Федеральное государственное бюджетное учреждение науки
МАТЕМАТИЧЕСКИЙ ИНСТИТУТ
им. В. А. Стеклова
Российской академии наук
(МИАН)

119991, Москва, ул. Губкина, д. 8
Тел.: (495) 984-81-41. Факс: (495) 984-81-39. Для телеграмм: Москва, 119333, математика
E-mail: steklov@mi.ras.ru http://www.mi.ras.ru
ОКПО 02699547 ОГРН 1027739665436 ИНН/КПП 7736029594/773601001

"УТВЕРЖДАЮ"

Зам. директора ФГБУН Математический
институт им. В. А. Стеклова
Российской академии наук
член-корр. РАН Д. В. Трещев

02 ноября 2015 г.

ОТЗЫВ
ведущей организации ФГБУН Математический институт
им. В. А. Стеклова Российской академии наук
о диссертационной работе
Есаулова Даниила Михайловича
«Робастные GM-тесты и оценки
в авторегрессионных схемах с выбросами»,
представленную на соискание ученой степени
кандидата физико-математических наук по специальности
01.01.05 — теория вероятностей и математическая статистика.

Диссертационная работа Д. М. Есаулова посвящена построению и исследованию свойств робастных статистических процедур в специальной схеме засорения данных в авторегрессионных моделях временных рядов.

Термин «робастный» (от англ. robust — крепкий, сильный) и связанная с ним область исследований появились в 1960-гг., после того, как была достаточно полно разработана статистическая теория, основанная на предположении нормальности (гауссовости) наблюдаемых случайных величин и использующая методы максимального правдоподобия и наименьших квадратов. В практике работы со статистическими данными было, однако, обнаружено, что эти данные далеко не всегда подчиняются нормальному закону, вопреки бытовавшему представлению о его «универсальности». Причиной могла быть как сама природа данных, определяющая иной закон распределения вероятностей, так и непредсказуемые искажения в процессе сбора, регистрации или передачи данных, получившие неформальное название их «засорения».

Первоначально вопросы робастности концентрировались вокруг оценки математического ожидания a по независимым одинаково распределенным наблюдениям $X_1, \ldots, X_n$. Нормальная теория рекомендует в качестве оценки для a среднее арифметическое $\bar{x} = \frac{1}{n}\sum_i X_i$, являющееся наилучшей несмещенной оценкой. Однако при нарушении гипотезы нормальности эта оценка теряет свои оптимальные свойства и уступает оценкам, менее эффективным при условии нормальности, но более устойчивым (робастным) к отклонениям от этого условия. В дальнейшем исследования робастности были распространены и на более сложные статистические модели.

В диссертации Д. М. Есаулова рассматриваются задачи построения робастных оценок параметров и критериев проверки гипотез в моделях временны́х рядов с определенного вида засорением данных. Основным инструментом и предметом изучения является так называемый «остаточный эмпирический процесс», с помощью которого строятся искомые оценки и критерии, свойства которых выводятся из свойств этого процесса.

Диссертация состоит из 4-х глав, первая из них — Введение. В нем дается обширный обзор результатов по робастным методам оценивания и проверки гипотез относительно параметров процессов линейной авторегрессии.

Введем обозначения, которые потребуются для изложения результатов диссертации. Процесс $\{u_t\}_{t=\ldots-1,0,1,2,\ldots}$ линейной авторегрессии порядка p (кратко: AR(p)), определяется рекуррентным соотношением

$$u_t = \beta_1 u_{t-1} + \cdots + \beta_p u_{t-p} + \varepsilon_t, \tag{1}$$

где $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ — вектор неизвестных параметров, а $\{\varepsilon_t\}$ — независимые одинаково распределенные случайные величины с функцией распределения G, имеющей плотность g. На вектор параметров $\boldsymbol{\beta}$ накладываются условия, обеспечивающие существование строго стационарного решения уравнения (1), которое и принимается за процесс AR(p).

Во Введении приводятся статистические методы, основанные на использовании «остаточного» эмпирического процесса, изучаемого далее в диссертации. Изложим кратко суть этого понятия, ограничиваясь для простоты случаем $p = 1$, когда уравнение (1) имеет вид $u_t = \beta u_{t-1} + \varepsilon_t$. «Инновации» ε_t суть разности $u_t - \beta u_{t-1}$, которые ненаблюдаемы, поскольку параметр β неизвестен. Разности $\varepsilon_t(\alpha) = u_t - \alpha u_{t-1}$, зависящие от «пробного» параметра α, называются «остатками» (совпадают с ε_t при $\alpha = \beta$). Значение α, удовлетворяющее определенным требованиям оптимизации, принимается за оценку $\hat{\beta}$ параметра β.

А именно, вводится случайная функция

$$L_n(\alpha) = n^{-1/2} \sum\nolimits_{t=1}^{n} \varphi(u_{t-j})\psi(u_t - \alpha u_{t-1}), \tag{2}$$

где φ и ψ — функции, выбираемые статистиком, и в качестве оценки для β принимается решение уравнения $L_n(\alpha) = 0$. Эта оценка обозначается $\hat{\beta}_{n,GM}$ и называется GM-оценкой (аббревиатура, возникшая в англоязычной литературе). Частными случаями GM-оценок являются оценки максимального правдоподобия (при $\varphi(x) = x$ и $\psi(x) = -g'(x)/g(x)$) и наименьших квадратов (при $\varphi(x) = \psi(x) = x$). Так же, как в упомянутой ранее классической модели, эти оценки могут не быть робастными, но этого свойства можно добиться выбором функций φ и ψ.

Важнейшим свойством функции $L_n(\alpha)$ является ее асимптотическая локальная линейность

$$\sup_{|\theta| \leq \Theta < \infty} |L_n(\beta + n^{-1/2}\theta) - (L_n(\beta) - C(\beta)\theta)| \xrightarrow{p} 0. \tag{3}$$

Отсюда легко выводится асимптотическая нормальность оценки $\hat{\beta}_{n,GM}$.

Разложение (3) справедливо при условии достаточной гладкости функции $\psi(x)$. Однако при построении робастных процедур используются и негладкие функции, например, $\psi(x) = x$ $(= c)$ при $|x| \leq c$ $(|x| > c)$ или $\psi(x) = \operatorname{sign} x$. Ослабление условий гладкости достигается путем использования случайно взвешенного остаточного эмпирического процесса

$$v_n(\alpha, x) = n^{-1/2} \sum\nolimits_{t=1}^{n} \varphi(u_{t-1}) I(\varepsilon_t(\alpha) \le x), \tag{4}$$

где $I(\cdot)$ — индикатор указанного события. Для $v_n(\alpha, x)$ выполняется асимптотическая локальная линейность, аналогичная свойству (3) для $L_n(\alpha)$, из которой ввиду соотношения

$$L_n(\alpha) = \int_{-\infty}^{\infty} \psi(x)\, dv_n(\alpha, x) \tag{5}$$

следует (3) и асимптотическая нормальность $\hat{\beta}_{n,GM}$.

Сформулированные выше результаты — известные ранее, они приведены для сопоставления с результатами автора.

В главах 2—4 диссертации рассматривается модель засорения процесса авторегрессии (1) (в гл. 2 — для $p = 1$, в гл. 3 — для $p > 1$), когда значения u_t подвергаются аддитивным искажениям, в результате которых наблюдаются

$$y_t = u_t + z_{t,\gamma_n} \xi_t, \tag{6}$$

где $\{u_t\}$ — определенный выше процесс AR(p), ξ_t — случайные величины с неизвестным распределением μ, а z_{t,γ_n} — случайные величины, принимающие значения 1 или 0 с вероятностями $\gamma_n = n^{-1/2}\gamma$ и $1 - \gamma_n$; все случайные величины $u_t, z_{t,\gamma_n}, \xi_t$ независимы между собой.

Основной интерес сосредоточен на задачах проверки гипотез о параметре β: в гл. 2, когда параметр β одномерный, — гипотезы, что β имеет данное значение β_0, а в гл. 3, когда $p > 1$, — гипотезы о том, что данный подвектор вектора $\boldsymbol{\beta}$ обращается в ноль (гипотеза о порядке авторегрессии). В этих главах строятся критерии проверки указанных гипотез и доказывается их «качественная робастность», означающая, что их мощность мало отличается от мощности при отсутствии засорения, если уровень засорения γ мал (точная формулировка будет дана ниже).

В главе 2 изучается мощность критериев проверки $H_0\colon \beta = \beta_0$ при односторонних локальных альтернативах $H_{1n}\colon \beta = \beta_{n,\tau} = \beta_0 + n^{-1/2}\tau$, $\tau > 0$. По формуле (2) с заменой u_t на y_t определяется случайная функция $L_n^Y(\alpha)$. В качестве статистики критерия предлагается

$$\Lambda_{n,Y} = \hat{s}_n^{-1} L_n^Y(\beta_0), \tag{7}$$

где $\hat{s}_n^2$ — состоятельная оценка дисперсии $L_n^Y(\beta_0)$. Исследуется мощность $W_n(\tau, \gamma, \mu)$ этого критерия, т.е. вероятность отвергнуть H_0 при альтернативе $\beta = \beta_{n,\tau}$, уровне засорения γ и распределении засорений μ.

При условиях ограниченности функций φ и ψ доказывается сходимость функции мощности $W_n(\tau, \gamma, \mu)$ к пределу $W(\tau, \gamma, \mu)$, выражаемому через функцию нормального распределения и функцию $\delta(\tau, \gamma, \mu)$, линейно зависящую от γ и τ. Сходимость к пределу выполняется равномерно по $(\tau, \gamma, \mu) \in B_{T,\Gamma}$, т.е. по ограниченным множествам значений τ и γ и, что особенно важно, по всевозможным распределениям μ. Отсюда выводится свойство качественной робастности, состоящее в сходимости

$$W_n(\tau, \gamma, \mu) - W_n(\tau, 0, \mu) \to 0 \quad \text{при} \quad \gamma \to 0,$$

равномерной по $\tau \le T$, всем достаточно большим n и всевозможным μ.

Статистическое значение этого свойства заключается в том, что при его невыполнении даже малое число засоренных наблюдений способно радикально изменить уровень значимости и мощность критерия, в то время как это свойство гарантирует отсутствие такого отрицательного эффекта.

Далее в § 3 этой главы рассматривается аналогичный критерий, основанный на оценке $\hat{\beta}_{n,GM}^Y$, определяемой как решение уравнения $L_n^Y(\alpha) = 0$. Для него доказывается несколько более слабое свойство предельной качественной робастности, состоящее в сходимости предельной мощности

$$W(\tau, \gamma, \mu) \to W(\tau, 0, \mu) \quad \text{при} \quad \gamma \to 0, \tag{8}$$

равномерной по $\tau \le T$.

Наконец, в § 4 главы 2 для той же оценки, что и в § 3, устанавливается асимптотическое распределение при иных условиях: функция ψ не предполагается ограниченной, а требуется существование ограниченной 2-й производной. В этом случае, однако, соответствующий критерий, вообще говоря, не обладает свойством робастности. Полученный результат об асимптотической нормальности оценки $\hat{\beta}_{n,GM}^Y$ является, тем не менее, новым и представляет самостоятельный интерес.

В главе 3 для модели AR(p), $p > 1$, с засорением вида (6) рассматривается задача проверки гипотезы H_0: $\boldsymbol{\beta}^{(2)} = 0$ при неизвестном $\boldsymbol{\beta}^{(1)}$, где $\boldsymbol{\beta}^{(1)}$ и $\boldsymbol{\beta}^{(2)}$ —компоненты разбиения $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})$ на подвекторы размерности m и $p - m$, $1 \le m < p$. Для этой задачи строятся многомерные аналоги критериев, построенных в §§ 1—3 главы 2. Один из них использует статистику $\mathbf{L}_n^Y(\hat{\boldsymbol{\beta}}_{n0,Y})$, где $\mathbf{L}_n^Y$ — многомерный аналог функции (2), примененной к y_t, а $\hat{\boldsymbol{\beta}}_{n0,Y}$ — оценка p-мерного параметра $\boldsymbol{\beta}$, составленная из $\sqrt{n}$-состоятельной оценки $\boldsymbol{\beta}^{(1)}$ и вектора $\mathbf{0}$ на месте $\boldsymbol{\beta}^{(2)}$. Другой критерий основан на GM-оценке $\hat{\boldsymbol{\beta}}_{n,Y}$ параметра $\boldsymbol{\beta}$, получаемой как решение системы уравнений $\mathbf{L}_n^Y(\boldsymbol{\alpha}) = 0$. Доказывается, что распределение той и другой статистики сходится к многомерному нормальному закону, что позволяет в каждом случае построить квадратичную форму, распределение которой сходится к распределению хи-квадрат с $p-m$ степенями свободы. Как следствие, оба критерия эквивалентны друг другу и обладают свойством предельной качественной робастности (8).

В главе 4 проведен численный эксперимент, демонстрирующий поведение различных критериев в модели AR(1) с засорением, изучавшейся в гл. 2. Рассматриваются критерии вида (7), где L_n^Y определяется формулой (2) с y_t на месте u_t и с $\psi(x) = \operatorname{sign} x$, $\varphi(x) = \varphi_b(x)$, где $\varphi_b(x)$, $0 < b \leq \infty$, — семейство функций специального вида, ограниченных за исключением $\varphi_\infty(x) = x$. Оптимальный выбор b^* параметра b зависит от неизвестных распределений инноваций и засорений, и в диссертации предлагается оценка $\hat{b}$ этого значения по результатам наблюдений. Методом численного моделирования сравнивается качество соответствующих критериев $\Lambda_{n,Y}(\hat{b})$ и $\Lambda_{n,Y}(\infty)$. Первый из них — робастный, он показывают достаточно хорошую устойчивость против засорений. Второй, в который входит неограниченная функция $\varphi = \varphi_\infty = x$, оказывается неробастным, в частности, его уровень значимости во многих случаях значительно отличается от номинального.

Численный эксперимент является важным дополнением к полученным автором математическим результатам. В условиях, когда характеристики задачи, такие как распределения инноваций и засорений и интенсивность засорений предполагаются неизвестными, формальная постановка задачи построения «наилучших» статистических решений крайне затруднительна. Кроме того, математические результаты имеют вид предельных теорем с неподдающейся оценке точностью аппроксимации. Поэтому численный эксперимент оказывается по существу единственным способом решения вопроса о границах применимости полученных результатов.

В качестве недостатка следует отметить отсутствие во многих случаях пояснений содержательного смысла полученных результатов и их взаимосвязи между собой. Например, указанный в отзыве факт, что оценка в § 4 главы 2 не является робастной, не отмечается явно автором, он просто не доказывает для нее этого свойства.

Имеется ряд мелких неточностей и опечаток. На стр. 32, строка 3 сн., говорится об условном среднем процесса без указания условия. На стр. 41 цепочка равенств в конце доказательства теоремы 4.1 заканчивается соотношением $(1-\gamma_n)^{n+1} \to 0$, $\gamma \to 0$, перед скобкой пропущено $1-$.

Указанные недостатки не влияют на высокую оценку полученных результатов, обоснование которых опирается на технически сложные математические доказательства.

Сформулируем кратко важнейшие результаты диссертации:

—Получены асимптотические равномерные разложения остаточных эмпирических процессов в схеме засорения. Они могут быть использованы при исследовании робастности не только GM-процедур, но и других процедур, при построении которых используют такие процессы.

—Построен новый робастный GM-тест для проверки гипотез о размерности авторегрессии без использования GM-оценок неизвестных параметров (находить которые обычно достаточно сложно).

—Описан численный алгоритм построения асипмтотически оптимальных GM-тестов. Численный эксперимент подтвердил теоретические результаты о построенных робастных GM-тестах в сравнении с другими процедурами.

Основные результаты диссертации опубликованы в изданиях из списка ВАК России и докладывались на научных конференциях и семинарах. Автореферат точно и полно отражает содержание диссертации.

Результаты диссертации могут быть использованы в исследованиях, проводимых в Институте математики им. С. Л. Соболева СО РАН, Математическом институте им. В. А. Стеклова РАН, Новосибирском государственном университете, Московском государственном университете им. М. В. Ломоносова.

Диссертация выполнена на высоком теоретическом уровне и представляет собой самостоятельное законченное исследование, в котором содержится решение научной задачи, имеющей важное значение для современной математической статистики. Таким образом, можно заключить, что диссертационная работа Д.М. Есаулова «Робастные GM-тесты и оценки в авторегрессионных схемах с выбросами» соответствует требованиям ВАК, предъявляемым к кандидатским диссертациям, а ее автор, Есаулов Даниил Михайлович, заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 01.01.05 - теория вероятностей и математическая статистика.

Отзыв обсужден и утвержден на заседании отдела теории вероятностей и математической статистики Математического института им. В. А. Стеклова РАН 02 ноября 2015 г. (протокол № 1).

Составитель отзыва
ведущий научный сотрудник
отдела теории вероятностей и математической статистики
Математического института им. В. А. Стеклова РАН,
доктор физико–математических наук Д. М. Чибисов
119991, Москва, ул. Губкина, д. 8
Тел. +7 (495) 984 81 41, доб. 37-73
E-mail: chibisov@mi.ras.ru

Заведующий отделом
теории вероятностей и математической статистики
Математического института им. В. А. Стеклова РАН,
доктор физико–математических наук А. С. Холево
119991, Москва, ул. Губкина, д. 8
Тел. +7 (499) 941 01 92
E-mail: holevo@mi.ras.ru