

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу **Шаталина Евгения Викторовича**
«ЭМПИРИЧЕСКИЙ МОСТ И ЗАДАЧИ ТЕСТИРОВАНИЯ АДЕКВАТНО-
СТИ РЕГРЕССИОННЫХ МОДЕЛЕЙ АНАЛИЗА ДАННЫХ»,

представленную на соискание ученой степени кандидата

физико-математических наук по специальности

05.13.17 — Теоретические основы информатики

В настоящее время для успешного развития производства, науки, техники, бизнеса необходимо решать широкий круг задач обработки данных, анализа зависимостей данных и их интерпретации. Широко используемым механизмом анализа зависимости данных выступают регрессионные модели. Применение таких моделей позволяет обосновать рекомендации при принятии коммерческих, управленческих и других решений и повысить их качество. В силу многообразия различных вариантов регрессионных зависимостей, а также ограниченности временных, вычислительных и других ресурсов набирает актуальность вопрос анализа адекватности предлагаемых моделей реальным данным.

Диссертационная работа Е.В. Шаталина посвящена построению и теоретическому обоснованию критериев (решающих правил) анализа адекватности регрессионных моделей. В основе построенных критериев лежит случайная ломаная – эмпирический мост, построенный по остаткам линейной регрессионных модели на порядковые статистики. С помощью конструкции эмпирического моста предлагается проводить анализ соответствия предложенной регрессионной модели наблюдаемому процессу. Важным отличием настоящей работы является изучение регрессионных моделей на порядковые статистики, чего не было представлено ранее в научной литературе. Рассмотрение порядковых статистик в качестве регрессора является весьма естественным и удобным способом анализа данных, так как в целом ряде приложений видится логичным графическое представление процесса, а следовательно, упорядочение одной из наблюдаемых переменных по убыванию или возрастанию. Таким образом, **тема диссертационной работы является, безусловно, актуальной и представляет серьезный научный интерес как с практической, так и с теоретической точек зрения.**

Рассмотрим содержание диссертации по главам.

Диссертация состоит из введения, двух глав, заключения, списка литературы и приложения. Во введении приводится обзор работ по теме диссертационного исследования, вводятся основные понятия, используемые в работе, описаны постановки задач, а также раскрыты вопросы апробации и структуры диссертации.

Глава 1 имеет теоретический характер и посвящена построению и обоснованию критериев анализа регрессионных моделей. В первом и втором параграфах главы 1 приводятся необходимые сведения из регрессионного анализа и теории случайных процессов соответственно. В параграфе 3 главы 1 сформулированы основные результаты диссертации (теоремы 1-3), а именно построен пошаговый алгоритм анализа адекватности линейных регрессионных моделей, а также приведено обоснование построенного алгоритма, для чего найдены предельные процессы для эмпирического моста, построенного по остаткам линейной регрессионной модели на порядковые статистики, в случаях трех регрессионных моделей: однопараметрической, двухпараметрической и двухпараметрической, остатки которой управляются цепью Маркова. В параграфах со четвертого по шестой приведены доказательства теорем 1-3. В конце первой главы (параграф 1.7) проведено сравнение предлагаемого в диссертации подхода анализа адекватности регрессионных моделей с известным в статистике F-тестом проверки гипотез. Приведен пример, на котором предлагаемый статистический критерий превосходит F-тест.

Глава 2 посвящена практической стороне исследуемого в диссертации вопроса. В параграфе 2.1 раскрыты вопросы практического применения построенного в первой главе критерия, для чего построено два статистических критерия на основе статистик типа хи-квадрат и омега-квадрат. Показано, что первый критерий наиболее универсален, однако не является состоятельным, второй критерий лишен данного недостатка, но требует определенных знаний об исходном распределении случайных величин. Три последующих параграфа главы 2 содержат прикладную иллюстрацию разработанных в диссертации алгоритмов. В параграфе 2 обсуждаются вопросы зависимости курсов американского доллара и евро, в параграфе 3 вопросы зависимости массы человеческого тела от его роста, и, наконец, в параграфе 4 рассмотрена зависимость длины прыжка человека от его роста. Проведен полный цикл анализа эмпирических данных в соответствии с построенным в диссертации алгоритмом, начиная от предложения регрессионных моделей и до численных расчетов, показывающих превосходство одной из моделей над остальными.

Остановимся на некоторых замечаниях к работе.

1. В нескольких частях диссертации опущено указание на то, что $n \rightarrow \infty$, в частности, на стр.33 в абзацах 2 и 3 при описании сходимости эмпирической обобщенной кривой Лоренца к теоретической, а также в формулировке теорем 1 и 2.
2. Хотя автор после доказательства теоремы 2 и приводит другое доказательство теоремы 1 (с помощью методов доказательства теоремы 2) не ясно зачем вообще приводится изначальное доказательство теоремы 1, когда известна более короткая и универсальная схема доказательства, используемая также и при доказательстве теоремы 3.
3. Упорядочение курсов валют не по времени, а по возрастанию значений одного из них является не совсем естественным, так как теряется часть исходной информации. Хотя, безусловно, данный пример иллюстрирует возможности разработанного в диссертации метода, можно было бы подобрать другой, более естественный и применимый на практике пример.

Указанные недостатки не влияют на высокую оценку работы. В целом диссертационная работа Е.В. Шаталина является **самостоятельным и законченным научным исследованием**. Все результаты прошли **успешную апробацию** на ряде российских и международных конференций, а также на семинарах Института математики им. С. Л. Соболева Сибирского отделения Российской академии наук и Национального исследовательского Томского государственного университета.

Защищаемые результаты по своей сути являются новыми решениями актуальных математических проблем, имеющих важное прикладное и теоретическое значение. Полученные результаты являются весомым вкладом, в частности, в решение проблем обработки, анализа данных, обнаружения закономерностей в данных. Полученные в диссертации **результаты являются новыми, достоверными, строго обоснованными и имеют высокую научную ценность и значимость** в разрешении проблем обработки данных, а также в области регрессионного анализа. Содержание диссертации своевременно опубликовано в четырех статьях. Автореферат достаточно полно и правильно отражает содержание диссертации.

Считаю, что рассматриваемая диссертационная работа удовлетворяет всем требованиям ВАК РФ, предъявляемым к кандидатским диссер-

тациям, и соответствует критериям действующего Положения о порядке присуждения ученых степеней, а ее автор **Шаталин Евгений Викторович** заслуживает присуждения ученой степени кандидата физико-математических наук по специальности **05.13.17 — Теоретические основы информатики**.

Официальный оппонент главный научный сотрудник научно-исследовательской группы вероятностных методов и системного анализа ФГБУН Институт прикладной математики ДВО РАН (г. Владивосток), доктор физико-математических наук, профессор

Цициани

Г. Ш. Цициашвили

Подпись д.ф.-м.н., профессора Г.Ш. Цициашвили заверяю

Ученый секретарь ИПМ ДВО РАН, к.ф.-м.н.

10 октября 2017

Святуха

В.А. Святуха

Цициашвили Гурами Шалвович, д.ф.-м.н. по специальности (системный анализ, управление и обработка информации), профессор, главный научный сотрудник научно-исследовательской группы вероятностных методов и системного анализа Федерального государственного бюджетного учреждения науки Института прикладной математики ДВО РАН (ИПМ ДВО РАН) 690041, г. Владивосток, ул. Радио 7, ИПМ ДВО РАН, тел. 8 (4232) 312375, e-mail: guram@iam.dvo.ru

