(x,\lambda^2)$ будет функция распределения нецентрального распределения хи-квадрат с k степенями свободы и параметром нецентральности λ^2.

Теорема 4. *Пусть* $g_0(x) = G_0'(x)$ *и* $\sup|g_0'(x)| < \infty$. *Тогда при* H_0

$$\sup_{x,\gamma\le\Gamma} |\,\mathsf{P}(\hat{\chi}_n^2 \le x) - F_{m-1}(x,\lambda^2(\gamma,\mu))| \to 0,\ n\to\infty,$$

где $\lambda^2(\gamma,\mu) = \sum_{j=1}^m \delta_j^2(\mu)/p_j^0$, *а* $\delta_j(\mu) = \Delta(x_j,\mu) - \Delta(x_{j-1},\mu)$.

Обозначим $\chi_k(\alpha)$ квантиль уровня α распределения хи-квадрат с k степенями свободы. Будем отвергать H_0, если $\hat{\chi}_n^2 > \chi_{m-1}(1-\alpha)$. Асимптотический уровень значимости такого теста равен

$$\hat{\alpha}(\gamma,\mu) = 1 - F(\chi_{m-1}(1-\alpha), \lambda^2(\gamma,\mu)).$$

Разумеется, в схеме без засорений с $\gamma = 0$ имееем $\hat{\alpha}(0,\mu) = \alpha$. Следующая теорема утверждает, что асимптотический уровень значимости $\hat{\alpha}(\gamma,\mu)$ качественно асимптотически устойчив к выбросам.

Теорема 5. $\sup_\mu |\hat{\alpha}(\gamma,\mu) - \alpha| \to 0, \gamma \to 0.$

Легко убедиться, что наш тест хи-квадрат состоятелен против стандартной (т.е. $p_j^0 \ne p_j^1$ хоть при одном j) фиксированной альтернативы.

4. Заключение

Исследование авторегрессионных схем с выбросами (засорениями) в наблюдениях – актуальная для приложений задача, она совершенно нетривиальна для теоретического исследования. Приведенные выше результаты – пример ее решения, допускающий многочисленные обобщения. В частности, на многопараметрические линейные и нелинейные авторегрессионные схемы. Мы сделаем это в отдельных публикациях.

Литература

1. М. В. Болдин, Г. И. Симонова, Ю. Н. Тюрин. *Знаковый статистический анализ линейных моделей.* Наука Физматлит Москва, 1997.
2. M. V. Boldin. On empirical processes in heteroscedastic time series and their use for hypothesis testing and estimation. *Mathematical Methods of Statistics*, 9(1):65–89, 2000.
3. D. M. Esaulov. Residual empirical processes and their application to GM-testing for the autoregression order. *Mathematical Methods of Statistics*, 22(4):333–349, 2013.
4. R. Douglas Martin and Victor J. Yohai. Influence functionals for time series. *Ann. Statist.*, 14(3):781–818, 09 1986.

UDC 519.233.33

Robust GM-estimators in the autoregression and Pearson's chi-square tests

M. V. Boldin*, M. N. Petriev*

* *Department of Probability Theory,*
Moscow State University,
Leninskie Gory 1, Moscow, 119991, Russia

This paper considers the situation where observations of the first-order autoregression contain gross errors (outliers or contaminations), and the distribution of the outliers is unknown and arbitrary. As an alternative to the least-squares estimator, which is not applicable in this situation, the generalized M-estimator for unknown parameter is proposed. The asymptotic qualitative robustness of this estimator in terms of limiting distribution functions equicontinuity is established. On the basis of GM-estimator the Pearson's chi-squared test for hypotheses about the distribution of the innovations is constructed and qualitative robustness of the asymptotic significance level is obtained.

Keywords: autoregression, hypotheses testing, GM-estimators, asymptotic qualitative robustness, Pearson's chi-squared tests.