Робастная регрессия с применением t-распределения и EM-алгоритма

Шведов А.С.

В работе рассматривается линейная регрессионная модель. EM-алгоритм представляет собой распространенный подход к оценке параметров таких моделей на основе общего принципа максимизации правдоподобия. Известно, что этот метод оценки параметров является робастным, если ошибки независимы, одинаково распределены и имеют многомерное t-распределение.

В предыдущих работах такой подход к оценке параметров регрессионных моделей применялся лишь при условии, что ошибки имеют многомерное t-распределение с числовым параметром степеней свободы. В настоящей работе рассматривается более общая ситуация, когда ошибки могут иметь многомерное t-распределение с векторным параметром степеней свободы. Ненаблюдаемые величины в EM-алгоритме при этом оказываются случайными матрицами.

На численных примерах при различных распределениях ошибок исследованы преимущества такого подхода по сравнению с методом наименьших квадратов.

Ключевые слова: робастная регрессия; многомерное t-распределение; EM-алгоритм.

1. Введение

Регрессионные модели являются основным инструментом для выявления зависимостей между различными показателями практически во всех областях экономической науки. Однако классические и наиболее распространенные методы построения регрессионных моделей не обладают свойством робастности, что, в ряде случаев, может приводить к неверным результатам. Робастные методы в эконометрике известны достаточно давно (см., например, [11]). Но все же в настоящее время, скорее, можно говорить об усилении тенденции к применению робастных методов, а не об обязательном требовании использования таких методов хотя бы наряду с классическими для контроля качества результатов.

Идея использовать регрессионные модели, у которых ошибки имеют не нормальные (гауссовские) распределения, а t-распределения, возникает, даже если не связы-

---

Шведов А.С. – д. физ.-мат. н., профессор кафедры математической экономики и эконометрики НИУ «Высшей школы экономики», e-mail: ashvedov@hse.ru

Статья поступила в Редакцию в декабре 2010 г.

вать этот подход с робастностью процедур оценки параметров. Еще в XIX в. ученые знали «об опасностях, порождаемых длинными хвостами функций распределения ошибок» (см. [2, с. 7]). (В настоящее время чаще используется термин не «длинные хвосты», а «тяжелые хвосты».) Изначальное использование при анализе предположения о нормальности ошибок включает и предположение о легких хвостах функций распределения. Это может не отвечать существу дела и приводить к искажениям в выводах. Одним из самых распространенных подходов к моделированию тяжелых хвостов является использование t-распределения.

Хотя для линейных регрессионных моделей, у которых ошибки имеют t-распределение, и не существует такой замкнутой и красивой теории, как для линейных регрессионных моделей, у которых ошибки имеют нормальное распределение, можно говорить и о преимуществах моделей с t-распределением. Так, фактически, регрессионные модели, у которых ошибки имеют t-распределение, включают в себя в качестве частного случая регрессионные модели, у которых ошибки имеют нормальное распределение, поскольку при стремлении числа степеней свободы к бесконечности t-распределения переходят в нормальное распределение. И результаты в предположении, что ошибки имеют t-распределение с достаточно большим числом степеней свободы, и в предположении, что ошибки имеют нормальное распределение, оказываются практически неотличимыми. Наконец, при оценке методом максимального правдоподобия параметров регрессионных моделей, у которых ошибки имеют t-распределения, можно использовать процедуры, обладающие свойством робастности.

Нередко статистические данные, по которым строится регрессионная модель, содержат резко выделяющиеся наблюдения (outliers). Эти наблюдения существенно отделены от основной части и не подчиняются общей структуре. В каких-то случаях такие выбросы являются просто следствием ошибок, допущенных при сборе или обработке информации, но могут отражать и реальные эффекты.

При использовании многих общепринятых процедур для оценки параметров даже одно резко выделяющееся наблюдение может оказать очень сильное и часто искажающее правильную картину действие. Это легко понять на примере выборочного среднего или выборочной дисперсии. То же относится и к методу наименьших квадратов при определении коэффициентов в линейной регрессионной модели.

Робастные процедуры оценки параметров претендуют на то, чтобы давать хорошее соответствие общей структуре и при наличии резко выделяющихся наблюдений, как и в случае, когда резко выделяющиеся наблюдения отсутствуют. Выявленная таким образом структура, в свою очередь, может быть использована для обнаружения резко выделяющихся наблюдений даже при работе с многомерными статистическими данными.

В какой мере можно говорить, что эти претензии соответствуют действительности? Существуют различные подходы к построению робастных алгоритмов. Иногда резко выделяющиеся наблюдения автоматически игнорируются. Для тех методов, которые изучаются в настоящей работе, вклад таких наблюдений только уменьшается. Для каждого класса алгоритмов слова «хорошее соответствие общей структуре и при наличии резко выделяющихся наблюдений» наполняются своим содержанием.

Среди предшествующих работ, в которых изучаются алгоритмы того же класса, что и у нас, назовем [7, 12, 19]. К перечисленным можно было бы добавить и интересную работу [13], однако в [8, с. 165] указывается на неправильные выводы, имеющиеся в этой работе. Подробнее о «подводных камнях», возникающих, если включать в состав аргументов функции правдоподобия число степеней свободы t-распределения, см. [8, 14].

Робастные алгоритмы других классов представлены, например, в книгах [15, 18]. Применяется и байесовский подход (см., например, [9]). Из работ прикладной направленности назовем [17, 22].

Содержание настоящей работы следующее. В параграфе 2 на примере множественной регрессии (наблюдения одномерные, объясняющих факторов несколько) обсуждается связь М-оценок и метода наименьших квадратов с итерационно модифицируемыми весами. В параграфе 3 приводится описание ЕМ-алгоритма, специализированного метода нахождения точки максимума именно функции правдоподобия. В параграфе 4 излагаются некоторые результаты, относящиеся к оценке параметров множественной регрессии с ошибками, имеющими одномерное t-распределение. Объясняется, почему применение ЕМ-алгоритма в данном случае дает робастный метод оценки параметров регрессии. Также в этом параграфе приводятся результаты численного исследования по методу Монте-Карло. В параграфе 5 устанавливаются две новые теоремы о матричном гамма-распределении. Затем эти теоремы используются в параграфе 6, где результаты, изложенные в параграфе 4, обобщаются на случай многомерной регрессии (наблюдения многомерные, объясняющих факторов несколько). При этом ошибки имеют t-распределение с векторным параметром степеней свободы (введенное в [3, 4]). Прием, применяемый в параграфе 6, когда в ЕМ-алгоритме в качестве ненаблюдаемых переменных берутся случайные матрицы, видимо, используется впервые. Также в этом параграфе приводятся результаты одного расчета.

2. Робастность М-оценок

Рассмотрим обычную линейную регрессию

$$y_i = \sum_{\alpha=1}^{q} x_{i\alpha}\beta_\alpha + e_i\,,\ i = 1,\dots,n\,. \tag{1}$$

Объясняющие переменные $x_{i\alpha}$ считаются известными числами. Через $y_1,\dots,y_n$ обозначаются и одномерные наблюдения, и случайные величины, представляющие собой вероятностную модель для этих наблюдений. Предполагается, что случайные величины $e_1,\dots,e_n$ независимы, одинаково распределены, и каждая из них имеет функцию плотности

$$\frac{1}{\sigma}\varphi\left(\frac{e_i}{\sigma}\right), \tag{2}$$

где $\varphi(x)$ – некоторая известная функция плотности; $\sigma > 0$ – масштабирующий множитель. (Как обычно, используется одно и то же обозначение e_i и для случайной величины, и для аргумента функции плотности.) Задача состоит в нахождении параметров $\beta_1,\dots,\beta_q$ и σ.

Если ввести в рассмотрение q-мерные вектора $x_i = (x_{i1},\dots,x_{iq})'$ и $\beta = (\beta_1,\dots,\beta_q)'$, штрих означает транспонирование, то сумму, входящую в правую часть (1), можно обозначить $x_i'\beta$.

Для определения вектора β может быть использован метод наименьших квадратов с весами, когда оценка вектора β строится путем минимизации выражения

(3) $$\sum_{i=1}^{n} w_i \left(y_i - x_i'\beta\right)^2 ,$$

где $w_1,...,w_n$ – заранее выбранные положительные числа. В частности, при $w_1 = ... = w_n = 1$ данный метод является обычным методом наименьших квадратов. (Мы сейчас не касаемся теоретических свойств метода наименьших квадратов с весами. Подчеркнем только, что речь не идет об обобщенном методе наименьших квадратов, см., например, [1, гл. 5], хотя там и возникают сходные уравнения.) Приравнивание к нулю частных производных функции (3) по $\beta_1,...,\beta_q$, что является необходимым условием минимума, дает систему уравнений

(4) $$\sum_{i=1}^{n} x_{i\alpha} w_i \left(y_i - x_i'\beta\right) = 0 , \quad \alpha = 1,...,q .$$

С другой стороны, оценка параметров β и σ может быть произведена методом максимального правдоподобия, когда ищется максимум функции

(5) $$-n \log\sigma + \sum_{i=1}^{n} \log\varphi\left(\frac{y_i - x_i'\beta}{\sigma}\right).$$

Если ввести в рассмотрение функцию

(6) $$w(x) = -\frac{1}{x}\frac{\varphi'(x)}{\varphi(x)} ,$$

то приравнивание к нулю частных производных функции (5) по $\beta_1,...,\beta_q$ дает систему уравнений

(7) $$\sum_{i=1}^{n} x_{i\alpha} w\left(\frac{y_i - x_i'\beta}{\sigma}\right)\left(y_i - x_i'\beta\right) = 0 , \quad \alpha = 1,...,q .$$

Уравнения (7), хотя и похожи на уравнения (4), отличаются от них тем, что веса зависят от искомого параметра β. Введем в рассмотрении $n \times q$ матрицу X, i-я строка которой – это x_i', и диагональную $n \times n$ матрицу W, у которой i-й элемент на главной диагонали – это w_i. Тогда уравнения (4) можно записать в форме $X' W\left(y - X\beta\right) = 0$, где $y = \left(y_1,...,y_n\right)'$. Отсюда

(8) $$\beta = \left(X' WX\right)^{-1} X' W y ,$$

если матрица $X' WX$ невырожденная. Этот же прием может быть использован и для решения уравнений (7).

Предположим, что функция $w(x)$ обладает следующими свойствами. Во-первых, она принимает только неотрицательные значения. Во-вторых, эта функция монотонно не убывает при $x<0$ и монотонно не возрастает при $x>0$. В-третьих, $w(x)$ стремится к нулю и при $x\to\infty$, и при $x\to-\infty$. Тогда можно говорить о робастности оценок параметра β, полученных при помощи системы уравнений (7), поскольку резко выделяющимся наблюдениям y_i, как правило, соответствуют большие по абсолютной величине разности $y_i - x_i'\beta$ и, соответственно, малые веса $w\left(\frac{y_i - x_i'\beta}{\sigma}\right)$.

Нетрудно увидеть, что если $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ – функция плотности стандартного нормального распределения, то $w(x)\equiv 1$. И в этом случае система уравнений (7) не приводит к робастным оценкам параметра β. Отсюда возникает идея рассмотреть так называемые М-оценки, которые включают в себя в качестве частного случая оценки максимального правдоподобия (см., например, [2, 6]). В этом случае функция $w(x)$, используемая в системе уравнений (7), не обязательно связана с функцией $\varphi(x)$ соотношением (6). Зато можно потребовать, чтобы функция $w(x)$ обладала тремя перечисленными выше свойствами. Или даже более сильными свойствами, например, обращалась в ноль при достаточно больших по абсолютной величине x.

Аргументация в пользу М-оценок может быть и такой. Если функция $\varphi(x)$ на практике все равно не известна, то почему нужно начинать с выбора функции $\varphi(x)$, а не с выбора функции $w(x)$? Может быть, правильнее начинать с выбора функции $w(x)$, а функцию $\varphi(x)$ определять из уравнения (6)?

Но мы все же начнем с выбора функции $\varphi(x)$, а не с выбора функции $w(x)$. При любом $a>0$ можно рассмотреть функцию

$$\varphi(x)=\frac{1}{\sqrt{2\pi a}}\frac{\Gamma(a+0{,}5)}{\Gamma(a)}\left(1+\frac{x^2}{2a}\right)^{-a-0{,}5} \tag{9}$$

– функцию плотности t-распределения с $2a$ степенями свободы. Известно, что при $a\to\infty$ функция $\varphi(x)$ переходит в функцию плотности стандартного нормального распределения. Нетрудно увидеть, что если функция $\varphi(x)$ задается формулой (9), то определяемая соотношением (6) функция $w(x)$ имеет вид $w(x)=\frac{2a+1}{2a+x^2}$. И в этом случае можно ожидать, что полученная путем решения системы уравнений (7) оценка параметра β будет обладать свойством робастности (даже если не переходить от оценок максимального правдоподобия к более общим М-оценкам), поскольку $w(x)$ стремится к нулю и при $x\to\infty$, и при $x\to-\infty$.

Если записать систему уравнений (7) в виде (8), то W – это диагональная $n\times n$ матрица, у которой i-й элемент на главной диагонали равен $w\left(\frac{y_i - x_i'\beta}{\sigma}\right)$. Решить уравнение (8) можно попытаться методом простой итерации. Выбрав начальное приближение $\beta^{(0)}$, например, при помощи метода наименьших квадратов, т.е. решив уравнение (8) с единичной матрицей W, затем принимаем

$$\beta^{(r+1)} = \left(X' W^{(r)} X\right)^{-1} X' W^{(r)} y, \tag{10}$$

где $W^{(r)}$ – это диагональная $n \times n$ матрица, у которой i-й элемент на главной диагонали равен $w\left(\frac{y_i - x_i'\beta^{(r)}}{\sigma^{(r)}}\right)$.

Приравнивание к нулю частной производной функции (5) по σ дает выражение

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right) w\left(\frac{y_i - x_i'\beta}{\sigma}\right)\left(y_i - x_i'\beta\right), \tag{11}$$

что равносильно соотношению

$$\sigma^2 = \frac{1}{n}\left(y - X\beta\right)' W\left(y - X\beta\right).$$

Последнее уравнение позволяет находить значения σ итерациями:

$$\sigma^{(r+1)^2} = \frac{1}{n}\left(y - X\beta^{(r+1)}\right)' W^{(r)}\left(y - X\beta^{(r+1)}\right). \tag{12}$$

Для определения $\sigma^{(0)^2}$ может быть использована единичная матрица W.

Хорошо известно, что последовательность значений, получаемых при помощи метода простой итерации, может быть как сходящейся, так и не быть сходящейся. На практике этот метод обычно применяют для небольших расчетов, руководствуясь принципом «раз сошлось, значит, решение получено». Хотя и существуют условия, гарантирующие сходимость метода простой итерации. Кроме того, в правой части (12) стоит не $\beta^{(r)}$, а $\beta^{(r+1)}$, т.е. в данном случае производится некоторое усложнение метода простой итерации.

В параграфе 4 показано, что для множественной линейной регрессии, когда ошибки имеют t-распределение, применение EM-алгоритма приводит к тому же итерационному процессу (10), (12) для определения $\beta^{(r+1)}$, $\sigma^{(r+1)^2}$. А тогда применимы теоремы о сходимости итерационного процесса, построенного на основе EM-алгоритма. В рамках EM-алгоритма также можно провести обобщение на случай, когда наблюдения $y_1,...,y_n$ не одномерные, а m-мерные (см. параграф 6).

Более подробно связь метода наименьших квадратов с итерационно модифицируемыми весами и робастных процедур освещается, например, в работе [21].

3. EM-алгоритм

EM-алгоритм предназначен для поиска точки, в которой достигает максимума функция правдоподобия, путем построения некоторого итерационного процесса. Каждый шаг итерационного процесса состоит из двух подшагов. E-подшаг заключается в нахождении ожидания (expectation) некоторой функции от случайных величин. При этом ожидание само оказывается функцией интересующего параметра. M-подшаг – это максимизация (maximization), определение того значения параметра, при котором данная функция достигает максимума. Первые буквы приведенных английских слов и дают название алгоритма. Обзор различных задач, для решения которых применяется EM-алгоритм, можно найти, например, в работе [16].

Пусть $y=(y_1,...,y_n)$ – это набор наблюдений, вообще говоря, многомерных. $z=(z_1,...,z_n)$ – набор ненаблюдаемых величин также, вообще говоря, многомерных. С одной стороны, предполагается, что ненаблюдаемые величины заметно влияют на наблюдаемые, и привлечение их для анализа отвечает существу дела. С другой стороны, нахождение точек максимума функций в рамках EM-алгоритма может оказаться значительно более простой и надежной с вычислительной точки зрения процедурой, чем непосредственное нахождение точки максимума исходной функции правдоподобия. В каких-то задачах не вызывает сомнений, что именно следует взять в качестве ненаблюдаемых величин. В других задачах ответ на этот вопрос не столь очевиден.

Обозначим через h совместную функцию плотности случайных векторов y и z, через g – условную функцию плотности случайного вектора z при заданном y, через f – маргинальную функцию плотности случайного вектора y. Все эти функции плотности считаются зависящими от некоторого параметра θ, вообще говоря, многомерного. Имеет место соотношение

$$f(y;\theta)=\frac{h(y,z;\theta)}{g(z\,|\,y;\theta)}.$$

Переходя к логарифмам, получаем соотношение для логарифмических функций правдоподобия

$$l(\theta\,|\,y)=l(\theta\,|\,y,z)-\log g(z\,|\,y;\theta)\,. \tag{13}$$

Задача состоит в нахождении точки θ, в которой достигает максимума функция $l(\theta\,|\,y)$. Пусть $\theta^{(r)}$ – значение параметра θ, найденное при r-й итерации. Умножим левую и правую части (13) на $g(z\,|\,y;\theta^{(r)})$ и проинтегрируем по z. Введем обозначения

$$U(\theta\,|\,y;\theta^{(r)})=\int l(\theta\,|\,y,z)\,g(z\,|\,y;\theta^{(r)})\,dz\,,$$

$$\rho(\theta\,|\,y;\theta^{(r)})=\int \log g(z\,|\,y;\theta)\,g(z\,|\,y;\theta^{(r)})\,dz\,.$$

Тогда

$$l(\theta\,|\,y)=U(\theta\,|\,y;\theta^{(r)})-\rho(\theta\,|\,y;\theta^{(r)})\,. \tag{14}$$

Заметим, что

$$\int \log \frac{g(z\,|\,y;\theta^{(r)})}{g(z\,|\,y;\theta)}\, g(z\,|\,y;\theta^{(r)})\, dz$$

– это расстояние Кульбака – Лейблера между функциями плотности $g(z\,|\,y;\theta^{(r)})$ и $g(z\,|\,y;\theta)$, которое, как известно, всегда неотрицательно. Поэтому при любом θ

(15) $$\rho(\theta\,|\,y;\theta^{(r)}) \le \rho(\theta^{(r)}\,|\,y;\theta^{(r)}).$$

Е-подшаг состоит в нахождении ожидаемого логарифмического правдоподобия $U(\theta\,|\,y;\theta^{(r)})$.

М-подшаг состоит в нахождении точки

$$\theta^{(r+1)} = \arg\max_{\theta}\ U(\theta\,|\,y;\theta^{(r)}).$$

Из (14), (15) и способа определения точки $\theta^{(r+1)}$ следует, что

(16) $$l(\theta^{(r+1)}\,|\,y) \ge l(\theta^{(r)}\,|\,y).$$

Соотношения (16) показывают, что движение идет «в правильном направлении», но еще не гарантируют, что последовательность $\theta^{(r)}$ сходится. Условия общего характера, из которых следует сходимость этой последовательности к точке максимума функции $l(\theta\,|\,y)$, даются в работе [20] теоремами 1 и 4.

4. Множественная линейная регрессия

В этом параграфе рассматривается набор одномерных наблюдений $y_1,\dots,y_n$, и будем предполагать, что ненаблюдаемые величины $z_1,\dots,z_n$ также одномерные и, кроме того, положительные. Двумерные случайные величины $(y_1,z_1),\dots,(y_n,z_n)$ считаем независимыми.

Предположим, что совместная функция плотности случайных величин y_i и z_i имеет вид

(17) $$\frac{z_i^{1/2}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z_i e_i^2}{2\sigma^2}\right) g(z_i),$$

где $e_i = y_i - x_i'\beta$ в соответствии с (1). Будем использовать обозначение θ для пары β,σ. Тогда $h(y,z;\beta,\sigma)$ – это произведение функций (17) при $i=1,\dots,n$. Следовательно,

$$\log h(y,z;\beta,\sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma + \frac{1}{2}\sum_{i=1}^{n}\log z_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n} z_i e_i^2 + \sum_{i=1}^{n}\log g(z_i).$$

Пренебрегая слагаемыми, не зависящими ни от β, ни от σ, получаем выражение для логарифмической функции правдоподобия

$$l(\theta \mid y, z) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} z_i e_i^2 . \qquad (18)$$

В соответствии с определением, данным в параграфе 3,

$$U\left(\theta \mid y; \theta^{(r)}\right) = E\left(l(\theta \mid y, z) \mid y; \theta^{(r)}\right),$$

где $l(\theta \mid y, z)$ рассматривается как функция случайного вектора z. Из (18) следует, что эта функция линейна относительно $z_1, \ldots, z_n$.

Распределение вероятностей с совместной функцией плотности (17) называется нормальным-гамма распределением, если

$$g(z) = \frac{A^a}{\Gamma(a)} \exp(-Az) z^{a-1} , \qquad (19)$$

где $a > 0$, $A > 0$. Тогда, как нетрудно увидеть, (выкладки для m-мерного случая приводятся в параграфе 6) маргинальная функция плотности случайной величины y_i имеет вид (2), где

$$\varphi(x) = (2\pi)^{-1/2} \frac{\Gamma(a + 0{,}5)}{\Gamma(a)} \frac{A^a}{\left(A + 0{,}5x^2\right)^{a+0{,}5}} . \qquad (20)$$

Из (6) следует, что в этом случае

$$w(x) = \frac{a + 0{,}5}{A + 0{,}5x^2} . \qquad (21)$$

Введем обозначение $e_i^{(r)} = y_i - x_i'\beta^{(r)}$ и воспользуемся тем, что

$$E\left(z_i \mid y_i; \theta^{(r)}\right) = w\left(\frac{1}{\sigma^{(r)}} e_i^{(r)}\right). \qquad (22)$$

Доказательство соотношения (22) для m-мерного случая приводится в параграфе 6. Для одномерного случая (22) доказано в работе [7], и в этом доказательстве не требуется, чтобы функция $g(z)$ обязательно имела вид (19). Из (18) и (22) следует, что

$$U\left(\theta \mid y; \theta^{(r)}\right) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} w\left(\frac{1}{\sigma^{(r)}} e_i^{(r)}\right) (y_i - x_i'\beta)^2 .$$

Таким образом, Е-подшаг ЕМ-алгоритма выполнен. Выполнение М-подшага аналогично процедуре, описанной в параграфе 2 (ср. (3), (5), (7), (11), (12)).

Пример 1. Сокращения М1 и М2 в этом примере используются для обозначения одного из двух методов, применяемых для определения параметра β и стандартного отклонения возмущений s.

М1 – метод максимального правдоподобия в предположении нормальности возмущений.

М2 – ЕМ-алгоритм в предположении, что возмущения имеют t-распределение с тремя степенями свободы.

Рассматриваются три различных вида сгенерированных рядов наблюдений.

Н0 – наблюдения соответствуют модели с нормальными возмущениями.

Н1 – наблюдения соответствуют модели с нормальными ε-засоренными возмущениями, $\varepsilon = 0{,}1$.

Н2 – наблюдения соответствуют модели с возмущениями, имеющими t-распределение с тремя степенями свободы.

Целью является исследовать поведение каждого из методов для «своих» и «чужих» рядов наблюдений. Для простоты мы ограничиваемся лишь возмущениями, имеющими конечные дисперсии.

Пусть $n = 15$, $q = 1$, $x_{i1} = 1$ при каждом i, $1 \le i \le n$. Случайная величина e_i имеет функцию плотности (2), где $\varphi(x)$ – либо функция плотности стандартного нормального распределения, либо функция плотности t-распределения при $2a = 3$; обе эти функции плотности приведены в параграфе 2. Для генерации наблюдений y_i используются значения $\beta = 1$, $s = 0{,}3$. При использовании нормальных возмущений $\sigma = s$. При использовании возмущений, имеющих t-распределение с ν степенями свободы, $\sigma = s\sqrt{\dfrac{\nu - 2}{\nu}}$.

При генерации ряда с ε-засоренными наблюдениями считается, что с вероятностью $(1-\varepsilon)$ стандартное отклонение возмущения равно s, и с вероятностью ε равно $5s$.

Для каждого из трех видов генерируется $L = 300$ рядов наблюдений длины n. Для l-го эксперимента, $l = 1,\ldots,L$ значения параметров β_l и s_l определяются и методом М1, и методом М2. Затем определяются средние значения

$$\overline{\beta} = \frac{1}{L}\sum_{l=1}^{L}\beta_l\,,\quad \bar{s} = \frac{1}{L}\sum_{l=1}^{L}s_l$$

и среднеквадратические отклонения

$$\left(\frac{1}{L}\sum_{l=1}^{L}\left(\beta_l - \overline{\beta}\right)^2\right)^{1/2},\quad \left(\frac{1}{L}\sum_{l=1}^{L}\left(s_l - \bar{s}\right)^2\right)^{1/2}.$$

Результаты для средних значений приведены в табл. 1 и 2. В скобках даются среднеквадратические отклонения.

Таблица 1.

Результаты для параметра β, истинное значение $\beta = 1$

	H0	H1	H2
M1	1,011 (0,071)	1,016 (0,133)	1,011 (0,067)
M2	1,017 (0,074)	1,019 (0,081)	1,012 (0,050)

Таблица 2.

Результаты для параметра s, истинное значение $s = 0{,}3$

	H0	H1	H2
M1	0,273 (0,052)	0,472 (0,228)	0,244 (0,093)
M2	0,370 (0,079)	0,454 (0,130)	0,276 (0,070)

Сравнивая результаты для рядов вида H0 и H2, мы видим, что «свой» метод (т.е. метод M1 для рядов H0 и метод M2 для рядов H2) дает лучшие результаты и в смысле меньшего разброса (т.е. среднеквадратического отклонения), и в смысле близости среднего значения найденных параметров к истинным значениям (за исключением значения $\overline{\beta} = 1{,}012$, которое несколько хуже, чем значение $\overline{\beta} = 1{,}011$).

На первый взгляд, результаты для параметра β для рядов вида H2 противоречат теореме Гаусса – Маркова. Метод M1 совпадает с методом наименьших квадратов, и, казалось бы, дисперсия оценки должна быть наименьшей. А среднеквадратическое отклонение 0,067 существенно больше, чем среднеквадратическое отклонение 0,050, полученное при использовании метода M2. На самом деле, этот пример всего лишь показывает, что требование линейности и несмещенности оценки, содержащееся в теореме Гаусса – Маркова, не может быть отброшено. Метод M2 не является линейным.

Для рядов вида H1 в результатах для параметра β видна робастность метода M2. Полученное среднеквадратическое отклонение примерно в 1,65 раза меньше, чем для метода M1. Различие в средних, 1,016 и 1,019, носит случайный характер. Так, для другой серии из 300 экспериментов для рядов вида H1 для параметра β при использовании метода M1 получены результаты $\genfrac{}{}{0pt}{}{1{,}023}{(0{,}140)}$, а при использовании метода M2 – результаты $\genfrac{}{}{0pt}{}{1{,}019}{(0{,}085)}$. Проявляется робастность метода M2 и в результатах для параметра s для рядов этого вида.

5. Две теоремы о матричном гамма-распределении

Теория матричных гамма-распределений излагается, например, в [10]. Наиболее известными из этих распределений являются, видимо, распределения Уишарта. Также в работе [10, с. 122–124] рассматриваются матричные гамма-распределения с векторным параметром – некоторое естественное обобщение распределений Уишарта. При том, что легкими и короткими формулировки здесь быть не могут, обозначения, используемые в [10] и других работах для распределений с векторным параметром, с нашей точки зрения, несколько избыточны, возможно, из-за этого остались неисследованными свойства этих распределений. Более прозрачные обозначения для матричных гамма-распределений с векторным параметром используются в работе [3]. Здесь мы повторим только самые необходимые определения.

Для $m\times m$ матрицы $C=\left\{c_{ij}\right\}_{i,j=1}^{m}$ при $k=1,\ldots,m$ рассматриваются подматрицы $C^{[k]}=\left\{c_{ij}\right\}_{i,j=1}^{k}$ и $C_{[k]}=\left\{c_{ij}\right\}_{i,j=m-k+1}^{m}$.

Рассматривается также вектор $a=\left(a_1,\ldots,a_m\right)$ такой, что $a_j>0{,}5(j-1)$ при $j=1,\ldots,m.$ Многомерная гамма-функция определяется следующим образом:

$$\Gamma_m^*(a)=\pi^{m(m-1)/4}\prod_{j=1}^{m}\Gamma\left(a_j-0{,}5(j-1)\right),$$

где $\Gamma(\cdot)$ – обычная гамма-функция. Дополнительно считается, что

$$a_0=0\,,\ \ a_{m+1}=0{,}5(m+1)\,.$$

Параметрами рассматриваемых гамма-распределений являются вектор a указанного вида и положительно определенная $m\times m$ матрица A. Функция плотности имеет вид

$$g(z)=\gamma_{a,A}\ etr(-Az)\prod_{j=1}^{m}|\,z_{[j]}\,|^{a_j-a_{j+1}}, \tag{23}$$

где z – положительно определенная $m\times m$ матрица; $etr(C)=\exp(tr\,C)$; $|\,C\,|$ – определитель матрицы C. Коэффициент $\gamma_{a,A}$ задается формулой

$$\gamma_{a,A}=\left(\Gamma_m^*(a)\prod_{j=0}^{m-1}|\,A^{[m-j]}\,|^{a_j-a_{j+1}}\right)^{-1} \tag{24}$$

(ср. (23) и (24) с (19)).

Пусть T – симметричная $m\times m$ матрица такая, что матрица $A-T$ положительно определенная. (Изначально симметричная матрица T может быть взята произвольно. При некотором $\varepsilon>0$ матрица $A-\varepsilon T$ будет положительно определенной. В этом смысле условие, что матрица $A-T$ положительно определенная, не является ограничительным.) Пусть Z – положительно определенная случайная матрица с функцией плотности (23). Определим производящую функцию моментов

$$M(T) = E\,etr(TZ).$$

Из формулы

$$tr(TZ) = \sum_{l=1}^{m} T_{ll} Z_{ll} + 2\sum_{k=1}^{m}\sum_{l>k} T_{kl} Z_{kl}$$

следует, что

$$\left.\frac{\partial M(T)}{\partial T_{ll}}\right|_{T=0} = E\left(Z_{ll}\right), \tag{25}$$

а при $l > k$

$$\left.\frac{\partial M(T)}{\partial T_{kl}}\right|_{T=0} = 2\,E\left(Z_{kl}\right). \tag{26}$$

Теорема 1.

$$M(T) = \frac{\gamma_{a,A}}{\gamma_{a,A-T}} = \prod_{j=0}^{m-1} |(A-T)^{[m-j]}|^{a_j - a_{j+1}} \left(\prod_{j=0}^{m-1} | A^{[m-j]} |^{a_j - a_{j+1}}\right)^{-1}.$$

Доказательство следует из выражения

$$M(T) = \gamma_{a,A} \int_{z>0} etr(-(A-T)z) \prod_{j=1}^{m} | z_{[j]} |^{a_j - a_{j+1}} dz$$

и из формулы (24).

При $j = 0,...,m-1$ определим $m \times m$ матрицу C_{m-j} такую, что

$$\left(C_{m-j}\right)^{[m-j]} = \left(A^{[m-j]}\right)^{-1},$$

остальные элементы матрицы C_{m-j} равны нулю.

Теорема 2.

$$E(Z) = \sum_{j=0}^{m-1} \left(a_{j+1} - a_j\right) C_{m-j}.$$

Доказательство. Чтобы найти ожидание каждого элемента Z_{kl}, $k \le l$, случайной матрицы Z, воспользуемся формулами (25), (26) и теоремой 1.

Зафиксируем j такое, что $l \le m-j$. Минор элемента $A_{kl} - T_{kl}$ матрицы $(A-T)^{[m-j]}$ обозначим M_{kl}. Через Σ обозначим определитель этой матрицы. Введем обозначение $s_{kl} = A_{kl} - T_{kl}$. Тогда

$$\Sigma = \sum_{u=1}^{m-j} (-1)^{u+l} s_{ul} M_{ul}. \tag{27}$$

Дифференцирование функции (27) по s_{ll} не вызывает затруднений, поскольку ни один из миноров от s_{ll} не зависит. Имеем

$$\frac{\partial \Sigma}{\partial s_{ll}} = M_{ll} = \left|(A-T)^{[m-j]}\right| \left((A-T)^{[m-j]}\right)^{-1}_{ll}. \tag{28}$$

Дифференцирование функции (27) по s_{kl} при $k < l$ несколько труднее. Все миноры кроме M_{ll} могут зависеть от s_{kl}, поскольку $s_{lk} = s_{kl}$. Имеем

$$\frac{\partial \Sigma}{\partial s_{kl}} = (-1)^{k+l} M_{kl} + \sum_{\substack{u=1 \\ u \ne l}}^{m-j} (-1)^{u+l} s_{ul} \frac{\partial}{\partial s_{kl}} M_{ul}. \tag{29}$$

Обозначим через $M_{ul,lv}$ определитель матрицы, получающейся из $(A-T)^{[m-j]}$ выкидыванием строк с номерами u и l и столбцов с номерами l и v. Тогда при $u < l$

$$M_{ul} = \sum_{v=1}^{l-1} (-1)^{v+l-1} s_{lv} M_{ul,lv} - \sum_{v=l+1}^{m-j} (-1)^{v+l-1} s_{lv} M_{ul,lv};$$

при $u > l$

$$M_{ul} = \sum_{v=1}^{l-1} (-1)^{v+l} s_{lv} M_{ul,lv} - \sum_{v=l+1}^{m-j} (-1)^{v+l} s_{lv} M_{ul,lv}.$$

Ни один из определителей $M_{ul,lv}$, входящих в две последние формулы, от s_{kl} не зависит. Вторыми суммами в правых частях также, очевидно, можно пренебречь, поскольку $k < l$. Поэтому при $u < l$

$$\frac{\partial}{\partial s_{kl}} M_{ul} = (-1)^{k+l-1} M_{ul,lk};$$

при $u > l$

$$\frac{\partial}{\partial s_{kl}} M_{ul} = (-1)^{k+l} M_{ul,lk}.$$

Из (29) при $k<l$ находим

$$\frac{\partial\Sigma}{\partial s_{kl}}=(-1)^{k+l}M_{kl}+\sum_{u=1}^{l-1}(-1)^{u+l}s_{ul}(-1)^{k+l-1}M_{lk,ul}+\sum_{u=l+1}^{m-j}(-1)^{u+l}s_{ul}(-1)^{k+l}M_{lk,ul}.$$

То есть при $k<l$

$$\frac{\partial\Sigma}{\partial s_{kl}}=2(-1)^{k+l}M_{kl}=2\left|(A-T)^{[m-j]}\right|\left((A-T)^{[m-j]}\right)^{-1}_{kl}. \tag{30}$$

Изменения, которые нужно внести в приведенные выкладки при $m-j\le 2$ или при $l=m-j$, очевидны.

Те же выражения (28) и (30) получаются при дифференцировании $\left|(A-T)^{[m-j]}\right|$ по T_{ll} и по T_{kl}, только в правые части добавляется знак минус.

Воспользовавшись теоремой 1, формулами (28) и (30), получаем

$$\left.\frac{\partial M(T)}{\partial T_{ll}}\right|_{T=0}=-\sum_{j=0}^{m-l}\left(a_j-a_{j+1}\right)\left(A^{[m-j]}\right)^{-1}_{ll},$$

и при $k<l$

$$\left.\frac{\partial M(T)}{\partial T_{kl}}\right|_{T=0}=-2\sum_{j=0}^{m-l}\left(a_j-a_{j+1}\right)\left(A^{[m-j]}\right)^{-1}_{kl}.$$

Воспользовавшись (25) и (26), при $k\le l$ получаем

$$E(Z_{kl})=\sum_{j=0}^{m-1}\left(a_{j+1}-a_j\right)\left(C_{m-j}\right)_{kl}.$$

Теорема 2 доказана.

Теорема 2 для случая $a_1=\ldots=a_m$ известна. Доказательство приводится, например, в [10]. При этом и в [10], и в других работах используется не производящая функция моментов, а характеристическая функция. Использование производящей функции моментов позволяет избежать рассмотрения матриц с комплексными элементами. Теорема 1, хорошо известная для одномерного случая, при $m>1$, по-видимому, является новой даже для случая $a_1=\ldots=a_m$.

6. Многомерная линейная регрессия

Рассмотрим уравнения, аналогичные (1):

$$y_i=\beta' x_i+e_i,\ i=1,\ldots,n. \tag{31}$$

Объясняющие переменные $x_{i\alpha}$ – известные числа, $x_i = (x_{i1},...,x_{iq})'$; β – $q \times m$ матрица. Через $y_1,...,y_n$ обозначаются и m-мерные наблюдения, и случайные вектора, представляющие собой вероятностную модель для этих наблюдений. Предполагается, что m-мерные случайные вектора $e_1,...,e_n$ независимы и одинаково распределены.

Ненаблюдаемые величины $z_1,...,z_n$ являются положительно определенными $m \times m$ случайными матрицами. Как и в одномерном случае, $(y_1, z_1),...,(y_n, z_n)$ независимы. Совместное распределение y_i и z_i будем считать нормальным-гамма, т.е. совместная функция плотности имеет вид

$$(2\pi)^{-m/2}\left|\frac{1}{\sigma^2}z_i\right|^{1/2}\exp\left(-\frac{1}{2\sigma^2}e_i' z_i e_i\right)g(z_i), \tag{32}$$

где $g(z_i)$ определяется формулой (23), и $e_i = y_i - \beta' x_i$ в соответствии с (31).

Рассмотрим вектор $b = (b_1,...,b_m)$, где $b_j = a_j + 0{,}5$ при $j = 1,...,m$. Пусть $b_0 = 0$, $b_{m+1} = 0{,}5(m+1)$.

При $x \in R^m$ рассмотрим функцию

$$\varphi(x) = (2\pi)^{-m/2}\frac{\Gamma_m^*(b)}{\Gamma_m^*(a)}|A|^{-1/2}\prod_{j=0}^{m-1}\left(1+\frac{1}{2}x^{[m-j]'}(A^{[m-j]})^{-1}x^{[m-j]}\right)^{b_j - b_{j+1}} \tag{33}$$

Она является функцией плотности многомерного t-распределения с векторным параметром степеней свободы (см. [3; 4]); ср. (20). Здесь для m-мерного вектора x через $x^{[k]}$ обозначается k-мерный вектор, состоящий из первых k компонент вектора x, $k = 1,...,m$.

Теорема 3. Маргинальная функция плотности случайного вектора y_i имеет вид

$$\frac{1}{\sigma^m}\varphi\left(\frac{1}{\sigma}e_i\right), \tag{34}$$

где функция φ задается формулой (33).

Доказательство. Маргинальная функция плотности случайного вектора y_i получается интегрированием по области $z_i > 0$ совместной функции плотности (32). Чтобы несколько сократить формулы, внутри доказательства теоремы будем использовать обозначение z вместо z_i и обозначение e вместо e_i. Во-первых, заметим, что

$$e'ze = tr(ee'z).$$

Используя (23) и (24), получаем следующее выражение для маргинальной функции плотности:

$$(2\pi)^{-m/2}\gamma_{a,A}\frac{1}{\sigma^m}\int\limits_{z>0} etr\left(-\left(A+\frac{1}{2\sigma^2}ee'\right)z\right)\prod_{j=1}^{m}\left|z_{[j]}\right|^{b_j-b_{j+1}}dz =$$

$$=(2\pi)^{-m/2}\gamma_{a,A}\frac{1}{\sigma^m}\Gamma_m^*(b)\prod_{j=0}^{m-1}\left|\left(A+\frac{1}{2\sigma^2}ee'\right)^{[m-j]}\right|^{b_i-b_{j+1}}.$$

Если воспользоваться тем, что

$$\left|A^{[m-j]}+\frac{1}{2\sigma^2}e^{[m-j]}e^{[m-j]'}\right| = \left|A^{[m-j]}\right|\left(1+\frac{1}{2\sigma^2}e^{[m-j]'}(A^{[m-j]})^{-1}e^{[m-j]}\right),$$

(см., например, лемму 5 в [3]), то получаем выражение

$$(2\pi)^{-m/2}\frac{\Gamma_m^*(b)}{\Gamma_m^*(a)}\frac{1}{\sigma^m}|A|^{-1/2}\prod_{j=0}^{m-1}\left(1+\frac{1}{2\sigma^2}e^{[m-j]'}(A^{[m-j]})^{-1}e^{[m-j]}\right)^{b_j-b_{j+1}}.$$

Теорема 3 доказана.

Теорема 4. Условная функция плотности случайного вектора z_i при условии y_i – это функция плотности матричного гамма-распределения с векторным параметром b и с матричным параметром $A+\frac{1}{2\sigma^2}e_ie_i{}'$.

Доказательство. Условная функция плотности случайного вектора z_i при условии y_i – это отношение совместной функции плотности (32) к маргинальной функции плотности (34). Как и в доказательстве предыдущей теоремы, будем использовать обозначение z вместо z_i и обозначение e вместо e_i. Искомая условная функция плотности представима в виде дроби с числителем

$$etr\left(-\left(A+\frac{1}{2\sigma^2}ee'\right)z\right)\prod_{j=1}^{m}\left|z_{[j]}\right|^{b_j-b_{j+1}}$$

и со знаменателем

$$\Gamma_m^*(b)\prod_{j=0}^{m-1}\left|A^{[m-j]}\right|^{b_i-b_{j+1}}\prod_{j=0}^{m-1}\left(1+\frac{1}{2\sigma^2}e^{[m-j]'}(A^{[m-j]})^{-1}e^{[m-j]}\right)^{b_j-b_{j+1}}.$$

Используя, как и в доказательстве предыдущей теоремы, лемму 5 из [3], получаем следующее выражение для знаменателя:

$$\Gamma_m^*(b)\prod_{j=0}^{m-1}\left|A^{[m-j]}+\frac{1}{2\sigma^2}e^{[m-j]}e^{[m-j]'}\right|^{b_j-b_{j+1}} = \left(\gamma_{b,A+\frac{1}{2\sigma^2}ee'}\right)^{-1}.$$

Теорема 4 доказана.

Будем использовать обозначение θ для пары β, σ. Из (32) следует, что и в многомерном случае логарифмическая функция правдоподобия имеет вид, сходный с (18):

(35) $$l(\theta \mid y, z) = -n \log \sigma^m - \frac{1}{2\sigma^2} \sum_{i=1}^{n} e_i' z_i e_i \, .$$

Ввиду линейности $l(\theta \mid y, z)$ по $z_1, \ldots, z_n$, чтобы построить функцию $U(\theta \mid y; \theta^{(r)})$, достаточно знать условное ожидание $E(z_i \mid y_i; \theta^{(r)})$.

При $j = 0, \ldots, m-1$ и при $x \in R^m$ определим $m \times m$ матрицу $C_{m-j}(x)$ такую, что

$$\left(C_{m-j}(x) \right)^{[m-j]} = \left(\left(A + \frac{1}{2} x x' \right)^{[m-j]} \right)^{-1},$$

остальные элементы матрицы $C_{m-j}(x)$ равны нулю. Положим

$$w(x) = \sum_{j=0}^{m-1} \left(b_{j+1} - b_j \right) C_{m-j}(x)$$

(ср. (21)). Тогда на основании теорем 2 и 4 получаем $E\left(z_i \mid y_i; \theta^{(r)} \right) = w\left(\frac{1}{\sigma^{(r)}} e_i^{(r)} \right)$ (ср. (22)).

Определим $m \times m$ матрицы

$$w_i^{(r)} = w\left(\frac{1}{\sigma^{(r)}} e_i^{(r)} \right), \; i = 1, \ldots, n \, .$$

В соответствии с определением функции $U(\theta \mid y; \theta^{(r)})$ в параграфе 3 (см. также параграф 4) и с (35) получаем

$$U(\beta, \sigma \mid y; \beta^{(r)}, \sigma^{(r)}) = -nm \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i' - x_i' \beta \right) w_i^{(r)} \left(y_i - \beta' x_i \right).$$

Дифференцирование функции U по $\beta_{\alpha k}$, $\alpha = 1, \ldots, q$, $k = 1, \ldots, m$ и приравнивание производной к нулю с учетом симметричности матрицы w дает уравнение

(36) $$\sum_{i=1}^{n} x_{i\alpha} \sum_{j=1}^{m} w_{i,jk}^{(r)} \left(y_{ij} - \sum_{\gamma=1}^{q} x_{i\gamma} \beta_{\gamma j} \right) = 0 \, .$$

Дифференцирование функции U по σ и приравнивание производной к нулю дает уравнение

(37) $$\sigma^2 = \frac{1}{nm} \sum_{i=1}^{n} \left(y_i' - x_i' \beta \right) w_i^{(r)} \left(y_i - \beta' x_i \right)$$

(ср. (11), (12)). Найденные из уравнений (36) и (37) значения β и σ принимаются в качестве $\beta^{(r+1)}$ и $\sigma^{(r+1)}$. Значения $\beta^{(0)}$ и $\sigma^{(0)}$ рассчитываются с единичными матрицами w_i.

Другой вариант применения EM-алгоритма в линейной регрессии, когда ошибки имеют многомерное t-распределение, представлен, например, в работе [12].

Пример 2. Пусть $m=3$, $n=1000$, $q=2$. При каждом $i=1,...,n$ вектор $x_i'=(1,i)$, а трехмерная случайная величина e_i имеет функцию плотности (33), (34) с параметрами $a=(2,5,9)$,

$$A=\begin{pmatrix} 1 & 1{,}3 & 1{,}5 \\ 1{,}3 & 2 & 2 \\ 1{,}5 & 2 & 4 \end{pmatrix},$$

$\sigma=12$. При генерации трехмерных возмущений с указанным t-распределением применяется алгоритм Метрополиса (см., например, [5]). Для генерации наблюдений используется матрица

$$\beta=\begin{pmatrix} 10 & 20 & -30 \\ 0{,}3 & -0{,}2 & 0{,}4 \end{pmatrix}.$$

Проводя аналогию с временными рядами, можно сказать, что рассматривается модель со свободным членом и с трендом, а возмущения представляют собой трехмерный белый шум, имеющий t-распределение с векторным параметром степеней свободы.

При применении EM-алгоритма, т.е. при использовании итерационного процесса, основанного на формулах (36), (37), сходимость в проведенном эксперименте была достигнута после 20 итераций. Получены значения параметров

$$\beta=\begin{pmatrix} 10{,}185 & 20{,}268 & -29{,}681 \\ 0{,}299 & -0{,}201 & 0{,}400 \end{pmatrix},$$

$\sigma=11{,}533$. Результаты являются удовлетворительными. Отметим, что при применении метода наименьших квадратов получено значение параметра

$$\beta=\begin{pmatrix} 10{,}191 & 20{,}399 & -29{,}634 \\ 0{,}299 & -0{,}201 & 0{,}400 \end{pmatrix}.$$

Оно же использовалось в качестве начального значения $\beta^{(0)}$ в EM-алгоритме. В данном расчете метод наименьших квадратов уступает EM-алгоритму.

* *
*

СПИСОК ЛИТЕРАТУРЫ

1. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс. М.: Дело, 2004.
2. *Хьюбер П.* Робастность в статистике. М.: Мир, 1984.
3. *Шведов А.С.* Бета-распределение случайной матрицы и его применение в модели состояние-наблюдение: препринт. WP2/2009/01. М.: ГУ ВШЭ, 2009.
4. *Шведов А.С.* t-распределение случайной матрицы и его применение в регрессионной модели: препринт. WP2/2010/01. М.: ГУ ВШЭ, 2010.
5. *Шведов А.С.* О методах Монте-Карло с цепями Маркова // Экономический журнал Высшей школы экономики. 2010. Т. 14. № 2. С. 227–243.
6. *Andrews D.F.* A Robust Method for Multiple Linear Regression // Technometrics. 1974. 16. P. 523–531.
7. Dempster A.P., Laird N.M., Rubin D.B. Iteratively Reweighted Least Squares for Linear Regression When Errors are Normal/Independent Distributed // Multivariate Analysis – V / ed. by P.R. Krishnaiah. Amsterdam: North-Holland, 1980. P. 35–57.
8. *Fernandez C., Steel M.F.J.* Multivariate Student-t Regression Models: Pitfalls and Inference // Biometrika. 1999. 86 (1). P. 153–167.
9. *Fonseca T.C.O., Ferreira M.A.R., Migon H.S.* Objective Bayesian Analysis for the Student-t Regression Model // Biometrika. 2008. 95 (2). P. 325–333.
10. *Gupta A.K., Nagar D.K.* Matrix Variate Distributions. N.Y.: Chapman & Hall, 1999.
11. *Koenker R.* Robust Methods in Econometrics // Econometric Reviews. 1982. 1. P. 213–255.
12. *Lange K.L., Little R.J.A., Taylor J.M.G.* Robust Statistical Modelling Using the t-distribution // Journal of the American Statistical Association. 1989. 84. P. 881–896.
13. *Liu C.H., Rubin D.B.* ML Estimation of the t-distribution Using EM and its Extensions, ECM and ECME // Statistica Sinica. 1995. 5. P. 19–39.
14. *Lucas A.* Robustness of the Student-t Based M-estimator // Communications in Statistics – Theory and Methods. 1997. 26 (5). P. 1165–1182.
15. *Maronna R.A., Martin R.D., Yohai V.J.* Robust Regression – Theory and Methods. N.Y.: Wiley, 2006.
16. *Meng X.L., van Dyk D.A.* The EM Algorithm – An Old Folk-song Sung to a Fast New Tune (with discussion) // Journal of the Royal Statistical Society. 1997. B. 59. P. 511–567.
17. *Preminger A., Franck R.* Forecasting Exchange Rates – A Robust Regression Approach // International Journal of Forecasting. 2007. 23(1). P. 71–84.
18. *Rousseeuw P.J., Leroy A.M.* Robust Regression and Outlier Detection. N.Y.: Wiley, 1987.
19. *Rubin D.B.* Iteratively Reweighted Least Squares // Encyclopedia of Statistical Sciences. N.Y.: Wiley, 1983. Vol. 4. P. 272–275.
20. *Wu C.F.J.* On the Convergence Properties of the EM Algorithm // Annals of Statistics. 1983. 11. P. 95–103.
21. *Yuan K.-H., Bentler P.M.* Robust Mean and Covariance Structure Analysis through Iteratively Reweighted Least Squares // Psychometrika. 2000. 65(1). P. 43–58.
22. *Zellner A., Ando T.* Bayesian and Non-Bayesian Analysis of the Seemingly Unrelated Regression Model with Student-t Errors, and its Application for Forecasting // International Journal of Forecasting. 2010. 26. P. 413–434.