ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы.

Непараметрические методы статистики - методы математической статистики, не предполагающие знания функционального вида генеральных распределений.

Одна из задач многомерного непараметрического анализа - задача многомерной линейной регрессии:

$$\boldsymbol{y}_i = B_0^T \boldsymbol{x}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

где $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iq})^T$ и $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$, $i = 1, 2, \ldots, n$ - значения отклика и фактора, случайные ошибки $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n$ - независимые одинаково распределенные $(q \times 1)$-векторы; задача - оценить неизвестную $(p \times q)$-матрицу регрессионных коэффициентов B_0.

Классическим предположением о случайных ошибках $\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n$ является отсутствие в них систематической составляющей, т.е. $\mathbf{E}\boldsymbol{\varepsilon}_i = 0$. В диссертационной работе предположение об отсутствии систематических ошибок принято в форме

$$\boldsymbol{\varepsilon}_i \overset{d}{=} -\boldsymbol{\varepsilon}_i. \tag{2}$$

Условие (2) о симметричности распределения ошибок относительно нуля можно рассматривать как ослабление другого классического предположения о нормально распределенных ошибках.

Наиболее известным методом решения задачи (1) является метод наименьших квадратов (МНК). В предположении, что $\mathbf{E}\varepsilon_i = 0$, этот метод дает для искомой матрицы B_0 несмещенную оценку. Если случайные ошибки имеют нормальное распределение (с нулевым средним), МНК-оценка в определенном смысле оптимальна (обладает наименьшей матрицей ковариаций). Другим важным свойством МНК-оценки является ее аффинная эквивариантность. Это означает, что при аффинных преобразованиях переменных (как факторов $\boldsymbol{x}$, так и откликов $\boldsymbol{y}$) оценка матрицы B_0 изменяется соответствующим образом. Эквивариантные статистические правила нужны тогда, когда статистическое решение основывается на всем пространстве признаков, а не на каждом признаке в отдельности.

Хорошо известно, что оценки наименьших квадратов крайне чувствительны к выбросам — единственное постороннее наблюдение может произвести на оценку неограниченное влияние. Робастные методы математической статистики в трудах Хьюбера[1] и Хампеля[2] возникли как попытка преодолеть этот недостаток метода наименьших квадратов.

Разработка робастных методов оценивания для многомерных и многофакторных линейных моделей привлекает внимание многих авторов. Пури и Сен[3] предложили покоординатные ранговые оценки. Рао[4] предложил использовать одномерный метод наименьших модулей отдельно для каждой координаты отклика. Конкер и Портной[5] обобщили метод Рао и предложили робастные М-оценки, заменив модуль на произвольную функцию. Оценка, предложенная Баи и др[6]. минимизирует среднее евклидовых норм остатков. Все эти методы, однако, не являются аффинно-эквивариантными. Руссиу и др[7]. в случае случайного фактора предложили робастную аффинно-эквивариантную оценку матрицы регрессионных коэффициентов, основанную на робастной оценке ковариационной матрицы вектора $\boldsymbol{z}_1 = (\boldsymbol{x}_1^T, \boldsymbol{y}_1^T)^T$ (но не исследовали ее асимптотические свойства). Оллила и др[8]. предложили аналогичный подход, использовав вместо оценки ковариационной матрицы Руссиу выборочную знаковую ковариационную матрицу вектора $\boldsymbol{z}_1$. Их оценка аффинно-эквивариантна, однако не робастна, хотя и более устойчива

[1] Хьюбер П. (1984) *Робастность в статистике.* Мир, Москва.

[2] Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

[3] Puri M.L. and Sen P.K. (1985) *Nonparametric methods in general linear models.* New York: Wiley.

[4] Rao C.R. (1988) Methodology based on L_1-norm in statistical inference. *Sankhya A*, **50**, 289 - 313.

[5] Koenker R. and Portnoy S. (1990) M-estimation of multivariate regressions. *J. Am. Statist. Ass.*, **85**, 1060 - 1068.

[6] Bai Z.D., Chen N.R., Miao B.Q. and Rao C.R. (1990) Asymptotic theory of least distances estimate in multivariate linear models. *Statistics*, **21**, 503 - 519.

[7] Rousseeuw P.J., Van Driessen K., Van Aelst S. and Agullo J. (2004) Robust multivariate regression. *Technometrics,* **46**, 293 - 305.

[8] Ollila E., Oja, H. and Hettmansperger T.P. (2002) Estimates of regression coefficients based on the sign covariance matrix. *J. R. Statist. Soc. Ser. B*, **64**, part 3, 447 - 466.

к выбросам, чем МНК-оценка.

Цель работы.

Построение робастных аффинно-эквивариантных непараметрических оценок и проверка гипотез для задачи многомерной линейной регрессии. Исследование асимптотических свойств предложенных оценок и статистических критериев.

Научная новизна.

Основные результаты работы являются новыми и состоят в следующем.

1. Предложены четыре новые робастные аффинно-эквивариантные оценки $\hat{B}_n$, $\tilde{B}_n$, $\hat{B}'_n$, $\tilde{B}'_n$ матрицы регрессионных коэффициентов для задачи многомерной линейной регрессии. Для этих оценок получены условия состоятельности и асимптотической нормальности, найдены функции влияния.

2. Для проверки простых гипотез о точных значениях матрицы регрессионных коэффициентов (в частности, о равенстве этой матрицы нулю, что означает независимость отклика от значения фактора) для задачи многомерной линейной регрессии предложены две новые тестовые статистики $\boldsymbol{T}_n$ и $\boldsymbol{T}'_n$, изучены их предельные распределения как при нулевой гипотезе, так и при последовательности близких альтернатив. Построены состоятельные оценки ковариационных матриц этих тестовых статистик при нулевой гипотезе.

3. Для задачи непараметрической проверки гипотезы о равенстве нулю матрицы регрессионных коэффициентов в многомерной линейной регрессионной модели рассмотрены две новые аффинно-инвариантные асимптотически свободные от исходных распределений данных тестовые статистики ψ_n и ψ'_n, найдены их предельные распределения при нулевой гипотезе и близких альтернативах, а также асимптотическая эффективность по Питману соответствующих критериев.

Методы исследования.

Методика исследования основана на общих методах теории вероятностей, математического анализа и математической статистики. Широко используется теория U-статистик.

Теоретическая и практическая ценность.

Работа носит теоретический характер. Предложенные в работе критерии и оценки могут быть использованы для статистической обработки регрессионного эксперимента. Рекомендуется их использование в задачах, где важно свойство аффинной инвариантности и где распределение случайных ошибок может иметь "тяжелые хвосты". Оценки $\hat{B}_n$, $\tilde{B}_n$ и тестовые статистики $\boldsymbol{T}_n$, ψ_n рекомендуется использовать в условиях активного эксперимента (когда экспериментатор сам выбирает план эксперимента); оценки $\hat{B}'_n$, $\tilde{B}'_n$ и тестовые статистики $\boldsymbol{T}'_n$, ψ'_n могут быть использованы также и в пассивном эксперименте.

Апробация работы.

Результаты диссертации докладывались на Большом семинаре кафедры теории вероятностей МГУ под руководством член-корр. РАН, проф. А.Н. Ширяева (2006 г.), на международной конференции "International Conference on Robust Statistics (ICORS) - 2005", Ювяскюля, Финляндия (2005 г.), на семинаре кафедры теории вероятностей МГУ "Непараметрическая статистика и временные ряды"под руководством проф. Ю.Н. Тюрина, проф. В.Н. Тутубалина, доц. М.В. Болдина (2005 г.), на семинаре "Многомерный статистический анализ и вероятностное моделирование реальных процессов"под руководством проф. С.А. Айвазяна в ЦЭМИ РАН (2006 г.), на семинаре под руководством профессора Х. Оя в университете Тампере, Финляндия (2004 г.).

Публикации.

Результаты диссертации опубликованы в 4 работах, список которых приведен в конце автореферата.

Структура диссертации.

Диссертация состоит из введения, двух глав (разбитых на разделы) и списка литературы, насчитывающего 35 наименований. Общий объем диссертации — 119 страниц.

КРАТКОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во **введении** приведен краткий обзор по тематике работы, изложены цели, методы и результаты исследования.

В **первой главе** исследуется задача робастного непараметрического оценивания для модели многомерной линейной регрессии. Общим для предложенных методов решения этой задачи является то, что все они основаны на понятии выборочной медианы Оя[9], аффинно-эквивариантны и, при некоторых условиях на распределения данных, обладают свойством асимптотической нормальности.

Сформулируем основные определения. Пусть $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$ - одинаково распределенные k-мерные случайные векторы с функцией распределения $F(\boldsymbol{x})$ и пусть $\boldsymbol{\theta} \in I\!R^k$. Обозначим $V(\boldsymbol{\theta}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k)$ - объем k-мерного симплекса, вершинами которого являются точки с координатами $\boldsymbol{\theta}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k$. Предположим, что для всех $\boldsymbol{\theta} \in I\!R^k$ определено математическое ожидание $\mathbf{E}V(\boldsymbol{\theta}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k)$ и обозначим $U(\boldsymbol{\theta}) = \mathbf{E}V(\boldsymbol{\theta}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k)$. Тогда медиана Оя распределения $F(\boldsymbol{x})$ - множество точек Θ_0, на котором функция $U(\boldsymbol{\theta})$ достигает свое наименьшее значение. Другими словами, для всех $\boldsymbol{\theta}_0 \in \Theta_0$ и $\boldsymbol{\theta} \in I\!R^k$ $U(\boldsymbol{\theta}_0) \leq U(\boldsymbol{\theta})$.

В случае, когда $\boldsymbol{\xi}_i - \boldsymbol{\theta}_* \stackrel{d}{=} -(\boldsymbol{\xi}_i - \boldsymbol{\theta}_*)$ медиана Оя совпадает с центром симметрии $\boldsymbol{\theta}_*$.

В одномерном случае $U(\theta) = \mathbf{E}|\xi_1 - \theta|$ и медиана Оя совпадает с обычной медианой распределения $F(x)$.

Пусть теперь $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ - произвольная выборка из распределения $F(\boldsymbol{x})$. Выборочной медианой Оя $\hat{\boldsymbol{\theta}}_n$ называется медиана Оя эмпирического распре-

[9]Oja H. (1983), Descriptive Statistics for Multivariate Distributions. *Stat. Probab. Lett.*, **1**, 327-332.

деления $F_n(\boldsymbol{x})$ выборки $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$. Другими словами,

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in I\!R^k} U_n(\boldsymbol{\theta}), \text{ где } U_n(\boldsymbol{\theta}) = \left(C_n^k\right)^{-1} \sum_{i_1<\ldots<i_k} V(\boldsymbol{\theta}, \boldsymbol{\xi}_{i_1}, \ldots, \boldsymbol{\xi}_{i_k}).$$

Четыре оценки, представленные в главе 1, являются многомерными обобщениями медианной оценки Тэйла[10] параметра наклона для модели простой линейной регрессии. Пусть $y_i = a + bx_i + \varepsilon_i, \quad i = 1, \ldots, n$. Медианная оценка Тэйла есть

$$\hat{\beta}_n = \mathrm{med}\left\{\frac{y_{i_2} - y_{i_1}}{x_{i_2} - x_{i_1}},\ 1 \le i_1 < i_2 \le n,\ x_{i_1} \ne x_{i_2}\right\}.$$

Заметим, что при $x_{i_1} \ne x_{i_2}$ величина $\dfrac{y_{i_2} - y_{i_1}}{x_{i_2} - x_{i_1}}$ есть вторая координата вектора

$$\boldsymbol{b}(i_1, i_2) = \begin{pmatrix} 1 & x_{i_1} \\ 1 & x_{i_2} \end{pmatrix}^{-1} \begin{pmatrix} y_{i_1} \\ y_{i_2} \end{pmatrix} = \begin{pmatrix} \dfrac{x_{i_2}y_{i_1} - x_{i_1}y_{i_2}}{x_{i_2} - x_{i_1}} \\ \dfrac{y_{i_2} - y_{i_1}}{x_{i_2} - x_{i_1}} \end{pmatrix}.$$

Обобщая подход Тэйла, для одновременного оценивания параметров a, b можно рассмотреть медиану Оя векторов $\boldsymbol{b}(i_1, i_2)$:

$$\hat{\boldsymbol{\beta}}_n = \mathrm{med}_{\mathrm{Oja}}\left\{\boldsymbol{b}(i_1, i_2),\ 1 \le i_1 < i_2 \le n,\ x_{i_1} \ne x_{i_2}\right\} = \arg\min_{\boldsymbol{\beta}\in I\!R^2} D_n(\boldsymbol{\beta}), \text{ где}$$

$$D_n(\boldsymbol{\beta}) = \frac{1}{C^2_{C^2_n}} \sum V(\boldsymbol{\beta}, \boldsymbol{b}(i_1, i_2), \boldsymbol{b}(i_3, i_4)).$$

Здесь $V(\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ - площадь треугольника с вершинами в точках $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$, и сумма берется по всем возможным парам векторов $\{\boldsymbol{b}(i_1, i_2), \boldsymbol{b}(i_3, i_4)\}$, $1 \le i_1 < i_2 \le n$, $1 \le i_3 < i_4 \le n$.

Перейдем теперь к многомерной многофакторной регрессионной модели (1). Определим понятие элементарной регрессии.

Определение 1. *Пусть* $I = \{i_1, i_2, \ldots, i_p\}$ *— произвольное подмножество размера* p *исходных наблюдений. Пусть, далее,* $Y(I)$ *есть* $(p \times q)$*-матрица* $(\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_2}, \ldots, \boldsymbol{y}_{i_p})^T$ *и* $X(I)$ *-* $(p \times p)$*-матрица* $(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_p})^T$*. Если* $rank(X(I)) = p$*, то* $(p \times q)-$*матрицу* $B(I) = X(I)^{-1}Y(I)$ *будем называть элементарной регрессией. Если же* $rank(X(I)) < p$*, то элементарная регрессия не определена и мы будем называть ее вырожденной.*

[10]Theil H. (1950) A rank-invariant method of linear and polynomial regression analysis (Parts 1 - 3). *Ned. Akad. Wetensch. Proc. Ser. A*, **53**, 386 - 392, 521 - 525, 1397 - 1412.

Следует упомянуть, что понятие элементарных регрессий было введено и применялось ранее (под разными именами) для модели множественной линейной регрессии с одномерным откликом ($q = 1$). В этом случае каждая невырожденная элементарная регрессия есть ($p \times 1$)-вектор. Было показано[11], что оценка методом наименьших квадратов есть взвешенное среднее элементарных регрессий:

$$\hat{\boldsymbol{B}}_{\text{МНК}} = \sum_I \omega(I)\boldsymbol{B}(I), \text{ где веса равны } \omega(I) = \frac{\det\bigl(X(I)^T X(I)\bigr)}{\det(X^T X)},$$

и $X = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^T$ обозначает матрицу факторов. Бассет[12] показал, что L_1-оценка, минимизирующая сумму модулей остатков $e_i(\boldsymbol{B}) = y_i - \boldsymbol{B}^T x_i, \ i = 1, ..., n$, либо совпадает с одной из элементарных регрессий $\boldsymbol{B}(I)$, либо принадлежит выпуклой оболочке не более чем $p + 1$ векторов $\boldsymbol{B}(I)$. Хоукинс[13] предложил основанный на элементарных регрессиях алгоритм для оценок, минимизирующих критериальную функцию от остатков $e_i(\boldsymbol{B}), \ i = 1, ..., n$. Его алгоритм состоит в вычислении совокупности $\{e_i(\boldsymbol{B}(I)), i = 1, \ldots, n\}$ для каждого возможного p-подмножества I и выборе в качестве искомой оценки той элементарной регрессии $\boldsymbol{B}^*(I)$, для которой совокупность $\{e_i(\boldsymbol{B}^*(I)), i = 1, \ldots, n\}$ минимизирует критерий. Методы, основанные на элементарных регрессиях исследуются также в работах Конкера и Бассета[14], Сигеля[15], Хоукинса и др[16]., а также Руссиу и Лероя[17].

Для построения многомерных модификаций оценки Тэйла в модели (1) превратим сначала каждую невырожденную элементарную регрессию в одностолбцовый вектор. Определим для этого операцию **vec**.

Определение 2. *Операция* ***vec*** *преобразует* $(p \times q)$*-матрицу в* (pq)*-вектор,*

[11] Sheynin O.B. (1973) R.J. Boscovich's work on probability. *Arch. Hist. Exact Sci.*, **9**, 306–324.

[12] Bassett G.W. (1988) A p-Subset property of L_1 and regression quantile estimates. *Computational Statistics and Data Analysis*, **6**, 297 - 304.

[13] Hawkins D.M. (1993) The accuracy of elemental set approximations for regression. *J. Am. Stat. Assoc.*, **88**, 580 - 589.

[14] Koenker R. and Bassett G.W. (1978) Regression quantiles. *Econometrica*, **46**, 33 - 50.

[15] Siegel A.F. (1982) Robust regression using repeated medians. *Biometrica*, **69**, 242 - 244.

[16] Hawkins D.M., Bradu D. and Kass G.V. (1984) Location of several outliers in multiple regression data using elemental sets. *Technometrics*, **26**, 197 - 208.

[17] Rousseeuw P.J. and Leroy A. (1987) *Robust Regression and Outlier Detection*. New York: Wiley.

помещая последовательно столбцы матрицы друг под другом:

$$\boldsymbol{vec}(B) = \boldsymbol{vec}\left((\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_q)\right) = \left(\boldsymbol{b}_1^T \boldsymbol{b}_2^T \ldots \boldsymbol{b}_q^T\right)^T.$$

Из свойств операции **vec** упомянем следующее:

$$\mathbf{vec}(ABC) = \left(C^T \otimes A\right) \mathbf{vec}(B),$$

где $\otimes$ - кронекеровское произведение матриц.

Определение 3. *Векторизованная невырожденная элементарная регрессия есть* $\boldsymbol{b}(I) = \boldsymbol{vec}(B(I)) = (I_{q\times q} \otimes X(I)^{-1})\, \boldsymbol{vec}(Y(I))$, *где* $I_{q\times q}$ *- единичная* $(q \times q)$*-матрица.*

Обозначим далее $\boldsymbol{\beta}_0 = \mathbf{vec}(B_0)$. Заметим, что совокупность векторизованных невырожденных элементарных регрессий позволяет перейти от задачи регрессии к многомерной задаче положения. В самом деле, вектор $\boldsymbol{\beta}_0$ может быть интерпретирован как центр распределения векторов $\boldsymbol{b}(I)$. Поэтому для аффинно-эквивариантного оценивания вектора $\boldsymbol{\beta}_0$ (а значит, и матрицы B_0) можно использовать любую аффинно-эквивариантную многомерную оценку положения для векторов $\boldsymbol{b}(I)$ в $\mathbb{R}^{pq}$.

Обобщая подход Тэйла, в **разделе 1.1** рассмотрим две оценки $\hat{B}_n = \mathbf{vec}^{-1}(\hat{\boldsymbol{\beta}}_n)$ и $\tilde{B}_n = \mathbf{vec}^{-1}(\tilde{\boldsymbol{\beta}}_n)$ (здесь $\mathbf{vec}^{-1}$ обозначает операцию, обратную к операции **vec**). Вектор $\tilde{\boldsymbol{\beta}}_n$ есть выборочная медиана Оя совокупности векторизованных невырожденных элементарных регрессий $\{\boldsymbol{b}(I)\}$. Вектор $\hat{\boldsymbol{\beta}}_n$ - немного модифицированная выборочная медиана Оя (для ослабления условий при которых выполнены асимптотические свойства) совокупности векторизованных невырожденных элементарных регрессий $\{\boldsymbol{b}(I)\}$.

Прежде чем сформулировать определения оценок, введем обозначения. Пусть, как и ранее, $I, I_1, I_2, ...$ обозначают различные подмножества размера p (p-подмножества) множества $1, 2, ..., n$. Обозначим $\mathcal{I} = \{\{I_1, \ldots, I_{pq}\}\}$, и пусть

$$\mathcal{I}_p = \left\{\{I_1, \ldots, I_{pq}\} \in \mathcal{I} : |I_1 \cup ... \cup I_{pq}| = p^2 q\right\}$$

есть множество из наборов pq p-подмножеств, таких, что все p^2q индексов различны. Очевидно, что $|\mathcal{I}| = C^{pq}_{C_n^p}$, $\quad |\mathcal{I}_p| = \dfrac{n!}{(pq)!(p!)^{pq}(n-p^2q)!}$.

Назовем элемент $\{I_1, \ldots, I_{pq}\}$ вырожденным, если хотя бы одно из входящих в него элементарных подмножеств вырождено. Пусть функция $\tau(I_1, ..., I_{pq})$ является индикатором этого; она равна 0, если элемент $\{I_1, \ldots, I_{pq}\}$ вырожден, и 1 иначе. Причем в вырожденном случае для любой функции $f(I_1, \ldots, I_{pq})$, принимающей значения из $I\!R \cup \{+\infty\} \cup \{-\infty\}$, будем считать, что $f(I_1, \ldots, I_{pq})\tau(I_1, ..., I_{pq}) = 0$.

Пусть, далее, $V(\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{k+1})$ обозначает объем k-мерного симплекса с вершинами $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{k+1} \in I\!R^k$.

Рассмотрим две целевые функции:

$$D_n(\boldsymbol{\beta}) = \operatorname{ave}_{\mathcal{I}}\Big\{V(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq})\Big\} \text{ и}$$

$$U_n(\boldsymbol{\beta}) = \operatorname{ave}_{\mathcal{I}_p}\Big\{V(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq})\Big\},$$

где средние берутся по всем элементам $\{I_1, \ldots, I_{pq}\}$ из $\mathcal{I}$ и $\mathcal{I}_p$, соответственно.

Определение 4. *Оценка $\tilde{B}_n$ матрицы коэффициентов B_0 определяется следующим образом:*

$$\tilde{B}_n = \boldsymbol{vec}^{-1}(\tilde{\boldsymbol{\beta}}_n), \quad \textit{где } \tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta} \in I\!R^{pq}} D_n(\boldsymbol{\beta}).$$

Определение 5. *Оценку $\hat{B}_n$ матрицы коэффициентов B_0 определим следующей формулой:*

$$\hat{B}_n = \boldsymbol{vec}^{-1}(\hat{\boldsymbol{\beta}}_n), \quad \textit{где } \hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta} \in I\!R^{pq}} U_n(\boldsymbol{\beta}).$$

Оценки $\hat{B}_n$ и $\tilde{B}_n$ X-аффинно, Y-аффинно и регрессионно эквивариантны в смысле следующего определения.

Определение 6. *Пусть $\hat{B}(X, Y)$ обозначает оценку матрицы регрессионных коэффициентов, вычисленную по матрице факторов $X = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^T$ и матрице откликов $Y = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)^T$, удовлетворяющих модели (1). Оценка $\hat{B}(X, Y)$ X-аффинно эквивариантна, если $\hat{B}(XA, Y) = A^{-1}\hat{B}(X, Y)$, Y-аффинно эквивариантна, если $\hat{B}(X, YC) = \hat{B}(X, Y)C$ и регрессионно эквивариантна, если $\hat{B}(X, Y + XD) = \hat{B}(X, Y) + D$ для любых невырожденных*

$(p\times p)$-матрицы A, $(q\times q)$-матрицы C и $(p\times q)$-матрицы D, соответственно.

При выполнении некоторых условий на распределения данных (см. теоремы 1, 2, 3) исследуемые оценки состоятельны, асимптотически нормальны и имеют ограниченные функции влияния. Прежде чем сформулировать эти условия, введем дополнительные обозначения.

Пусть $I_k = \{(k-1)p+1,\dots,kp\}$, $k = 1,\dots,pq$. Предположим, что для всех $\boldsymbol{\beta} \in I\!\!R^{pq}$ существует математическое ожидание $\mathbf{E}\Big(V(\boldsymbol{\beta},\boldsymbol{b}(I_1),\dots,\boldsymbol{b}(I_{pq}))\tau(I_1,...,I_{pq})\Big)$ и обозначим за

$$U(\boldsymbol{\beta}) = \mathbf{E}\Big(V(\boldsymbol{\beta},\boldsymbol{b}(I_1),\dots,\boldsymbol{b}(I_{pq}))\tau(I_1,...,I_{pq})\Big)$$

теоретическую целевую функцию.

Заметим, что $U(\boldsymbol{\beta})$ - непрерывная ограниченная снизу выпуклая функция. Пусть множество, где функция $U(\boldsymbol{\beta})$ достигает своего минимума непусто, и $\boldsymbol{\beta}^* \in I\!\!R^{pq}$ - одна из точек этого множества.

Далее, пусть вектор $\boldsymbol{d}(I_1,\dots,I_{pq}) = (d_1(I_1,\dots,I_{pq}),\dots,d_{pq}(I_1,\dots,I_{pq}))^T$ и скаляр $d_0(I_1,\dots,I_{pq})$ определяются из следующего разложения:

$$V(\boldsymbol{\beta},\boldsymbol{b}(I_1),\dots,\boldsymbol{b}(I_{pq}))\tau(I_1,...,I_{pq}) =$$
$$= \frac{1}{(pq)!}\left|\det\begin{pmatrix} 1 & 1 & \dots & 1 \\ \boldsymbol{\beta} & \boldsymbol{b}(I_1) & \dots & \boldsymbol{b}(I_{pq})\end{pmatrix}\right|\tau(I_1,...,I_{pq}) =$$
$$= |d_0(I_1,\dots,I_{pq}) + \boldsymbol{d}^T(I_1,\dots,I_{pq})\boldsymbol{\beta}|.$$

Или, более подробно,

$$d_0(I_1,\dots,I_{pq}) = \frac{1}{(pq)!}\det\Big(\boldsymbol{b}(I_1)...\boldsymbol{b}(I_{pq})\Big)\tau(I_1,...,I_{pq}),$$

$$d_s(I_1,\dots,I_{pq}) = \frac{1}{(pq)!}c_s(I_1,\dots,I_{pq})\tau(I_1,...,I_{pq}),$$

где $c_s(I_1,\dots,I_{pq})$ - алгебраическое дополнение к β_s в вышеприведенной матрице, $s = 1,...,pq$. Пусть теперь $\boldsymbol{z}_1 = (\boldsymbol{x}_1^T,\boldsymbol{y}_1^T)^T$ и определим случайный вектор

$$\boldsymbol{\Lambda}(\boldsymbol{z}_1) = \mathbf{E}\Big(\operatorname{sgn}\left(d_0(I_1,\dots,I_{pq}) + \boldsymbol{d}^T(I_1,\dots,I_{pq})\boldsymbol{\beta}^*\right)\boldsymbol{d}(I_1,\dots,I_{pq})\,\Big|\,\boldsymbol{z}_1\Big),$$

векторную функцию

$$\boldsymbol{a}(\boldsymbol{T},\boldsymbol{z}) = \mathbf{E}\Big(\operatorname{sgn}\big(d_0(I_1,\ldots,I_{pq}) + \boldsymbol{d}^T(I_1,\ldots,I_{pq})\boldsymbol{T}\big)\, \boldsymbol{d}(I_1,\ldots,I_{pq}) \,\Big|\, \boldsymbol{z}_1 = \boldsymbol{z}\Big),$$

где $\boldsymbol{T} \in I\!R^{pq}$ и $\boldsymbol{z} \in I\!R^{p+q}$, и матрицу $\Gamma = \operatorname{cov}(\boldsymbol{\Lambda}(\boldsymbol{z}_1), \boldsymbol{\Lambda}(\boldsymbol{z}_1))$.

Теоремы о состоятельности, асимптотической нормальности и о функции влияния оценки $\hat{B}_n$ сформулируем в терминах векторов $\hat{\boldsymbol{\beta}}_n = \mathbf{vec}(\hat{B}_n)$ и $\boldsymbol{\beta}_0 = \mathbf{vec}(B_0)$.

Теорема 1 (о состоятельности оценки $\hat{B}_n$). *Пусть выполнены следующие условия:*

(a) $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n$ *- н.о.р.с.в.,* $\boldsymbol{\varepsilon}_1 \stackrel{d}{=} -\boldsymbol{\varepsilon}_1$*;* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ *- н.о.р.с.в.; совокупности* $\{\boldsymbol{x}_i\}_{i=1}^n$ *и* $\{\boldsymbol{\varepsilon}_i\}_{i=1}^n$ *независимы,*

(b) $\boldsymbol{E}\|\boldsymbol{\varepsilon}_1\| < \infty$,

(c) $\boldsymbol{E}\Big(\big\|X(I)^{-1}\big\|\, I\big(\det X(I) \neq 0\big)\Big) < \infty$, *где* $\|\cdot\|$ *обозначает евклидову матричную норму,*

(d) функция $U(\boldsymbol{\beta})$ *достигает минимум в единственной точке* $\boldsymbol{\beta} = \boldsymbol{\beta}^*$.

Тогда $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$ *и при* $n \to \infty$ $\hat{\boldsymbol{\beta}}_n$ *сходится почти всюду к* $\boldsymbol{\beta}_0$.

Теорема 2 (об асимптотической нормальности оценки $\hat{B}_n$). *Пусть выполнены условия (a), (d) теоремы 1, а также*

(b′) $\boldsymbol{E}\|\boldsymbol{\varepsilon}_1\|^2 < \infty$,

(c′) $\boldsymbol{E}\Big(\|X(I)^{-1}\|^2 I\big(\det X(I) \neq 0\big)\Big) < \infty$,

(e) определены частные производные функции $U(\boldsymbol{\beta})$ *в точке* $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ *и возможно дифференцирование под знаком математического ожидания*

$$\boldsymbol{\nabla} U(\boldsymbol{\beta})\,\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} = \boldsymbol{E}\Big(\boldsymbol{\nabla}|d_0(I_1,\ldots,I_{pq}) + \boldsymbol{d}^T(I_1,\ldots,I_{pq})\boldsymbol{\beta}|\,\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\Big),$$

(f) в окрестности точки $\boldsymbol{\beta}^*$ *верно следующее разложение:*

$$U(\boldsymbol{\beta}) = U(\boldsymbol{\beta}^*) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T W(\boldsymbol{\beta} - \boldsymbol{\beta}^*) + o(\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|^2),$$

где W *- некоторая положительно определенная матрица,* *(g) при* $\{I_1, \ldots, I_{pq}\} \in \mathcal{I}_p$

$$P\left(\Big\{\det\big(\boldsymbol{b}(I_1) - \boldsymbol{\beta}^* \ldots \boldsymbol{b}(I_{pq}) - \boldsymbol{\beta}^*\big) = 0\Big\} \cap \Big\{\tau(I_1, ..., I_{pq}) = 1\Big\}\right) = 0.$$

Тогда при $n \to \infty$ *распределение случайной величины* $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ *слабо сходится к* pq*-мерному нормальному распределению с вектором средних* $\mathbf{0}$ *и ковариационной матрицей* $p^4q^2W^{-1}\Gamma W^{-1}$.

Робастность оценок, представленных в работе будем исследовать с точки зрения их функций влияния.

Теорема 3 (о функции влияния оценки $\hat{B}_n$). *Пусть выполнены условия (a), (b), (c), (d), (f) теорем* 1, 2, *а также*

(e′) $\forall \boldsymbol{\beta}'$ *из некоторой окрестности точки* $\boldsymbol{\beta}^*$ *выполнено*

$$\boldsymbol{\nabla} U(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}'} = \boldsymbol{E}\Big(\boldsymbol{\nabla}|d_0(I_1, \ldots, I_{pq}) + \boldsymbol{d}^T(I_1, \ldots, I_{pq})\boldsymbol{\beta}|\Big|_{\boldsymbol{\beta}'}\Big),$$

(h) функция $\boldsymbol{a}(\boldsymbol{T}, \boldsymbol{z})$ *непрерывна по* $\boldsymbol{T}$ *в окрестности точки* $\boldsymbol{T} = \boldsymbol{\beta}^*$.

Тогда функция влияния $IF(\boldsymbol{z}, F_{\boldsymbol{\beta}_0}, \hat{B}_n)$ *оценки* $\hat{B}_n$ *ограничена по* $\boldsymbol{z} \in I\!R^{p+q}$ *и равна*

$$IF(\boldsymbol{z}, F_{\boldsymbol{\beta}_0}, \hat{B}_n) = -p^2qW^{-1}\boldsymbol{a}(\boldsymbol{\beta}_0, \boldsymbol{z}),$$

где $F_{\boldsymbol{\beta}_0}$ *обозначает совместную функцию распределения векторов* $\boldsymbol{x}_1$, $\boldsymbol{y}_1$ *в нашей модели.*

Состоятельность, асимптотическая нормальность и ограниченность функции влияния оценки $\tilde{B}_n$ доказаны в более сильных предположениях.

Как известно[18], в условиях активного эксперимента оптимальными являются планы с конечным числом значений. Такие планы (т. е. распределения $\boldsymbol{x}_1$) удовлетворяют условиям (c), (c′) теорем $1-3$. В то же время эти условия сильно ограничивают множество удовлетворяющих им законов распределения, что ставит под вопрос целесообразность применения оценок $\hat{B}_n$ и $\tilde{B}_n$,

[18] Ермаков С. М., Жиглявский А. А. (1982), *Математическая теория оптимального эксперимента.* Наука, Москва.

например, в условиях пассивного эксперимента. В связи с этим в **разделе 1.2** рассматриваются еще две оценки матрицы регрессионных коэффициентов вида $\hat{B}'_n = \mathbf{vec}^{-1}(\hat{\boldsymbol{\beta}}'_n)$ и $\tilde{B}'_n = \mathbf{vec}^{-1}(\tilde{\boldsymbol{\beta}}'_n)$. Здесь вектор $\tilde{\boldsymbol{\beta}}'_n$ есть особым образом "взвешенная"выборочная медиана Оя совокупности векторизованных невырожденных элементарных регрессий $\{\boldsymbol{b}(I)\}$, и вектор $\hat{\boldsymbol{\beta}}'_n$, как и в разделе 1.1, модифицирован для ослабления условий при которых выполнены асимптотические свойства. Сформулируем определения оценок:

Определение 7. *Пусть*

$$D'_n(\boldsymbol{\beta}) = \ ave_{\mathcal{I}}\Big\{V'(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq})\Big\},$$

где $V'(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq})) = V(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))|\det X(I_1)\ldots\det X(I_{pq})|$ *обозначает "взвешенный"объем.*

Тогда оценка $\tilde{B}'_n$ матрицы коэффициентов B_0 определяется следующим образом:

$$\tilde{B}'_n = \ \boldsymbol{vec}^{-1}(\tilde{\boldsymbol{\beta}}'_n), \ \text{где } \tilde{\boldsymbol{\beta}}'_n = \arg\min_{\boldsymbol{\beta}\in I\!\!R^{pq}} D'_n(\boldsymbol{\beta}).$$

Определение 8. *Оценку $\hat{B}'_n$ матрицы коэффициентов B_0 определим следующей формулой:*

$$\hat{B}'_n = \ \boldsymbol{vec}^{-1}(\hat{\boldsymbol{\beta}}'_n), \quad \text{где } \hat{\boldsymbol{\beta}}'_n = \arg\min_{\boldsymbol{\beta}\in I\!\!R^{pq}} U'_n(\boldsymbol{\beta}),$$

$$U'_n(\boldsymbol{\beta}) = \ ave_{\mathcal{I}_p}\Big\{V'(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq})\Big\}.$$

Оценки $\hat{B}'_n$ и $\tilde{B}'_n$ X-аффинно, Y-аффинно и регрессионно эквивариантны и, при выполнении некоторых условий на распределения данных (см. теоремы 4, 5, 6), робастны, состоятельны и асимптотически нормальны. Для формулировки этих условий введем дополнительные обозначения.

Пусть $I_k = \{(k-1)p+1, \ldots, kp\}, \ k = 1, \ldots, pq.$

Предположим, что для всех $\boldsymbol{\beta} \in I\!\!R^{pq}$ существует математическое ожидание $\mathbf{E}\Big(V'(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq})\Big)$ и обозначим за

$$U'(\boldsymbol{\beta}) = \ \mathbf{E}\Big(V'(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq})\Big)$$

теоретическую целевую функцию. $U'(\boldsymbol{\beta})$ - непрерывная ограниченная снизу выпуклая функция, и пусть $\boldsymbol{\beta}_* \in I\!R^{pq}$ - одна из точек множества, на котором она достигает своего минимума. Пусть вектор $\boldsymbol{d}'(I_1, \ldots, I_{pq}) = \big(d'_1(I_1, \ldots, I_{pq}), \ldots, d'_{pq}(I_1, \ldots, I_{pq})\big)^T$ и скаляр $d'_0(I_1, \ldots, I_{pq})$ определяются из разложения:

$$V'(\boldsymbol{\beta}, \boldsymbol{b}(I_1), \ldots, \boldsymbol{b}(I_{pq}))\tau(I_1, ..., I_{pq}) = |d'_0(I_1, \ldots, I_{pq}) + \boldsymbol{d}'^T(I_1, \ldots, I_{pq})\boldsymbol{\beta}|.$$

Таким образом,

$$d'_s(I_1, \ldots, I_{pq}) = d_s(I_1, \ldots, I_{pq}) \det X(I_1) \ldots \det X(I_{pq}), \ s = 0, ..., pq.$$

Пусть $\boldsymbol{z}_1 = (\boldsymbol{x}_1^T, \boldsymbol{y}_1^T)^T$ и, аналогично разделу 1.1, определим случайный вектор

$$\boldsymbol{\Lambda}'(\boldsymbol{z}_1) = \mathbf{E}\Big(\operatorname{sgn}\big(d'_0(I_1, \ldots, I_{pq}) + \boldsymbol{d}'^T(I_1, \ldots, I_{pq})\boldsymbol{\beta}_*\big)\, \boldsymbol{d}'(I_1, \ldots, I_{pq} \,\Big|\, \boldsymbol{z}_1\Big),$$

векторную функцию

$$\boldsymbol{a}'(\boldsymbol{T}, \boldsymbol{z}) = \mathbf{E}\Big(\operatorname{sgn}\big(d'_0(I_1, \ldots, I_{pq}) + \boldsymbol{d}'^T(I_1, \ldots, I_{pq})\boldsymbol{T}\big)\, \boldsymbol{d}'(I_1, \ldots, I_{pq}) \,\Big|\, \boldsymbol{z}_1 = \boldsymbol{z}\Big),$$

где $\boldsymbol{T} \in I\!R^{pq}$ и $\boldsymbol{z} \in I\!R^{p+q}$, и матрицу $\Gamma' = \operatorname{cov}(\boldsymbol{\Lambda}'(\boldsymbol{z}_1), \boldsymbol{\Lambda}'(\boldsymbol{z}_1))$.

Сформулируем теперь теоремы о состоятельности, асимптотической нормальности и о функции влияния оценки $\hat{B}'_n$ (в терминах векторов $\hat{\boldsymbol{\beta}}'_n = \mathbf{vec}(\hat{B}'_n)$ и $\boldsymbol{\beta}_0 = \mathbf{vec}(B_0)$).

Теорема 4 (о состоятельности оценки $\hat{B}'_n$). *Пусть выполнены следующие условия:*

(a) $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n$ *- н.о.р.с.в.,* $\boldsymbol{\varepsilon}_1 \stackrel{d}{=} -\boldsymbol{\varepsilon}_1$*;* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ *- н.о.р.с.в.; совокупности* $\{\boldsymbol{x}_i\}_{i=1}^n$ *и* $\{\boldsymbol{\varepsilon}_i\}_{i=1}^n$ *независимы,*

(b) $\boldsymbol{E}\|\boldsymbol{\varepsilon}_1\| < \infty$,

(c) $\boldsymbol{E}\|\boldsymbol{x}_1\| < \infty$,

(d) функция $U'(\boldsymbol{\beta})$ *достигает минимум в единственной точке* $\boldsymbol{\beta} = \boldsymbol{\beta}_*$.

Тогда $\boldsymbol{\beta}_* = \boldsymbol{\beta}_0$ *и при* $n \to \infty$ $\hat{\boldsymbol{\beta}}'_n$ *сходится почти всюду к* $\boldsymbol{\beta}_0$.

Теорема 5 (об асимптотической нормальности оценки $\hat{B}'_n$). *Пусть выполнены условия (a), (d) теоремы 4, а также*

(b′) $\boldsymbol{E}\|\boldsymbol{\varepsilon}_1\|^2 < \infty$,

(c′) $\boldsymbol{E}\|\boldsymbol{x}_1\|^2 < \infty$,

(e) определены частные производные функции $U'(\boldsymbol{\beta})$ в точке $\boldsymbol{\beta} = \boldsymbol{\beta}_*$ и возможно дифференцирование под знаком математического ожидания

$$\boldsymbol{\nabla} U'(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_*} = \boldsymbol{E}\Big(\boldsymbol{\nabla}|d'_0(I_1,\dots,I_{pq}) + \boldsymbol{d}'^T(I_1,\dots,I_{pq})\boldsymbol{\beta}|\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_*}\Big),$$

(f) в окрестности точки $\boldsymbol{\beta}_*$ верно следующее разложение:

$$U'(\boldsymbol{\beta}) = U'(\boldsymbol{\beta}_*) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_*)^T W'(\boldsymbol{\beta} - \boldsymbol{\beta}_*) + o(\|\boldsymbol{\beta} - \boldsymbol{\beta}_*\|^2),$$

где W' - некоторая положительно определенная матрица,

(g) при $\{I_1,\dots,I_{pq}\} \in \mathcal{I}_p$

$$P\left(\Big\{\det\big(\boldsymbol{b}(I_1) - \boldsymbol{\beta}_* \dots \boldsymbol{b}(I_{pq}) - \boldsymbol{\beta}_*\big) = 0\Big\} \cap \Big\{\tau(I_1,...,I_{pq}) = 1\Big\}\right) = 0.$$

Тогда при $n \to \infty$ распределение случайной величины $\sqrt{n}(\hat{\boldsymbol{\beta}}'_n - \boldsymbol{\beta}_0)$ слабо сходится к pq-мерному нормальному распределению с вектором средних $\mathbf{0}$ и ковариационной матрицей $p^4q^2W'^{-1}\Gamma'W'^{-1}$.

Теорема 6 (о функции влияния оценки $\hat{B}'_n$). *Пусть выполнены условия (a), (b), (c), (d), (f) теорем 4, 5, а также*

(e′) $\forall\boldsymbol{\beta}'$ из некоторой окрестности точки $\boldsymbol{\beta}^*$ выполнено

$$\boldsymbol{\nabla} U'(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}'} = \boldsymbol{E}\Big(\boldsymbol{\nabla}|d'_0(I_1,\dots,I_{pq}) + \boldsymbol{d}'^T(I_1,\dots,I_{pq})\boldsymbol{\beta}|\Big|_{\boldsymbol{\beta}'}\Big),$$

(h) функция $\boldsymbol{a}'(\boldsymbol{T},\boldsymbol{z})$ непрерывна по $\boldsymbol{T}$ в окрестности точки $\boldsymbol{T} = \boldsymbol{\beta}_*$.

Тогда функция влияния $IF(\boldsymbol{z}, F_{\boldsymbol{\beta}_0}, \hat{\boldsymbol{\beta}}'_n)$ оценки $\hat{B}'_n$ при фиксированном $\boldsymbol{x}$ ограничена по $\boldsymbol{y}$ на $\mathbb{R}^q$, при фиксированном $\boldsymbol{y}$ ограничена по $\boldsymbol{x}$ на любом компакте в $\mathbb{R}^p$ и равна

$$IF(\boldsymbol{z}, F_{\boldsymbol{\beta}_0}, \hat{\boldsymbol{\beta}}'_n) = -p^2qW'^{-1}\boldsymbol{a}'(\boldsymbol{\beta}_0, \boldsymbol{z}).$$

Функция влияния оценки $\hat{B}'_n$, вообще говоря, неограниченна по $\boldsymbol{x}$ на $I\!R^p$. Но на практике, как в пассивном, так и активном экспериментах, наблюдаемые значения фактора $\boldsymbol{x}$ принадлежат некоторому компакту. Поэтому важна ограниченность функции влияния по $\boldsymbol{x}$ на любом компакте, а это свойство оценки $\hat{B}'_n$ выполнено по теореме 6.

Робастность, состоятельность и асимптотическая нормальность оценки $\tilde{B}'_n$ доказаны в более сильных предположениях.

В **разделе 1.3** получены формулы для асимптотических эффективностей (через обобщенные дисперсии) всех вышеперечисленных оценок. Теория проиллюстрирована примером, в котором вычислены (симулированы) асимптотические эффективности для случаев, когда вектор случайных ошибок имеет нормальное распределение, распределение Лапласа и распределение Стьюдента с $3, 5, 10, 20$ степенями свободы. На этом примере показана высокая асимптотическая эффективность представленных оценок в случаях распределений вектора случайных ошибок с тяжелыми хвостами. Кроме того, в разделе перечислены способы вычисления указанных оценок матрицы регрессионных коэффициентов.

Вторая глава диссертационной работы посвящена проверке гипотез о точных значениях матрицы регрессионных коэффициентов для модели многомерной линейной регрессии: есть n наблюдений $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$, подчиняющихся модели (1), и требуется проверить гипотезу $H_0 : B_0 = B^\star$ против альтернативы $H_1 : B_0 \neq B^\star$. Без ограничения общности можно считать, что $B^\star = 0$, то есть проверять гипотезу о независимости значений отклика от значений фактора. Ниже будем предполагать, что вектор $\boldsymbol{\Lambda}(\boldsymbol{z}_1)$ ($\boldsymbol{\Lambda}'(\boldsymbol{z}_1)$) и матрица Γ (Γ') определены так же, как и выше, но с $\boldsymbol{\beta}^* = \mathbf{0}$ ($\boldsymbol{\beta}_* = \mathbf{0}$).

В **разделе 2.1** предлагаются четыре тестовые статистики: $\boldsymbol{T}_n = \boldsymbol{\nabla} U_n(\mathbf{0})$ (т. е. вектор частных производных функции $U_n(\boldsymbol{\beta})$ в точке $\boldsymbol{\beta} = \mathbf{0}$), $\boldsymbol{T}'_n = \boldsymbol{\nabla} U'_n(\mathbf{0})$,

$$\psi_n := n\frac{1}{p^4q^2}\boldsymbol{T}_n^T\hat{\Gamma}_n^{-1}\boldsymbol{T}_n \quad \text{и } \psi'_n := n\frac{1}{p^4q^2}(\boldsymbol{T}'_n)^T(\hat{\Gamma}'_n)^{-1}\boldsymbol{T}'_n,$$

где функции U_n и U'_n заданы определениеми 4 и 7, а $\hat{\Gamma}_n$ и $\hat{\Gamma}'_n$ - некоторые состоятельные оценки матриц Γ и Γ', соответственно.

Статистики ψ_n и ψ'_n асимптотически свободны от распределений исходных данных. Кроме того, они обладают свойством аффинной инвариантности в смысле следующего определения:

Определение 9. *Пусть $\psi_n(X, Y)$ обозначает тестовую статистику, вычисленную по матрице факторов X и матрице откликов Y.*
Статистика $\psi_n(X, Y)$ X-аффинно инвариантна, если $\psi_n(XV, Y) = \psi_n(X, Y)$ и Y-аффинно инвариантна, если $\psi_n(X, YW) = \psi_n(X, Y)$ для любых невырожденных $(p\times p)$-матрицы V и $(q\times q)$-матрицы W, соответственно.

В работе получены следующие результаты об асимптотическом распределении статистик $\boldsymbol{T}_n$, $\boldsymbol{T}'_n$, ψ_n и ψ'_n при нулевой гипотезе:

Теорема 7. *Пусть выполнены условия*

(a) $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ *- н.о.р.с.в.,* $\boldsymbol{\varepsilon}_1 \overset{d}{=} -\boldsymbol{\varepsilon}_1$*;* $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ *- н.о.р.с.в.; совокупности* $\{\boldsymbol{x}_i\}_{i=1}^n$ *и* $\{\boldsymbol{\varepsilon}_i\}_{i=1}^n$ *независимы,*

(b) $\boldsymbol{E}\|\boldsymbol{\varepsilon}_1\|^2 < \infty$*,*

(c) $\boldsymbol{E}\Big(\|X(I)^{-1}\|^2 I\big(\det X(I) \neq 0\big)\Big) < \infty$*, где* $\|\cdot\|$ *- евклидова матричная норма.*

Тогда при нулевой гипотезе $H_0 : \boldsymbol{\beta}_0 = \mathbf{0}$ предельное распределение статистики $\sqrt{n}\boldsymbol{T}_n$ - pq-мерное нормальное с нулевым вектором средних и ковариационной матрицей $p^4q^2\Gamma$.

Теорема 8. *Пусть выполнены условия*

(a) $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n$ *- н.о.р.с.в.,* $\boldsymbol{\varepsilon}_1 \overset{d}{=} -\boldsymbol{\varepsilon}_1$*;* $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ *- н.о.р.с.в.; совокупности* $\{\boldsymbol{x}_i\}_{i=1}^n$ *и* $\{\boldsymbol{\varepsilon}_i\}_{i=1}^n$ *независимы,*

(b) $\boldsymbol{E}\|\boldsymbol{\varepsilon}_1\|^2 < \infty$*,*

(c) $\boldsymbol{E}\|\boldsymbol{x}_1\|^2 < \infty$.

Тогда при нулевой гипотезе $H_0 : \boldsymbol{\beta}_0 = \mathbf{0}$ $\sqrt{n}\boldsymbol{T}'_n \sim AN_{pq}(\mathbf{0}, p^4q^2\Gamma')$.

Теорема 9. *Пусть выполнены условия теоремы* 7 *и матрица* Γ *невырождена. Тогда при нулевой гипотезе* $H_0 : \boldsymbol{\beta}_0 = 0$ *предельное распределение статистики* ψ_n *- центральное* χ^2*-распределение с* pq *степенями свободы.*

Теорема 10. *Пусть выполнены условия теоремы* 8 *и матрица* Γ' *невырождена. Тогда при нулевой гипотезе* $H_0 : \boldsymbol{\beta}_0 = 0 \ \ \psi'_n \xrightarrow{d} \chi^2_{pq}$.

В **разделе 2.2** получены предельные распределения статистик $\boldsymbol{T}_n$, $\boldsymbol{T}'_n$, ψ_n и ψ'_n при последовательности близких альтернатив $H_n : \boldsymbol{\beta}_0 = \frac{\boldsymbol{\delta}}{\sqrt{n}}, \boldsymbol{\delta} \neq \mathbf{0}$:

Теорема 11. *Пусть выполнены условия теоремы* 7, *а также*

(d) *в окрестности нуля верно следующее разложение:*

$$\boldsymbol{T}(\boldsymbol{\beta}) = \boldsymbol{T}(\mathbf{0}) + A\boldsymbol{\beta} + \boldsymbol{o}(\|\boldsymbol{\beta}\|),$$

где функция $\boldsymbol{T}(\boldsymbol{\beta})$ *определяется при* $\{I_1, \ldots, I_{pq}\} \in \mathcal{I}_p$ *как*

$$\boldsymbol{T}(\boldsymbol{\beta}) = \boldsymbol{E}_0 \left(sgn\Big(d_0(I_1, \ldots, I_{pq}) + \boldsymbol{d}^T(I_1, \ldots, I_{pq})\boldsymbol{\beta}\Big)\boldsymbol{d}(I_1, \ldots, I_{pq})\right),$$

и A *- некоторая* $(pq \times pq)$*-матрица,*

(e) *при гипотезе* H_0 *для* $\{I_1, \ldots, I_{pq}\} \in \mathcal{I}_p$

$$P\left(\Big\{ \det\big(\boldsymbol{b}(I_1)\ldots\boldsymbol{b}(I_{pq})\big) = 0\Big\} \cap \Big\{\tau(I_1, ..., I_{pq}) = 1\Big\}\right) = 0.$$

Тогда при последовательности альтернатив $H_n : \boldsymbol{\beta}_0 = \frac{\boldsymbol{\delta}}{\sqrt{n}}$ *предельное распределение статистики* $\sqrt{n}\boldsymbol{T}_n$ *есть* pq*-мерное нормальное с вектором математических ожиданий* $-A\boldsymbol{\delta}$ *и ковариационной матрицей* $p^4q^2\Gamma$.

Если матрица Γ *невырождена, то при последовательности альтернатив* H_n *предельное распределение статистики* ψ_n *есть нецентральное* χ^2*-распределение с* pq *степенями свободы и параметром нецентральности* $\frac{1}{p^4q^2}\boldsymbol{\delta}^T A\Gamma^{-1}A\boldsymbol{\delta}$.

Теорема 12. *Пусть выполнены условия теоремы* 8, *а также*

(d) *в окрестности нуля верно следующее разложение:*

$$\boldsymbol{T}'(\boldsymbol{\beta}) = \boldsymbol{T}'(\mathbf{0}) + A'\boldsymbol{\beta} + \boldsymbol{o}(\|\boldsymbol{\beta}\|),$$

где функция $\boldsymbol{T}'(\boldsymbol{\beta})$ определяется при $\{I_1, \ldots, I_{pq}\} \in \mathcal{I}_p$ как

$$\boldsymbol{T}'(\boldsymbol{\beta}) = \boldsymbol{E}_0 \Big(sgn\Big(d'_0(I_1, \ldots, I_{pq}) + \boldsymbol{d}'^T(I_1, \ldots, I_{pq})\boldsymbol{\beta}\Big)\boldsymbol{d}'(I_1, \ldots, I_{pq})\Big),$$

и A' - некоторая $(pq \times pq)$-матрица,

(e) при гипотезе H_0 для $\{I_1, \ldots, I_{pq}\} \in \mathcal{I}_p$

$$P\left(\Big\{ \det\left(\boldsymbol{b}(I_1)\ldots\boldsymbol{b}(I_{pq})\right) = 0\Big\} \cap \Big\{\tau(I_1, ..., I_{pq}) = 1\Big\}\right) = 0.$$

Тогда при последовательности альтернатив $H_n : \boldsymbol{\beta}_0 = \frac{\boldsymbol{\delta}}{\sqrt{n}}$ имеем $\sqrt{n}\boldsymbol{T}'_n \xrightarrow{d} N_{pq}\left(-A'\boldsymbol{\delta}, p^4q^2\Gamma'\right)$, и если матрица Γ' невырождена, то $\psi'_n \xrightarrow{d} \chi^2_{pq}\left(\frac{1}{p^4q^2}\boldsymbol{\delta}^T A'\Gamma'^{-1}A'\boldsymbol{\delta}\right)$.

Как следствие, получается следующий результат.

Теорема 13. *Асимптотическая эффективность по Питману критерия на основе статистики ψ_n есть*

$$e = \frac{\boldsymbol{\delta}^T A\Gamma^{-1}A\boldsymbol{\delta}}{p^4q^2\boldsymbol{\delta}^T I(0)\boldsymbol{\delta}},$$

где $I(0)$ - информационная матрица Фишера, которая определяется (при существовании плотности $f(\boldsymbol{\varepsilon})$ распределения случайного вектора $\boldsymbol{\varepsilon}_1$) как $I(0) = \boldsymbol{E}_0\Big(\left(\boldsymbol{\nabla} ln(f(\boldsymbol{y}_1)) \otimes \boldsymbol{x}_1\right)\left(\boldsymbol{\nabla} ln(f(\boldsymbol{y}_1)) \otimes \boldsymbol{x}_1\right)^T\Big)$. Асимптотическая эффективность по Питману критерия на основе статистики ψ'_n есть

$$e' = \frac{\boldsymbol{\delta}^T A'\Gamma'^{-1}A'\boldsymbol{\delta}}{p^4q^2\boldsymbol{\delta}^T I(0)\boldsymbol{\delta}}.$$

В разделе 2.2.3 приведен пример, для которого вычислены (симулированы) асимптотические эффективности по Питману критериев на основе статистик ψ_n и ψ'_n для случаев, когда вектор случайных ошибок имеет нормальное распределение, распределение Лапласа и распределение Стьюдента с $3, 5, 10, 20$ степенями свободы. На этом примере показана высокая асимптотическая эффективность данных критериев в случаях распределений вектора случайных ошибок с тяжелыми хвостами.

По причинам, упомянутым выше, критерии на основе статистик $\boldsymbol{T}_n$ и ψ_n рекомендуется применять в условиях активного эксперимента, в то время как критерии на основе статистик $\boldsymbol{T}'_n$ и ψ'_n применимы и в пассивном эксперименте.

Автор выражает глубокую благодарность Юрию Николаевичу Тюрину за постоянное внимание, искреннюю заинтересованность, многочисленные обсуждения и ценные советы. Автор благодарит Ханну Оя за предложенные идеи, а также Валерия Николаевича Тутубалина и Михаила Васильевича Болдина за интерес к работе.

Работы автора по теме диссертации

[1] Бусарова Д. А. (2006) Проверка гипотез о матрице коэффициентов многомерной линейной регрессии. *Вестн. Моск. ун-та. Сер. 1, Математика. Механика*, **4**, с. 8 - 14.

[2] Бусарова Д. А. (2006) Робастное оценивание матрицы коэффициентов в многомерной линейной регрессионной модели. *Успехи мат. наук*, **61**, Вып. 3, с. 169 - 170.

[3] Busarova D., Tyurin Y., Mottonen J. and Oja H. (2006) Multivariate Theil estimator with the corresponding test. *Mathematical methods of statistics*, **15**, №1, pp. 1 - 19.

Ю.Н. Тюрину принадлежат постановка задачи и формулировки основных определений. Д.А. Бусаровой принадлежат формулировки и доказательства теорем, а также, совместно с Ю. Мотоннен и Х. Оя, проведение численных экспериментов.

[4] Busarova D. (2005) Robust multivariate regression. *ICORS-2005 (International Conference on Robust Statistics), Abstracts*, p. 9.