DOI <https://doi.org/10.18551/rjoas.2017-07.03>

**РАСШИРЕННАЯ ПАРАМЕТРИЧЕСКАЯ РЕГРЕССИЯ:
ОСНОВНЫЕ ИДЕИ И ИХ РЕАЛИЗАЦИЯ**
AUGMENTED PARAMETRIC REGRESSION: BASIC IDEAS AND THEIR REALIZATION

Прилуков А.Н., кандидат социологических наук

Prilukov A.N., Candidate of Sociological Sciences

Институт горного дела Дальневосточного отделения РАН, Хабаровск, Россия
Mining Institute, Far Eastern Branch of Russian Academy of Sciences, Khabarovsk, Russia
E-mail: a_prilukov@mail.ru

АННОТАЦИЯ

В статье анализируются состояние, виды и направления развития методов регрессионного анализа, начиная с его канонических (классических) форм и заканчивая такими современными идеями и реализующими их методами и приёмами, как цензурирование и взвешивание выборок, робастные методы, гребневая (ридж-) регрессия, бутстреп-анализ и другие. Подробно рассматривается влияние, оказываемое на методы регрессионного анализа априорными ограничениями, принятыми в его классическом варианте. Обосновываются идеи “расширенной параметрической регрессии”, фигурировавшие в более ранних авторских работах под названием “толерантного регрессионного анализа”. В противовес двум содержательно взаимоисключающим, но в равной мере неоправданно абстрактным концепциям “непараметрической регрессии” и “всепараметрического моделирования” содержание обосновываемых идей заключается в упорядочении и расширении состава параметров, позволяющем в значительной мере унифицировать и оптимизировать компьютерную реализацию регрессионного анализа. В разработанных автором программах реализация этих идей достигается за счёт пополнения состава используемых параметров признаками, определяющими собой виды реализуемых формализованных моделей, значения фигурирующих в них постоянных и переменных величин, особенности применяемых способов интерполяции и экстраполяции исходных и промежуточных зависимостей. Принципиальную новизну описываемому методу математической статистики придаёт включение в расширенный состав его параметров варьируемой *метрики Минковского*, позволяющей диверсифицировать в динамике механизм агрегирования регрессионных остатков, обеспечивая тем самым отыскание оптимального варианта этого механизма и получение расчётных результатов, максимально точно отражающих особенности исходных данных применительно к выбранной регрессионной модели.

ABSTRACT

Assessment results are presented in the article concerning regression analysis current state, as well as its methods' development directions, beginning from its canonic (classical) forms and ending at such modern ideas and methods of their realization as censoring and weighing of data samplings, robust methods, ridge regression, bootstrap analysis, and the like. Impact is thoroughly overviewed, imposed on regression analysis methods by a priori restrictions inherent to its canonic version. Ideas are substantiated with respect to the Augmented Parametric Regression conception being developed, which was presented in the previous author's publications under the name of Tolerant Regression Analysis. In contrast to two formally mutually exclusive but equally and unreasonably abstract conceptions of “Nonparametric Regression” and “Omniparametric Simulation”, the proposed conception's essence consists in the set of parameters streamlining and widening, which enables significant unification and optimization of regression analyses computerized performance. In computer programs developed by the author this idea is realized through the set of applied parameters augmenting by those indicators that specify the applied formalized models' types, the values of constants and variables characteristic to the models, the peculiarities of

interpolation and extrapolation methods applied to both input and interim dependencies. Described mathematical statistics method's principal novelty consists in including variable *Minkowski metric* into the method's augmented parameters set. This enables the possibility of dynamic modification concerning regression residuals aggregation mechanism, leading to establishing its optimal configuration and ultimately procuring calculation outcomes that precisely fit to input data peculiarities with regard to chosen regression model.

КЛЮЧЕВЫЕ СЛОВА

Канонический регрессионный анализ, толерантный регрессионный анализ, непараметрическая регрессия, расширенная параметрическая регрессия, гребневая регрессия, бутстреп-анализ, метрика Минковского.

KEY WORDS

Canonic regression analysis, tolerant regression analysis, nonparametric regression, augmented parametric regression, ridge regression, bootstrap analysis, Minkowski metric.

Регрессионный анализ входит в число наиболее широко применяемых и интенсивно развивающихся разделов математической статистики. Согласно сложившимся представлениям [1], его предназначением является оценка взаимоотношений между переменными величинами, имеющими различную природу. При этом одна из названных величин считается результирующей зависимой переменной и именуется откликом либо регрессандом, а другая или множество других исходных величин, называемых предикторами, факторами либо регрессорами, рассматриваются как независимые переменные. Наличие у одного регрессанда многих регрессоров является признаком, отделяющим двумерную (парную) регрессию от многомерной (множественной).

Поскольку все участвующие в регрессионном анализе переменные могут принимать форму числовых, ранговых (порядковых, ординальных) либо номинальных (категориальных) величин, соответственно различаются и виды регрессионного анализа. Аналогичное размежевание происходит в результате выбора при проведении регрессионного анализа детерминированных (функциональных), статистических (вероятностных) либо логико-математических (категориальных) моделей.

Существенным основанием для выделения разновидностей регрессионного анализа является класс, а часто и конкретный вид используемых в нём формализованных моделей. В этом отношении максимально простыми и пользующимися наибольшей популярностью являются линейные математические модели, представляемые алгебраическими многочленами первой степени и позволяющие с максимальной эффективностью и наименьшими усилиями использовать теоретико-прикладной потенциал, накопленный в регрессионном анализе. Существенным ограничивающим фактором для повсеместного использования линейных моделей является их плохое сочетание с нелинейными формами многих зависимостей, изучаемых с помощью регрессионного анализа.

В отношении способов обобщённого оценивания регрессионных ошибок, характеризующих отклонения расчётных значений отклика от соответствующих им исходных значений регрессанда, базовыми считаются метод наименьших квадратов (МНК) или, в англоязычной терминологии, Ordinary Least Squares (OLS), и метод наименьших модулей (МНМ). Встречаются также “мультипликативный” и “гармонический” способы агрегирования регрессионных остатков. В первом обобщённая оценка получается за счёт многократного воспроизведения операции умножения, в основе второго лежит использование величин, обратных к регрессанду и регрессионным остаткам.

Довольно часто появляются также модели, а наряду с ними и способы проведения регрессионного анализа, построенные на сочетаниях их “базовых” вариантов. Показательным в этом отношении является пример так называемого мета-регрессионного анализа, предназначенного для решения конкретной

исследовательской задачи и построенного на совмещённом использовании нескольких десятков разновидностей моделей [2].

Развитие регрессионного анализа идёт как в направлении расширения его функциональных и процедурных возможностей, так и в отношении улучшения надёжности, скорости, универсальности и простоты расчётов при одновременном повышении достоверности и точности получаемых результатов. Названные теоретико-прикладные направления реализуются в десятках программных пакетов, таких как CatReg, Dap, Explorer, JASP, JMP, MATLAB, MicrOsiris, Minitab, NCSS, R, Regress, S, SaiStat, SAS, Simulink, SPSS, SSP, Statext, Statist, Tanagra и многих других [3]. Многие из продолжающих совершенствоваться пакетов могут работать в интерактивном режиме, позволяющем пользователям оперативно управлять процессом вычислений, причём в некоторых пакетах предусмотрено интерактивное функционирование как off-line, так и on-line [4–10]. В наиболее развитых и мощных программных пакетах предусматривается возможность создания пользователями своих уникальных, в том числе *ad hoc* моделей регрессионного анализа, для чего пакеты снабжаются собственными языками программирования, строящимися на основе универсальных компьютерных языков. Типичными недостатками подобных пакетов, осложняющими их широкое использование, являются высокая специализированность и громоздкость. Пользовательская инструкция далеко не самого сложного программного пакета R [5], например, занимает 3538 страниц, заполненных справочной информацией, требующей для своего понимания особых профессиональных знаний и навыков.

Среди направлений совершенствования программных пакетов выделяются те, в которых предусматривается хотя бы частичное смягчение жёстких и трудновыполнимых исходных требований, предъявляемых постулатами так называемого *канонического* (классического) регрессионного анализа. Благодаря этому удаётся полнее удовлетворять критериям достоверности и точности получаемых решений при охвате всё более широкого круга исследовательских задач. В качестве соответствующих характерных примеров можно назвать приёмы масштабирования, цензурирования и взвешивания выборок, робастные методы, гребневую регрессию, бутстреп-анализ. Продолжают усложняться и совершенствоваться методы и вычислительные алгоритмы для “восстановления регрессионных уравнений” – поиска конкретного вида аналитических зависимостей, наиболее полно и точно отражающих рассматриваемые явления. При этом состав используемых регрессионных моделей пополняется их новыми разновидностями, далеко продвинувшимися по своей сложности относительно традиционных линейных либо полиномиальных моделей классического регрессионного анализа [11–26].

В последнее время не вполне оправданную популярность приобрела идея обособления математико-статистического раздела, называемого *непараметрической регрессией* [27–32]. Встречаются даже утверждения, приписывающие “непараметрическим” регрессионным моделям больший удельный вес по сравнению с параметрическими [31]. Издаётся выходящий в открытом доступе журнал Journal of Nonparametric Statistics, в публикациях которого появляются статьи по тематике непараметрической и полу-параметрической регрессии – Semi/Nonparametric Regression [29].

Однако ближайшее ознакомление с имеющимися определениями “непараметрической регрессии” оставляет ощущение, по крайней мере, их неполноты и слабой согласованности с действительностью. В популярной англоязычной Википедии, например, непараметрическая регрессия определяется как “*категория регрессионного анализа, в которой предиктор не имеет заданной формы, а конструируется в соответствии с информацией, почерпнутой из <исходных> данных*” [30]. Приведённое определение почти дословно повторяется и в других специализированных публикациях.

Первое недоумение, провоцируемое процитированным текстом, связано с тем, что в таком виде определение может быть распространено в целом на регрессионный анализ, поскольку пользовательские инструкции по применению многих его видов

начинаются с рекомендаций предварять реальные расчёты выбором их конкретных моделей. При этом данный выбор рекомендуется основывать на ознакомлении с особенностями исходных данных, выявляемых, например, посредством визуального просмотра графиков и диаграмм, характеризующих исходные данные. Помимо этого во многих программных пакетах, предназначенных для проведения регрессионного анализа, предусматриваются рекуррентные интерактивные процедуры, с помощью которых обеспечивается возможность осуществления многократных корректировок, а при необходимости и оперативных замен используемых в расчётах моделей, оказавшихся не вполне подходящими при решении конкретной задачи.

К этому следует добавить, что практические случаи и целиком области применения термина “непараметрическая регрессия” также уязвимы для обоснованной критики. Так, пример “ядерной регрессии” (Kernel Regression), приводимый в качестве классического образца рассматриваемого вида анализа [30], на поверку не является таковым, поскольку реализация алгоритмов “ядерной регрессии” жёстко управляется такими параметрами, как вид применяемого варианта сглаживания (гауссовского, равномерного и других); толерантность сглаживания, под которой понимается ширина его интервала (bandwidth of smoothing parameter); характер поведения функции сглаживания на начальном и конечном участках регрессоров.

Приведённый пример, который может быть подкреплён и другими аналогичными примерами, развенчивает бытующие мифы об отсутствии необходимости в применении параметров при обращении к “непараметрической регрессии” и её якобы автоматической самореализации в процессе вычислений. Помимо этого, попытки придания “непараметрической регрессии” особой значимости, обосновываемые её предполагаемыми простотой, надёжностью и широкой распространённостью, опровергаются типичной практикой выполнения регрессионных анализов, в которой учёт множества привходящих и сопутствующих обстоятельств, а также творческий подход со стороны исследователей к решению задачи в большинстве случаев играют ключевые роли в успешности достижения ожидаемых результатов.

Исходя из перечисленного, понятие *непараметрическая регрессия* нуждается, во-первых, в сужении и конкретизации областей его применения и, во-вторых, в уточнении названия и формулировок, специфицирующих сущность группы приёмов и методов, в большой степени условно объединяемых данным не вполне адекватным термином.

Наряду с распространёнными утверждениями о существовании “непараметрической регрессии”, якобы избавляющей исследователей от необходимости использования определяющих вычисления параметров, идея создания особого *всепараметрического моделирования* (Omniparametric Simulation [33–34]) выглядит со стороны как призыв к “логическому кульбиту” в противоположную сторону. Сомнения в реализуемости данной идеи подкрепляются тем, что набор задач и моделей регрессионного анализа, уже сейчас достигший вместе с присущими им параметрами внушительных размеров, продолжает пополняться при отсутствии даже намёков на замедление этой тенденции. Поэтому серьёзная дискуссия о применимости “всепараметрического моделирования” по отношению к стохастическим процессам представляется попросту бессмысленной. Возможное же появление упреков в адрес автора публикаций [33–34] устраняется тем обстоятельством, что речь в них идёт всего лишь о предлагаемом способе решения двух конкретных методически однотипных задач (моделировании поведения систем элементарных частиц и распространения биологических инфекций), исчерпывающим образом описываемых одной и той же формализованной моделью. Фактически под громким названием “всепараметрического моделирования” в указанных публикациях подразумевается способ решения названных задач, в котором участвуют все варьируемые параметры применяемой модели. При этом в компьютерных программах, обеспечивающих вычисления, предусматриваются либо сплошное сканирование всех возможных сочетаний дискретно изменяемых параметров, либо их более экономичный выборочный перебор, выполняемый методом Монте-Карло.

В противовес двум рассмотренным диаметрально противоположным по существу, но одинаково “запредельным” концепциям, принципиальная и практическая реализуемость обсуждаемой в настоящей статье идеи максимально возможного расширения состава используемых в регрессионном анализе параметров вряд ли может быть опровергнута вескими доводами. В статье описывается пример реализации этой идеи, задуманной и начавшей осуществляться более 30 лет назад. В разработанной и реализованной в компьютерных программах теоретико-прикладной концепции *расширенной параметрической регрессии*, обозначавшейся в предыдущих авторских публикациях [35–38] как *толерантный регрессионный анализ*, расширение номенклатуры используемых параметров достигается за счёт включения в их состав признаков, определяющих собой виды реализуемых математических моделей, значения фигурирующих в них постоянных и переменных величин, виды и особенности применяемых способов интерполяции и экстраполяции исходных и промежуточных зависимостей.

Принципиальную новизну описываемому методу математической статистики придаёт включение в состав его параметров *варьируемой метрики Минковского*, позволяющей диверсифицировать в динамическом режиме механизм агрегирования регрессионных остатков, обеспечивая тем самым отыскание оптимального варианта этого механизма и получение расчётных результатов, максимально точных применительно к особенностям исходных данных и выбранной математической модели.

Реализуемая в данной публикации замена первоначально выбранного наименования *толерантный регрессионный анализ* новым обусловлена прежде всего тем, что предлагаемый вариант существенно полнее и точнее отражает особенности разрабатываемого метода. Дополнительным побудительным мотивом внесения изменений в терминологию послужила происходящая в последнее время политизация изначально научно-профессионального термина “толерантность” и существенное смещение его содержания в область социально-бытовых отношений, а также внутри- и межгосударственной политики, зачастую сочетающихся с негативными эмоционально окрашенными оценками.

ОБОСНОВАНИЕ ОСНОВНЫХ КОНЦЕПТУАЛЬНЫХ ИДЕЙ

Математический аппарат “базового” варианта регрессионного анализа, называемого классическим или каноническим, строится на нескольких априорных постулатах, накладывающих жёсткие ограничения на характер используемых исходных данных и специфику решаемых задач. Здесь, например, постулируется наличие взаимной статистической независимости как значений результирующей переменной величины (так называемых откликов), так и отклонений последних от искомого тренда (регрессионных ошибок). Однако при решении реальных задач “восстановления” регрессионных зависимостей, под которым понимается отыскание их конкретного вида, справедливость априорных ограничений по ряду причин нередко оказывается недоказуемой, что часто ведёт к их игнорированию или принятию вариантов “по умолчанию”. Подобным же умолчанием нередко сопровождается и заведомый обход ограничений при наличии осведомлённости о них.

Например, при регрессионном анализе временных рядов [39; 40] и многих других типов зависимостей, как правило, заведомо нарушается требование взаимной статистической независимости как откликов, так и регрессионных ошибок. Похожим образом часто оказываются нарушенными канонические условия, требующие подчинения случайного распределения регрессионных ошибок нормальному статистическому закону, независимости ошибок от регрессоров (аргументов регрессионного уравнения) и от откликов, аддитивности ошибок.

Подобные нарушения, являющиеся скорее правилом, чем исключением, побуждают к поиску альтернативных вариантов традиционным схемам регрессионного анализа. В своё время разными авторами были предложены:

а) приёмы цензурирования выборки [41,с.288], заключающиеся в отбрасывании некоторого количества крайних по своим значениям членов вариационного ряда, соответствующего результирующей переменной;

б) методы взвешивания [41,с.287; 42,с.99; 43,с.145; 44], состоящие в использовании систем весовых коэффициентов, налагаемых на элементы исходного вариационного ряда;

в) использование разновидностей выбранного функционала, которыми при знании закона распределения ошибок обеспечивается наиболее адекватное суммирование последних [45–46];

г) робастные методы [28; 47; 48,с.58], характеризующиеся использованием различного рода нестандартных регрессионных моделей, чем предположительно обеспечивается получение более устойчивых и точных результатов регрессионного анализа;

д) гребневая (ридж-) регрессия [42,с.118; 47; 49], заключающаяся во введении в исходное регрессионное уравнение дополнительных остаточных членов, обеспечивающих повышение устойчивости решения в случае плохой обусловленности исходной матрицы;

е) бутстреп-анализ [50–53], ключевыми процедурами которого являются многократное тиражирование исходной выборки с последующим формированием серии случайных подвыборок и статистической оценкой с помощью последних искомым расчётных параметров регрессии;

ж) так называемый анализ данных [54], состоящий в целенаправленном варьировании вариантов предобработки исходных данных и выборе такой их совокупности, которая обеспечивает получение наиболее привлекательных, с точки зрения исследователя, результатов регрессионного анализа.

К числу особенностей, характерных для перечисленных направлений совершенствования регрессионного анализа, относится их ориентированность на устранение или ослабление влияния, как правило, одного-двух из названных нарушений априорных постулатов. Существуют и более радикальные способы нейтрализации негативного влияния нарушений классических предпосылок, заключающиеся, например, в переходе от собственно регрессионного к дисперсионному, ковариационному или конъюнктному анализам, а также к непараметрическому оцениванию [42,с.9]. Однако такой переход неизбежно сопровождается потерей точности и содержательности получаемых результатов, что далеко не всегда бывает приемлемым. Названными обстоятельствами объясняется одна из ведущих тенденций в развитии регрессионного анализа, а именно настойчивый поиск средств, которые давали бы возможность отказа одновременно от нескольких ограничивающих постулатов при возможно более полном сохранении строгости формальных процедур анализа [48,с.6].

В описываемом авторском варианте регрессионного анализа задача ослабления влияния ограничений, предписываемых классическим регрессионным анализом, решается с помощью следующих обеспечивающих большую или меньшую эффективность средств.

1. Отказ от априорного постулата аддитивности регрессионных ошибок. Как демонстрируется далее, в известных пределах отрицательное влияние на точность результатов неаддитивности ошибок может быть уменьшено за счёт выбора подходящего способа их нормирования. Более радикальный путь состоит, очевидно, в поиске универсального способа нахождения итогового кумулятивного значения регрессионных остатков, предусматривающего в качестве одного из частных случаев аддитивность последних. Ниже намечается один из возможных путей решения этой задачи.

2. Отказ от требования нормальности закона распределения ошибок. Это направление активно развивается в робастных и близких к ним методах регрессионного анализа. В авторском варианте произвольность закона распределения компенсируется вариацией в широких пределах коэффициента (метрики) Минковского

с последующим выбором наилучшего его значения, отвечающего одному или нескольким объективным критериям оптимальности. В качестве дополнительных приёмов предлагается взвешивание, в том числе и его частный вариант – цензурирование, а также выбор оптимального способа нормирования регрессионных ошибок.

3. *Учёт статистической взаимосвязанности откликов или, иными словами, инерционности моделируемых зависимостей.* Как показывает относительно несложный логико-математический анализ, для осуществления такого учёта на строгом формализованном уровне существующий математический аппарат оказывается не вполне пригодным и, следовательно, требуется обращение к математическому аппарату, родственному интегро-дифференциальному исчислению, но в отличие от последнего основывающемуся на мультипликативных идеях [55]. В качестве временного способа предлагается моделировать инерционность зависимостей подбором оптимального способа аккумуляции регрессионных ошибок при решении регрессионного уравнения, что, как показывают проведённые статистические эксперименты, достаточно аффективно осуществляется упоминавшейся вариацией метрики Минковского.

ФОРМАЛИЗОВАННАЯ ПОСТАНОВКА ЗАДАЧИ

Рассмотрим регрессионную модель вида:

$$y_i = f(X_i; \Theta) + \varepsilon_i, \quad (1)$$

где y_i – i -е значение отклика (объясняемой переменной); $X_i = x_{1i}, x_{2i}, \dots, x_{ki}$ – набор i -х значений векторов аргументов (объясняющих переменных или факторов); ε_i – i -е значение регрессионной ошибки (случайной компоненты уравнения); $i = 1, 2, \dots, n$ – номер наблюдения; $\Theta = \theta_1, \theta_2, \dots, \theta_m$ – коэффициенты уравнений и другие изменяемые параметры формализованной модели вычислений.

Модель (1) является достаточно общей, поскольку, во-первых, относится к классу многофакторных и, во-вторых, в ней не накладывается ограничений на вид функциональной зависимости f , которая может быть как линейной, так и нелинейной по параметрам Θ .

Как указывается в [35–36], а также в готовящемся к отдельной публикации описании вычислительного алгоритма, в состав параметров Θ включаются непрерывно либо дискретно варьируемые признаки, определяющие собой особенности процесса вычислений. Поскольку с точки зрения реализации вычислительного процесса между названными подмножествами переменных величин, т.е. коэффициентов уравнения и управляющих параметров, нет принципиальной разницы, имеет смысл объединить их в единый массив, оставив за ним название “массива параметров Θ ”. Если, как это обычно бывает, в качестве исходной рабочей гипотезы принимается априорное предположение о взаимной некоррелированности входящих в Θ элементов, их совокупность можно считать принадлежащей к *ортогональному признаковому пространству*, размерность которого совпадает с количеством элементов m в массиве Θ .

Таким образом, в описываемой разновидности расширенной параметрической регрессии формально решается задача отыскания минимума некоего m -мерного функционала, найденным скалярным значением которого характеризуется достигнутая степень согласованности реальных исходных данных и их приближённой интерпретации с помощью функции f .

Как отмечалось выше, классический регрессионный анализ базируется на нескольких основных постулатах, носящих априорный характер.

Постулат 1. Регрессионные ошибки ε_i имеют аддитивную природу. Этот постулат непосредственно соотносится с видом уравнения (1), в котором ε_i

фигурируют в качестве слагаемых.

Постулат 2. Ошибки ε_i не зависят (некоррелированы) ни от откликов y_i , ни от какого-либо из аргументов x_j .

Постулат 3. Ошибки ε_i случайны, их отдельные значения не коррелируют друг с другом:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0; \quad i \neq j; \quad i, j = 1, 2, \dots, n \quad (2)$$

Из совокупности постулатов (2) и (3) также вытекает, что дисперсия σ^2 случайной величины ε_i постоянна и не зависит от номера i .

Постулат 4. Ошибки ε_i распределены по нормальному закону. Используя векторные обозначения и учитывая только что изложенное, можем записать:

$$\varepsilon \sim N(0, 1\sigma^2) \quad (3)$$

Здесь выражение в скобках представляет собой квадратную матрицу ранга n с одинаковыми диагональными элементами, равными σ^2 , и нулевыми остальными элементами.

В перечень исходных постулатов может быть включён ряд дополнительных пунктов [42, с.22-26; 47, с.6-7].

Нужно заметить, что в постановке, соответствующей выражению (1), задача может рассматриваться как выходящая за жёсткие границы канонического регрессионного анализа, поскольку снятие ограничений относительно линейности функции f неизбежно затрудняет аналитические выкладки и зачастую оставляет в качестве единственно возможного численный способ решения задачи приближёнными методами. Отрицательное свойство функционала (1) состоит также в том, что при его использовании затрудняется проверка статистических гипотез, связанных с оценкой результатов решения. В частности, становится трудновыполнимым установление доверительных границ вычисляемых коэффициентов регрессионного уравнения, а вследствие этого и целенаправленный выбор конкретного вида последнего. С другой стороны, неоспоримым достоинством функционала (1) является обеспечиваемая им возможность максимально полного отражения в регрессионной модели специфики исследуемых зависимостей, что часто существенно перевешивает связанные с выбором этого функционала неудобства.

Отказ от перечисленных постулатов – каждого в отдельности и тем более нескольких одновременно – ведёт к дальнейшему усложнению задачи. При отказе от требования аддитивности ошибок (постулат I) функционал (1) должен быть приведён к более общему виду:

$$y_i = f(X_i, \varepsilon_i; \Theta), \quad (4)$$

или, в том частном случае, когда ошибки воздействуют на отклик мультипликативно [39, с.74], к виду

$$y_i = f(X_i; \Theta) \times (1 + \varepsilon_i) = f(X_i; \Theta) \times \gamma_i. \quad (5)$$

Здесь γ_i – измеряемая относительно единицы мультипликативная регрессионная ошибка; ε_i – обычная, измеряемая относительно нуля, регрессионная ошибка аддитивного типа.

Очевидно, что функционалами (1) и (5) описываются существенно более простые схемы наложения регрессионных ошибок по сравнению с функционалом общего вида (4), где ошибки ε_i входят в число аргументов произвольной функции f . Но в то же время два дискретных способа, описываемые функционалами (1) и (5), заведомо не исчерпывают всего многообразия способов обобщения ошибок, которые могут

потребуется при решении реальных задач. Один из наиболее простых выходов из положения может быть найден путём обращения к известным идеям обобщённых сумм [56] или к близким к ним идеям мер в пространстве L_p [57].

Расширяя, в частности, понятие обобщённых сумм на область возможных отрицательных значений аргументов, можем записать:

$$y_i = \text{sign } z_i \times |z_i|^{1/p}, \quad (6)$$

где $z_i = \text{sign } \hat{y}_i \times |\hat{y}_i|^p + \text{sign } \eta_i \times |\eta_i|^p$; η_i – обобщённая регрессионная ошибка; $\hat{y}_i = f(X_i; \Theta)$, – расчётное значение отклика.

Пользуясь обозначениями для так называемых континуальных алгебраических операций [55, с.44-48], можно выражение (6) представить в более компактном виде:

$$y_i = (\hat{y}_i \uparrow p + \varepsilon_i \uparrow p) \uparrow (1/p). \quad (6a)$$

Нетрудно убедиться, что функционал (6) или эквивалентный ему (6a) охватывают оба рассмотренных дискретных случая, а именно, при $p = 1$ они сводятся к функционалу (1), а при $p = 0$ – к функционалу (5). Вдобавок к этому, функционалами (6) и (6a) описываются все промежуточные между (1) и (5) способы обобщения ошибок, т.е. варианты, занимающие промежуточное положение между суммированием и умножением, а также схемы, выходящие за границы промежутка, охватываемого (1) и (5). Например, при $p = 2$ имеет место весьма популярное в математической статистике квадратичное суммирование, а при $p = -1$ – известное гармоническое агрегирование ошибок, определяющееся выражением:

$$1/y_i = 1/\hat{y}_i + 1/\eta_i. \quad (7)$$

Преобразуя необходимым образом функционал (6), получим выражение для приведения регрессионных ошибок к обобщённому виду, необходимому для их последующего агрегирования:

$$\eta_i = \text{sign } w_i \times |w_i|^{1/p}, \quad (8)$$

где $w_i = \text{sign } y_i \times |y_i|^p - \text{sign } \hat{y}_i \times |\hat{y}_i|^p$.

Из (8) непосредственно вытекает, что поиск наилучшего приближения исходной зависимости y расчётной функцией $\hat{y}$ связан с минимизацией нормы (абсолютной величины) w_i (но не η_i , которая при $p < 0$ и при уменьшении расхождения между y и $\hat{y}$ стремится возрасти).

Абстрагируясь от характера статистического распределения ошибок η_i , можно принять в качестве одного из возможных критериев аппроксимации минимизацию суммы модулей величин w_i , использующихся при вычислении η_i :

$$D_1(\Theta) = |w_1| + |w_2| + \dots + |w_n| \Rightarrow \min, \quad (9)$$

где $\Theta = \Theta_1, \Theta_2, \dots, \Theta_m$ – отыскиваемые параметры аппроксимирующей функции $\hat{y}$. Функционал (9), как явствует из его вида, соответствует методу, основанному на суммировании модулей регрессионных ошибок и носящему в регрессионном анализе

название метода наименьших модулей. Если же при использовании функционала (8) есть основания предположить, что w_i распределены по нормальному статистическому закону, тогда наилучшим способом агрегирования ошибок становится метод наименьших квадратов:

$$D_2(\Theta) = \sum_{i=1}^n w_i^2 \Rightarrow \min. \quad (9a)$$

При обращении ко второму из перечисленных выше постулатов канонического регрессионного анализа обнаруживается немало прикладных областей, где этот постулат явным образом нарушается. В качестве примеров можно назвать эконометрическое моделирование, связанное с инерционными экономическими процессами, а также математическое описание процессов, параметры которых измеряются опять-таки инерционными механическими и электрическими приборами. Наиболее типичной в таких случаях является ситуация, когда абсолютное отклонение ошибок характеризуется выраженной корреляционной связью со значениями откликов y_i . Естественным способом борьбы с подобного рода нарушениями априорных предпосылок является применение взвешивания. В научных и технических публикациях за этим способом закрепилось название взвешенных методов, из которых наиболее популярными являются уже упоминавшиеся методы наименьших квадратов и наименьших модулей.

Не менее частые нарушения постулата (3) формально обнаруживаются в том, что ковариационная матрица (2) перестаёт быть диагональной и принимает общий вид:

$$E(\varepsilon, \varepsilon') = \Omega, \quad (10)$$

где E – знак математического ожидания; Ω – матрица размерности $n \times n$; n – количество членов вариационного ряда.

В исследованиях, в том числе проведённых способом имитационного моделирования, показано, что применение к подобным задачам обычного метода наименьших квадратов может приводить к получению ложных результатов. Может, например, утверждаться существование статистически значимой зависимости между факторами и откликом там, где на самом деле она отсутствует [40, с.21].

В случае использования линейной регрессионной модели проблема учёта взаимной коррелированности ошибок имеет строгое решение, заключающееся в приведении исходного регрессионного уравнения к такому виду, когда матрица регрессионных ошибок становится диагональной. Соответствующий способ вычислений носит название обобщённого метода наименьших квадратов – ОМНК. Основная трудность при применении ОМНК связана с нахождением ковариационной матрицы (10), которая обычно бывает неизвестна. Если же довольствоваться нахождением приближённой оценки матрицы (10), процедура поиска решения становится итеративной и, следовательно, существенно усложняется [42, с.154].

Вместе с тем в теоретических исследованиях показано, что несмотря на наличие корреляционных связей между регрессионными ошибками при применении линейной регрессионной модели оценки её параметров, полученные на основе обычного невзвешенного метода наименьших квадратов, остаются несмещёнными, хотя и бывают при этом неэффективными [42, с.153]. Следовательно, основной задачей становится повышение эффективности оценок, что достигается, в частности, переходом к использованию ОМНК.

Рассматриваемая задача может решаться и иначе, а именно, переходом от ОМНК к альтернативным способам оценивания регрессионных остатков. Выяснилось, например, что повышенной устойчивостью и эффективностью обладают оценки параметров, полученные при минимизации суммы модулей ошибок либо

максимального модуля ошибки.

Наиболее существенным является то, что перечисленные способы оценивания могут быть описаны единым функционалом, основанным на использовании варьируемой метрики Минковского:

$$D_r = \left(\frac{1}{n} \sum_{i=1}^n |\varepsilon_i|^r \right)^{1/r}. \quad (11)$$

Здесь r – метрика (показатель) Минковского; n – количество членов вариационного ряда.

Функционал (11) относится к так называемым обобщённым средним величинам. Варьируя в нём r можно получать различные варианты расчётов. В частности, при $r=1$ минимизация D_r равнозначна реализации метода наименьших модулей, а при $r=2$ – метода наименьших квадратов. При значительном увеличении используемой метрики Минковского ($r \rightarrow \infty$) расчётное значение D_r приближается к выбору максимального по значению модуля ошибок $|\varepsilon_i| \rightarrow \max$, которым в конечном счёте и определяются результаты проводимых вычислений.

Выполненные автором многочисленные статистические эксперименты с различными типами задач позволили выявить эффективные значения метрики Минковского. Выяснилось, к примеру, что при $r=1$ решение в большинстве случаев оказывается более устойчивым, а его результаты имеют меньшую дисперсию, т.е. оценки параметров более эффективны по сравнению с решением при $r=2$. Найдено и наиболее эффективное для целого ряда задач значение показателя: $r=0,75$.

В математической теории меры, иногда привлекаемой для обоснования методов и процедур регрессионного анализа, значения $r < 1$ считаются недопустимыми из-за нарушения при этом одной из аксиом этой теории – так называемого правила треугольника [57, с.18]. Учитывая недостаточную логическую строгость апелляции в данном случае к теории меры, можно, по-видимому, признать оправданным применение представлений об обобщённых средних величинах [56], в которых для r допустимыми считаются любые действительные значения, включая отрицательные.

Нужно отметить, что отличаясь простотой реализации, предлагаемый здесь способ учета взаимосвязанности откликов, по-видимому, лишь с оговорками может считаться отвечающим требованиям теоретической строгости. В любом случае этот вопрос требует более детального исследования. О том, насколько вопрос статистической взаимосвязанности откликов является трудным в теоретическом плане, говорит тот факт, что даже в рамках наиболее простой линейной регрессионной модели с аддитивными ошибками теоретически исследованными можно считать лишь простую лаговую и авторегрессионную схемы взаимосвязи откликов [40; 42]. В отношении же регрессионных моделей общего вида (1) и тем более (6) вопрос практически остается открытым. Наиболее радикальным способом решения проблемы, по-видимому, является моделирование инерционности исследуемых зависимостей средствами функционального анализа, в частности его вновь вводимого раздела, дополняющего представления известного интегрально-дифференциального исчисления и обозначенного по аналогии с названными пигральным исчислением [55].

Случаи нарушения четвёртого из перечисленных выше постулатов канонического регрессионного анализа достаточно подробно освещены в литературе [45–46]. В частности, имеются доказательства того, что при двустороннем экспоненциальном законе распределения ошибок:

$$f_1(\varepsilon, \bar{\varepsilon}) = \frac{1}{2\lambda} \exp[-(\varepsilon - \bar{\varepsilon})/\lambda] \quad (12)$$

наилучшим, дающим наиболее эффективные оценки, является метод наименьших

модулей.

Если же ориентироваться на обобщённую форму закона статистического распределения [46,с.160]

$$f(\varepsilon, \bar{\varepsilon}) = A \times \exp \left[-(|\varepsilon - \bar{\varepsilon}| / B)^p \right], \quad (13)$$

то несложно получить и соответствующий обобщённый функционал для нахождения агрегированного значения регрессионных остатков, минимизация которого даёт наиболее эффективные оценки параметров Θ регрессионной модели:

$$D_r(\Theta) = \left(\sum_{i=1}^n |\varepsilon_i - \bar{\varepsilon}|^r \right)^{1/r} \Rightarrow \min. \quad (14)$$

Также как и выражение (6), функционал (14) является воплощением представлений об обобщённых суммах, позволяющих свести существующие способы оценивания и агрегирования регрессионных ошибок к единому виду, а также значительно расширить их за счёт промежуточных и выходящих за пределы общепринятых вариантов. К сожалению, в математическом отношении обобщённая сумма (14) менее удобна, чем обобщённая средняя величина вида (11) [56]. Очевидно, например, что при использовании функционала (14) получающаяся при вычислениях расчётная величина регрессионных остатков D_r будет в сильной степени зависеть от количества имеющихся дискретных точек n . Если же в довершение всего учесть, что при постановке и решении регрессионных задач $\bar{\varepsilon}$ обычно предполагается равным нулю, выбор требуемой формы функционала склоняется в пользу функционала (11).

Подводя промежуточный итог сказанному, следует отметить особую роль метрики Минковского в рассматриваемой разновидности регрессионного анализа. Можно утверждать, как это достаточно убедительно подтверждено многочисленными выполненными автором расчётами, что варьированием p или r в функционалах (8) и (11) в значительной степени нейтрализуются отрицательные эффекты канонического регрессионного анализа, вызываемые неаддитивностью ошибок (постулат 1) и произвольностью закона их распределения (постулат 4).

Гипотетически можно предположить реализуемость частичной нейтрализации с помощью варьируемой метрики Минковского также эффектов, связанных с взаимной коррелированностью регрессионных ошибок (постулат 3).

ВИДЫ РЕАЛИЗОВАННЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Описанные идеи расширенной параметрической регрессии и алгоритмы решения соответствующих задач реализованы в серии программ нарастающей сложности, составленных на алгоритмическом языке высокого уровня PL-1 и предназначенных для выполнения на крупных стационарных отечественных ЭВМ марок ЕС-1020 и ЕС-1060. Соответствующие алгоритмы и программы разрабатывались как универсальные, обеспечивающие широкий выбор видов и параметров регрессионных уравнений, режимов выполнения расчётов и форм выдачи результатов.

В одной из последних программных версий реализованы 14 видов регрессионных моделей: многофакторная аддитивная (15); многофакторная мультипликативно-степенная (16); однофакторная полиномиальная (17); однофакторная линейно-экспоненциальная (18); однофакторная биполиномиальная (19); расширенная экспоненциальная (20); обобщённый закон распределения случайных величин (23); разновидности производственных функций (21) – (22), (24), (25); разновидности функций пигрального исчисления (26) – (28).

$$\hat{y} = e^{at} \times \left(a_0 + \sum_{i=1}^5 a_j x_j \right). \quad (15)$$

$$\hat{y} = e^{at} \times a_0 \times \prod_{j=1}^5 x_j^{a_j}. \quad (16)$$

$$\hat{y} = a_0 + \sum_{i=1}^5 a_j t_j. \quad (17)$$

$$\hat{y} = a_0 + a_1 t + a_2 e^{a_3 t}. \quad (18)$$

$$\hat{y} = \frac{a_0 + a_1 t + a_2 t^2}{1 + a_3 t + a_4 t^2 + a_5 t^3}. \quad (19)$$

$$\hat{y} = \frac{a_0 + a_1 t}{1 + a_2 e^{a_3 t}}. \quad (20)$$

$$\hat{y} = a_0 e^{at} \left[a_1 x_1^{-a_2} + (1 - a_1) x_2^{-a_2} \right]. \quad (21)$$

$$\hat{y} = a_0 e^{at} \left(x_1^{a_1} \times x_2^{1-a_1} \right)^{-a_3}, \quad (\text{при } a_2=0). \quad (21a)$$

$$\hat{y} = a_0 + a_1 e^{a_3 t} x_1 + a_2 e^{a_4 t} x_2. \quad (22)$$

$$\hat{y} = a_0 \times \exp(-a_1 t^{a_2}) \times t^{a_3}. \quad (23)$$

$$\hat{y} = a_0 x_1^{a_1} \left(1 - a_2^{x_2} \right). \quad (24)$$

$$\hat{y} = a_0 x_1^{a_1} \left(1 - a_2^{x_2/x_3} \right) x_3. \quad (25)$$

$$\hat{y} = e^{at} \times a_0 \times \prod_{j=1}^5 a_j \uparrow x_j. \quad (26)$$

$$\hat{y} = a_0 \uparrow \left(\frac{a_1^2}{2a} + a_1 t + \frac{at^2}{2} \right). \quad (27)$$

$$\hat{y} = a_0 \uparrow (a_1 t), \quad (\text{при } a=0). \quad (27a)$$

$$\hat{y} = \left| \frac{at}{a_0} + 1 \right|^{a_0/a} \times \left(\frac{at + a_0}{e} \right)^{\uparrow t}. \quad (28)$$

В приведённом перечне выражения (15) – (25) соответствуют функциям, широко

известным в теоретических и прикладных разделах математики, и поэтому не нуждаются в особых пояснениях. Функционалы (26) – (28) представляют собой результаты выполнения упоминавшихся выше нетрадиционных математических операций пигрального исчисления [55] по отношению к некоторым типичным исходным функциям. Реализация этих операций в компьютерных программах позволяет расширить состав используемых регрессионных моделей за счёт тех их разновидностей, которые соответствуют функциональным и статистическим зависимостям мультипликативного типа.

Ограниченность приведённого перечня объясняется главным образом особенностями программной реализации вычислительного алгоритма. Одна из них состоит в том, что в целях максимального сокращения времени работы программ, реализованных на ЭВМ старых образцов, многократно уступавших по технической оснащённости и производительности их современным аналогам, было принято решение о предельно ограниченном использовании внешних программных модулей, существенно замедлявших вычисления. За счёт самостоятельной программной разработки всех массовых вычислительных процедур и их встраивания в ядро вычислительного алгоритма удалось значительно ускорить вычисления, обеспечив тем самым возможность выполнения больших серий вычислительных экспериментов и параллельно с этим последовательной доработки вычислительного алгоритма.

Очевидно, что при разработке новых версий предложенного алгоритма, предназначенных для реализации на современных ЭВМ, включая настольные ПК и ноутбуки, названное ограничение полностью утрачивает свою актуальность, чем обеспечиваются возможности для дальнейшего совершенствования реализуемого вычислительного алгоритма и расширения состава решаемых с его помощью задач.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ РАЗРАБОТКИ

В статье описывается теоретико-прикладная разработка, инициированная более 30 лет назад. Потребность в ней возникла при решении исследовательских задач, относящихся к промышленной, аграрной и другим областям деятельности и основывающихся на математико-статистической обработке больших массивов первичной социально-экономической, экологической и производственно-технической информации.

По мере решения названных исследовательских задач обнаружились ограниченности и слабые места в применявшихся при этом математико-статистических разделах корреляционного и регрессионного анализов. В частности, в очередной раз нашёл подтверждение тот факт, что применение известных методов “наименьших квадратов”, “наименьших модулей”, а также вновь появляющихся их вариантов, различающихся зачастую лишь разновидностями выбираемых математических моделей (линейная, логлинейная, экспоненциальная и т.п.) и нюансами априорно задаваемых параметров вычислительных процессов, не избавляет исследователей от сомнений в правильности избранных методов и достоверности получаемых результатов.

Поэтому одна из основных идей, воплощённых в описываемом методе, состояла в его *универсализации*. Осуществляется она за счёт по возможности максимально полного охвата вычислительным алгоритмом исходных математических моделей и разнообразных параметров, выбираемых при реализации методов корреляционного и регрессионного анализов. Принципиальную новизну несёт в себе использование в качестве одного из ключевых расчётных параметров метрики Минковского, варьирование которой позволяет объединить в одном вычислительном процессе различные способы агрегирования регрессионных ошибок, что способствует получению наиболее достоверных и точных результатов вычислений.

Несмотря на то, что в отношении названных и ряда других математико-статистических методов задача универсализации расчётов в настоящее время уже в значительной степени решается при обращении к таким современным программным

комплексам, как MATLAB, Simulink, R, SAS, JMP, Minitab, NCSS, JASP, PSPP, SPSS, GUI, BMDP, S-PLUS и другим, вопросы упрощения, унификации и повышения надёжности расчётов продолжают оставаться актуальными.

Нужно учесть, что в момент инициализации и начала разработки описываемого метода существовали лишь единицы исходных версий современных пакетов, которые значительно уступали нынешним. Например, в существовавшем уже в то время пакете BMDP набор реализованных регрессионных моделей ограничивался их наиболее распространёнными вариантами, а взаимодействие с исследователем предполагалось при посредничестве специалистов, посвящённых в особенности работы пакета и освоивших специфический язык взаимодействия с ним, включая расшифровку результатов работы программ пакета.

В отличие от этого описываемый метод реализован в user-friendly (ориентированной на пользователя) комплексной программе, не требующей обращения к внешним программам, за исключением непредвиденных и нестандартных случаев. Насколько известно автору, использование варьируемой метрики Минковского, несмотря на свою наглядно продемонстрированную результативность [35; 36; 38], обойдено вниманием разработчиков современных программных пакетов, предназначенных для выполнения математико-статистических и функциональных анализов, что существенно снижает аналитические возможности последних.

В качестве обобщающих выводов можно отметить следующее.

1. Регрессионный анализ, охватывающий своими приложениями различные области производственной и научной деятельности, развивается и пополняется новыми методами и приёмами. Вместе с тем не преодолены и продолжают оказывать сдерживающее, а нередко и негативное влияние априорные ограничения и условия, сформулированные на начальных этапах становления данного вида функционально-статистического анализа. Несмотря на продолжающиеся создание и совершенствование соответствующих программных пакетов, вопросы упрощения, унификации и повышения надёжности обработки и функционально-статистического анализа разнообразной информации продолжают оставаться актуальными.

2. Многие появляющиеся примеры решения конкретных исследовательских задач, обозначаемые подчас как новые методы регрессионного анализа, по существу не являются таковыми, поскольку различаются лишь нюансами применяемых функциональных моделей и/или наборами задействуемых параметров расчётов. Примерами этого являются так называемая “непараметрическая регрессия” и, с другой стороны, “всепараметрическое моделирование”, оказывающиеся на поверку абстракциями, не подтверждаемыми необходимыми обоснованиями и поэтому лишёнными значимого практического смысла.

3. Приведённое в статье обоснование и успешная программная реализация *расширенной параметрической регрессии*, в которой все отыскиваемые переменные величины вместе с принимаемыми априорно либо определяемыми в ходе вычислений параметрами представляются как элементы единого многомерного *признакового пространства*, могут рассматриваться как “строительные блоки” перспективного направления регрессионного анализа, заслуживающего “встраивания” в существующие программные пакеты функционально-статистического анализа и создания новых подобных пакетов.

БИБЛИОГРАФИЯ

1. Regression analysis. – URL: https://en.wikipedia.org/wiki/Regression_analysis
2. Irsova Z., Havranek T. Measuring Bank efficiency: A meta-regression analysis // Prague Economic Papers. 2010. No.4. P. 307-328.
3. Interactive Statistical Calculation Pages. Free Statistical Software. – URL: <http://statpages.info/javasta2.html>
4. CatReg Software User Manual (R-Version). – URL: http://oaspub.epa.gov/eims/eimscomm.getfile?p_download_id=500572

5. Language and Environment for Statistical Computing. 3538 p. – URL: <https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf>
6. Multiple Regression Diagnostics – SPSS. – URL: http://psych.unl.edu/psycrs/statpage/regdiag_eg.pdf
7. MATLAB & Simulink – App Building. – URL: <https://www.mathworks.com/help/matlab/gui-development.html>
8. Minitab 17 – Topic Library Overview. – URL: <http://support.minitab.com/en-us/minitab/17/topic-library/topic-library-overview/>
9. Regression Analysis in NCSS. – URL: <https://www.ncss.com/software/ncss/regression-analysis-in-ncss/>
10. Faraway J.J. Practical Regression and Anova using R. 2002. 213 p. – URL: <https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
11. Deijfen M., Nonmonotonic coexistence regions for the two-type Richardson model on graphs // *Electronic Journal of Probability*. 2006. Vol.11. Art.13. P. 331-344.
12. Goyal S., Goyal G.K. Linear layer and generalized regression computational intelligence models for predicting shelf life of processed cheese // *Russian Journal of Agricultural and Socio-Economic Sciences (RJOAS)*. 2012. No.3. P. 28-32.
13. Goyal S., Goyal G.K. Soft computing single hidden layer models for shelf life prediction of burfi // *RJOAS*. 2012. No.5. P. 28-32.
14. Acquah De-Graft H. A threshold cointegration analysis of asymmetric adjustments in the Ghanaian maize markets // *RJOAS*. 2012. No.8. P. 21-25.
15. Oyakhilomen O., Omadachi U.O., Zibah R.G. Cocoa production – agricultural credit guarantee scheme fund nexus in Nigeria: a cointegration approach // *RJOAS*. 2012. No.9. P. 28-32.
16. Acquah De-Graft H., Kyei C.K. The effects of climatic variables and crop area on maize yield and variability in Ghana // *RJOAS*. 2012. No.10 (10). P. 10-13.
17. Nazarifar M.H., Momeni R., Kanani M.H. Agriculture drought risk management using standardized precipitation index and AEZ model // *RJOAS*. 2013. No.12(24). P. 3-12.
18. Frank N.N., Umoh G.S. Participation in contract fishing in a developing economy: a qualitative response model analysis // *RJOAS*. 2015. No.2(38). P. 20-30.
19. Latifa Djebbari. Sale modeling in an Algerian industrial firm: looking into possible applications using Box and Jenkins methodology // *RJOAS*. 2015. No.7(43). P. 22-28.
20. Sallawu H., Tanko L., Nmadu J.N., Ndanitsa A.M. Determinants of income diversification among farm households in Niger State, Nigeria // *RJOAS*. 2016. No.2(50). P. 55-65.
21. Danso-Abbeam G., Cobbina M.T., Antwi R.A. Agricultural credit utilization among farmers in Bole District of Northern Region, Ghana // *RJOAS*. 2016. No.3(51). P. 70-80
22. Elkhatim A.K., Alobeid H.A. Assessing the economic impact of climate change (maximum and minimum temperature) on productivity of sorghum crop in Gadarif State, Sudan // *RJOAS*. 2016. No.6 (54). P. 29-38.
23. Shinta A. The influence of technical inefficiency level that involve farmer's behaviour on risk towards profit in rice production of Indonesia // *RJOAS*. 2016. No.10(58). P. 3-12.
24. Искендеров Р., Гюльалиев М., Насирова О. Оценка основных факторов продовольственной безопасности // *RJOAS*. 2016. No.10(58). P. 128-138.
25. Moriasi D.N., et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations // *Transactions of the ASABE* 2007. Vol. 50(3). P. 885-900.
26. What's New in Base SAS 9.4: Details. – URL: <http://documentation.sas.com/api/collections/pgmmvacdc/9.4/docsets/basewn/content/basewn.pdf?locale=ru#nameddest=helpcenterfeedback>
27. Levit B. Optimal methods of interpolation in Nonparametric Regression // *Mathematical Methods of Statistics*. 2016. Vol.25. No.4. P. 235–261.
28. Acquah De-Graft H. Rank-based estimation for asymmetric price transmission modelling // *Russian Journal of Agricultural and Socio-Economic Sciences (RJOAS)*. 2017. No.3(63). P. 86-92.
29. Mahmoud H.F.F. Parametric versus Semi/Nonparametric Regression Models / Virginia Polytechnic Institute and State University. Department of Statistics, July 23, 2014. 23 p.

30. Nonparametric regression. – URL: https://en.wikipedia.org/wiki/Nonparametric_regression
31. Fan J. Huang L.S. Goodness-of-Fit Tests for Parametric Regression Models // Journal of the American Statistical Association. 2001. V.96, No.454. P. 640-652.
32. Stephenson D. Parametric and non-parametric regression. – URL: <http://folk.uib.no/ngbnk/kurs/notes/node77.html>
33. Lundin F. Omniparametric simulation of the two-type Richardson model. – URL: <http://www.math.chalmers.se/Math/Research/Preprints/2003/6.pdf>
34. Lundin F. Case studies in omniparametric simulation. – URL: <http://www.math.chalmers.se/Stat/Research/Preprints/Doctoral/2006/1.pdf>
35. Прилуков А.Н. Толерантный регрессионный анализ и пример его программной реализации. – Хабаровск, 1988. 40 с. Деп. в ВИНТИ. № 8527-B88.
36. Прилуков А.Н. Толерантный регрессионный анализ: предпосылки, концепции, программная реализация. Препринт. – Владивосток, 1989. 48 с.
37. Прилуков А.Н. Новые формализованные методы анализа объектов и процессов минералопользования // Проблемы комплексного освоения георесурсов. – Хабаровск, ИГД ДВО РАН, 2010. С. 197-202.
38. Прилуков А.Н. Идеи толерантного регрессионного анализа и их реализация // Успехи современной науки и образования. 2017. №1. Т. 1. С.134-136.
39. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. 755 с.
40. Песаран М., Слейтер Л. Динамическая регрессия: Теория и алгоритмы. М.: Финансы и статистика, 1984. 310 с.
41. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. 471 с.
42. Вучков И., Бояджиева Д., Солаков Е. Прикладной линейный регрессионный анализ. – М.: Финансы и статистика, 1987. 239 с.
43. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Кн.1. – М.: Финансы и статистика, 1986. 366 с.
44. Schaeffli B., Gupta H.V. Do Nash values have value? – DOI: 10.1002/hyp
45. Мудров В.И., Кушко В.Л. Метод наименьших модулей. – М.: Знание, 1971. 61 с.
46. Мудров В.И., Кушко В.Л. Методы обработки измерений. (Квази-правдоподобные оценки). – М.: Радио и связь, 1983. 304 с.
47. Демиденко Е.З. Линейная и нелинейная регрессии. – М.: Финансы и статистика, 1981. 304 с.
48. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Кн.2. – М.: Финансы и статистика, 1987. 351 с.
49. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Статистика, 1973. 392с.
50. Эфрон В. Нетрадиционные методы многомерного статистического анализа. – М.: Финансы и статистика, 1988. 263 с.
51. Acquah De-Graft H. A bootstrap approach to testing for symmetry in the Granger and Lee asymmetric error correction model // RJOAS. 2012. No.11. P. 33-36.
52. Acquah De-Graft H., Acquah J. On the comparison of Bayesian information criterion and Draper's information criterion in selection of an asymmetric price relationship: bootstrap simulation results // RJOAS. 2013. No.3(15). P. 73-78.
53. Acquah De-Graft H. A comparison of bootstrap and Monte Carlo approaches to testing for symmetry in the Houck's model // RJOAS. 2013. No.5(17). P. 3-6.
54. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. – М.: Финансы и статистика, 1982. Вып.1. 320 с.; Вып.2. 239 с.
55. Прилуков А.Н. Элементарное введение в пигральное исчисление. Препринт / ИГД ДВО АН СССР. – Владивосток, 1987. 66 с..
56. Харди Г., Литтлвуд Дж., Полиа Г. Неравенства. – М.: 1948. С. 40.
57. Шрейдер Ю.А. Что такое расстояние? – М.: Физматгиз, 1963. С. 27.