

УТВЕРЖДАЮ

Проректор по научной работе
Национального исследовательского
Томского государственного университета,
доктор физико-математических наук



И.В. Ивонин

12 сентября 2017 г.

ОТЗЫВ

ведущей организации на диссертационную работу
Шаталина Евгения Викторовича
«Эмпирический мост и задачи тестирования адекватности
регрессионных моделей анализа данных»,
представленную на соискание ученой степени кандидата физико-математических наук
по специальности 05.13.17 – Теоретические основы информатики

Диссертационная работа Шаталина Е.В. посвящена построению и обоснованию статистических критериев, используемых для анализа адекватности линейных регрессионных моделей эмпирических данных. В качестве регрессора в работе рассматриваются порядковые статистики. В основе предлагаемого алгоритма лежит случайная ломаная – так называемый эмпирический мост, который представляет собой процесс нормированных и центрированных частичных сумм регрессионных остатков.

В первой главе диссертации предлагается пошаговый алгоритм для изучаемых в работе критериев и доказываются предельные теоремы для эмпирического моста, обосновывающие построенные критерии для однопараметрических и двух типов двухпараметрических линейных регрессионных моделей на порядковые статистики. Предлагаемые доказательства предельных теорем используют классические методы доказательства слабой сходимости в функциональных пространствах, однако требуют специальных модификаций, учитывающих специфику изучаемых задач. Дополнительные трудности возникают, в частности, из-за наличия зависимости между элементами регрессора, а также из-за отсутствия гомоскедастичности регрессионных ошибок. Автор успешно справляется с асимптотическим анализом предлагаемой модели. Устанавливается, что предельными процессами для эмпирического моста являются центрированные гауссовские процессы с ковариационными функциями, полученными в явном виде.

Суть предлагаемого в диссертации алгоритма можно изложить следующим образом: подбирается подходящий функционал от эмпирического моста, предельное распределение которого известно, и предлагается сравнивать фактическое и теоретическое (предельное) значения указанного функционала от эмпирического моста. На основании проведенного сравнения принимается решение об адекватности той или иной регрессионной модели анализируемым данным. Эмпирический мост как бы «накапливает» в себе информацию о регрессионных остатках (отклонениях эмпирических данных от регрессионной кривой), что позволяет на основании анализа его поведения делать выводы о соответствии модели и

данных. Предлагаемое решающее правило на основе поведения эмпирического моста представляется логичным и вполне естественно для изучения регрессионных моделей. Приведенный в конце первой главы пример показывает определенные преимущества использования метода эмпирического моста над известным F-тестом.

Вторая глава работы носит прикладной характер и иллюстрирует практические возможности применения построенного метода анализа данных. Для практического применения предложено два статистических критерия на основе статистик хи-квадрат и омега-квадрат. Критерий на основании статистики омега-квадрат оказывается более предпочтительным, поскольку в отличие от критерия хи-квадрат он является состоятельным. Однако стоит отметить, что его применение требует дополнительных ограничений на распределение исходной выборки. Отметим, что основанный на статистике хи-квадрат критерий хотя и не является состоятельным, может применяться без ограничений на распределение выборки, что позволяет решать задачи при большей степени неопределенности в исходных данных. В диссертации приводятся примеры, связанные с анализом зависимости между курсами американского доллара и евро, с изучением зависимости массы человеческого тела от его роста, а также длины прыжка спортсмена в зависимости от его роста. Указанные примеры достаточно хорошо иллюстрируют потенциальные возможности применения разработанных в диссертации критериев, доводя их до практически работающих алгоритмов.

Тема диссертации является вполне **актуальной**, поскольку разрастание массивов требующих обработки и анализа данных, обгоняющее по своей скорости рост вычислительных мощностей, выводит на первый план вопросы быстрых и, главное, надежных методов анализа предлагаемых моделей описания данных. Предлагаемый в работе алгоритм дает эффективное решение указанной задачи, что подкрепляется реальными практическими примерами. Полученные в работе теоретические результаты вносят заметный вклад в прикладной анализ регрессионных моделей.

Все полученные в диссертации **результаты являются новыми, строго математически обоснованными** и имеют высокую теоретическую и **практическую значимость**. Практическая значимость иллюстрируется на «нестандартных» регрессионных моделях с зависимым между собой элементами регрессора (порядковыми статистиками), а также моделях, в которой отсутствует одно из основных предположений регрессионного анализа – гомоскедастичность.

Все основные результаты диссертации опубликованы с достаточной полнотой в российских и иностранных изданиях и прошли апробацию на конференциях и семинарах. Автореферат правильно отражает содержание диссертации.

Полученные в диссертации результаты представляют интерес для научно-исследовательских и учебных организаций (таких, как ИМ СО РАН, МИАН, ТГУ, ИСИ СО РАН, ИПМ РАН, НГУ, МГУ, МФТИ, ИВМиМГ СО РАН и др.) а также для специалистов в области статистического анализа данных, обработки информации, регрессионного анализа и других задач математической статистики. Результаты рекомендуется использовать в указанных организациях, а также в других университетах РФ при подготовке специалистов в указанных областях науки.

Диссертация изложена на 102 страницах, состоит из введения, двух глав, заключения, списка литературы и приложения с графиками. Работа написана ясным языком, хорошо структурирована и логично оформлена.

Замечания по диссертационной работе:

1. В текстах автореферата и диссертации содержится значительное количество опечаток (например, в диссертации на с. 47, строка 2 снизу; с. 50, строка 5 снизу; с. 52, строка 12 сверху и др.).
2. С. 49, строка 5: пропущено слово «асимптотически», поскольку замена при конечных значениях n неправомерна.
3. С. 51, строка 6 сверху: недостаточно точно объясняется величина $o(1)$.
4. В работе не приводятся примеры использования модели (3) с «Markov-modulated noise», исследование которой является одним из наиболее интересных результатов работы.
5. В работе не затрагиваются вопросы применения полученных результатов к моделям, содержащим нелинейности. На основе полученных результатов вполне можно было предложить линейные аппроксимации в некоторых нелинейных моделях, используемых в прикладных задачах.

Перечисленные замечания не снижают общее положительное впечатление о диссертационной работе. В целом диссертационная работа Е.В. Шаталина представляет собой законченную научно-квалификационную работу, выполненную на высоком математическом уровне.

Диссертация удовлетворяет всем требованиям, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук Положением о присуждении ученых степеней, а ее автор Шаталин Евгений Викторович заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.17 – Теоретические основы информатики.

Отзыв на диссертацию обсужден и единогласно одобрен на заседании кафедры высшей математики и математического моделирования Института прикладной математики и компьютерных наук Томского государственного университета 11 сентября 2017 года, протокол № 9.

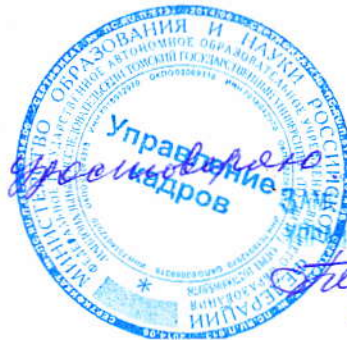
Заведующий кафедрой
высшей математики и математического моделирования
Томского государственного университета,
доктор физико-математических наук,
профессор

Конев Виктор Васильевич

12.09.2017

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский
Томский государственный университет»,
634050, г. Томск, пр. Ленина, 36, (3822) 529-852,
www.tsu.ru, rector@tsu.ru

Подпись В.В. Конев



Заместитель начальника
управления кадров
И. П. Де Глина