МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи
УДК 519.233.2, 519.233.3

Есаулов Даниил Михайлович

РОБАСТНЫЕ GM-ТЕСТЫ И ОЦЕНКИ В АВТОРЕГРЕССИОННЫХ СХЕМАХ С ВЫБРОСАМИ

01.01.05 — теория вероятностей и математическая статистика

Автореферат

диссертации на соискание учёной степени
кандидата физико-математических наук

Москва 2015

Работа выполнена на кафедре теории вероятностей
механико-математического факультета
Московского государственного университета имени М. В. Ломоносова.

Научный руководитель:

кандидат физико–математических наук,
доцент Болдин Михаил Васильевич

Официальные оппоненты:

доктор физико-математических наук,
профессор Смородина Наталья Васильевна,
Санкт-Петербургский Государственный Университет
физический факультет,
профессор кафедры математики и
математической физики;

кандидат физико-математических наук,
старший научный сотрудник
Мартынов Геннадий Владимирович,
Институт проблем передачи информации
им. А.А. Харкевича, РАН,
Лаборатория № 1 им. М.С.Пинскера;

Ведущая организация:

ФГБУ РАН Математический институт им. В.А. Стеклова

Защита диссертации состоится 27 ноября 2015 года в 16 часов 45 минут на заседании диссертационного совета Д 501.001.85 на базе Московского государственного университета имени М. В. Ломоносова по адресу: Российская Федерация, 119991, Москва, ГСП-1, Ленинские горы, д. 1, МГУ, механико-математический факультет, аудитория 16-24.
С диссертацией можно ознакомиться в Фундаментальной библиотеке МГУ имени М. В. Ломоносова.

Автореферат разослан   октября 2015 г.

Учёный секретарь
диссертационного совета
Д 501.001.85 на базе МГУ,
доктор физико-математических наук   В. В. Власов.

Общая характеристика работы.

Актуальность темы.

Классический статистический анализ конечнопараметрических авторегрессионных схем основан на процедурах наименьших квадратов и родственных им. Для линейных моделей в случае гауссовских инноваций процедуры (оценки и тесты) наименьших квадратов эквивалентны процедурам максимального правдоподобия, а потому являются асимптотически оптимальными. См., например, монографию Brockwell и Davis[1]. Если же инновации имеют негауссовские распределения (например, обладают тяжелыми хвостами), то известны непараметрические асимптотически более эффективные способы оценивания неизвестных параметров и проверки гипотез о них. Например, обобщенные M-процедуры (GM-процедуры), процедуры минимального расстояния (MD-процедуры), ранговые, знаковые процедуры.

Упомянутые процедуры обладают еще одним существенным достоинством: возможны такие их варианты, которые устойчивы к грубым ошибкам (засорениям) в данных. Напротив, оценки и тесты наименьших квадратов весьма чувствительны к засорениям.

Построение новых непараметрических процедур оценивания и проверки гипотез, асимптотически более эффективных, чем процедуры наименьших квадратов, является актуальной теоретической задачей. Важной содержательной задачей также является исследование устойчивости известных и вновь предлагаемых процедур к грубым ошибкам в данных.

Цель настоящей работы — исследовать известные GM-процедуры в линейной авторегрессионной модели на устойчивость к выбросам, а также предложить новые способы проверки гипотез в такой модели в рамках GM-подхода.

Важно, что все перечисленные процедуры (GM, MD, знаково-ранговые) могут быть исследованы в рамках единого подхода — с использованием так называемых остаточных эмпирических процессов (о.э.п.). Интересующие нас статистики могут быть представлены как функционалы от о.э.п. Это позволяет свести изучение асимптотических свойств данных статистик к изучению свойств (равномерных линейных разложений, слабой сходимости в подходящих метрических пространствах и т.д.) соответствующих процессов.

Упомянутые процедуры детально исследовались в случае, когда авторегрессионный процесс наблюдается непосредственно. Так, в AR(p) модели GM-оценки и MD-оценки изучал Koul[2]. Ранговые оценки в модели авторе-

[1] Brockwell P.J., Davis R.A., *Time Series: Theory and Methods,* New York, Springer-Verlag, 1987, 519 p.

[2] Koul H.L., *Weighted empiricals and linear models,* IMS Lecture Notes — Monograph Series, Hayward, CA, Vol. 21, 1992.

грессии исследовались в работах Koul и Ossiander[3], Mukherjee и Bai[4]. Kreiss[5] строил ранговые и M-тесты для проверки линейных гипотез в AR(p) модели. Ранговые тесты в ARMA моделях детально изучались в работах Hallin et al.[6,7]. Знаковые тесты исследовались в монографии Болдина и др.[8] для линейной регрессии и авторегрессии, в работе Болдина и Штутте[9] для ARMA моделей.

В диссертации с помощью о.э.п. мы строим и исследуем GM-процедуры для модели авторегрессии в схеме засорения данных аддитивными одиночными выбросами интенсивности $O(n^{-1/2})$, n — объем данных. Рассматриваемая схема является локальным вариантом общеупотребительной схемы засорения для временных рядов, см. работу Martin, Yohai[10]. Нас интересуют достаточные условия качественной робастности исследуемых процедур против выбросов. Особое внимание уделяется робастности GM-тестов. Робастность GM-оценок в линейных и нелинейных авторегрессионных схемах в отличие от тестов рассматривалась давно. Обычно ее характеризуют функционалом влияния. В настоящее время функционалы влияния для GM-оценок вычислены в ARMA моделях в работе [10], в моделях типа ARCH и GARCH в работах Boldin[11] и Sorokin[12], в нелинейных моделях с авторегрессионными ошибками в статье Sinha et. al[13].

Для GM-тестов в диссертации мы используем две характеризации качественной робастности. Они формулируются в терминах равностепенной непрерывности допредельной мощности (локальная качественная робастность или LQ-робастность) и предельной мощности тестов (предельная качественная робастность). Болдин[14,15] использовал такие характеризации при

[3]Koul H.L., Ossiander M.,Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression, *Ann. Statist.*, Vol. 22, p. 540–562, 1994.

[4]Mukherjee K., Bai Z.D., R-estimation in Autoregression with Square-Integrable Score Function *J. Multivar. Anal.*, Vol. 81, 167–186, 2002.

[5]Kreiss J.-P., Testing linear hypotheses in autoregressions, *Ann. Statist.*, Vol. 18, No. 3, p. 1470–1482, 1990.

[6]Hallin M., Ingenbleek J.-F., Puri M.L., Linear serial rank tests for randomness against ARMA alternatives, *Ann. Statist.*, Vol. 13, No. 3, p. 1156–1181, 1985.

[7]Hallin M., Puri M.L., Aligned Rank Tests for Linear Models with Autocorrelated Error Terms, *J. Multivar. Anal.*, Vol. 50, No. 2, p. 175–237, 1994.

[8]Болдин М.В., Симонова Г.И., Тюрин Ю.Н., *Знаковый статистический анализ линейных моделей,* М.: ФИЗМАТЛИТ, 1997.

[9]Болдин М.В., Штуте В., О знаковых тестах в ARMA модели с возможно бесконечной дисперсией ошибок, *Теория вероятн. и ее примен.*, Т. 49, №3, с. 436–460, 2004.

[10]Martin R.D., Yohai V.J., Influence Functionals for Time Series, *Ann. Statist.*, Vol. 14, p. 781–818, 1986.

[11]Boldin M.V. On empirical processes in heteroscedastic time series and their use for hypothesis testing and estimation, *Math. Methods Statist.*, Vol. 9. p. 65–89, 2000.

[12]Sorokin A.A. On parameter estimation and testing hypotheses on dimension in ARCH(p) model, *Math. Methods Statist.*, Vol. 15, No. 3, p. 327–348., 2006.

[13]Sinha S.K, Field C., Smith B. Robust estimation of nonlinear regression with autoregressive errors, *Statist. and Probab. Letters.*, Vol. 63., p.49–59, 2002.

[14]Boldin M.V., Local robustness of sign tests in AR(1) against outliers, *Math. Methods of Statist.*, 20, 1, 1–13, 2011.

[15]Болдин М.В., Робастность знаковых тестов для гипотез о порядке авторегрессии, *Теория вероятн.*

построении качественно робастных знаковых тестов в авторегрессии. Отметим, что общеупотребительные тесты наименьших квадратов робастными в этих смыслах не являются.

Рассматриваемые определения робастности тестов родственны определению Reider[16] качественной робастности тестов в схеме независимых данных, но Rieder исследовал качественную робастность ранговых тестов в нелокальной схеме.

Цель работы.

1. Получить равномерные линейные разложения остаточных эмпирических процессов в AR(p) модели в локальной схеме засорения данных независимыми одиночными выбросами.

2. С помощью разложений остаточных эмпирических процессов получить достаточные условия качественной робастности известных GM-тестов, основанных на GM-оценках.

3. Построить новый GM-тест (без использования GM-оценок) для проверки линейных гипотез в AR(p) модели и исследовать это тест на качественную робастность.

Научная новизна.

Основные результаты работы являются новыми и состоят в следующем:

1. Получены асимптотические равномерные разложения остаточных эмпирических процессов для авторегрессии в локальной схеме засорения данных аддитивными одиночными выбросами.

2. Установлены достаточные условия качественной робастности известных GM-процедур в локальной схеме засорения.

3. Построен новый GM-тест для проверки гипотез о размерности AR(p) модели без использования GM-оценок неизвестных параметров, и установлены достаточные условия его качественной робастности.

4. Предложен численный алгоритм построения асимптотически оптимальных GM-тестов. Теоретические результаты подтверждены численным экспериментом.

Методы исследования.

В диссертации используются методы математического и функционального анализа, методы теории вероятностей и математической статистики. Метод исследования основан на использовании остаточных эмпирических процессов. При доказательстве основных теорем используется равномерные линейные разложения таких процессов.

Теоретическая и практическая значимость.

Диссертация носит теоретический характер. Её результаты могут быть полезны специалистам по математической статистике, теории временных рядов и эконометрике. Также они могут быть использованы специалистами,

---

и ее примен., 57, 4, 1–10, 2012.

[16] Reider H., A Robust Asymptotic Testing Model, *Ann. Statist.*, Vol. 6, p.1080–1094, 1978.

применяющими модели временных рядов на практике.

Апробация работы.

Основные результаты диссертации неоднократно докладывались на семинаре "Непараметрическая Статистика и Временные Ряды" под руководством проф. Ю.Н. Тюрина, доц. М.В. Болдина и проф. В.Н. Тутубалина в МГУ (2011–2015 гг.). Также были сделаны доклады на нескольких конференциях: Международной конференции студентов, аспирантов и молодых учёных "Ломоносов" в МГУ (Москва, 2010-2011 гг.), Международной конференции "Теория вероятностей и ее приложения" в МГУ, посвященной столетию со дня рождения Б.В.Гнеденко (Москва, 2012 г.), Ломоносовских чтениях (Москва, 2012 г.), X Международной конференции "Компьютерный анализ данных и моделирование" (Минск, 2013 г.), Городском семинаре по теории вероятностей и математической статистике в ПОМИ (Санкт-Петербург, 2015);

Публикации.

Результаты диссертации опубликованы в пяти работах, из которых две — в журналах из перечня ВАК. Список работ приведен в конце автореферата [1]-[5].

Структура и объём работы.

Диссертация состоит из четырех глав, первая из которых - введение, списка обозначений и списка используемой литературы, насчитывающего 76 наименований. Формулы, леммы, теоремы и утверждения будут иметь номер, состоящий из двух чисел. Первое из них соответствует номеру главы, а второе – номеру формулы (леммы, теоремы, утверждения) в данной главе. Ссылки на работы других авторов нумеруются по алфавиту, согласно фамилии первого из них. Общий объем диссертации — 115 страниц.

Краткое содержание диссертации.

В автореферате сохранены оригинальные номера теорем, но номера условий отличны от диссертации.

Первая глава диссертации является введением. Она содержит необходимые определения, обозначения, обзор известных результатов и результатов работы.

Из всего введения приведем основные определения и обозначения.

Линейная авторегрессионная AR(p) модель имеет вид

$$u_t = \beta_1 u_{t-1} + \dots + \beta_p u_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}. \tag{1}$$

В (1) $\{\varepsilon_t\}$ — независимые одинаково распределенные случайные величины (н.о.р.с.в.) с неизвестными функцией распределения G и Лебеговой плотностью g, $\mathbf{E}\varepsilon_1 = 0$, $\mathbf{E}\varepsilon_1^2 < \infty$; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ — вектор неизвестных параметров, для которых характеристическое уравнение, соответствующее (1), имеет корни по модулю меньше единицы. Эти условия гарантируют (см. монографию [1]) существование п.н. единственного строго стационарного решения уравнения (1).

Будем предполагать, что наблюдения содержат грубые выбросы и имеют вид

$$y_t = u_t + z_t^{\gamma_n}\xi_t, \quad t = 1-p, 2-p, \ldots, n. \tag{2}$$

В (2) $\{u_t\}$ — выборка из стационарного решения (1); $\{z_t^{\gamma_n}\}$ — н.о.р.с.в. с распределением Бернулли $\mathrm{Br}\,(\gamma_n)$, $\gamma_n = \min(1, n^{-1/2}\gamma)$, параметр $\gamma \geq 0$ неизвестен; $\{\xi_t\}$ — н.о.р.с.в. с неизвестным распределением μ; последовательности $\{u_t\}$, $\{z_t^{\gamma_n}\}$, $\{\xi_t\}$ независимы между собой. Последовательность $\{\xi_t\}$ интерпретируется как последовательность грубых выбросов (засорений), γ_n — уровень засорения.

Опишем, как строятся GM-оценки и необходимые нам эмпирические процессы сразу для схемы (2) с засорениями.

Для априори выбранных функций φ, ψ и параметра $\boldsymbol{\alpha} \in \mathbb{R}^p$ введем вектор

$$\boldsymbol{L}_n^Y(\boldsymbol{\alpha}) := \left(L_{n1}^Y(\boldsymbol{\alpha}), L_{n2}^Y(\boldsymbol{\alpha}), \ldots, L_{np}^Y(\boldsymbol{\alpha})\right)^T,$$

где

$$L_{nj}^Y(\boldsymbol{\alpha}) := n^{-1/2}\sum_{t=1}^{n}\varphi(y_{t-j})\psi(y_t - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p}).$$

Индекс „Y“ здесь и далее означает, что статистика строится по засоренным данным $\{y_t\}$.

GM-оценка определяется как подходящее (т.е. $n^{1/2}$-состоятельное) решение уравнения $\boldsymbol{L}_n^Y(\boldsymbol{\alpha}) = \mathbf{0}$. Обозначим ее $\widehat{\boldsymbol{\beta}}_{n,GM}^Y$. GM-оценки для авторегрессионных моделей в отсутствии засорений были предложены в работах Denby и Martin[18] и Martin[19,20].

По наблюдениям $\{y_t\}$ построим векторы $\boldsymbol{Y}_{t-1} := (y_{t-1}, y_{t-2}, \ldots, y_{t-p})^T$, $t = 1, \ldots, n$. Введем векторную функцию

$$\boldsymbol{\varphi}(\boldsymbol{Y}_{t-1}) := (\varphi(y_{t-1}), \varphi(y_{t-2}), \ldots, \varphi(y_{t-p}))^T.$$

[18] Denby L., Martin R.D., Robust estimation of the first-order autoregressive parameter, *J. Amer. Statist. Assoc.*, Vol. 74, p. 140–146, 1979.

[19] Martin R.D., Robust estimation of autoregressive models, *Direct. Time Ser.*, Haywood, CA: Institute of Mathematical Statistics, p.228–254, 1980.

[20] Martin R.D., *Robust methods for time series,* Applied Time Series Analysis II, New York: Academic Press, p.683–759, 1981.

Остатками в схеме (2) называются величины $\varepsilon_t^Y(\boldsymbol{\alpha}) := y_t - \boldsymbol{\alpha}^T \boldsymbol{Y}_{t-1}$, $\boldsymbol{\alpha} \in \mathbb{R}^p$, $t = 1, \ldots, n$. Определим остаточный взвешенный эмпирический процесс

$$\boldsymbol{v}_n^Y(\boldsymbol{\alpha}, x) := n^{-1/2} \sum_{t=1}^{n} \boldsymbol{\varphi}(\boldsymbol{Y}_{t-1}) I(\varepsilon_t^Y(\boldsymbol{\alpha}) \le x),$$

$I(\cdot)$— индикатор события, $x \in \mathbb{R}^1$. Тогда

$$\boldsymbol{L}_n^Y(\boldsymbol{\alpha}) = \int_{-\infty}^{\infty} \psi(x) \mathrm{d}\boldsymbol{v}_n^Y(\boldsymbol{\alpha}, x). \tag{3}$$

Введем σ-алгебры

$$\mathcal{F}_{t-1} := \sigma\{\varepsilon_i, i \le t-1; (\xi_j, z_j^{\gamma_n}), 0 \le j \le t\}, \quad t = 1, \ldots, n.$$

Рассмотрим процесс $\boldsymbol{v}_n^Y(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{\theta}, x)$. Условное среднее каждого слагаемого в данной статистике относительно $\mathcal{F}_{t-1}$ равно

$$\boldsymbol{\varphi}(\boldsymbol{Y}_{t-1}) G\big(x + n^{-1/2}\boldsymbol{\theta}^T \boldsymbol{Y}_{t-1} - \eta_t(\boldsymbol{\beta}, \gamma)\big),$$

где

$$\eta_t(\boldsymbol{\beta}, \gamma) := z_t^{\gamma_n}\xi_t - \beta_1 z_{t-1}^{\gamma_n}\xi_{t-1} - \beta_2 z_{t-2}^{\gamma_n}\xi_{t-2} - \cdots - \beta_p z_{t-p}^{\gamma_n}\xi_{t-p}.$$

Введем условно-центрированный процесс

$$\boldsymbol{u}_n^Y(\boldsymbol{\theta}, x) := n^{-1/2} \sum_{t=1}^{n} \boldsymbol{\varphi}(\boldsymbol{Y}_{t-1}) \Big[I\big(\varepsilon_t \le x + n^{-1/2}\boldsymbol{\theta}^T \boldsymbol{Y}_{t-1} - \eta_t(\boldsymbol{\beta}, \gamma)\big) - G\big(x + n^{-1/2}\boldsymbol{\theta}^T \boldsymbol{Y}_{t-1} - \eta_t(\boldsymbol{\beta}, \gamma)\big) \Big]. \tag{4}$$

Слагаемые в (4) образуют мартингал-разность относительно $\{\mathcal{F}_t\}$.

Пусть $\boldsymbol{\beta}$ разбит на подвекторы $\boldsymbol{\beta}^T = (\boldsymbol{\beta}^{(1)T}, \boldsymbol{\beta}^{(2)T})$, где $\boldsymbol{\beta}^{(i)}$, $i = 1, 2$, имеют размерности m и $p - m$ соответственно, $1 \le m < p$. Одной из основных задач работы является исследование тестов для проверки гипотезы о размерности модели (1) $H_0\colon \boldsymbol{\beta}^{(2)} = \mathbf{0}$. Альтернативой возьмем $H_1\colon \boldsymbol{\beta}^{(2)} \ne \mathbf{0}$. Таким образом, $\boldsymbol{\beta}^{(1)}$ является мешающим параметром. Если гипотеза H_0 верна, то размерность (1) не превышает m. Мощность теста будем исследовать при локальных альтернативах $H_{1n}(\boldsymbol{\tau})\colon \boldsymbol{\beta} = \boldsymbol{\beta}_n := \boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{\tau}$, где $\boldsymbol{\beta}_0^T = (\boldsymbol{\beta}^{(1)T}, \mathbf{0}^T)$, $\boldsymbol{\tau}^T = (\boldsymbol{\tau}^{(1)T}, \boldsymbol{\tau}^{(2)T}) \in \mathbb{R}^p$ — постоянный вектор с подвекторами размерности m и $p - m$ соответственно.

Мощность теста в схеме (2) на локальной альтернативе $H_{1n}(\boldsymbol{\tau})$ обозначим $W_n(\boldsymbol{\tau}, \gamma, \mu)$. Тест называется *LQ-робастным,* если семейство $\{W_n(\boldsymbol{\tau}, \gamma, \mu)\}$ равностепенно непрерывно по γ в точке $\gamma = 0$. То есть,

$$\sup |W_n(\boldsymbol{\tau}, \gamma, \mu) - W_n(\boldsymbol{\tau}, 0, \mu)| \to 0,\ \gamma \to 0. \tag{5}$$

Супремум в (5) берется по произвольным μ, $\|\boldsymbol{\tau}\| \le T < \infty$ и $n \ge n_0(T)$. Здесь $\|\cdot\|$— евклидова норма вектора, а $n_0(T)$ — наименьшее целое число такое, что корни характеристического многочлена, соответствующего (1), по модулю меньше единицы для всех $\|\boldsymbol{\tau}\| \le T$.

Соотношение (5) означает, что для малых γ равномерно по μ, $\|\boldsymbol{\tau}\| \le T$ и $n \ge n_0$ близки уровни значимости и мощности тестов в схемах с засорениями и без засорений, т.е. при $\gamma = 0$.

Предположим, что существует $W(\boldsymbol{\tau}, \gamma, \mu) := \lim_{n\to\infty} W_n(\boldsymbol{\tau}, \gamma, \mu)$ — предельная мощность теста. Будем называть тест *предельно качественно робастным,* если

$$\sup_{\mu, \|\boldsymbol{\tau}\| \le T} |W(\boldsymbol{\tau}, \gamma, \mu) - W(\boldsymbol{\tau}, 0, \mu)| \to 0,\ \gamma \to 0. \tag{6}$$

Вторая глава диссертации состоит из пяти параграфов. Она посвящена построению GM-процедур для стационарной AR(1) модели в случае, когда наблюдения содержат выбросы. Таким образом, рассматривается модель (1) в схеме засорений (2) при $p = 1$. В этой главе мы будем писать β_n вместо $\boldsymbol{\beta}_n$, $L_n(\alpha)$ вместо $\boldsymbol{L}_n(\boldsymbol{\alpha})$ и т.п.

По засоренным наблюдениям $\{y_t\}$ мы будем проверять гипотезу $H_0\colon \beta = \beta_0$ против правосторонних альтернатив $H_1^+\colon \beta > \beta_0$. Мощность тестов будем исследовать при локальных альтернативах $H_{1n}(\tau)\colon \beta = \beta_n := \beta_0 + n^{-1/2}\tau, \tau > 0$. Удобно дальше полагать $\tau \ge 0$, так что $H_{1n}(0)$ есть H_0. Кроме того, будем предполагать, что $n \ge n_\tau$, где n_τ есть наименьшее натуральное число, при котором $|\beta_n| < 1$ для $n \ge n_\tau$. Это условие гарантирует существование п.н. единственного строго стационарного решения уравнения (1) при $p = 1$.

Непараметрическая GM-оценка в рассматриваемой схеме определяется как корень уравнения

$$L_n^Y(\alpha) := \int_{-\infty}^{\infty} \psi(x) dv_n^Y(\alpha, x) = n^{-1/2} \sum_{t=1}^{n} \varphi(y_{t-1})\psi(y_t - \alpha y_{t-1}) = 0 \tag{7}$$

для априори выбранных функций φ и ψ.

Нам понадобятся следующие условия.

Условие 1. G дважды дифференцируема с производной g, $g(x) > 0$ при всех x, $\sup_x |g'(x)| < \infty$.

Условие 2. $\sup_x |\varphi(x)| < \infty$.

Пусть $B_{T,\Gamma} := \{(\tau, \gamma, \mu)\colon 0 \le \tau \le T < \infty,\ 0 \le \gamma \le \Gamma < \infty,\ \mu - \text{любое}\}$. Обозначим $\{u_t^0\}$ — стационарное решение (1) при H_0, $p = 1$.

Основной результат первого параграфа составляет теорема о линейном по τ и γ асимптотическом разложении эмпирического процесса $v_n^Y(\alpha, x)$, равномерном по $0 \le \tau \le T$, $0 \le \gamma \le \Gamma$ и любым μ. А именно,

Теорема 2.2. *Пусть выполнены Условия 1–2 и функция φ п.в. непрерывна. Пусть верна $H_{1n}(\tau)$. Тогда при любых $\delta > 0$, $0 \le \Gamma < \infty$, $0 \le T < \infty$*

$$\sup_{(\tau,\gamma,\mu)\in B_{T,\Gamma}} P_{\beta_n}\Big(\sup_{x\in\mathbb{R}}\big|v_n^Y(\beta_0,x) - v_n(\beta_n,x) \\ + g(x)\mathbf{E}[u_1^0\varphi(u_1^0)]\tau - \mathbf{E}[\rho(x,\mu)]\gamma\big| > \delta\Big) \to 0,\ n\to\infty,$$

где $\rho(x,\mu) := \varphi(u_1^0+\xi_1)G(x+\beta_0\xi_1) + \varphi(u_1^0)G(x-\xi_2) - 2\varphi(u_1^0)G(x)$.

Доказательство этого утверждения использует Теорему 2.1 об асимптотическом равномерном разложении соответствующего условно-центрированного эмпирического процесса.

Используя полученные результаты, во втором параграфе мы строим тест для проверки гипотезы $H_0\colon \beta = \beta_0$. В качестве тестовой статистики берется

$$\Lambda_{n,Y} = \Lambda_{n,Y}(\beta_0) := \hat{s}_n^{-1} L_n^Y(\beta_0), \tag{8}$$

где $\hat{s}_n^2$ — состоятельная оценка параметра $\sigma^2(\beta_0) := \mathbf{E}\varphi^2(u_1^0)\psi^2(\varepsilon_1)$.

Сформулируем условие на функцию ψ.

Условие 3. Вариация $\mathrm{Var}\,|_{-\infty}^{\infty}\,[\psi] < \infty$, $\mathbf{E}\psi(\varepsilon_1) = 0$.

Обозначим

$$\Delta_1 = \Delta_1(\beta_0) := \mathbf{E}[u_1^0\varphi(u_1^0)]\int_{-\infty}^{\infty} g(x)d\psi(x),$$
$$\Delta_2 = \Delta_2(\beta_0,\mu) := \mathbf{E}\varphi(u_1^0)\mathbf{E}\psi(\varepsilon_1+\xi_1) + \mathbf{E}\varphi(u_1^0+\xi_1)\psi(\varepsilon_2-\beta_0\xi_1).$$

Положим $\delta(\tau,\gamma,\mu) := \sigma^{-1}(\beta_0)[\Delta_1\tau + \Delta_2\gamma]$.

Теорема 2.2 и соотношение (3) влекут асимптотическую равномерную линейность (AUL) статистики $L_n^Y(\beta_0)$ (Следствие 2.1). Используя Следствие 2.1, доказывается следующая теорема о равномерной слабой сходимости статистики (8) к нормальному закону.

Теорема 2.3. *Пусть верна $H_{1n}(\tau)$, $\tau \ge 0$. Пусть выполнены Условия 1–3 и функция φ п.в. непрерывна. Тогда*

$$\sup|P_{\beta_n}(\Lambda_{n,Y} \le x) - \Phi(x - \delta(\tau,\gamma,\mu))| \to 0, \quad n\to\infty. \tag{9}$$

Супремум в (9) *берется по $x\in\mathbb{R}$ и $(\tau,\gamma,\mu)\in B_{T,\Gamma}$.*

В силу Теоремы 2.3 при $\gamma = 0$ при гипотезе H_0 $\Lambda_{n,Y} \xrightarrow{d_{\beta_0}} N(0,1)$, $n\to\infty$. Отвергать H_0 будем при $\Lambda_{n,Y} > t_{1-\alpha}$, $t_{1-\alpha}$ — квантиль $\Phi(x)$ уровня $1-\alpha$. В схеме без засорений такой тест имеет асимптотический уровень α. При произвольном γ его мощность на $H_{1n}(\tau)$ есть $W_n(\tau,\gamma,\mu) := P_{\beta_n}(\Lambda_{n,Y} > t_{1-\alpha})$.

Простым следствием Теоремы 2.3 является равномерная по $0 \le \tau \le T < \infty$, $0 \le \gamma \le \Gamma < \infty$ и произвольным μ сходимость

$$W_n(\tau,\gamma,\mu) \to W(\tau,\gamma,\mu) := 1 - \Phi(t_{1-\alpha} - \delta(\tau,\gamma,\mu)), \quad n \to \infty. \qquad (10)$$

Очевидно, при ограниченных φ и ψ и любого $T \ge 0$ выполнено соотношение вида (6), т.е семейство предельных мощностей $\{W(\tau,\gamma,\mu)\}$ равностепенно непрерывно по γ в точке $\gamma = 0$.

Соотношения (6) и (10) позволяют установить (5), т.е. LQ-робастность построенного теста (Теорема 2.4.).

В отличие от второго параграфа, где GM-тесты строятся без использования оценок неизвестного параметра, в третьем параграфе тесты основаны на заранее построенных GM-оценках. Сформулируем сначала теорему о существовании и асимптотической гауссовости таких оценок в схеме (2) при $p = 1$. Нам понадобятся дополнительные условия на функции φ, ψ и распределение выбросов.

Условие 4. $\int_{-\infty}^{\infty} g(x)d\psi(x)\, \mathbf{E}[\varphi(u_1^0)\, u_1^0] \ne 0$.

Условие 5. $\mathbf{E}\xi_1^2 < \infty$.

Теорема 2.5. *Пусть выполнены Условия 1–5. Пусть, кроме того, функции $\varphi(x)$ и $\psi(x)$ непрерывны. Пусть верна альтернатива $H_{1n}(\tau)$. Тогда с вероятностью, стремящейся к 1, уравнение (7) имеет $n^{1/2}$-состоятельное решение $\widehat{\beta}_{n,GM}^{Y}$, для которого*

$$n^{1/2}(\widehat{\beta}_{n,GM}^{Y} - \beta_n) \xrightarrow{d} N\big(\gamma\, \Delta_1^{-1}\Delta_2, \Delta_1^{-2}\sigma^2(\beta_0)\big).$$

Для доказательства Теоремы 2.5 используется асимптотическая равномерная по $|\theta| \le \Theta$ $(0 \le \Theta < \infty)$ линейность $L_n^Y(\beta_n + n^{-1/2}\theta)$ (Теорема 2.7). Теорема 2.7, в свою очередь, является следствием соотношения (3) и Теоремы 2.6 об асимптотическом равномерном по $|\theta| \le \Theta$ разложении о.э.п. (4) при $p = 1$.

Пусть $\hat{d}_n^2$ — произвольная состоятельная оценка $\Delta_1^{-2}\sigma^2(\beta_0)$. Тестовой статистикой для H_0 возьмем $T_{n,Y} := \hat{d}_n^{-1} n^{1/2}(\widehat{\beta}_{n,GM}^{Y} - \beta_0)$. Очевидным следствием Теоремы 2.5 является асимптотическая нормальность $T_{n,Y}$ на альтернативе $H_{1n}(\tau)$, а именно $T_{n,Y} \xrightarrow{d} N(\delta(\tau,\gamma,\mu), 1)$, $n \to \infty$. Таким образом, в силу (9) GM-тесты, основанные на $T_{n,Y}$ и $\Lambda_{n,Y}$, асимптотически эквивалентны. Это, в частности, означает, что для предельной мощности теста, основанного на $T_{n,Y}$, выполняется соотношение (6). Отсюда следует, что построенный тест также является предельно качественно робастным.

Четвертый параграф посвящен случаю, когда целевая функция ψ является гладкой. Это важно, т.к. в этом случае удается отказаться от условия ограниченности ψ. Такая ситуация включает в себя, например, случай

$\varphi(x) = \psi(x) = x$, при котором уравнение (7) определяет оценку наименьших квадратов, и случай $\varphi(x) = x, \psi(x) = -\frac{g'(x)}{g(x)}$, при котором (7) становится уравнением максимального правдоподобия.

Итак, сформулируем новое условие для функции ψ вместо Условия 3.

Условие 6. $\psi(x) \in C^2$, $\sup_x\{|\psi'(x)|, |\psi''(x)|\} < \infty$, $\mathbf{E}\psi(\varepsilon_1) = 0$.

При выполненных Условиях 2, 5–6 доказательство AUL статистки $L_n^Y(\beta_n + n^{-1/2}\theta)$ (Утверждение 2.2) проводится с помощью обычной формулы Тейлора для самой функции $\psi(x)$ без использования свойства AUL остаточных процессов $v_n^Y(\alpha, x)$. Утверждение 2.2 в свою очередь влечет аналог Теоремы 2.5 для случая гладкой ψ.

В пятый параграф для удобства вынесены громоздкие доказательства утверждений первого и второго параграфов.

Третья глава диссертации состоит из трех параграфов. В ней мы исследуем GM-процедуры в общей линейной авторегрессионной AR(p) модели (1).

Основными задачами главы является построение нового непараметрического GM-теста для проверки гипотезы о порядке этой модели в схеме (2) и исследование его робастности против грубых выбросов.

Напомним, что проверяется гипотеза H_0: $\boldsymbol{\beta}^{(2)} = 0$, где $\boldsymbol{\beta}^{(2)} \in \mathbb{R}^{p-m}$ — компонента векторного параметра $\boldsymbol{\beta}^T = (\boldsymbol{\beta}^{(1)T}, \boldsymbol{\beta}^{(2)T})$, $1 \le m < p$. Отметим, что тест, который мы строим, может быть основан на любой предварительной $n^{1/2}$-состоятельной оценке $\widehat{\boldsymbol{\beta}}_n$ параметра $\boldsymbol{\beta}$, например, на оценке наименьших квадратов.

Первый параграф посвящен построению нового теста в схеме (1) без засорений. Для краткости мы приведем результаты сразу для более общего случая из второго параграфа, в котором рассматривается модель (1) в схеме засорений (2). Выводы первого параграфа могут быть получены из второго, если взять $\gamma = 0$. Отметим, однако, что результаты в случае без засорений (Теоремы 3.1-3.3) получены при более слабых условиях на функцию φ.

Основным результатом первого параграфа является асимптотическое равномерное разложение процесса $\boldsymbol{u}_n^Y(\boldsymbol{\theta}, x)$. Соответствующее утверждение в схеме без засорений является следствием более общей Теоремы 2 из Boldin[21].

Теорема 3.4. *Пусть выполнены Условия 1, 2 и 5. Пусть верна* $H_{1n}(\boldsymbol{\tau})$. *Тогда для любого* $0 \le \Theta < \infty$

$$\sup_{\|\boldsymbol{\theta}\| \le \Theta,\ x \in \mathbb{R}^1} \|\boldsymbol{u}_n^Y(\boldsymbol{\theta}, x) - \boldsymbol{u}_n^Y(\mathbf{0}, x)\| \xrightarrow{P} 0, \quad n \to \infty.$$

Этот результат является обобщением Теоремы 2.6 на случай многопараметрической авторегрессии.

[21] Boldin M.V., On sequential residual empirical processes in heteroscedastic time series, *Math. Methods of Statist.*, Vol. 11, No. 4, 453–464, 2002.

Мы используем Теорему 3.4 и соотношение (3) для доказательства AUL процесса $\boldsymbol{L}_n^Y(\boldsymbol{\alpha})$ (Теорема 3.5).

Пусть $u_{1-p}^0, \ldots, u_n^0$ — выборка из строго стационарного решения уравнения (1) при $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Введем статистику $\widetilde{\boldsymbol{L}}_n(\boldsymbol{\beta}_0) := n^{-1/2} \sum_{t=1}^n \boldsymbol{\varphi}(\boldsymbol{U}_{t-1}^0)\psi(\varepsilon_t)$ и матрицу $\mathbf{C} := \int_{-\infty}^{\infty} g(x) \mathrm{d}\psi(x)\, \mathbf{E}\boldsymbol{\varphi}(\boldsymbol{U}_0^0)(\boldsymbol{U}_0^0)^T$.

Введем вектор

$$\boldsymbol{\Delta}(\mu) := \big(\Delta_1(\mu), \Delta_2(\mu), \ldots, \Delta_p(\mu)\big)^T, \text{ где}$$
$$\Delta_j(\mu) := \mathbf{E}\varphi(u_{2-j}^0 + \xi_{2-j})\psi(\varepsilon_2 - \beta_{0j}\xi_{2-j}) + \mathbf{E}\varphi(u_{2-j}^0)\mathbf{E}\psi(\varepsilon_1 + \xi_1)$$
$$+ \sum_{i=1,\ i\neq j}^{p} \mathbf{E}\varphi(u_{2-j}^0)\mathbf{E}\psi(\varepsilon_2 - \beta_{0i}\xi_{2-i}),$$

Вектор $\boldsymbol{\Delta}(\mu)$ будет характеризовать асимптотическое влияние засорений на вектор $\boldsymbol{L}_n^Y(\boldsymbol{\alpha})$ и тестовую статистику.

В силу Теоремы 3.5 для произвольной $n^{1/2}$-состоятельной оценки $\widehat{\boldsymbol{\beta}}_{n,Y}$ при $H_{1n}(\boldsymbol{\tau})$ справедливо следующее разложение:

$$\boldsymbol{L}_n^Y(\widehat{\boldsymbol{\beta}}_{n,Y}) = \widetilde{\boldsymbol{L}}_n(\boldsymbol{\beta}_0) - \mathbf{C}n^{1/2}(\widehat{\boldsymbol{\beta}}_{n,Y} - \boldsymbol{\beta}_n) + \gamma\boldsymbol{\Delta}(\mu) + \boldsymbol{o}_p(1). \tag{11}$$

Пусть $\widehat{\boldsymbol{\beta}}_{n0,Y}^T := (\widehat{\boldsymbol{\beta}}_{n0,Y}^{(1)T}, \mathbf{0}^T)$, где $\widehat{\boldsymbol{\beta}}_{n0,Y}^{(1)} \in \mathbb{R}^m$ — $n^{1/2}$–состоятельная оценка параметра $\boldsymbol{\beta}^{(1)}$, построенная по $\{y_t\}$. Введем вектор $\boldsymbol{a}(\mu) := \mathbf{C}^{-1}\boldsymbol{\Delta}(\mu)$.

Обозначим $\pi\circ\boldsymbol{\alpha}$ — проекцию на последние $(p-m)$ координат вектора $\boldsymbol{\alpha} \in \mathbb{R}^p$. Используя (11) и $\widehat{\mathbf{C}}_{n,Y}$, состоятельную оценку невырожденной матрицы $\mathbf{C}$, получаем

$$\pi \circ \widehat{\mathbf{C}}_{n,Y}^{-1}\boldsymbol{L}_n^Y(\widehat{\boldsymbol{\beta}}_{n0,Y}) = \pi \circ \mathbf{C}^{-1}\widetilde{\boldsymbol{L}}_n(\boldsymbol{\beta}_0) + \boldsymbol{\tau}^{(2)} + \gamma\boldsymbol{a}^{(2)} + \boldsymbol{o}_p(1), \tag{12}$$

где $\boldsymbol{a}^{(2)}$ — вектор размерности $(p-m)$ такой, что $\boldsymbol{a}^T(\mu) = (\boldsymbol{a}^{(1)T}, \boldsymbol{a}^{(2)T})$.

Введем матрицу $\mathbf{K} := \mathbf{E}\psi^2(\varepsilon_1)\, \mathbf{E}\boldsymbol{\varphi}(\boldsymbol{U}_0^0)\boldsymbol{\varphi}^T(\boldsymbol{U}_0^0)$. В силу ЦПТ для мартингал-разностей имеет место сходимость $\widetilde{\boldsymbol{L}}_n(\boldsymbol{\beta}_0) \xrightarrow{d} N(0, \mathbf{K}), n \to \infty$. Для матриц $\mathbf{K}$ и $\mathbf{C}$ нам потребуется выполнение следующего условия.

Условие 7. $\det \mathbf{C} \neq 0$, $\det \mathbf{K} > 0$.

Пусть $\mathbf{J}$ — такая матрица размерности $(p-m)\times(p-m)$, что

$$\mathbf{C}^{-1}\mathbf{K}(\mathbf{C}^{-1})^T := \begin{pmatrix} \mathbf{Q} & \mathbf{B} \\ \mathbf{B^T} & \mathbf{J} \end{pmatrix}.$$

В качестве статистики нового теста для H_0 рассматривается

$$\Lambda_{n,Y}^{\pi} := \left[\pi \circ \widehat{\mathbf{C}}_{n,Y}^{-1}\boldsymbol{L}_n^Y(\widehat{\boldsymbol{\beta}}_{n0,Y})\right]^T \widehat{\mathbf{J}}_{n,Y}^{-1} \left[\pi \circ \widehat{\mathbf{C}}_{n,Y}^{-1}\boldsymbol{L}_n^Y(\widehat{\boldsymbol{\beta}}_{n0,Y})\right], \tag{13}$$

$\widehat{\mathbf{J}}_{n,Y}$ — произвольная состоятельная при $H_{1n}(\boldsymbol{\tau})$ оценка $\mathbf{J}$.

Обозначим нецентральное распределение хи-квадрат с $p-m$ степенями свободы и параметром нецентральности λ^2 как $\chi^2(p-m, \lambda^2)$. Центральное распределение будем обозначать $\chi^2(p-m)$. Используя разложение (12), доказывается следующая

Теорема 3.6. *Пусть выполнены Условия 1–3, 5, 7 и φ п.в. непрерывна. Тогда при альтернативе $H_{1n}(\boldsymbol{\tau})$*

$$\Lambda^{\pi}_{n,Y} \xrightarrow{d} \chi^2(p-m, \lambda^2),$$

где параметр нецентральности $\lambda^2 = \left\| \mathbf{J}^{-1/2}(\boldsymbol{\tau}^{(2)} + \gamma \boldsymbol{a}^{(2)}) \right\|^2$.

Отвергать H_0 будем при $\Lambda^{\pi}_{n,Y} > \chi^{p-m}_{1-\alpha}$, где $\chi^{p-m}_{1-\alpha}$ — $(1-\alpha)$–квантиль $\chi^2(p-m)$. Мощность теста, основанного на $\Lambda^{\pi}_{n,Y}$, при альтернативе $H_{1n}(\boldsymbol{\tau})$ есть $W_n(\boldsymbol{\tau}, \gamma, \mu) = P_{\boldsymbol{\beta}_n}(\Lambda^{\pi}_{n,Y} > \chi^{p-m}_{1-\alpha})$.

Обозначим функцию распределения $\chi^2(p-m, \lambda^2)$ через $F_{p-m}(x, \lambda^2)$. В силу Теоремы 3.6

$$\lim_{n \to \infty} W_n(\boldsymbol{\tau}, \gamma, \mu) = W(\boldsymbol{\tau}, \gamma, \mu) = 1 - F_{p-m}(\chi^{p-m}_{1-\alpha}, \lambda^2),$$

и тест имеет асимптотический уровень значимости α. $W_n(\boldsymbol{\tau}, 0, \mu)$ — мощность статистического теста в схеме (1) без засорений. Обозначим $\mathfrak{M}_2$ класс распределений с конечным вторым моментом. Для предельной мощности оказывается верна

Теорема 3.7. *Пусть выполнены условия Теоремы 3.6. Тогда*

$$\sup_{\mu \in \mathfrak{M}_2} |W(\boldsymbol{\tau}, \gamma, \mu) - W(\boldsymbol{\tau}, 0, \mu)| \to 0, \quad \gamma \to 0. \tag{14}$$

Таким образом, семейство предельных мощностей $\{W(\boldsymbol{\tau}, \gamma, \mu)\}_{\mu \in \mathfrak{M}_2}$ равностепенно непрерывно по γ в точке $\gamma = 0$. Свойство (14) характеризует предельную качественную робастность теста (13) против выбросов.

Отметим, что предложенные в [5] „general score tests" также были основаны на специальным образом преобразованной статистике $\boldsymbol{L}_n(\widehat{\boldsymbol{\beta}}_n)$, где $\widehat{\boldsymbol{\beta}}_n$ — произвольная $n^{1/2}$-состоятельная оценка. Причем при $\varphi(x) = x$ наша тестовая статистика (13) в схеме без засорений совпадает со статистикой из [5]. Однако преобразование $\boldsymbol{L}_n(\widehat{\boldsymbol{\beta}}_n)$ из [5] в случае произвольной φ неприменимо. Наш способ позволяет строить тесты, в частности, для ограниченных φ. Кроме того, в отличие от [5] мы рассматриваем случай, вообще говоря, негладкой ψ.

В третьем параграфе в схеме засорений (2) исследуется GM-тест, основанный на GM-оценке $\widehat{\boldsymbol{\beta}}^{Y}_{n,GM}$ — $n^{1/2}$-состоятельном решении нелинейной системы уравнений $\boldsymbol{L}^{Y}_n(\boldsymbol{\alpha}) = \mathbf{0}$. Основным утверждением параграфа является

Теорема 3.8. *Пусть выполнены условия 1–3, 5, 7, функции* $\varphi(x)$ *и* $\psi(x)$ *непрерывны. Пусть верна альтернатива* $H_{1n}(\boldsymbol{\tau})$. *Тогда с вероятностью, стремящейся к 1, система уравнений* $\boldsymbol{L}_n^Y(\boldsymbol{\alpha}) = \mathbf{0}$ *имеет* $n^{1/2}$*-состоятельное решение* $\widehat{\boldsymbol{\beta}}_{n,GM}^Y$, *для которого*

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_{n,GM}^Y - \boldsymbol{\beta}_n) \xrightarrow{d} N\big(\gamma \boldsymbol{a}(\mu), \mathbf{C}^{-1}\mathbf{K}(\mathbf{C}^{-1})^T\big), \ n \to \infty.$$

Теперь в качестве последовательности тестовых статистик возьмем

$$V_{n,Y} := n\,\widehat{\boldsymbol{\beta}}_{n,GM}^{Y,(2)T}(\widehat{\mathbf{J}}_{\mathbf{n},\mathbf{Y}})^{-1}\widehat{\boldsymbol{\beta}}_{n,GM}^{Y,(2)}. \tag{15}$$

В силу Теоремы 3.8 $V_{n,Y} \xrightarrow{d} \chi^2(p-m, \lambda^2)$, $n \to \infty$, где параметр нецентральности λ^2 такой же, как и в Теореме 3.6. Таким образом, статистические тесты, основанные на $\Lambda_{n,Y}^{\pi}(\widehat{\boldsymbol{\beta}}_{n0})$ из (13) и $V_{n,Y}$ из (15) являются асимптотически эквивалентными. Это означает, что последний тест также является предельно качественно робастным.

Четвертая глава диссертации состоит из трех параграфов.

В первом параграфе, который является вспомогательным, излагаются известные результаты об асимптотической оптимальности тестов, построенных в Главе 2 для AR(1) модели в схеме засорений (2). Они принадлежат М.В. Болдину и содержатся в совместной публикации [4] (раздел 3). Они необходимы нам для второго параграфа, в котором излагаются результаты численного эксперимента на моделированных данных.

Дополнительно к Условиям 1–3 полагаем, что выполнено

Условие 8. $\mathbf{E}\varphi(u_1^0) = 0$, $\mathbf{E}[u_1^0\varphi(u_1^0)] > 0$, $\int_{-\infty}^{\infty} g(x)d\psi(x) > 0$.

Тогда в силу (9) на альтернативе $H_{1n}(\tau)$ верна равномерная слабая сходимость статистики $\Lambda_{n,Y}$ из (8) к нормальному закону с дисперсией 1 и средним $\delta_1(\tau,\beta_0) + \delta_2(\gamma,\mu,\beta_0)$, где

$$\delta_1 := \frac{\mathbf{E}[u_1^0\varphi(u_1^0)]}{[\mathbf{E}\varphi^2(u_1^0)]^{1/2}} \frac{\int g d\psi}{[\mathbf{E}\psi^2(\varepsilon_1)]^{1/2}}\,\tau, \quad \delta_2 := \frac{\mathbf{E}\varphi(u_1^0+\xi_1)\psi(\varepsilon_2 - \beta_0\xi_1)}{[\mathbf{E}\varphi^2(u_1^0)]^{1/2}\,[\mathbf{E}\psi^2(\varepsilon_1)]^{1/2}}\,\gamma.$$

Сдвиг $\delta_1(\tau,\beta_0)$ определяется альтернативой $H_{1n}(\tau)$, а $\delta_2(\gamma,\mu,\beta_0)$ характеризует асимптотическое влияние засорений на $\Lambda_{n,Y}$.

Для $b > 0$ введем $\varphi_b(x) := (x/b)\min(1, b/|x|)$ — усечение функции x/b на уровнях 1 и -1. Пусть $\varphi_0(x) := \operatorname{sign} x$. Для $b \geq 0$ положим $e_\varphi(b) := 1/[\mathbf{E}\varphi_b^2(u_1^0)]^{1/2}$. Введем для $c \geq 0$ функции $\psi_c(x) := \varphi_c\big(-\frac{g'}{g}(x)\big)$ и $l_\psi(c) := 1/[\mathbf{E}\psi_c^2(\varepsilon_1)]^{1/2}$.

Поиск оптимальных φ и ψ, удовлетворяющих описанным выше условиям, ведется на классах функций

$$\mathcal{K}_\varphi(b) := \Big\{\varphi\colon \frac{\sup_x|\varphi(x)|}{[\mathbf{E}\varphi^2(u_1^0)]^{1/2}} \leq e_\varphi(b)\Big\}, \mathcal{K}_\psi(c) := \Big\{\psi\colon \frac{\sup_x|\psi(x)|}{[\mathbf{E}\psi^2(\varepsilon_1)]^{1/2}} \leq l_\psi(c)\Big\}.$$

Параметры $b, c \geq 0$ заданы априори (алгоритм выбора их значений описан во втором параграфе). Заметим, что при $\varphi \in \mathcal{K}_\varphi(b)$, $\psi \in \mathcal{K}_\psi(c)$ сдвиг $\delta_2(\gamma, \mu, \beta_0)$ ограничен для фиксированных γ. Сдвиг $\delta_1(\tau, \beta_0)$ достигает максимума по φ, ψ на паре функций φ_b, ψ_c (Лемма 4.2).

Предельную мощность $W(\tau, \gamma, \mu)$ теста со статистикой $\Lambda_{n,Y}$ переобозначим через $W(\tau, \gamma, \mu|\varphi, \psi)$. Основными характеристиками теста являются $\sup_\mu W(0, \gamma, \mu|\varphi, \psi)$ — его наибольший асимптотический объем и $\inf_\mu W(\tau, \gamma, \mu|\varphi, \psi)$ — его наименьшая асимптотическая локальная мощность. Следствием Леммы 4.2 является Теорема 4.1, которая и описывает оптимальность GM-теста основанного на $\Lambda_{n,Y}$ с целевыми функциями φ_b и ψ_c. А именно, для любых $\varphi \in \mathcal{K}_\varphi(b)$, $\psi \in \mathcal{K}_\psi(c)$ при $\tau > 0$ и $0 \leq \gamma < \gamma_0(\varphi, \psi)$ выполнены соотношения

$$\sup_\mu W(0, \gamma, \mu|\varphi, \psi) \leq \Phi(t_\alpha + e_\varphi(b) l_\psi(c)\gamma); \tag{16}$$
$$\inf_\mu W(\tau, \gamma, \mu|\varphi_b, \psi_c) \geq \inf_\mu W(\tau, \gamma, \mu|\varphi, \psi).$$

Во втором параграфе приводится алгоритм выбора подходящих значений параметров b и c для статистики $\Lambda_{n,Y}(\varphi_b(x), \psi_c(x))$.

Отметим, что уравнение $\hat{s}_n^{-1} n^{-1/2} \sum_{t=1}^n \varphi_b(y_{t-1}) \operatorname{sign}(y_t - \theta y_{t-1}) \div 0$, соответствующее $\Lambda_{n,Y}(\varphi_b(x), \psi_0(x))$, определяет взвешенную оценку наименьших модулей (LDW-оценку). Здесь символ $\div$ означает переход через ноль. LDW-оценки были введены в [8] (§5.5.3). Тест со статистикой $\Lambda_{n,Y}(b) := \Lambda_{n,Y}(\varphi_b(x), \psi_0(x))$ по аналогии будем называть LDW-тестом. В случае $\gamma = 0$ соответствующую статистику обозначим $\Lambda_n(b)$. Тест наименьших модулей (LAD-тест) соответствует статистике

$$\Lambda_{n,Y}(\infty) := \hat{s}_{n,LAD}^{-1} n^{-1/2} \sum_{t=1}^n y_{t-1} \operatorname{sign}(y_t - \beta_0 y_{t-1}),$$

где $\hat{s}_{n,LAD}^2 = n^{-1} \sum_{t=1}^n y_{t-1}^2$.

Зафиксировав $c = 0$ мы, таким образом, рассматриваем класс асимптотически оптимальных LDW-тестов. Прежде чем для этого класса описывать численный метод поиска подходящего значения b при неизвестном распределении $\{\varepsilon_t\}$, мы рассмотрим вспомогательный случай, когда это распределение известно. Оптимальное $b = b^*$ в этом случае выбирается среди всех значений, для которых верхняя граница наибольшего асимптотического объема GM-теста из (16) при всех $\gamma \leq \gamma_0$ (γ_0 известно) не превосходит априори выбранного числа α_0, $\alpha_0 > \alpha$: $\Phi(t_\alpha + e_\varphi(b)\gamma) \leq \alpha_0$. Это условие эквивалентно неравенству

$$e_\varphi(b) \leq (t_{\alpha_0} - t_\alpha)/\gamma_0. \tag{17}$$

Наша цель — среди всех b, удовлетворяющих (17), найти то, при котором наименьшая асимптотическая локальная мощность теста

$\inf_\mu W(\tau, \gamma, \mu|\varphi_b, \psi_0)$ достигает наибольшего значения. Для этого воспользуемся тем, что для b_1, b_2, $b_1 < b_2$

$$\sup_\mu W(\tau, \gamma, \mu|\varphi_{b_1}, \psi_0) < \inf_\mu W(\tau, \gamma, \mu|\varphi_{b_2}, \psi_0), \tag{18}$$

где $\tau > 0$ и $0 \leq \gamma \leq \gamma_0(b_1, b_2, \tau)$ (Утверждение 4.1). Используя (18) и возрастание функции $e_\varphi(b)$ (Лемма 4.1), имеем, что искомым b^* является (Следствие 4.1) решение уравнения

$$e_\varphi(b) = (t_{\alpha_0} - t_\alpha)/\gamma_0. \tag{19}$$

Значение b^* характеризуется еще одним важным свойством. Пусть $\gamma = 0$ (засорения отсутствуют). Тогда АОЭ по Питмену теста со статистикой $\Lambda_n(b)$ относительно LAD-теста, которому соответствует статистика $\Lambda_n(\infty)$, достигает в $b = b^*$ своего максимума (Замечание 4.1).

В семипараметрическом случае, когда распределение $\{\varepsilon_t\}$ неизвестно, $e_\varphi(b)$ в (19) оценивается с помощью выборки $\{y_t\}$. В качестве оптимального b теперь берется оценка $\widehat{b}_n$ — единственное решение уравнения

$$\frac{1}{n}\sum_{t=1}^{n} \varphi_b^2(y_t) = \gamma_0^2/(t_{\alpha_0} - t_\alpha)^2. \tag{20}$$

Утверждение 4.2. *Пусть выполнено Условие 1. Пусть верна альтернатива $H_{1n}(\tau)$, $\tau \geq 0$. Тогда уравнение* (20) *имеет единственное решение $\widehat{b}_n$, причем $\widehat{b}_n \xrightarrow{P} b^*$, $n \to \infty$.*

Значения оценок $\widehat{b}_n$ для различных β_0 и объемов выборки n получены в работе численно методом дихотомии.

Ниже в Таблице 1 частично описаны результаты обширного численного эксперимента, проведенного для данной диссертации. А именно, представлены значения уровней значимости и мощностей тестов, основанных на статистиках $\Lambda_{n,Y}(b^*)$, $\Lambda_{n,Y}(\infty)$ и $\Lambda_{n,Y}(\widehat{b}_n)$ для проверки гипотез $H_0\colon \beta = -0.5$, $H_0\colon \beta = 0.5$, $H_0\colon \beta = 0$ и $H_0\colon \beta = 0.9$ в случае гауссовских $N(0, 1)$ инноваций и различных истинных значений параметра β. Обозначим мощности тестов $W_n(b^*)$, $W_n(\infty)$ и $W_n(\widehat{b}_n)$ соответственно.

Рассматривается случай $\xi_i \sim N(10, 5)$. При этом для краткости изложения здесь мы приводим подробные результаты только для случая $\gamma = 0.1$, $\alpha_0 = 0.0615$. Значения мощностей получены методом Монте-Карло, а именно, было смоделировано 10000 выборок различных объемов. Здесь мы приводим только случай $n = 1000$.

Результаты Таблицы 1 и других вычислений, проведенных для диссертации, в частности, показывают, что при больших значениях n мощности при альтернативе и гипотезе (т.е. уровни значимости) тестов, построенных по $\widehat{b}_n$ и b^*, близки. То есть, при достаточно больших n в качестве тестовой статистики можно брать оценку $\Lambda_{n,Y}(\widehat{b}_n)$.

Таблица 1: Мощности тестов для $H_0\colon \beta = \beta_0$, $\gamma_0 = 0.1$

	$\beta_0 = -0.5$, $\mathbf{E}\varepsilon_1^2 = 1$, $b^* = 0.107$					$\beta_0 = 0.5$, $\mathbf{E}\varepsilon_1^2 = 1$, $b^* = 0.107$				
β	**-0.5**	**-0.48**	**-0.4**	**-0.25**	**0**	**0.5**	**0.52**	**0.6**	**0.75**	**0.9**
$W_n(b^*)$	0.064	0.143	0.759	1	1	0.042	0.105	0.811	1	1
$W_n(\infty)$	0.185	0.356	0.959	1	1	0.01	0.034	0.648	1	1
$W_n(\widehat{b}_n)$	0.062	0.141	0.75	1	1	0.043	0.106	0.767	1	1
	$\beta_0 = 0$, $\mathbf{E}\varepsilon_1^2 = 1$, $b^* = 0.092$					$\beta_0 = 0.9$, $\mathbf{E}\varepsilon_1^2 = 1$, $b^* = 0.212$				
β	**0**	**0.02**	**0.1**	**0.25**	**0.5**	**0.9**	**0.92**	**0.99**	**0.8**	**0.7**
$W_n(b^*)$	0.05	0.109	0.661	1	1	0.006	0.066	1	0.962	1
$W_n(\infty)$	0.05	0.127	0.81	1	1	0	0	0.999	1	1
$W_n(\widehat{b}_n)$	0.0505	0.105	0.642	1	1	0.044	0.247	1	1	1

В случае $\beta_0 = 0$ присутствие засорений не оказывает влияние на предельное распределение рассматриваемых статистик. Далее, в случаях $\beta_0 = 0$ и $\beta_0 = -0.5$ тест, основанный на $\Lambda_{n,Y}(b^*)$, является менее мощным, чем LAD-тест. Однако при $\beta_0 = -0.5$ уровень значимости построенного теста равен 0.064, что гораздо ближе к асимптотическому значению $\alpha = 0.05$, чем значение 0.185, соответствующее LAD-тесту. Отметим, что в случае большой доли выбросов ($\gamma = 1$) LAD-тест оказывается вообще неприменим, т.к. ошибка первого рода для него равна 1 против примерно 0.25 у теста со статистикой $\Lambda_{n,Y}(b^*)$.

В случае положительных β_0 ситуация меняется на противоположную. А именно, при $\beta_0 = 0.5$ и $\beta_0 = 0.9$ значения мощности GM-теста со статистикой $\Lambda_{n,Y}(b^*)$ больше соответствующих значений для LAD-теста. При этом несмотря на то, что уровень построенного GM-теста больше уровня значимости LAD-теста, его значение не превышает 0.05 (оно равно 0.042 для $\beta_0 = 0.5$ и 0.006 для $\beta_0 = 0.9$).

Доказательства представленных результатов получены автором диссертации самостоятельно и вынесены для удобства в третий параграф.

Заключение.

В данной диссертации исследована качественная робастность GM-процедур в авторегрессионных схемах с засорениями. Но полученные в работе результаты, в частности, асимптотические равномерные разложения остаточных эмпирических процессов в схемах с засорениями, могут быть использованы при исследовании качественной робастности и других статистических процедур, например, ранговых, знаково-ранговых и процедур минимального расстояния.

Таким образом, рассмотренные в данной работе задачи являются частью более общей проблемы, решение которой будет продолжено в дальнейшем.

Автор выражает глубокую признательность своему научному руководителю кандидату физико-математических наук, доценту Михаилу Васильевичу Болдину за постановку задачи и постоянное внимание к работе.

Работы автора по теме диссертации

[1] Есаулов Д.М., Робастность GM-тестов в авторегрессии против выбросов, *Вестник МГУ. Математика. Механика*, №2, с. 47-50, 2012.

[2] Esaulov D., Residual empirical processes and its application to GM-testing for the autoregression order, *Mathematical Methods of Statistics*, Vol. 22, No. 4, pp. 333-349, 2013.

[3] Esaulov D., Application of residual empirical processes to robust linear hypotheses testing in autoregression, *Proceedings of 10th International Conference CDAM: Theoretical and applied stochastic*, Vol. 1, pp. 153-156, 2013.

[4] Болдин М.В., Есаулов Д.М., Остаточные эмпирические процессы и качественно робастные GM-тесты в авторегрессии, *Вестник МГУ. Математика. Механика*, №1, с. 46-50, 2014.

Постановка задач и результаты раздела 3 об асимптотически оптимальных тестах принадлежат М.В. Болдину. Результаты раздела 2 о линейных разложениях о.э.п. принадлежат Д.М. Есаулову.

[5] Болдин М.В., Есаулов Д.М. Эмпирические процессы в авторегрессионных схемах с выбросами. Робастные GM-тесты, *Тезисы докладов Международной конференции „Теория Вероятностей и ее Приложения“, посвященной 100-летию со дня рождения Б.В. Гнеденко*, Москва, с. 22-23, 2012.

Постановка задачи принадлежит М.В. Болдину. Результаты о разложениях о.э.п. принадлежат Д.М. Есаулову.