УДК 519.234 EDN: QCRHHH

ВЫЧИСЛЕНИЕ АСИМПТОТИЧЕСКОЙ КОВАРИАЦИОННОЙ МАТРИЦЫ ОЦЕНКИ ХЬЮБЕРА В ОБОБЩЕННОЙ ЭКСПОНЕНЦИАЛЬНОЙ АВТОРЕГРЕССИОННОЙ МОДЕЛИ

В.Б. Горяинов vb-goryainov@bmstu.ru
М.М. Масягин masyagin@bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

При решении задач регрессионного анализа наиболее часто в качестве функции потерь используют функции наименьших квадратов и наименьших модулей. Тем не менее они не лишены недостатков. Первая функция крайне чувствительна к выбросам в обучающих данных, а вторая не является дифференцируемой в нуле, что осложняет ее использование с широким классом оптимизационных методов. Для того чтобы нивелировать эти недостатки, было разработано несколько функций, комбинирующих возможности оценок наименьших квадратов и наименьших модулей, например, функция Хьюбера. До определенного порога они ведут себя схоже с квадратичной функцией, после него — схоже с функцией модуля числа. Тем самым они являются и дифференцируемыми на всей области определения, и малочувствительными к выбросам (превышающим заданный порог). В настоящей работе вычислена асимптотическая ковариационная матрица обобщенной экспоненциальной авторегрессионной модели — модель Озаки — для оценки функции Хьюбера. В нее входит обобщение полученной формулы на более широкий класс оценок, таких как оценка наименьших квадратов, оценка наименьших модулей и другие оценки, объединяемые вместе в множество M-оценок

Ключевые слова

Обобщенная экспоненциальная авторегрессионная модель, функция потерь Хьюбера, M-оценки, асимптотическая ковариационная матрица

Поступила 06.06.2024
Принята 04.02.2025

Введение. Современное машинное обучение включает в себя широкий спектр различных задач. Одной из классических и наиболее часто встречающихся является задача регрессии, а ее частным случаем — задача авторегрессии. В качестве функции потерь в них активно используются функции наименьших квадратов и наименьших модулей, а также их различные комбинации, например функция Хьюбера [1]. Функция Хьюбера характеризуется тем, что до определенного порога представляет собой квадратичную функцию, а после — функцию модуля. В отличие от квадратичной функции потерь она более устойчива к выбросам в данных, и в то же время, в отличие от модульной функции потерь, она дифференцируема на всей области определения и наследует высокую эффективность метода наименьших квадратов (МНК) в отсутствие выбросов.

Если речь идет о задаче авторегрессии, то нельзя не упомянуть экспоненциальной авторегрессионной модели, также известной как модель Озаки [2], и несколько ее модификаций [3–5]. Предложенная в 1981 г., она и по сей день активно используется при моделировании различных физических и экономических процессов [6–9]. Кроме того, с развитием искусственных нейронных сетей модель Озаки нашла применение в генерации тестовых и обучающих данных для архитектур [10].

Цель работы — нахождение асимптотической ковариационной матрицы оценки Хьюбера в модели Озаки и ее обобщение на другие оценки.

Постановка задачи. Функция Хьюбера задается соотношением на базе квадратичной функции и функции абсолютного значения числа:

$$\rho_\delta(x) = \begin{cases} \dfrac{1}{2}x^2, \text{ если } |x| < \delta; \\ \delta\left(|x| - \dfrac{1}{2}\delta\right), \text{ иначе.} \end{cases} \tag{1}$$

где δ — пороговое значение функции Хьюбера.

Экспоненциальная авторегрессионная модель порядка p может быть выражена по формуле:

$$X_t = \sum_{i=0}^{p-1}\left(a_i + b_i e^{-cX_{t-1}^2}\right)X_{t-(i+1)} + \xi_t, \tag{2}$$

где $a_0, b_0, a_1, b_1, \ldots, a_{p-2}, b_{p-2}, a_{p-1}, b_{p-1}, c$ — действительные коэффициенты; ξ_t, $t = 0, 1, \ldots$ — последовательность случайных величин, независимых друг от друга, каждая из которых удовлетворяет условиям:

– $E\xi_t = 0$, т. е. каждая случайная величина в последовательности имеет нулевое математическое ожидание;

– $D\xi_t = E\xi_t^2 = \sigma^2 < +\infty$, т. е. каждая случайная величина в последовательности имеет конечную дисперсию.

Для дальнейшего использования введем следующие обозначения: $a=\left(a_0,\, a_1,\, \ldots,\, a_{p-1}\right)$; $b=\left(b_0,\, b_1,\, \ldots,\, b_{p-1}\right)$.

Пусть даны процесс X_t, удовлетворяющий (2), и n его наблюдений $X_0, \ldots, X_{n-1}$. Рассмотрим задачу оценивания параметров $(a,\, b,\, c)$ модели Озаки по $X_0, \ldots, X_{n-1}$ при оценке Хьюбера, т. е. задачу минимизации функции нескольких переменных:

$$g(a,b,c)=\sum_{t=p}^{n}\rho_\delta\left(X_t-\sum_{i=0}^{p-1}\left(a_i+b_i e^{-cX_{t-1}^2}\right)X_{t-(i+1)}\right).$$

Вычислим асимптотическую ковариационную матрицу оценки Хьюбера в модели Озаки (ковариационную матрицу этой оценки в рассматриваемой модели при $n\to\infty$).

Вычисление асимптотической ковариационной матрицы. Для удобства введем обозначение

$$x_t(a,b,c)=\xi_t=X_t-\sum_{i=0}^{p-1}\left(a_i+b_i e^{-cX_{t-1}^2}\right)X_{t-(i+1)},$$

тогда $g(a,b,c)=\sum_{t=p}^{n}\rho_\delta\left(\xi_t\right)=\sum_{t=p}^{n}\rho_\delta\left(x_t(a,b,c)\right)$.

Для вычисления асимптотической ковариационной матрицы используем следующий алгоритм.

1. Разложим функцию $g(a,b,c)$ по формуле Тейлора в окрестности некоторой фиксированной точки $\left(\tilde{a},\tilde{b},\tilde{c}\right)$ вплоть до второго порядка. Обоснованно исключим бесконечно малый член разложения и используем полученное выражение в качестве аппроксимации исходной функции $g(a,b,c)$.

2. Предположим, что модель Озаки с оценкой Хьюбера является стационарной и эргодической. Это предположение является корректным и не сильно сужает сферу практического применения модели, поскольку в инженерных задачах, как правило, рассматривают именно стационарные и эргодичные процессы или процессы, приводимые к ним (квазистационарные процессы). С учетом наложенных ограничений выразим предельные значения вектора и матрицы коэффициентов разложения по формуле Тейлора из п. 1 при $n\to\infty$.

3. Вычислим асимптотическое распределение точки минимума квадратичной формы, являющейся основной частью разложения Тейлора из п. 1. Докажем, что оно совпадает с асимптотическим распределением точки $\left(\hat{a},\hat{b},\hat{c}\right)$ минимума исходной функции $g(a,b,c)$. Выразим асимпто-

тическую ковариационную матрицу оценки Хьюбера в экспоненциальной авторегрессионной модели в общем виде.

4. Явно вычислим первые и вторые производные функции $g(a, b, c)$ по всем ее переменным (парам ее переменных), а также их предельные значения при $n \to \infty$. Подставим их значения в выражение, полученное в п. 3, и запишем итоговую формулу асимптотической ковариационной матрицы оценки Хьюбера в модели Озаки.

5. Обобщим результаты, полученные в пп. 1–4 для широкого класса функций потерь — функций М-оценок.

Асимптотическая ковариационная матрица в п. 3 найдена в общем виде, поэтому вместо оценки Хьюбера могут быть легко рассмотрены и любые другие оценки, например оценки наименьших квадратов или наименьших модулей, оценки Тьюки, Эндрюса или Уэлша [11].

Аппроксимация разложением по формуле Тейлора. Зафиксируем некоторую точку $\left(\tilde{a}, \tilde{b}, \tilde{c}\right)$ и разложим функцию $g(a, b, c)$ в ее окрестности по формуле Тейлора второго порядка:

$$g(a, b, c) = g\left(\tilde{a}, \tilde{b}, \tilde{c}\right) + A^{\mathrm{T}}\Theta + (1/2)\Theta^{\mathrm{T}} B\Theta + \varphi(a, b, c).$$

Здесь A — блочный вектор-столбец размером $(2p+1)\times 1$, элементами которого являются первые производные функции $g(a, b, c)$ в точке $\left(\tilde{a}, \tilde{b}, \tilde{c}\right)$, он состоит из двух подвекторов размером $p\times 1$ и одного числа; $\Theta = \sqrt{n}\left(a-\tilde{a},\ b-\tilde{b},\ c-\tilde{c}\right)$ — блочный вектор-столбец такого же размера, что и вектор-столбец A; B — блочная матрица Гессе размером $(2p+1)\times(2p+1)$ функции $g(a, b, c)$ в точке $\left(\tilde{a}, \tilde{b}, \tilde{c}\right)$, он состоит из четырех подматриц размером $p\times p$, двух подвекторов размером $p\times 1$, двух подвекторов размером $1\times p$ и одного числа; $\varphi(a, b, c)$ — бесконечно малая величина при $(a, b, c) \to \left(\tilde{a}, \tilde{b}, \tilde{c}\right)$.

Формулы, задающие блочное представление вектора-столбца A и матрицы Гессе B:

$$A = \frac{1}{\sqrt{n}}\begin{pmatrix} \dfrac{\partial g(a,b,c)}{\partial a} \\ \dfrac{\partial g(a,b,c)}{\partial b} \\ \dfrac{\partial g(a,b,c)}{\partial c} \end{pmatrix}; \quad B = \frac{1}{n}\begin{pmatrix} \dfrac{\partial g(a,b,c)}{\partial a^2} & \dfrac{\partial g(a,b,c)}{\partial a\partial b} & \dfrac{\partial g(a,b,c)}{\partial a\partial c} \\ \dfrac{\partial g(a,b,c)}{\partial b\partial a} & \dfrac{\partial g(a,b,c)}{\partial b^2} & \dfrac{\partial g(a,b,c)}{\partial b\partial c} \\ \dfrac{\partial g(a,b,c)}{\partial c\partial a} & \dfrac{\partial g(a,b,c)}{\partial c\partial b} & \dfrac{\partial g(a,b,c)}{\partial c^2} \end{pmatrix}.$$

Несложно доказать, что порядок малости функции $\varphi(a,b,c)$ выше порядка малости функции

$$\tau(a,b,c)=(a-\tilde{a})^{\mathrm{T}}(a-\tilde{a})+\left(b-\tilde{b}\right)^{\mathrm{T}}\left(b-\tilde{b}\right)+(c-\tilde{c})^2.$$

Таким образом, ее вкладом в разложение по формуле Тейлора далее можно пренебречь.

Ограничения эргодичности и стационарности. Запишем характеристическое уравнение исследуемой модели

$$\lambda^{p}-a_0\lambda^{p-1}-\ldots-a_{p-1}=0. \tag{3}$$

Предположим, что исследуемая модель (2) является стационарной и эргодической. Согласно [12], это накладывает на ее коэффициенты следующие ограничения:

– $c>0,$ так как иначе множитель $e^{-cX_{t-1}^2}$ будет бесконечно большим;

– $|a_i|<1,\ i=0,\ 1,\ \ldots,\ p-1,$ т. е. все корни (3) лежат внутри единичного круга с центром в точке (0, 0).

Модель стационарна и эргодична, поэтому существует предел $\lim\limits_{n\to\infty}B.$ Обозначим его как $\left(E\rho''_{\delta}\left(x_t(a,b,c)\right)\right)K,$ где K — матрица коэффициентов размером $(2p+1)\times(2p+1);$ $E\rho''_{\delta}\left(x_t(a,b,c)\right)$ — математическое ожидание случайной величины, являющейся второй производной функции $\rho_{\delta}\left(x_t(a,b,c)\right)$, причем $E\rho'_{\delta}\left(x_t(a,b,c)\right)=0;$ $E\rho''_{\delta}\left(x_t(a,b,c)\right)<+\infty.$

Непосредственные вычисления первых и вторых производных функции $g(a,b,c)$ в п. 4 алгоритма покажут корректность выбранного представления $\lim\limits_{n\to\infty}B.$ Таким образом, $\lim\limits_{n\to\infty}B=\left(E\rho''_{\delta}\left(x_t(a,b,c)\right)\right)K,$ следовательно,

$$g(a,b,c)=g\left(\tilde{a},\tilde{b},\tilde{c}\right)+A^{\mathrm{T}}\Theta+\frac{1}{2}\Theta^{\mathrm{T}}\left(E\rho''_{\delta}\left(x_t(a,b,c)\right)\right)K\Theta.$$

Поиск асимптотического распределения точки минимума целевой функции. Покажем, что точка $\left(\hat{a},\,\hat{b},\,\hat{c}\right)$ минимума функции $g(a,b,c)$ и точка минимума квадратичной формы $A^{\mathrm{T}}\Theta+(1/2)\Theta^{\mathrm{T}}\left(E\rho''_{\delta}\left(x_t(a,b,c)\right)\right)\times$ $\times K\Theta$ обладают одинаковыми асимптотическими распределениями. Пусть

$$E_t=\sum_{i=0}^{p-1}\left(\left(a_i+b_ie^{-cX_{t-1}^2}\right)X_{t-(i+1)}+\xi_t\right)-\left(\hat{a}_i+\hat{b}_ie^{-\hat{c}X_{t-1}^2}\right)X_{t-(i+1)}.$$

Тогда функция ошибки задается как $L\left(\hat{a},\, \hat{b},\, \hat{c}\right) = \sum\limits_{t=p}^{n} \left(\rho_\delta(E_t) - \rho_\delta(\xi_t)\right)$.

Выполним подстановку $\widehat{\Theta} = \sqrt{n}\left(\hat{a}-a,\, \hat{b}-b,\, \hat{c}-c\right)$ и разложение $L\left(\hat{a},\, \hat{b},\, \hat{c}\right)$ по формуле Тейлора. Для этого используем вид разложения в ряд Тейлора с остаточным членом в форме Лагранжа. Очевидно, что производные функции $L(\widehat{\Theta})$ до второго порядка включительно с точностью до обозначений совпали с соответствующими производными функции $g\left(\tilde{a},\, \tilde{b},\, \tilde{c}\right)$. Таким образом,

$$L(\widehat{\Theta}) = A^{\mathrm{T}}\widehat{\Theta} + \frac{1}{2}\widehat{\Theta}^{\mathrm{T}} B\widehat{\Theta} + \beta(\widehat{\Theta}) = A^{\mathrm{T}}\widehat{\Theta} + \frac{1}{2}\widehat{\Theta}^{\mathrm{T}}\left(E\rho''_\delta\left(x_t\left(a,b,c\right)\right)\right)K\widehat{\Theta} + \beta(\widehat{\Theta}),$$

$$\beta(\widehat{\Theta}) = \widehat{\Theta}^{\mathrm{T}}\left(\frac{\partial^2 L}{\partial\widehat{\Theta}^2}(\widetilde{\Theta}) - \frac{\partial^2 L}{\partial\widehat{\Theta}^2}(0)\right)\widehat{\Theta},$$

где $\widetilde{\Theta} \in \left[0,\, \widehat{\Theta}\right]$.

Осталось показать, что $\beta(\widehat{\Theta}) \to 0$ при $n \to \infty$. Рассчитаем покоординатные значения элементов матрицы $\left(\frac{\partial^2 L}{\partial\widehat{\Theta}^2}(\widetilde{\Theta}) - \frac{\partial^2 L}{\partial\widehat{\Theta}^2}(0)\right)$ с учетом того, что X_t и ξ_t являются стационарными и эргодичными последовательностями. На базе полученных результатов сделаем вывод, что все они равны нулю, а величины элементов векторов $\widehat{\Theta}$ и $\widehat{\Theta}^{\mathrm{T}}$ конечны. Это доказывает, что $\beta(\widehat{\Theta}) \to 0$ при $n \to \infty$. Таким образом, $E\,|\beta(\widehat{\Theta})| \to 0$ при $n \to \infty$, следовательно, $\beta(\widehat{\Theta}) \to 0$ при $n \to \infty$ по вероятности.

Пусть $\widetilde{\Theta}$ — минимум функции $\widetilde{L}(\Theta) = A^{\mathrm{T}}\Theta + (1/2)\Theta^{\mathrm{T}} W\Theta$, где $W = \left(E\rho''_\delta\left(x_t\left(a,b,c\right)\right)\right)K$. Учитывая, что $W < +\infty$ и $\widetilde{L}(\Theta)$ — квадратичная форма, очевидно следующее:

$$\widetilde{\Theta} = -W^{-1}A = -\frac{1}{E\rho''_\delta\left(x_t(a,b,c)\right)}K^{-1}A.$$

Рассмотрим μ_t — σ-алгебру событий, порожденную множеством случайных величин ξ_s, $s \le t$. По определению $x_t(a,b,c)$ независима от μ_t, X_{t-1} измерима относительно μ_{t-1}. Кроме того, $E\rho'_\delta\left(x_t(a,b,c)\right) = 0$. Следовательно, $E\left[\rho'_\delta\left(x_t(a,b,c)\right)X_t \,|\, \mu_{t-1}\right] = 0$.

Последовательность $\rho'_\delta\left(x_t(a,b,c)\right)X_t$, $t=1, 2, \ldots,$ образует мартингал-разность относительно последовательности σ-алгебр. В силу центральной предельной теоремы для мартингалов [13] вектор A является асимптотически нормальным, $EA=0$, его ковариационная матрица равна $KE\left[\left(\rho'_\delta\left(x_t(a,b,c)\right)\right)^2\right]$. Вследствие этого случайная величина $\widetilde{\Theta}=-W^{-1}A$ является асимптотически нормальной, $E\widetilde{\Theta}=0$, причем ее ковариационная матрица имеет вид [14]

$$\frac{E\left[\left(\rho'_\delta(x_t(a,b,c))\right)^2\right]}{\left[E\rho''_\delta\left(x_t(a,b,c)\right)\right]^2}K^{-1}. \tag{4}$$

Осталось доказать, что $\widehat{\Theta}-\widetilde{\Theta}\to 0$ по вероятности. Для доказательства фиксируем некоторый параметр $\delta>0$ и покажем, что $P\left[\left|\widehat{\Theta}-\widetilde{\Theta}\right|\le\delta\right]\to 1$ [15, 16].

Случайная величина $\widetilde{\Theta}$ является асимптотически нормальной и $E\widetilde{\Theta}=0$, поэтому она также ограничена по распределению. Следовательно, должен существовать такой компакт $C\in\mathbb{R}^n$, который содержит сразу для всех n шары с центром в $\widetilde{\Theta}$ и радиусом δ. Согласно закону больших чисел, $B\xrightarrow{n\to\infty}W$, значит $L(\Theta)-\tilde{L}(\Theta)\xrightarrow{n\to\infty}0$ по вероятности. Поэтому $L(\Theta)+A^{\mathrm{T}}\xrightarrow{n\to\infty}(1/2)\Theta^{\mathrm{T}}W\Theta$. Таким образом, $\beta(\widetilde{\Theta})\to 0$ равномерно на любом компакте. Из [17] следует, что $\Delta=\sup_{\Theta\in C}\left|\beta(\widetilde{\Theta})\right|\xrightarrow{n\to\infty}0$ по вероятности.

Зададим произвольный нормированный вектор e, такой что $\Theta^*= =\widetilde{\Theta}+\delta e$, $\Theta=\widetilde{\Theta}+te$, $t>\delta$. Тогда из выпуклости $L(\Theta)$ получим

$$L(\Theta^*)\ge\left(1-\frac{\delta}{t}\right)L(\widetilde{\Theta})+\frac{\delta}{t}L(\Theta),$$

следовательно,

$$L(\Theta)\ge L(\widetilde{\Theta})+\frac{t}{\delta}\left(L(\Theta^*)-L(\widetilde{\Theta})\right)\ge L(\widetilde{\Theta})+\frac{t}{\delta}\left(\tilde{L}(\Theta^*)-\tilde{L}(\widetilde{\Theta}))+\beta(\Theta^*)-\beta(\widetilde{\Theta}).$$

Таким образом, матрица W является положительно определенной, и $\tilde{L}(\Theta^*)-\tilde{L}(\widetilde{\Theta})=\frac{1}{2}\delta^2e^{\mathrm{T}}We>0.$ Тогда

$$\inf_{\Theta\notin C}L(\Theta)=L(\widetilde{\Theta})+\inf_{t>\delta}\frac{t}{\delta}\left(L(\Theta^*)-L(\widetilde{\Theta})\right)\ge L(\widetilde{\Theta})+\frac{1}{2}\delta^2e^{\mathrm{T}}We-2\Delta.$$

Следовательно, минимум $L(\Theta)$ гарантированно лежит внутри компакта C, а минимумы функции $g(a, b, c)$ и рассматриваемой квадратичной формы совпадают и являются асимптотически нормально распределенными.

С учетом изложенного случайный вектор A является асимптотически нормальным с нулевым математическим ожиданием и ковариационной матрицей $KE\left[\left(\rho'_\delta\left(x_t(a, b, c)\right)\right)^2\right]$; ковариационная матрица асимптотически нормальная с нулевым математическим ожиданием случайной величины $\widetilde{\Theta} = -W^{-1}A$ имеет вид (4). В результате совпадения минимумов функции $g(a, b, c)$ и рассматриваемой квадратичной формы и того, что они асимптотически нормальные, получаем, что асимптотическая ковариационная матрица оценки Хьюбера в обобщенной экспоненциальной авторегрессионной модели Озаки имеет вид (4).

Вычисление производных и их подстановка в формулу (4). Вычислим первую и вторую производные функции Хьюбера (1):

$$\rho'_\delta(x) = \begin{cases} \dfrac{\partial\left((1/2)x^2\right)}{\partial x}, \text{ если } |x| < \delta; \\ \dfrac{\partial\left(\delta\left(|x| - (1/2)\delta\right)\right)}{\partial x}, \text{ иначе,} \end{cases} = \begin{cases} x, \text{ если } |x| < \delta; \\ \delta \operatorname{sign} x, \text{ иначе,} \end{cases}$$

$$\rho''_\delta(x) = \begin{cases} \dfrac{\partial x}{\partial x}, \text{ если } |x| < \delta; \\ \dfrac{\partial\left(\delta \operatorname{sign} x\right)}{\partial x}, \text{ иначе,} \end{cases} = \begin{cases} 1, \text{ если } |x| < \delta; \\ 0, \text{ иначе.} \end{cases}$$

Далее определим первые и вторые производные функции $x_t(a, b, c)$:

$$\frac{\partial x_t(a, b, c)}{\partial a_i} = -X_{t-(i+1)}, \quad \frac{\partial x_t(a, b, c)}{\partial b_i} = -X_{t-(i+1)} e^{-cX_{t-1}^2},$$
$$\frac{\partial x_t(a, b, c)}{\partial c} = \sum_{i=0}^{p-1} b_i X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2};$$

$$\frac{\partial^2 x_t(a, b, c)}{\partial a_i \partial a_j} = 0, \quad \frac{\partial^2 x_t(a, b, c)}{\partial a_i \partial b_j} = 0, \quad \frac{\partial^2 x_t(a, b, c)}{\partial a_i \partial c} = 0;$$

$$\frac{\partial^2 x_t(a, b, c)}{\partial b_i \partial a_j} = 0, \quad \frac{\partial^2 x_t(a, b, c)}{\partial b_i \partial b_j} = 0, \quad \frac{\partial^2 x_t(a, b, c)}{\partial b_i \partial c} = X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2};$$

$$\frac{\partial^2 x_t(a,b,c)}{\partial c \partial a_j} = 0, \ \frac{\partial^2 x_t(a,b,c)}{\partial c \partial b_j} = X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2},$$

$$\frac{\partial^2 x_t(a,b,c)}{\partial^2 c} = -\sum_{i=0}^{p-1} b_i X_{t-(i+1)} X_{t-1}^4 e^{-cX_{t-1}^2}.$$

Обозначим по очереди a_i, b_i, c за p_i и используем правило вычисления сложной производной:

$$\frac{\partial g(a,b,c)}{\partial p_i} = \frac{\partial \left(\sum_{t=p}^{n} \rho_\delta \left(x_t(a,b,c) \right) \right)}{\partial p_i} = \sum_{t=p}^{n} \frac{\partial \rho_\delta \left(x_t(a,b,c) \right)}{\partial x_t(a,b,c)} \frac{\partial x_t(a,b,c)}{\partial p_i} =$$

$$= \sum_{t=p}^{n} \rho'_\delta \left(x_t(a,b,c) \right) \frac{\partial x_t(a,b,c)}{\partial p_i};$$

$$\frac{\partial^2 g(a,b,c)}{\partial p_i \partial p_j} = \frac{\partial \left(\sum_{t=p}^{n} \rho'_\delta \left(x_t(a,b,c) \right) \frac{\partial x_t(a,b,c)}{\partial p_i} \right)}{\partial p_j} =$$

$$= \sum_{t=p}^{n} \frac{\partial \left(\rho'_\delta \left(x_t(a,b,c) \right) \frac{\partial x_t(a,b,c)}{\partial p_i} \right)}{\partial p_j} =$$

$$= \sum_{t=p}^{n} \left(\frac{\partial \rho'_\delta \left(x_t(a,b,c) \right)}{\partial p_j} \frac{\partial x_t(a,b,c)}{\partial p_i} + \rho''_\delta \left(x_t(a,b,c) \right) \frac{\partial^2 x_t(a,b,c)}{\partial p_i \partial p_j} \right) =$$

$$= \sum_{t=p}^{n} \left(\rho''_\delta \left(x_t(a,b,c) \right) \frac{\partial x_t(a,b,c)}{\partial p_i} \frac{\partial x_t(a,b,c)}{\partial p_j} + \rho'_\delta \left(x_t(a,b,c) \right) \frac{\partial^2 x_t(a,b,c)}{\partial p_i \partial p_j} \right).$$

Непосредственно вычислим первые и вторые производные первоначальной функции $g(a,b,c)$:

$$\frac{\partial g(a,b,c)}{\partial a_i} = -\sum_{t=p}^{n} \rho'_\delta \left(x_t(a,b,c) \right) X_{t-(i+1)}, \ \frac{\partial g(a,b,c)}{\partial b_i} =$$

$$= -\sum_{t=p}^{n} \rho'_\delta \left(x_t(a,b,c) \right) X_{t-(i+1)} e^{-cX_{t-1}^2},$$

$$\frac{\partial g(a,b,c)}{\partial c} = -\sum_{t=p}^{n} \rho'_\delta \left(x_t(a,b,c) \right) \sum_{i=0}^{p-1} b_i X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2};$$

$$\frac{\partial^2 g(a,b,c)}{\partial a_i \partial a_j} = \sum_{t=p}^{n} \rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} X_{t-(j+1)},$$

$$\frac{\partial^2 g(a,b,c)}{\partial a_i \partial b_j} = \sum_{t=p}^{n} \rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} X_{t-(j+1)} e^{-cX_{t-1}^2},$$

$$\frac{\partial^2 g(a,b,c)}{\partial a_i \partial c} = -\sum_{t=p}^{n} \rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} \sum_{j=0}^{p-1} b_j X_{t-(j+1)} X_{t-1}^2 e^{-cX_{t-1}^2};$$

$$\frac{\partial^2 g(a,b,c)}{\partial b_i \partial a_j} = \sum_{t=p}^{n} \rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} X_{t-(j+1)} e^{-cX_{t-1}^2},$$

$$\frac{\partial^2 g(a,b,c)}{\partial b_i \partial b_j} = \sum_{t=p}^{n} \rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} X_{t-(j+1)} e^{-2cX_{t-1}^2},$$

$$\frac{\partial^2 g(a,b,c)}{\partial b_i \partial c} = \sum_{t=p}^{n} \left(-\rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} e^{-cX_{t-1}^2} \sum_{j=0}^{p-1} b_j X_{t-(j+1)} X_{t-1}^2 e^{-cX_{t-1}^2} + \right.$$

$$\left. + \rho'_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2} \right);$$

$$\frac{\partial^2 g(a,b,c)}{\partial c \partial a_j} = -\sum_{t=p}^{n} \rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} \sum_{j=0}^{p-1} b_j X_{t-(j+1)} X_{t-1}^2 e^{-cX_{t-1}^2},$$

$$\frac{\partial^2 g(a,b,c)}{\partial c \partial b_j} =$$

$$= \sum_{t=p}^{n} \left(-\rho''_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} e^{-cX_{t-1}^2} \sum_{j=0}^{p-1} b_j X_{t-(j+1)} X_{t-1}^2 e^{-cX_{t-1}^2} + \right.$$

$$\left. + \rho'_\delta\left(x_t(a,b,c)\right) X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2} \right),$$

$$\frac{\partial^2 g(a,b,c)}{\partial c^2} =$$

$$= \sum_{t=p}^{n} \left(\rho''_\delta\left(x_t(a,b,c)\right) \left(\sum_{i=0}^{p-1} b_i X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2} \right)^2 - \right.$$

$$\left. - \rho''_\delta\left(x_t(a,b,c)\right) \left(\sum_{i=0}^{p-1} b_i X_{t-(i+1)} X_{t-1}^2 e^{-cX_{t-1}^2} \right)^2 \right).$$

Вычислим предельные значения полученных выражений первых и вторых производных. Для этого предварительно покажем, что $E\rho'_\delta\left(x_t(a,b,c)\right) = 0$. Отметим, что функция $\rho'_\delta\left(x_t(a,b,c)\right)$ является нечетной:

– $\rho'_\delta(x) = x = -(-x) = -\rho'_\delta(-x)$ на интервале $(-\delta, \delta)$;

– $\rho'_\delta(x) = \delta \operatorname{sign} x = \delta(-\operatorname{sign}(-x)) = -\delta \operatorname{sign}(-x) = -\rho'_\delta(-x)$ на остальной числовой прямой.

Математическое ожидание нечетной функции от случайной величины с четной функцией плотности вероятности на симметричном относительно нуля отрезке равно нулю [18]. Действительно, $E\rho'_\delta(x_t(a,b,c)) = 0$. Согласно [19], произведение двух стационарных и эргодических последовательностей является стационарной и эргодической последовательностью. В вычислениях внесем множитель $1/n$ в пределы для того, чтобы можно было использовать свойства центральной предельной теоремы:

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial a_i \partial a_j} = E\left[\rho''_\delta(x_t(a,b,c))\right] E\left[X_0 X_{i-j}\right],$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial a_i \partial b_j} = E\left[\rho''_\delta(x_t(a,b,c))\right] E\left[X_0 X_{i-j} e^{-cX_i^2}\right],$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial a_i \partial c} = -E\left[\rho''_\delta(x_t(a,b,c))\right] \sum_{j=0}^{p-1} b_j E\left[X_j^2 X_0^2 e^{-cX_j^2}\right];$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial b_i \partial a_j} = E\left[\rho''_\delta(x_t(a,b,c))\right] E\left[X_0 X_{i-j} e^{-cX_i^2}\right],$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial b_i \partial b_j} = E\left[\rho''_\delta(x_t(a,b,c))\right] E\left[X_0 X_{i-j} e^{-2cX_i^2}\right],$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial b_i \partial c} = -E\left[\rho''_\delta(x_t(a,b,c))\right] \sum_{j=0}^{p-1} b_j E\left[X_j^2 X_0^2 e^{-2cX_j^2}\right];$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial c \partial a_j} = -E\left[\rho''_\delta(x_t(a,b,c))\right] \sum_{j=0}^{p-1} b_j E\left[X_j^2 X_0^2 e^{-cX_j^2}\right],$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial c \partial b_j} = -E\left[\rho''_\delta(x_t(a,b,c))\right] \sum_{j=0}^{p-1} b_j E\left[X_j^2 X_0^2 e^{-2cX_j^2}\right],$$

$$\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g(a,b,c)}{\partial c^2} = E\left[\rho''_\delta(x_t(a,b,c))\right] \sum_{t=p}^{n} \left(\sum_{j=0}^{p-1} b_j^2\right) E X_0^2 X_j^4 e^{-2cX_j^2}.$$

Следовательно, $\lim\limits_{n\to\infty} B = \left(E\rho''_\delta(x_t(a,b,c))\right)K$, где K — блочная матрица, составленная из четырех подматриц размером $p\times p$, двух подвекторов размером $p\times 1$, двух подвекторов размером $1\times p$ и одного числа аналогично исходной матрице B.

Действительно, асимптотическая ковариационная матрица оценки Хьюбера в обобщенной экспоненциальной авторегрессионной модели Озаки выражается по (4).

Обобщение результатов. В пп. 1–3 вычисление асимптотической ковариационной матрицы оценки Хьюбера в модели Озаки выполнено в общем виде. В п. 4 показано, что $\lim\limits_{n\to\infty} B = \left(E\rho''_\delta\left(x_t(a,b,c)\right)\right)K.$ За счет того, что значения $E\rho''\left(x_t(a,b,c)\right)$ нигде явно не использовались, можно утверждать, что формула (4) истинна для любой четной функции оценки, первая производная которой будет нечетной.

Следовательно, асимптотическая ковариационная матрица М-оценки $\rho\left(x_t(a,b,c)\right)$ в обобщенной экспоненциальной модели может быть вычислена по формуле:

$$\frac{E\left[\left(\rho'\left(x_t(a,b,c)\right)\right)^2\right]}{\left[E\rho''\left(x_t(a,b,c)\right)\right]^2}K^{-1}. \tag{5}$$

В качестве примеров функций М-оценок $\rho(x)$ могут быть использованы квадратичная функция

$$\rho(x) = x^2, \tag{6}$$

функция абсолютных значений

$$\rho(x) = |x|, \tag{7}$$

рассмотренная функция Хьюбера (1) и ее аналоги — функция Коши

$$\rho(x) = \frac{\delta}{2}\log\left(1+\left(\frac{x}{\delta}\right)^2\right), \tag{8}$$

функция Джемана — МакКлюра

$$\rho(x) = \frac{x^2}{2}\frac{1}{1+x^2}, \tag{9}$$

функция Уэлша

$$\rho(x) = \frac{\delta^2}{2}\left(1-e^{-(x/\delta)^2}\right), \tag{10}$$

функция Тьюки

$$\rho(x) = \begin{cases} \dfrac{\delta^2}{6}\left(1-\left(1-(x/\delta)^2\right)^3\right), \text{ если } |x|<\delta; \\ \delta^2/6, \text{ иначе,} \end{cases} \tag{11}$$

функция Эндрюса

$$\rho(x) = \begin{cases} \delta^2 \left(1 - \cos(x / \delta)\right), \text{ если } |x| < \delta\pi; \\ 2\delta^2, \text{ иначе.} \end{cases} \qquad (12)$$

Выводы. Получена формула вычисления асимптотической ковариационной матрицы для заданной функции М-оценки в обобщенной модели Озаки (5). Ее частным случаем является формула для оценки функции Хьюбера (4).

В следующих работах планируется провести масштабное сравнение эффективности оценок (1), (6)–(12) для различных распределений обновляющего процесса с использованием методов компьютерного моделирования, а также построить алгоритмы вычисления асимптотических ковариационных матриц в модели Озаки для непараметрических оценок [20].

ЛИТЕРАТУРА

[1] Hastie T., Tibshirani R., Friedman Jr. The elements of statistical learning. Springer, 2009.

[2] Ozaki T. Non-linear phenomena and time series models. *43 Session of the International Statistical Institute*, 1981.

[3] Hesamian Gh., Torkian F., Johannssen A., et al. An exponential autoregressive time series model for complex data. *Mathematics*, 2023, vol. 11, iss. 19, art. 4022.
DOI: https://doi.org/10.3390/math11194022

[4] Xu H., Wan L., Ding F., et al. Fitting the exponential autoregressive model through recursive search. *J. Frankl. Inst.*, 2019, vol. 356, iss. 11, pp. 5801–5818.
DOI: https://doi.org/10.1016/j.jfranklin.2019.03.016

[5] Harvill J.L., Ray B.K. Functional coefficient autoregressive models for vector time series. *Comput. Stat. Data An.*, 2006, vol. 50, iss. 12, pp. 3547–3566.
DOI: https://doi.org/10.1016/j.csda.2005.07.016

[6] Patrick D.J., Harvill J.L., Hansen C.W. A semiparametric spatio-temporal model for solar irradiance data. *Renew. Energy*, 2016, vol. 87-1, pp. 15–30.
DOI: https://doi.org/10.1016/j.renene.2015.10.001

[7] Kulik R., Bilayi-Biakana Cl., Soulier Ph. Statistical inference for heavy tailed series with extremal independence. *Extremes*, 2020, vol. 23, no. 1, pp. 1–33.
DOI: https://doi.org/10.1007/s10687-019-00365-z

[8] Galvao A., Kato K. Estimation and inference for linear panel data models under misspecification when both *n* and *T* are large. *J. Bus. Econ. Stat.*, 2014, vol. 32, iss. 2, pp. 285–309. DOI: https://doi.org/10.1080/07350015.2013.875473

[9] Sánchez C.J., Torvisco J.A.F., Rodriguez-Arias M. TAC method for fitting exponential autoregressive models and others: applications in economy and finance. *Mathematics*, 2021, vol. 9, iss. 8, pp. 862–882. DOI: https://doi.org/10.3390/math9080862

[10] Parker P.A., Holan S.H., Ravishanker N. Nonlinear time series classification using bispectrum-based deep convolutional neural networks. *Appl. Stoch. Models Bus. Ind.*, 2020, vol. 36, iss. 5, pp. 877–890. DOI: https://doi.org/10.1002/asmb.2536

[11] van de Geer S.A. Empirical processes in M-estimation. Cambridge Univ. Press, 2009.

[12] Chan K.S., Tong H. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.*, 1985, vol. 17, iss. 3, pp. 666–668. DOI: https://doi.org/10.2307/1427125

[13] Häusler E., Luschgy H. Stable convergence and stable limit theorems. *Probability Theory and Stochastic Modelling*, vol. 74. Cham, Springer, 2015.
DOI: https://doi.org/10.1007/978-3-319-18329-9

[14] Magnus J.R., Neudecker H. Matrix differential calculus with applications in statistics and econometrics. Wiley, 2019.

[15] Goryainov V.B. Least-modules estimates for spatial autoregression coefficients. *J. Comput. Syst. Sci. Int.*, 2011, vol. 50, no. 4, pp. 565–572.
DOI: https://doi.org/10.1134/S1064230711040101

[16] Huber P.J., Ronchetti E.M. Robust statistics. Wiley, 2009.

[17] Andersen P.K., Gill R.D. Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 1982, vol. 10, iss. 4, pp. 1100–1120.
DOI: https://doi.org/10.1214/aos/1176345976

[18] Billingsley P. Convergence of probability measures. Willey, 1999.

[19] White H. Asymptotic theory for econometricians. Academic Press, 2014.

[20] Corder G.W., Foreman D.I. Nonparametric statistics: a step-by-step approach. Willey, 2014.

Горяинов Владимир Борисович — д-р физ.-мат. наук, профессор кафедры «Математическое моделирование» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Масягин Михаил Михайлович — аспирант кафедры «Математическое моделирование» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Просьба ссылаться на эту статью следующим образом:

Горяинов В.Б., Масягин М.М. Вычисление асимптотической ковариационной матрицы оценки Хьюбера в обобщенной экспоненциальной авторегрессионной модели. *Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки*, 2025, № 4 (121), с. 23–39. EDN: QCRHHH

CALCULATION OF THE ASYMPTOTIC COVARIANCE MATRIX OF THE GENERALIZED EXPONENTIAL AUTOREGRESSIVE MODEL FOR THE HUBER LOSS FUNCTION

V.B. Goryainov vb-goryainov@bmstu.ru
M.M. Masyagin masyagin@bmstu.ru

BMSTU, Moscow, Russian Federation

Abstract

When solving regression analysis problems, the functions of least squares and least modules are most often used as a loss function. Nevertheless, they are not without drawbacks. The first function is extremely sensitive to outliers in the training data, and the second is not differentiable at zero, which complicates its use with a wide class of optimization methods. In order to offset these shortcomings, several functions have been developed that combine the capabilities of least squares and least modules estimates, for example, the Huber function. Up to a certain threshold, they behave similarly to a quadratic function; after that, they behave similarly to a modulus function. Thus, they are both differentiable across the entire definition area and insensitive to outliers (exceeding a given threshold). This paper presents the calculation of the asymptotic covariance matrix of a generalized exponential autoregressive model, the Ozaki model, for estimating the Huber function. It includes generalizing the resulting formula to a broader class of estimates, such as the least squares estimate, the least modules estimate, and other estimates combined together into a set of M-estimates

Keywords

Generalized exponential autoregressive model, Huber loss function, asymptotic covariance matrix, Taylor series expansion

Received 06.06.2024
Accepted 04.02.2025

REFERENCES

[1] Hastie T., Tibshirani R., Friedman Jr. The elements of statistical learning. Springer, 2009.

[2] Ozaki T. Non-linear phenomena and time series models. *43 Session of the International Statistical Institute*, 1981.

[3] Hesamian Gh., Torkian F., Johannssen A., et al. An exponential autoregressive time series model for complex data. *Mathematics*, 2023, vol. 11, iss. 19, art. 4022.
DOI: https://doi.org/10.3390/math11194022

[4] Xu H., Wan L., Ding F., et al. Fitting the exponential autoregressive model through recursive search. *J. Frankl. Inst.*, 2019, vol. 356, iss. 11, pp. 5801–5818.
DOI: https://doi.org/10.1016/j.jfranklin.2019.03.016

[5] Harvill J.L., Ray B.K. Functional coefficient autoregressive models for vector time series. *Comput. Stat. Data An.*, 2006, vol. 50, iss. 12, pp. 3547–3566.
DOI: https://doi.org/10.1016/j.csda.2005.07.016

[6] Patrick D.J., Harvill J.L., Hansen C.W. A semiparametric spatio-temporal model for solar irradiance data. *Renew. Energy*, 2016, vol. 87-1, pp. 15–30.
DOI: https://doi.org/10.1016/j.renene.2015.10.001

[7] Kulik R., Bilayi-Biakana Cl., Soulier Ph. Statistical inference for heavy tailed series with extremal independence. *Extremes*, 2020, vol. 23, no. 1, pp. 1–33.
DOI: https://doi.org/10.1007/s10687-019-00365-z

[8] Galvao A., Kato K. Estimation and inference for linear panel data models under misspecification when both *n* and *T* are large. *J. Bus. Econ. Stat.*, 2014, vol. 32, iss. 2, pp. 285–309. DOI: https://doi.org/10.1080/07350015.2013.875473

[9] Sánchez C.J., Torvisco J.A.F., Rodriguez-Arias M. TAC method for fitting exponential autoregressive models and others: applications in economy and finance. *Mathematics*, 2021, vol. 9, iss. 8, pp. 862–882. DOI: https://doi.org/10.3390/math9080862

[10] Parker P.A., Holan S.H., Ravishanker N. Nonlinear time series classification using bispectrum-based deep convolutional neural networks. *Appl. Stoch. Models Bus. Ind.*, 2020, vol. 36, iss. 5, pp. 877–890. DOI: https://doi.org/10.1002/asmb.2536

[11] van de Geer S.A. Empirical processes in M-estimation. Cambridge Univ. Press, 2009.

[12] Chan K.S., Tong H. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.*, 1985, vol. 17, iss. 3, pp. 666–668. DOI: https://doi.org/10.2307/1427125

[13] Häusler E., Luschgy H. Stable convergence and stable limit theorems. *Probability Theory and Stochastic Modelling*, vol. 74. Cham, Springer, 2015.
DOI: https://doi.org/10.1007/978-3-319-18329-9

[14] Magnus J.R., Neudecker H. Matrix differential calculus with applications in statistics and econometrics. Wiley, 2019.

[15] Goryainov V.B. Least-modules estimates for spatial autoregression coefficients. *J. Comput. Syst. Sci. Int.*, 2011, vol. 50, no. 4, pp. 565–572.
DOI: https://doi.org/10.1134/S1064230711040101

[16] Huber P.J., Ronchetti E.M. Robust statistics. Wiley, 2009.

[17] Andersen P.K., Gill R.D. Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 1982, vol. 10, iss. 4, pp. 1100–1120.
DOI: https://doi.org/10.1214/aos/1176345976

[18] Billingsley P. Convergence of probability measures. Willey, 1999.

[19] White H. Asymptotic theory for econometricians. Academic Press, 2014.

[20] Corder G.W., Foreman D.I. Nonparametric statistics: a step-by-step approach. Willey, 2014.

Goryainov V.B. — Dr. Sc. (Phys.-Math.), Professor, Department of Mathematical Simulation, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Masyagin M.M. — Post-Graduate Student, Department of Mathematical Simulation, BMSTU (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Goryainov V.B., Masyagin M.M. Calculation of the asymptotic covariance matrix of the generalized exponential autoregressive model for the Huber loss function. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2025, no. 4 (121), pp. 23–39 (in Russ.). EDN: QCRHHH