Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Байкальский государственный университет»

На правах рукописи

БАЕНХАЕВА
Аюна Валерьевна

АЛГОРИТМИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ МНОЖЕСТВЕННОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ ЛИНЕЙНОЙ РЕГРЕССИИ

05.13.18 – Математическое моделирование, численные методы и комплексы программ

диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
доктор технических наук, профессор
Носков Сергей Иванович

Иркутск – 2018

Оглавление

ВВЕДЕНИЕ

Актуальность темы. Несомненно, бесспорным является тезис о том, что обязательным и одним из основных этапов анализа любой сложной системы независимо от ее характера, структуры и масштаба является построение соответствующей математической модели. Тщательное исследование свойств этой модели позволяет, как правило, получать новые знания об объекте анализа и использовать их для повышения эффективности его функционирования.

Весьма высока прикладная значимость методов математического моделирования. Традиционно одной из наиболее широких сфер их применения является экономика, которая вследствие своей специфики особенно восприимчива к новым научным результатам, постоянно появляющимся в этой области.

В настоящее время существует много подходов к моделированию сложных систем. Один из наиболее эффективных из них основан на методах современной прикладной статистики, в частности, на регрессионном анализе, занимающимся решением проблем оценивания (идентификации) неизвестных параметров математических моделей статистического типа. Известны различные классы таких оценок. Наиболее широко используемым в силу своей эффективности и хорошей интерпретируемости классом оценок являются так называемые L_ν-оценки, где ν задает метрику, в которой производится минимизация ошибок модельной аппроксимации. Следует отметить, что вопросами разработки новых методов оценивания параметров моделей в рамках регрессионного анализа активно занимались, в частности, такие известные зарубежные и российские ученые, как Дрейпер Н., Смит Г., Джонстон Дж., Афифи А., Эйзен С., Винн Р., Холден И., Кади Дж., Литтл Р., Бард Я., Поллард Дж., Фишер Ф., Хьюбер П., Айвазян С. А., Енюков И. С., Мешалкин Л. Д., Демиденко Е. З., Дубровский С. А., Носков С. И., Вапник В. Н., Ершов А. А., Иванов А. В., Мудров В. И., Кушко В. Л. и др.

Очевидно, что чем шире арсенал методов оценивания параметров регрессионных моделей, имеющихся в распоряжении исследователя, тем более точную (адекватную) модель анализируемого объекта он может построить. Настоящая работа как раз и посвящена некоторым способам расширения этого арсенала посредством так называемого множественного оценивания параметров, основанного на использовании L_v-оценок.

Цель работы состоит в разработке в рамках прикладной статистики алгоритмического и программного обеспечения множественного оценивания регрессионных параметров. Реализация этой цели предполагает необходимость решения следующих **задач**:

1) проведения глубокого содержательного анализа существующих методов оценивания параметров регрессионных уравнений;

2) построения множества Парето в многокритериальной задаче оценивания параметров линейной регрессии;

3) разработки способов выделения из паретовского множества точечных оценок, обладающих некоторыми специальными свойствами;

4) разработка программного комплекса множественного оценивания параметров линейного регрессионного уравнения;

5) построения на основе использования аппарата множественного оценивания параметров регрессионной модели динамики валового регионального продукта Иркутской области.

Объект исследования — линейное регрессионное уравнение.

Предмет исследования — методы параметрической идентификации, средства разработки методо-ориентированных программных комплексов.

Методы исследования. Для решения поставленных в работе задач использовались методы регрессионного анализа, исследования операций, теории принятия решений, линейного программирования.

Научную новизну диссертации составили следующие результаты:

1. Проведен критический анализ методов оценивания параметров регрессионных уравнений с упором на методы получения L_v-оценок посредством сведения задач минимизации соответствующих функций потерь либо к итерационным процедурам, либо к задачам математического программирования.

2. Разработана алгоритмическая схема формирования множества паретовских оценок регрессионных параметров в двухкритериальной задаче их оценивания, представляющего собой объединение областей совместности систем линейных ограничений.

3. Предложены способы выделения из множества недоминируемых оценок его точечного представления и построения описанного m-мерного параллелепипеда.

4. Разработан программный комплекс множественного и точечного оценивания параметров линейной регрессии.

5. Построена регрессионная модель динамики регионального валового продукта Иркутской области с множественной оценкой параметров, предназначенная для решения задач его среднесрочного интервального прогнозирования.

Достоверность полученных результатов обусловлена корректным применением апробированного математического аппарата.

Теоретическая и практическая значимость результатов диссертации состоит в возможности построения регрессионных моделей с множественным представлением оценок параметров для широкого спектра социально-эколого-экономических и технических объектов с целью интервального прогнозирования их функционирования и развития. Имеется акт о внедрении алгоритмов, методов и программного средства в учебный процесс ФГБОУ ВО «ИрГУПС» в дисциплинах: «Математическое моделирование, численные методы и комплексы программ», «Моделирование», «Моделирование систем». Программный комплекс МОРМ используется при выполнении тем НИР: «Разработка методик моделирования оценки эффективности и надежности функционирования, поддержки принятия решений в сложных социально-экономических, технических информационных систем», № 116011510035, «Разработка методики принятия решений на основе моделей правдоподобных рассуждений», № 115121810005.

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на: 5, 6, 7-й Международных научно-практических конференциях «Транспортная инфраструктура Сибирского региона» (Иркутск, 2014, 2015, 2016 гг.), на многочисленных семинарах в Иркутском государственном университете путей сообщения.

Публикации. Основные результаты исследований опубликованы в 9 работах, в том числе 2 – в изданиях, рекомендованных ВАК. Получено свидетельство о регистрации программы на ЭВМ.

Структура и объем работы. Диссертационная работа состоит из введения, трех глав, заключения, списка использованной литературы, включающего 108 наименований. Общий объем диссертации составляет 115 страниц машинописного текста, содержит 27 рисунков и 14 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы исследования, сформулирована его цель и основные задачи, определена научная новизна.

В первой главе кратко рассмотрены основные понятия и определения регрессионного анализа, его цель и задачи. Проведен анализ наиболее часто используемых на практике методов оценивания неизвестных параметров регрессионных моделей: метод максимального правдоподобия, наименьших квадратов и модулей, антиробастного оценивания. Установлены достоинства и недостатки рассмотренных методов.

Проанализировано существующее программное обеспечение для оценивания неизвестных параметров регрессионных моделей: Eviews, Gretl, Wolfram Mathematica, MATLAB (Statistics Toolbox), Minitab, R, SAS, SPSS, Stata, Statgraphics, Statistica, SYSTAT и т. д. Установлено, что все рассмотренные программные продукты содержат лишь стандартные процедуры и методы оценивания неизвестных параметров регрессионных моделей.

На основе анализа математического и программного обеспечения оценивания неизвестных параметров регрессионных моделей уточнена цель диссертации и задачи исследования.

Во второй главе описываются алгоритмическое обеспечение множественного оценивания параметров линейного регрессионного уравнения и разработанный на этой основе программный комплекс МОРМ с обоснованием выбора среды программирования и формулировкой функциональных требований, а также представлением его архитектуры, информационной схемы и общего алгоритма работы.

Интерфейс программного комплекса множественного оценивания регрессионных моделей (ПК МОРМ) разработан в среде программирования Delphi, а его математическим ядром является пакет решения задач линейного программирования LPSolve.

В третьей главе с использованием разработанного программного комплекса МОРМ решена задача моделирования динамики валового регионального продукта (ВРП) Иркутской области.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИОННОЙ РАБОТЫ

1. Выполнен критический анализ наиболее часто используемых в рамках регрессионного анализа методов численного оценивания параметров уравнений, главным образом сводящихся к минимизации функций потерь, приводящих к получению L_v-оценок.

2. Разработана алгоритмическая схема построения множества параметрических оценок Парето в двухкритериальной задаче идентификации на основе одновременного использования полярных по отношению к выбросам методов наименьших модулей и антиробастного оценивания.

3. Предложены способы конкретизации множества паретовских оценок, облегчающие оперирование им.

4. Разработан программный комплекс МОРМ множественного оценивания параметров линейного регрессионного уравнения и проведения прогнозных расчетов на основе полученных оценок.

5. Построена регрессионная модель динамики валового регионального продукта Иркутской области с множественной оценкой параметров и на ее основе проведены прогнозные расчеты на ближнюю перспективу.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

В изданиях, рекомендованных ВАК:

1. Баенхаева А. В. Множественное оценивание параметров линейного регрессионного уравнения / С. И. Носков // Современные технологии. Системный анализ. Моделирование. – 2016. – № 3(51) – С. 133 – 140.

2. Баенхаева А. В. Моделирование валового регионального продукта Иркутской области на основе применения методики множественного оценивания / М. П. Базилевский, С. И. Носков // Фундаментальные исследования. – 2016. – № 10 (часть 1). – С. 9 – 14.

Свидетельства о государственной регистрации:

3. Баенхаева А. В. Программный комплекс множественного оценивания регрессионных моделей (ПК МОРМ) / Базилевский М. П., Носков С. И. // Свидетельство о государственной регистрации программы для ЭВМ № 2017611338 от 1 февраля 2017 г.

В других изданиях:

4. Баенхаева А. В. Прогнозирование валового регионального продукта // Экономика и бизнес. Теория и практика. – 2016. – Вып. 11. – С. 5 – 11– 0,5 п. л..

5. Баенхаева А. В. Выбор структурной спецификации регрессионной модели валового регионального продукта Иркутской области / М. П. Базилевский, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – Вып. 16. – С. 30 – 37– 0,43 п. л.

6. Баенхаева А. В. Программный комплекс множественного оценивания регрессионных моделей / М П Базилевский, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – Вып. 17. – С. 38 – 44– 0.5 п. л.

7. Баенхаева А.В. Один подход к исследованию эффективности вузов // Мы продолжаем традиции Российской статистики. Сборник докладов I Открытого российского статистического конгресса. Российская ассоциация статистиков; Федеральная служба государственной статистики РФ; Новосибирский государственный университет экономики и управления «НИНХ». – 2016. – С. 160-166.

8. Баенхаева А.В. К вопросу о методике проведения мониторинга эффективности деятельности вузов / И.А. Слободняк, А.В. Баенхаева // Экономический анализ: теория и практика. – 2015. – №36(435). – С.50-60.

9. Баенхаева А.В. Один подход к прогнозированию валового регионального продукта Иркутской области//Материалы научной конференции «Высокие технологии и инновации в науке» ГНИИ «Нацразвитие». Июль 2018: Сборник избранных статей.– СПб.: ГНИИ «Нацразвитие», 2018.– С. 178-185.

1 АНАЛИЗ МАТЕМАТИЧЕСКОГО И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ОЦЕНИВАНИЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ

1.1 Регрессионный анализ как инструмент построения математических моделей

Практически все экономические системы — это сложные объекты, в которых присутствуют десятки, а может и сотни экономических, технических и социальных процессов, подверженных изменению со стороны, в том числе и научно-технического прогресса [28]. В свою очередь, моделирование — это один из способов изучения таких систем, которое позволяет рассматривать их в специальных условиях, но с непременной оговоркой об одностороннем взгляде на проблему. Системы, сложные для изучения, рассматриваются не сами по себе, а через более доступную, подобную им модель.

Под моделью будем понимать логико-математическое описание звеньев и их назначений, представляющих основные свойства моделируемого объекта или процесса.

Экономико-математическая модель — это набор связанных математических соотношений (уравнений и неравенств), примерно отражающих состояние деятельности экономического объекта.

Итак, под экономико-математическим моделированием будем подразумевать процесс построения, изучения за счет средств вычислительной техники модели, которая аналогична исследуемому объекту. Причем построенная модель по возможности имитирует структуру экономического объекта, а если она неизвестна, то хотя бы его характерное поведение, беря за основу принцип «черного ящика» [28, 103]. Далеко не всегда и не обо всех свойствах объекта можно вести речь, а только о тех, которые подобны и в модели, и в объекте, и при этом важны для исследования. Такие свойства будем называть существенными.

Актуальность экономико-математического моделирования не вызывает сомнения, поскольку такая модель оказывается ключевым моментом исследования экономики и имеет следующие преимущества:

- воспроизводит реальный экономический процесс (или поведение объекта);
- имеет достаточно низкую стоимость;
- не один раз может использоваться;
- позволяет рассматривать различные условия функционирования объекта.

Построение модели состоит из трех элементов:

- субъекта,
- объекта исследования,
- модели, через которую субъект изучает объект.

Сущность процесса моделирования схематически может быть представлена на рис. 1.1.

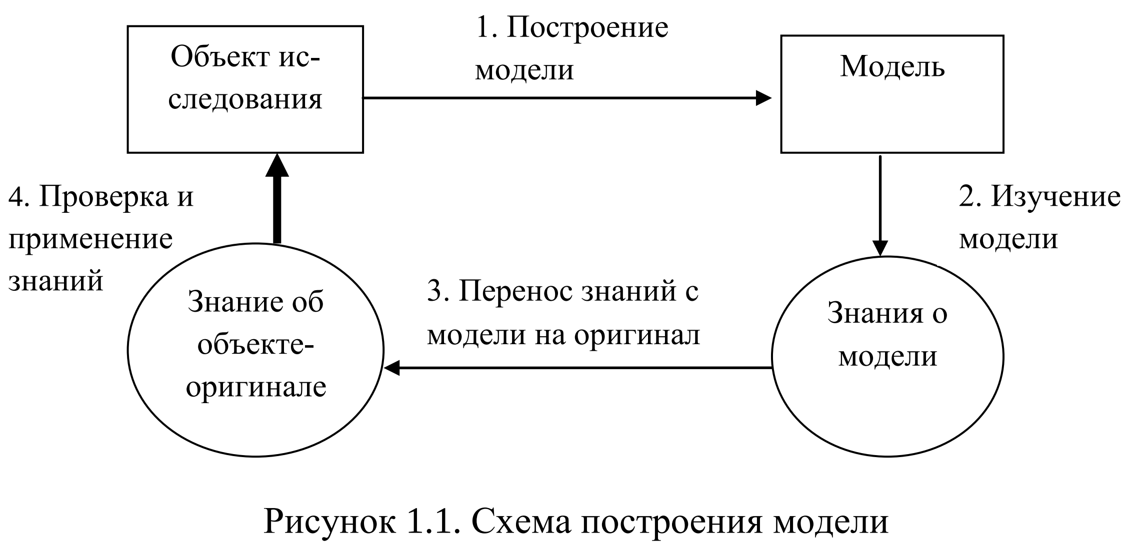


Рисунок 1.1. Схема построения модели

Конструирование модели начинается с предварительной идентификации объекта и выделения его существенных характеристик. Проведение экспериментов, теоретического анализа модели, сравнение результатов с данными об объекте, корректировка модели и т. д. составляют суть моделирования.

Благодаря наличию модели появляется возможность рассмотреть процесс управления экономическим объектом с нахождением наилучших решений без непосредственного экспериментирования с самой системой. Процесс управления с помощью модели представлен на рис. 1.2.

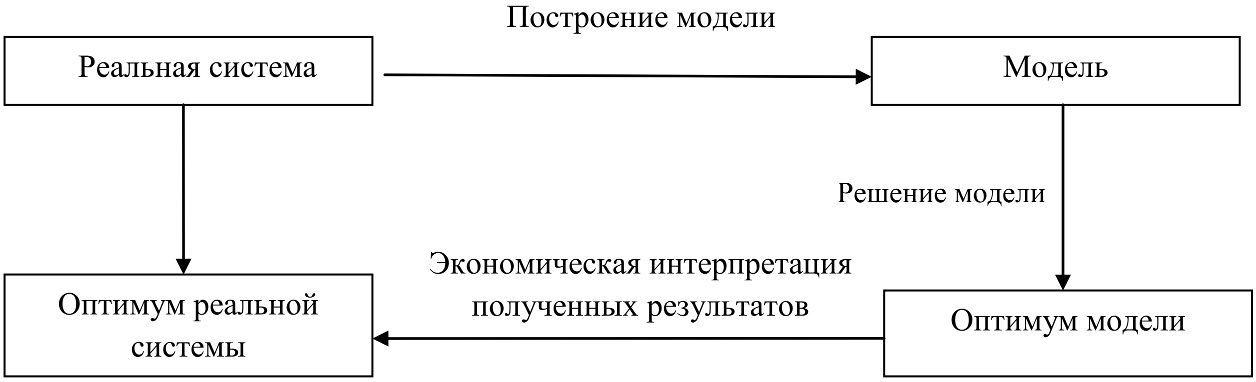


Рисунок 1.2. Процесс управления с использованием модели

Как видно из рисунка 1.2, прямой ход, ведущий к оптимальному решению, заменяется альтернативным, включающим построение и оптимизацию рассматриваемой модели. Поэтому, экономико-математическое моделирование представляет возможность находить истину дешевле, нежели дорогостоящим методом «проб и ошибок», и формулировать рекомендации по управлению экономикой, опираясь на научное предвидение.

Невзирая на сложность многих экономических систем, таких как: национальная экономика, экономика региона, а также их взаимодействие, они остаются объектом усиленного внимания исследователей [7, 8, 15, 28, 31, 48, 56, 57, 60, 90], что способствует развитию инструментария математического моделирования экономики и разработки методики использования современных вычислительных методов и технических систем.

При исследовании экономических систем математические модели имеет смысл разделить по свойствам моделируемых объектов, по направлениям приложения, по степени описания и инструментам моделирования [27, 28, 42, 50, 53].

Выделяют две большие группы методов моделирования: материальное (предметное) моделирование и идеальное моделирование.

Под материальным моделированием будем понимать исследование на основе модели, воспроизводящей основные физические, геометрические, функциональные, пространственные, динамические характеристики изучаемого объекта [28]. А идеальное моделирование применяется в экономических исследованиях, поскольку основывается не на материальной аналогии моделируемого объекта модели, а мыслимой аналогии. В свою очередь идеальное моделирование разделяется на два подкласса: интуитивное моделирование и числовое моделирование.

В основе числового моделирования положены «знаковые» образования какого-либо вида: графики, формулы, схемы, чертежи и т.д., причем знаковые образования и их элементы, всегда задаются, вместе с ними. Главный вид знакового моделирования — это математическое моделирование, осуществляемое средствами языка математики и логики [28].

Зато при интуитивном моделировании не пользуются четко фиксированными знаковыми системами, и оно осуществляется, как принято говорить, «на модельном уровне». Если познавательный процесс находится еще на начальной стадии, то такое моделирование очень актуально.

Также классифицировать модели нужно по следующим признакам [28]:

- Если описывается моментное состояние экономики, то такие модели называются статическими, а те, что отражают изменение объекта во времени, называются динамическими.
- Модели могут быть в форме таблиц (матричное представление), числовых примеров (численное представление), особого рода графов (сетевое представление).

Процесс экономико-математического моделирования состоит из следующих этапов:

- Изучение объекта или процесса;
- Спецификация модели;
- Оценка параметров модели;

– Определение зависимостей между параметрами модели;
– Верификация модели.

Изучение объекта включает: определение характеристик объекта, приложенных к нему воздействий, его реакций с помощью наблюдения за его входами и выходами и последующей статистической обработкой полученных данных.

Под спецификацией модели понимается выбор переменных модели, вида и значений параметров ее уравнений, а затем их оценкой на основе статистических данных, уже полученных в результате наблюдений или эксперимента. Иногда при формировании модели допускается построение одного или нескольких аналитических выражений, которые определяют взаимосвязь переменных и параметров модели и отражают моделируемые процессы. Поэтому модель может содержать более одного уравнения или неравенства, или вообще систему уравнений и (или) неравенств.

На основании этих действий составляется спецификация модели. Это главный этап построения экономико-математической модели.

Обнаруженные связи и соотношения отображаются в математической форме, а значит, параметры и переменные, которые представляются существенными для данного исследования.

Верификация модели – это процедура эмпирического подтверждения рассматриваемой модели, путем сопоставления результатов расчетов по модели с опытными данными.

Экономико-математические модели классифицируются по типам, которые отличаются следующими характеристиками [28]:

– способом отражения действительности;
– целью модели и сферой ее применения;
– способами логико-математического описания;
– структурой описания системы;
– местом в хозяйственной иерархии;
– временными признаками;
– пространственным признаком.

Например, модели разделяются на глобальные, макроэкономические и микроэкономические. В первой из них описываются модели, отражающие процессы глобального характера, масштабные социальные и экологические процессы, во-второй – макромодели оперируют крупноагрегированными показателями (валовой национальный продукт, инвестиции). Функционирование и развитие экономики страны или региона нуждается в макромодели, которую можно использовать для теоретического анализа наиболее общих закономерностей. В практической деятельности они применяются для построения прогнозов экономических процессов. В микроэкономических моделях для отдельного элемента экономической системы отражается функционирование и структура его взаимодействие с другими элементами системы. Причем четкое разграничение между микро- и макромоделями отсутствует.

Главная миссия экономико-математических моделей на практике заключается в прогнозировании поведения эндогенных переменных при определенных допущениях относительно поведения экзогенных переменных [28].

Применение математического моделирования в экономике позволяет расширить экономический анализ, углубить область экономической информации, улучшить экономические расчеты [103].

Наполнение разработанных моделей конкретной и качественной информацией – это основа для применения математического моделирования в экономике, причем определяющим для выбора модели является точность, непротиворечивость и полнота первичной информации.

Начальная информация может иметь принципиально различный характер и происхождение, и логично ее разделить на две категории:

1) прошлое развитие, современное состояние объектов (экономические наблюдения и их обработка);

2) перспективное развитие объектов (прогнозы).

Экономические процессы носят массовый характер, но при этом обязательно включают случайные (стохастические) составляющие. Эти составляющие могут быть вызваны природными явлениями, изменениями в политической обстановке, научно-техническим прогрессом и другими субъективными факторами. Поэтому экономические закономерности часто имеют стохастический характер.

В работах [2, 8, 27] по прогнозированию и планированию в экономике различают два типа неопределенности: «истинная», как свойство экономического процесса, и «информационная», связанная с неполнотой и неточностью имеющейся информации об этих процессах.

Для определения закономерностей, свойственных большим совокупностям однородных объектов, используют методы математической статистики посредством рассмотрения некоторой выборки [23, 52, 85]. Полученные характеристики такой выборки будут использоваться для сравнения совокупностей, их качественных характеристик, а также для определения связей между рассматриваемыми величинами и прогнозирования развития системы в будущем.

Методы математической прикладной статистики включают: корреляционный, регрессионный, дисперсионный, факторный анализ и др. Подробно рассмотрим основы регрессионного анализа, как базу данной работы.

Регрессионный анализ (РА) – раздел математической статистики, включающий практические методы изучения зависимости между величинами по статистическим данным [3, 41]. Суть РА состоит в выводе уравнения регрессии (включая оценку его параметров), благодаря которому оценивается величина случайной переменной, если величина другой (или других в случае множественной или многофакторной регрессии) задана, т. е. неслучайна и фиксирована [83]. А сама регрессия – это зависимость среднего значения случайной величины от некоторой другой величины или нескольких величин [23, 30, 35, 65].

История развития регрессионного анализа началась с 1805 года, когда А. М. Лежандром был предложен метод наименьших квадратов. Понятие «регрессия» было впервые упомянуто в 1885 г. Френсисом Гальтоном.

Регрессионный анализ является самым популярным математико-статистическим инструментарием и занимает центральное место в математической статистике, машинном обучении и эконометрике [73, 74].

Начало использованию эконометрических методов при описании экономических систем положили работы Калецкого, Тинбергена, Фриша, Кобба и Дугласа, опубликованные в 30-е гг. Начиная с 50-х гг. эконометрические модели на Западе стали разрабатываться особенно интенсивно благодаря исследованиям в этой области Клейна, Голдбергера, Брауна, де Вольфа, Эванса, Фромма. Развитие эконометрических методов в странах Восточной Европы связано с именем Михалевского Б. Н., Емельянова А. С., Четыркина Е. М., Дадаяна В. С., Колека Ю., Шуяна И., Фундарека М., Вольфе В., Ормоша З. и других. Широкое распространение эконометрического подхода объясняется рядом причин [8, 33, 36, 42, 66, 85, 97].

Во-первых, использование эконометрических моделей представляет детальный анализ систем, причем помимо основных вариантов прогнозов предлагает множество альтернатив, в основу которых положены другие гипотезы относительно перспектив развития рассматриваемых объектов.

Во-вторых, эти модели хорошо отражают ход реализации управленческих решений, поскольку показывают структурные и динамические изменения.

В-третьих, этот подход разрешает достаточно просто вносить в модели большой диапазон изменений и дополнений, тем самым, улучшая отображение исследуемых процессов.

В-четвертых, в качестве исходной информации эконометрические модели используют статистическую отчетность в форме динамических рядов наблюдений показателей.

Помимо несомненных достоинств, эконометрическим моделям присущ ряд характерных недостатков, отмеченных в литературе [24, 42, 57, 69, 80, 82,].

Во-первых, иногда использование некоторых моделей регрессионного анализа приводит к неадекватным результатам, так называемой «ложной» корреляции. Поэтому обязательным является тщательный предварительный анализ взаимосвязей с позиций эконометрической теории.

Во-вторых, длина временных рядов, на основе которых конструируются зависимости, зачастую очень сокращает число входящих в них показателей, что влечет за собой неучет существенных с экономической точки зрения моментов.

В-третьих, нет единого подхода к сопоставлению экономических показателей, полученных в разных эконометрических моделях.

Оценивая перспективы развития экономико-математического моделирования Моисеев Н. Н. отмечает необходимость разработки теории математических моделей экономических процессов, занимающейся вопросами исследования соответствия моделей изучаемым экономическим проблемам [70]. Использование эконометрического подхода при построении модели экономических систем включает выполнение тщательного анализа соответствия математической модели рассматриваемому процессу на основе уже имеющихся в эконометрике методов и возможности оценки соответствия параметров эконометрической модели экономическому содержанию входящих в нее переменных.

Применяемые в настоящее время эконометрические модели можно подразделить на:

1. Аналитические, имитационные и прогностические. Эти модели отличаются по целям их применения.
2. Комплексные или некомплексные (отличие заключается в том, насколько детально отражают связи между макроэкономическими показателями на всех стадиях процесса воспроизводства);
3. Статические и динамические – по характеру используемых данных (первые основаны на одновременных данных по совокупности объектов, вторые – на временных рядах). Между ними находятся модели панельных данных, основанные на данных по одной и той же совокупности за ряд лет;

Построение любой эконометрической модели проходит следующие этапы (см. рис. 1.3) [43]:

1) теоретический, в течение, которого определяется цель исследования, число участвующих в модели экономических характеристик, формулируется априорное описание связей между ними;

2) информационный, вовремя, которого осуществляется поиск данных, ведется проверка данных на достоверность, делаются необходимые пересчеты, со-

поставимость, используются пространственные и временные данные, формируется матрица исходных данных и последующие действия с ней;

3) при спецификации модели определяются экзогенные (внешние) и эндогенные (внутренние) переменные, выявляются соотношения и связи. Наличие общей спецификации модели позволяет перейти к последующему весьма ответственному этапу – численному оцениванию параметров с привлечением широкого арсенала методов и программных средств, число которых в анализе данных весьма велико [43];

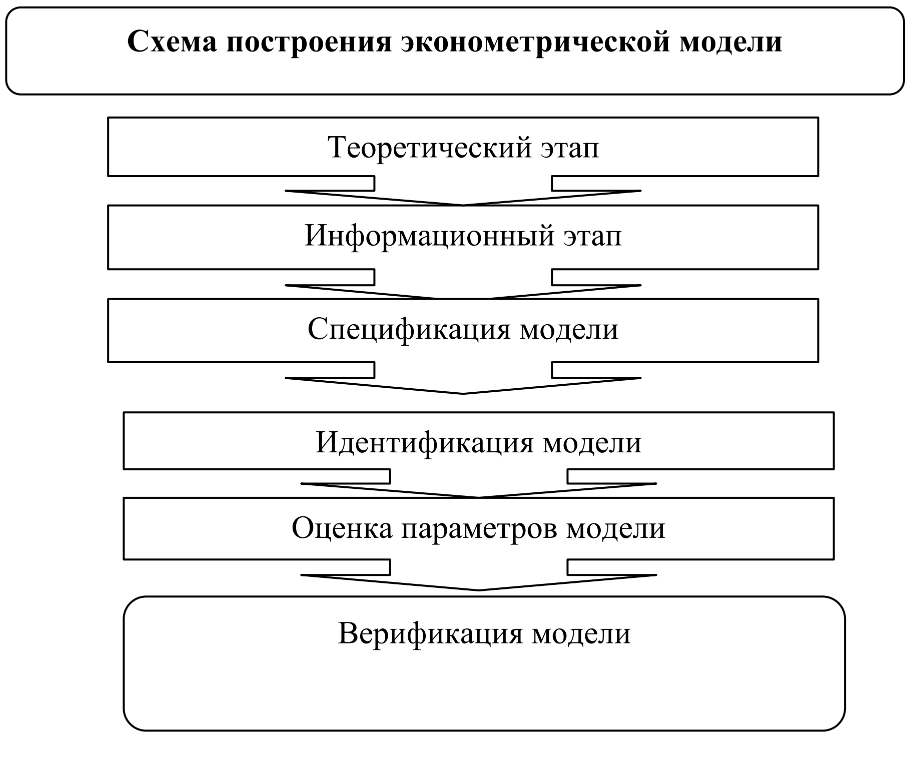


Рисунок 1.3. Схема построения эконометрической модели

4) идентификация модели – это выявление условий более точного оценивания параметров модели на основе количества переменных и связей между ними;

5) оценка параметров модели – количественный расчет оцениваемых параметров, который может быть точечным и интервальным;

6) верификация модели – это вывод о точности расчетов на основе модели получаемых прогнозных оценок, а также анализ аппроксимационных ошибок. Если предъявляемые требования к модели удовлетворены, то она может быть ис-

пользована либо для прогнозирования, либо для объяснения (анализа) скрытых механизмов исследуемых процессов.

Прогнозные расчеты по модели выполняются посредством расчета значений внутренних переменных после фиксации значений внешних факторов. При обнаружении явного несоответствия прогнозных траекторий развития объекта исследований содержательному смыслу показателей следует вернуться на один из предыдущих этапов и произвести соответствующие корректировки.

Для формулировки задачи статистического исследования зависимостей функционирование исследуемого объекта (системы, процесса, явления) будем описывать с помощью набора переменных (см. рис. 1.4) [4]:

$(x_1, x_2, ..., x_m)$ – объясняющие переменные, описывающие условия функционирования. В некоторых математических моделях их называют независимыми, экзогенными, факторами-аргументами, предикторами, объясняющими (часть из них, как правило, поддается регулированию или частичному управлению);

$(y_1, y_2, ..., y_n)$ – результирующие переменные, характеризующие поведение или результат функционирования объекта, так называемые отклики, эндогенные, результирующие, зависимые или объясняемые;

$(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ – внешние факторы (ошибки в измерении анализируемых показателей), не поддающиеся учету, отражающие влияние на $y_1, y_2, ..., y_n$;

Результирующая переменная y является откликом функции, значение которой определяется значениями объясняющих переменных $X = (x_1, x_2, ..., x_m)^T$, выступающих в роли аргументов этой функции. Эту связь можно представить математически в виде уравнения регрессионной связи:

$$y = f(X) + \varepsilon, \tag{1.1}$$

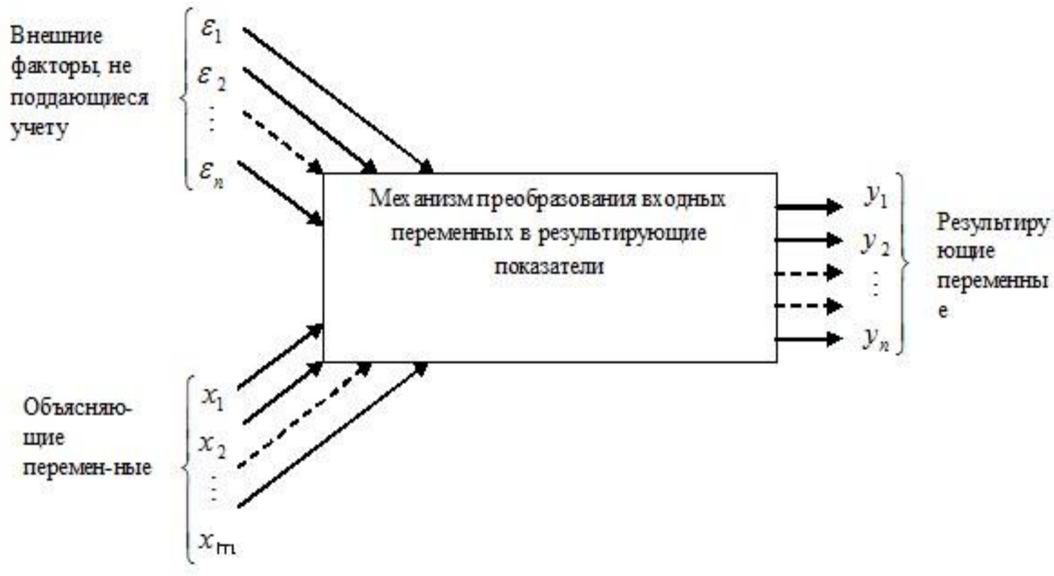


Рисунок 1.4. Схема взаимодействия переменных при статистическом исследовании зависимостей

Объясним появление ошибки ε. Напомним, что есть два направления в интерпретации изначальных данных и их статистической обработке [75]. Следуя первому из них, все наблюдения трактуются как выборка из соответствующей генеральной совокупности, по свойствам которой будем говорить о свойствах всей совокупности. При такой постановке будем основывать свои методы исследования на вероятностной природе исходных данных, применяя для этого соответствующую вероятностную модель. В этом случае ε в (1.1) будем считать случайной величиной, распределенной по заданному закону и строить на этом анализ ее свойств.

При втором подходе совокупность рядов наблюдений – это одна уникальная выборка наблюдений за некоторым реальным объектом (что часто и бывает при моделировании социально-экономических, а также других систем), а любые априорные сведения о вероятностной природе исходных данных отсутствуют.

В работе будем придерживаться второго подхода, при котором ε в (1.1) интерпретируется только как ошибка аппроксимации. Вот почему на ε не налагается заранее требований принадлежности к какому-либо распределению с фиксированными характеристиками и, потому не исследуют традиционные свойства получаемых оценок – состоятельность, несмещенность, эффективность и другие, тем более что они, как правило, имеют асимптотический характер. В конечном счете, правильность выбора подхода определяется, как справедливо отмечено в [3], тем, позволяет ли он исследователю решить поставленную перед ним конкретную задачу. В нашем случае необходимо, как правило, минимизировать ошибку аппроксимации в некотором наперед заданном смысле.

Основные цели регрессионного анализа

1. Регрессионный анализ позволяет устанавливать факт наличия и форму связи между важнейшими показателями с учетом возможной стохастичности этих связей между y и X и сопровождается характеристикой степени тесноты изучаемой зависимости. Подбор формы связи, т. е. вида функции $f(X)$ и состава объясняющих переменных X, играет ключевую роль, цель которого – максимизация величины измерителя степени тесноты связи [2].

2. Разработка прогноза неизвестных значений исследуемых результирующих показателей по заданным значениям X, а также описание интервала вероятных значений прогнозируемого показателя $y(X)$ и величины доверительной вероятности P, с которой гарантируется справедливость прогноза.

3. Выявление причинных связей между объясняющими переменными X и результирующими показателями y, а также отчасти управление значениями y за счет регулирования величин объясняющих переменных X.

Основные задачи регрессионного анализа

1. Для заданных значений независимых переменных $X=(x_1, x_2, ..., x_m)^T$ найти наилучшее в определенном смысле точечные и интервальные (с доверительной вероятностью P) прогнозы $\hat{f}(X)$ и $\Delta[f(X)]_P$ для неизвестной регрессионной функции $f(X)$.

2. По заданным значениям независимых переменных $X=(x_1, x_2, ..., x_m)^T$ найти наилучшие в определенном смысле точечные и интервальные (с доверительной вероятностью P) прогнозы $\hat{y}(X)$ и $\Delta\left[y(X)\right]_P$ для неизвестного значения результирующей переменной $y(X)$.

3. Если известно, что искомая функция регрессии принадлежит некоторому параметрическому семейству функций $\left\{f(X;a)\right\}$, где $a=(a_0, a_1, ..., a_m)^T$ – вектор параметров, причем все или часть компонент которого допускают определенную экономическую интерпретацию. Необходимо, построить наилучшие в определенном смысле точечные и интервальные оценки для неизвестных значений этих параметров.

4. Определить удельный вес (влияние) каждой из объясняющих переменных $(x_1, x_2, ..., x_m)$ в выходном показателе $y(X)$ и, в особенности, определить, какие из объясняющих переменных можно исключить из модели (1.1) как практически не влияющие на процесс формирования значений результирующего показателя [2].

Основной проблемой регрессионного анализа является выбор соотношения между всеми переменными. При построении множественной нелинейной регрессии рассматриваются модели как нелинейные по параметрам, так и по переменным. Таких соотношений необычайно много, и они могут быть чрезвычайно сложны, вот только часть из них:

– уравнение, линейное по параметрам и нелинейное по факторам

$$y_i = a_0 + a_1\varphi_1(x_{i1}, x_{i2}, ..., x_{im}) + a_2\varphi_2(x_{i1}, x_{i2}, ..., x_{im}) + ... + a_k\varphi_k(x_{i1}, x_{i2}, ..., x_{im}) + \varepsilon_i,$$

– степенное уравнение $y_i = a_0 \cdot x_{i1}^{a_1} \cdot x_{i2}^{a_2} \cdot ... \cdot x_{im}^{a_m} \cdot \varepsilon_i$;

– экспоненциальное уравнение $y_i = e^{a_0 + a_1x_{i1} + a_2x_{i2} + ... + a_px_{im} + \varepsilon_i}$;

– гиперболическое уравнение $y_i = \dfrac{1}{a_0 + a_1x_{i1} + a_2x_{i2} + ... + a_px_{im} + \varepsilon_i}$.

При исследовании зависимости между переменными чаще всего используется все-таки линейная форма связи. Это связано с двумя обстоятельствами:

1) возможность четкой экономической интерпретации параметров модели;

2) в большинстве случаев нелинейные модели регрессии преобразуются к линейному виду.

Методы определения общей спецификации модели и оценивания ее параметров в рамках традиционного статистического подхода «устроены» таким образом, что предполагают «встраивание» в модель в неявном виде тех тенденций функционирования объекта, которые имели место в прошлом, на предыстории, к которой относится сформированная статистическая информация. Тем самым при проведении прогнозных расчетов предполагается, что прошлые тенденции «в основном» сохранятся и в будущем. Это свойство статистического подхода не всегда является обременительным, поскольку многие сложные системы обладают сильной инерционностью функционирования (в частности, на небольших временных отрезках) [24].

Наиболее распространенной в эконометрических приложениях формой представления регрессионной зависимости является аддитивная линейная форма, которая и будет главным предметом исследования.

Рассмотрим линейное соотношение между зависимой переменной y и объясняющими факторами $x_1, x_2, ..., x_m$ в виде линейной функции [67]

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + ... + a_mx_m \ . \quad (1.2)$$

Коэффициенты $a_k \quad (k = 0,...,m)$ называются коэффициентами множественной регрессии. В уравнении регрессии постоянная a_0 выполняет роль выравнивания. Она является точкой пересечения гиперповерхности регрессии с осью ординат. Значения $a_1, a_2,...,a_m$ представляют собой оценки коэффициентов регрессии. Причем, $a_k \ (k = \overline{1,m})$ указывает среднюю величину изменения y при изменении x_k на одну единицу при условии, что другие переменные остаются без изменения. Эмпирические значения y можно представить следующим образом:

$$y = \hat{y} + \varepsilon \, , \quad (1.3)$$

В выражении функции (1.3) $\hat{y}_i \quad (i=1,...,n)$ – рассчитанные по модели значения объясняемой переменной. Они равны средним значениям переменной y в точке i при фиксированных значениях x_{ik} объясняющих переменных $x_k \ (k=1,...,m)$ причем, только эти m переменных являются причиной изменения переменной y. Значения $\hat{y}$ представляют собой оценки средних значений y для фиксированных значений переменных x_k в точке i.

Функция регрессии (1.2) может быть представлена компактно в матричной форме

$$\hat{y} = Xa, \tag{1.4}$$

а функция (1.3) – соответственно

$$y = Xa + \varepsilon, \tag{1.5}$$

где

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \qquad X = \begin{bmatrix} 1 & x_{11} & x_{1m} \\ 1 & x_{21} & x_{2m} \\ & & \\ 1 & x_{n1} & x_{nm} \end{bmatrix}; \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}; \qquad a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, \tag{1.6}$$

результаты наблюдений $y_1, y_2,..., y_n$, записываются в форме вектор-столбца размерности $n \times 1$. Значения объясняющих переменных $x_1, x_2,..., x_m$ записываются в виде матрицы X размерности $n \times (m+1)$, а остатки ε – в виде вектор-столбца размерности $n \times 1$. Параметры регрессии $a_k (k=0,...,m)$ образуют вектор-столбец a размерности $(m+1) \times 1$ [67].

1.2 Методы оценивания параметров регрессионных моделей

Методы оценивания параметров регрессионных моделей разделяются на: классические методы, такие как метод наименьших квадратов, метод наименьших модулей и другие, а также устойчивые методы – оценки Хьюбера, М-оценки, знаковый метод и др., и адаптивные методы: *Lv*-оценки, адаптивные оценки на основе универсальных распределений [75, 101, 102].

В зависимости от анализа ошибок измерений выбирается тот или иной способ статистического решения системы уравнений (1.4). Представим самые известные подходы для нахождения оценок вектора неизвестных параметров уравнений регрессии: метод максимального правдоподобия (ММП), метод наименьших квадратов (МНК), метод наименьших модулей (МНМ), метод антиробастного оценивания (МАО).

1.2.1 Метод максимального правдоподобия

Метод максимального правдоподобия (ММП) был предложен Р. Фишером в 1937 году [42, 73, 99]. Для рассмотренной задачи (1.5) величина ε_i называется остаточной невязкой *i*-го измерения. Решение системы (1.5) ищут в статистическом смысле, а именно, подбирают значения неизвестных параметров $a_0, a_1, ..., a_m$ так, чтобы совокупность невязок обладала бы каким-то минимальными свойствами.

Иными словами, решение задачи предполагает построение некоторой функции $G(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ и ее минимизацию в области искомых параметров. Поскольку невязки $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ зависят не только от величины определяемых параметров, $a_0, a_1, ..., a_m$, но и от результатов измерений $y_1, y_2, ..., y_n$, то совокупность значений $\hat{a}_0, \hat{a}_1, ..., \hat{a}_m$, доставляющих минимум функции $G(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, будет зависеть как от вида этой функции, так и от результатов измерений

$$\hat{a}_i = \hat{a}_i(G, y_1, y_2, ... y_n). \tag{1.7}$$

Меняя вид функции $G(\varepsilon_1,\varepsilon_2,....,\varepsilon_n)$, мы будем получать различные наборы $\tilde{a}_0, \tilde{a}_1,...,\tilde{a}_m$. Выбираемое в качестве подходящего значения $\tilde{a}_i$ неизвестного параметра a_i называется оценкой этого параметра.

Такая оценка является некоторой функцией от измерений и также является случайной величиной, обладающей функцией плотности $h(\hat{a})$ и относящейся к закону распределения $H(\hat{a}) = \int\limits_{-\infty}^{\hat{a}} h(\hat{a})d\hat{a}$.

Пусть закон распределения ошибок измерения ε_i известен. Обозначим с помощью $g_i(\varepsilon_i)$ функцию плотности *i*-ой ошибки и с помощью $G(\varepsilon_1,\varepsilon_2,....,\varepsilon_n)$ – многомерную функцию плотности всей совокупности ошибок $\varepsilon_1,\varepsilon_2,....,\varepsilon_n$. В случае, когда ошибки измерений независимы,

$$G(\varepsilon_1,\varepsilon_2,....,\varepsilon_n) = g_1(\varepsilon_1)\cdot g_2(\varepsilon_2),....,g_n(\varepsilon_n) = \prod_{i=1}^{n} g_i(\varepsilon_i)\,. \qquad (1.8)$$

Функция плотности совместного распределения ошибок измерения $G(\varepsilon_1,\varepsilon_2,....,\varepsilon_n)$ называется функцией правдоподобия.

Из выражения (1.8) следует, что каждая невязка есть функция определяемых параметров и имеющихся измерений. Набор $a_0, a_1,...,a_m$, максимизирующий функцию правдоподобия, называется оценкой максимального правдоподобия. При некоторых достаточно общих условиях ММП-оценки являются асимптотически-несмещенными, состоятельными и асимптотически-эффективными. То есть для больших выборок оценки максимального правдоподобия имеют наибольшую возможную точность.

Функция правдоподобия зависит от функции плотности, что видно из выражения (1.8), т. е. от распределения ошибок измерения. Предположим, что ошибки измерения в (1.8) подчинены нормальному закону, т. е.

$$g_i(\varepsilon_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{\varepsilon_i^2}{2\sigma_i^2}}\,. \qquad (1.9)$$

Пусть среднеквадратическая ошибка σ_i каждого измерения одинакова для всех измерений. В этом случае функция правдоподобия имеет вид

$$G = \prod_{i=1}^{n} g_i(\varepsilon_i) = \left(\frac{1}{\sigma_i \sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma_i^2}\sum_{i=1}^{n}\varepsilon_i^2} , \tag{1.10}$$

и для поиска максимально-правдоподобной оценки нужно минимизировать взвешенную сумму квадратов невязок $\Phi = \sum_{i=1}^{n} p_i \varepsilon_i^2$, где $p_i = 1/\sigma_i^2$ и является весом i-го измерения.

1.2.2 Метод наименьших квадратов

Метод наименьших квадратов является самым популярным классическим методом для получения оценок параметров регрессионных моделей. К. Гаусс применил его для расчета орбиты астеройда Церера – для борьбы с ошибками астрономических наблюдений [72]. Данный метод применим, если для случайной ошибки уравнения регрессии (при его вероятностной трактовке) выполняются следующие условия:

1) $M(\varepsilon) = 0$;

2) $M(\varepsilon\varepsilon^T) = \sigma^2 E$;

3) $\sigma^2 < \infty$;

4) $rang(X) = m$, где ε – ошибки регрессии, σ^2 – дисперсия ошибок, X – значения объясняющих переменных, записываются в виде матрицы размерности $n \times (m+1)$.

Чаще всего считают, что ошибки измерений распределены по нормальному закону (закону Гаусса). Причин для этого достаточно много. Например, весьма веским доводом в пользу предположения о нормальном законе распределения ошибок является центральная предельная теорема, которая утверждает, что сумма бесконечно большого числа независимых случайных величин с одинаковыми математическими ожиданиями, распределенными по произвольному закону с равномерно ограниченными дисперсиями, является случайной величиной, распределенной по нормальному закону. Желательно, чтобы регрессия хорошо согласовывалась с эмпирическими данными. Для этого сумма квадратов разностей между наблюдаемыми значениями зависимой переменной и значениями вычисленных по уравнению регрессии должна быть минимальна, в векторной записи имеет вид

$$S(a) = (y - \hat{y})'(y - \hat{y}) = \hat{\varepsilon}'\hat{\varepsilon} \to \min_a, \tag{1.11}$$

Или, подставляя вместо $\hat{y}$ его выражение,

$$S(a) = \hat{\varepsilon}'\hat{\varepsilon} = (y - Xa)'(y - Xa) \to \min_a, \tag{1.12}$$

$$S(a) = y'y - 2a'X'y + a'X'Xa \to \min_a.$$

Продифференцировав (1.12) по элементам *a*, приравниваем полученное выражение к нулю:

$$\frac{\partial S(a)}{\partial a} = -2X'y + 2X'Xa = 0. \tag{1.13}$$

Отсюда получаем нормальные уравнения, которым должен удовлетворять вектор a при соблюдении требования (1.13):

$$X'Xa = X'y, \tag{1.14}$$

откуда

$$\hat{a} = (X'X)^{-1}X'y. \tag{1.15}$$

Основное преимущество МНК в том, что существует аналитическое решение (1.15) для определения неизвестных параметров уравнения регрессии [38, 42, 45, 67, 68, 82, 88]. При вероятностной трактовке МНК-оценки обладают следующими оптимальными свойствами: несмещенность, эффективность, состоятельность. Если случайная ошибка уравнения регрессии распределена по нормальному закону, то МНК-оценки совпадают с оценками максимального правдоподобия, и потому являются наилучшими оценками. Метод наименьших квадратов вследствие возможности аналитического описания оценок особенно популярен. Этот метод имеет еще ряд преимуществ: простой, наиболее точный для определения оценок с ошибкой, подчиненной нормальному закону, который является наиболее распространенным. Кроме того, этот метод обладает свойством транзитивности, т. е. независимости результатов оценки от способа разбиения всей совокупности уравнений, образующих фундаментальную систему, на подгруппы, обрабатываемые независимо с последующим усреднением результатов обработки. Оптимальность метода связана с допущением о нормальном характере распределения ошибок, а также с предположением о точно известных параметрах законов распределения числовых характеристик. Если хотя бы одно из этих двух предположений нарушается, так сразу же возникает вопрос о качестве оценок и сравнении МНК-оценок с другими методами. Необходимо учитывать сложность при получении хорошего «первого приближения». Наконец, определенные сложности появляются при отборе грубых, посторонних и других аномальных измерений. Такие измерения способны отвести полученные оценки от истинных значений параметров настолько, что все дальнейшие измерения, действительно относящиеся к данному процессу, будут выбракованы как неспецифические. Поэтому необходимо иметь в своем арсенале подстраховывающие средства на случай расходимости итерационных процессов, связанных с уточнением начального приближения [73]. Когда дисперсия каждого измерения заранее известна и не известно ничего другого о характере ошибки измерения, наиболее естественным предположением о законе распределения ошибок является предположение о нем, как о нормальном законе, и наиболее правильным методом обработки является обработка измерений по методу наименьших квадратов. Когда же дисперсия измерений известна только в среднем, наиболее логичным предположением о законе распределения ошибок измерения этой группы является предположение о нем, как о законе Лапласа, и наиболее правильным методом обработки в этом случае является обработка измерения по методу наименьших модулей [4, 72].

1.2.3 Метод наименьших модулей

При нарушении ряда предпосылок эффективность МНК сильно снижается [2, 36, 41], например, обрабатываемая выборка может содержать наблюдения, не очень согласующиеся с оставшимися, т. е. имеющая выбросы. В такой ситуации можно использовать метод оценивания, менее чувствительный, чем МНК, к ошибкам спецификации и разрешающий получать «робастные» оценки [101]. К числу таких методов относится метод наименьших модулей, реализация которого приводит к следующей задаче кусочно-линейного программирования:

$$\hat{a} = \arg \min_a \sum_{k=1}^{n} \left| y_k - a_0 - \sum_{i=1}^{m} a_i x_{ki} \right|. \tag{1.16}$$

Есть два способа решить задачу(1.16) [36, 73,75]: с помощью сведения к задаче линейного программирования (ЛП) и метода вариационно-взвешенных квадратичных приближений. Опишем только первый из них. Необходимо ввести неотрицательные вещественные переменные u_k и v_k, $k = \overline{1,n}$ представленные следующим образом:

$$u_k = \begin{cases} y_k - a_0 - \sum_{i=1}^{m} a_i x_{ki}, & при \ \ y_k - a_0 - \sum_{i=1}^{m} a_i x_{ki} > 0 \\ 0, \ в \ противном \ случае \end{cases} \qquad v_k = \begin{cases} a_0 + \sum_{i=1}^{m} a_i x_{ki} - y_k, & при \ a_0 + \sum_{i=1}^{m} a_i x_{ki} - y_k > 0 \\ 0, \ в \ противном \ случае \end{cases}$$

(1.17)

Из чего следует, что имеют место тождества

$$a_0 + \sum_{i=1}^{m} a_i x_{ki} + u_k + v_k = y_k \ , \qquad k = \overline{1,n}. \tag{1.18}$$

Функционал (1.16) преобразуется в линейную форму

$$J_1 = \sum_{k=1}^{n} (v_k + u_k) \ \rightarrow \ \min . \tag{1.19}$$

Тогда, задача (1.16) сводится к задаче (1.19) с ограничениями (1.18) и условиями (1.17). Можно увидеть, что условие $u_k v_k = 0$ для всех k, следующее из определения переменных u_k и v_k, выполняется на оптимальном решении этой задачи.

Этот способ определения оценки привлекателен тем, что позволяет накладывать на a различные линейные ограничения – как равенства, так и неравенства.

Применение МНМ для идентификации зависимостей, описывающих динамику ряда, зачастую показывает лучшие оценки при прогнозировании по сравнению с МНК-оценками [33].

1.2.4 Метод антиробастного оценивания

Антиробастное оценивание параметров в некотором смысле противоположно по отношению к МНМ и базируется на придании заведомо больших «весов» «выбросам» и ориентированное, поэтому на большие отклонения ε в регрессии (1.3).

Данное оценивание следует осуществлять тогда, когда каждое наблюдение выборки уникально, а длина ее ограничена, поэтому игнорировать какие-то наблюдения, используя робастные или близкие к ним методы, может оказаться нерациональным.

Антиробастная оценка (или L_v-оценка при $v = \infty$) является решением задачи

$$J_\infty = \max_k \left| y_k - a_0 - \sum_{i=1}^{m} a_i x_{ki} \right| \to \min. \qquad (1.19)$$

Оценка может быть также определена с помощью задачи линейного программирования [36, 75, 107]. Дополним тождество (1.17) ограничениями:

$$u_k + \upsilon_k \le g, \qquad k = \overline{1,n}, \qquad (1.20)$$

где новая переменная g представляет собой величину, превышающую ошибки аппроксимации ε_k для всех наблюдений выборки и равную максимальной ошибке для такого номера k_0, для которого (1.20) обращается в равенство.

Функционал (1.18) заменим на следующий

$$g \to \min. \tag{1.21}$$

Задача ЛП (1.17), (1.20), (1.21) с ограничениями неотрицательности переменных u_k, υ_k, $k = \overline{1,n}$, эквивалентна задаче (1.19). Неотрицательность переменной g обеспечивается условиями (1.20). Если в задаче (1.17), (1.20), (1.21) функционал (1.21) заменить на (1.18), то получившаяся в результате оценка a будет МНМ-оценкой, то есть дополнение ограничений (1.17) и функционала (1.18) условиями (1.20) не меняет решения задачи.

1.3. Программное обеспечение для оценивания параметров регрессионных моделей

В настоящее время разработано немало программных средств для обработки статистических данных. Поэтому перед тем, как приступить к разработке нового математического и программного обеспечения для оценивания параметров регрессионных моделей, необходимо провести анализ уже существующих для решения данных задач программ. Были рассмотрены следующие программные продукты: Analyse-it, EViews, GenStat, Gretl, LIMDEP, Wolfram Mathematica, MATLAB, Minitab, NCSS, R, RATS, SageMath, SAS, SHAZAM, SPSS, Stata, Statgraphics, Statistica, SYSTAT. Краткое описание этих систем с акцентом на методы анализа данных представлено ниже.

Программа Analyse-it представляет собой надстройку для табличного процессора Microsoft Excel. Эта программа обеспечивает ряд стандартных параметрических и непараметрических процедур, в том числе корреляционный анализ, линейная и полиномиальная регрессия, логистическая регрессия, анализ главных компонент и факторный анализ.

EViews представляет собой статистический пакет для Windows. EViews может быть использован для эконометрического анализа пространственных и панельных данных, а также временных рядов. Пакет позволяет оценивать параметры регрессионных моделей практически любым известным способом.

GenStat представляет собой статистический пакет с возможностями анализа данных, в частности, в области сельского хозяйства. Позволяет оценивать модели линейной и нелинейной регрессии, обобщенные линейные модели, логистические регрессии.

Gretl – кросс-платформенный эконометрический пакет, обладающий Открытой Публичной Лицензией GNU, которая гарантирует любому пользователю бесплатный доступ к программе. В пакете реализованы все основные эконометрические процедуры и функции, а также в нем содержатся встроенные наборы статистических данных.

LIMDEP – эконометрический пакет с различными инструментами оценки. LIMDEP поддерживает методы анализа пространственных и панельных данных, временных рядов. Пакет имеет встроенный язык программирования, для оценивания моделей, не содержащихся во встроенном меню типовых форм. Позволяет работать с линейными и нелинейными регрессиями, робастными оценками, логистическими регрессиями и т.д.

Wolfram Mathematica – математическая программа символьных расчетов, называемая также системой компьютерной алгебры. Программа позволяет оценивать параметры линейных и нелинейных регрессионных моделей, обобщенные модели регрессии, логит и пробит модели.

MATLAB (матричная лаборатория) является многопрофильной парадигмой численной вычислительной среды и языком программирования четвертого поколения. MATLAB позволяет взаимодействовать с программами, написанными на других языках, включая C, C ++, C #, Java, Fortran и Python. Пакет Statistics Toolbox содержит процедуры и функции для анализа данных. Пакет позволяет использовать описательную статистику и строить графики для разведочного анализа данных, пригодных для вероятностных распределений данных, генерировать случайные числа для моделирования методом Монте-Карло, а также выполнять проверки гипотез. Алгоритмы регрессии и классификации позволяют делать выводы и осуществлять прогнозирование. Для многомерного анализа данных Toolbox обеспечивает набор функций для ступенчатой регрессии, анализа главных компонент, регуляризации и других методов снижения размерности.

Minitab – статистический пакет, позволяющий строить линейные регрессии, логистические регрессии, нелинейные регрессии, ортогональные регрессии, регрессии Пуассона, формировать различную графическую информацию, в том числе, строить графики остатков моделей, для выбора лучшей модели осуществлять пошаговые процедуры.

Программное обеспечение NCSS обеспечивает полную и легкую в использовании коллекцию сотен статистических и графических инструментов для анализа и визуализации данных. NCSS позволяет оценивать линейную регрессию, робастную линейную регрессию, множественную регрессию, ридж-регрессию, логистическую регрессию, регрессию Кокса, регрессию Пуассона, геометрическую регрессию, регрессию Деминга, нелинейную регрессию, гармоническую регрессию.

NLOGIT является продолжением эконометрического и статистического программного пакета LIMDEP. В дополнение к инструментам оценки в LIMDEP, NLOGIT предоставляет возможности для оценки моделей полиномиальных данных и для данных обследований и рынка, в котором потребители выбирают среди множества конкурирующих альтернатив. В дополнение к экономическим наукам, NLOGIT имеет применения в биостатистике, неэкономических социальных науках, физических науках.

R – это язык программирования с открытым исходным кодом и программная среда для статистической обработки данных. Содержит широкий набор встроенных функций для оценивания регрессионных моделей с помощью методов наименьших квадратов, модулей, максимального правдоподобия и т.д.

Программа RATS, позволяющая выполнять ряд эконометрических и статистических операций: строить линейную регрессию, регрессию с гетероскедастичностью и последовательной корреляционной коррекцией, нелинейные регрессии, системы уравнений, регрессии с дискретными зависимыми переменными, в том числе логистические регрессии. Все эти методы могут быть использованы для прогнозирования, а также для проведения анализа данных.

SageMath – математическое программное обеспечение с функциями, охватывающими многие аспекты математики. Позволяет оценивать параметры регрессий практически всеми известными методами.

SAS (статистическая аналитическая система) – программный пакет, разработанный SAS Institute для расширенной аналитики, многомерного анализа, бизнес-аналитики, управления данными и интеллектуального анализа. Одной из компонент системы является SAS/ETS – компонент для построения эконометрических моделей и анализа временных рядов. Эта компонента позволяет оценивать параметры регрессионных моделей практически любым известным способом.

SHAZAM – это комплексный эконометрический пакет статистических данных для оценки, тестирования, моделирования и прогнозирования многих типов эконометрических и статистических моделей. В этой программе можно оценивать модели ARCH, GARCH, ARIMA, модели с автокорреляцией, проверять значимость параметров и коэффициентов моделей, прогнозировать, строить многие виды регрессионных моделей.

SPSS Statistics – это известный программный пакет, используемый для логического пакетного статистического анализа.

Stata – это универсальный статистический программный пакет. Большинство его пользователей работают в области исследований, особенно в области экономики, социологии, политологии, биомедицины и эпидемиологии. Позволяет оценивать линейные модели методами наименьших квадратов, максимального правдоподобия и т.д.

Statgraphics – это пакет статистики, который выполняет базовые и расширенные статистические функции.

Statistica – это усовершенствованный пакет аналитического программного обеспечения. Statistica обеспечивает анализ данных, управление данными, статистику, интеллектуальный анализ данных, машинное обучение, текстовую аналитику и процедуры визуализации данных.

SYSTAT – пакет статистических и графических программ.

Результаты анализа рассмотренных систем с точки зрения оценивания регрессий по методам наименьших квадратов, модулей, максимального правдоподобия и антиробастного оценивания представлены в табл. 1.1. В ней символ «+» означает наличие в пакете соответствующих функция, а символ «–» – отсутствие функций.

Таблица 1.1

Программное обеспечение для оценивания регрессионных моделей

№	**Пакет**	**МНК**	**МНМ**	**ММП**	**МАО**
1	Analyse-it	+	–	–	–
2	EViews	+	+	+	+
3	GenStat	+	+	+	+
4	Gretl	+	+	+	–
5	LIMDEP	+	+	+	+
6	Wolfram Mathematica	+	+	+	+
7	MATLAB (Statistics Toolbox)	+	+	+	+
8	Minitab	+	–	+	–
9	NCSS	+	+	+	–
10	NLOGIT	+	+	+	+

11	R	+	+	+	+
12	RATS	+	+	+	+
13	SageMath	+	+	+	+
14	SAS	+	+	+	+
15	SHAZAM	+	+	+	+
16	SPSS	+	–	–	–
17	Stata	+	+	+	+
18	Statgraphics	+	–	+	–
19	Statistica	+	+	+	+
20	SYSTAT	+	+	+	–

1.4 Выводы

В первой главе представлены этапы процесса моделирования, сложных систем, таких как: национальная экономика, экономика региона, а также их взаимодействие. Являясь объектом усиленного внимания исследователей, данного вида системы способствуют развитию инструментария математического моделирования экономики и разработки методики использования современных вычислительных методов и технических систем. Показано, что при исследовании экономических систем математические модели имеет смысл разделить по свойствам моделируемых объектов, по направлениям приложения, по степени описания и инструментам моделирования.

Представлены основные этапы построения экономико-математической модели.

Сформулирована постановка задачи построения математической модели статистического типа путем оценивания параметров входящих в ее состав регрессионных уравнений. Описаны основные цели и задачи регрессионного анализа, как одного из методов построения математических моделей.

Дано подробное описание аддитивной, линейной формы регрессии, которая и является главным предметом исследования.

Представлены самые известные подходы для нахождения оценок вектора неизвестных параметров уравнений регрессии: метод максимального правдоподобия (ММП), метод наименьших квадратов (МНК), метод наименьших модулей (МНМ), метод антиробастного оценивания (МАО). В зависимости от анализа ошибок измерений выбирается тот или иной способ статистического решения. Установлены достоинства и недостатки каждого и описаны перспективные подходы к их реализации.

Представлено краткое описание программных продуктов: Analyse-it, EViews, GenStat, Gretl, LIMDEP, Wolfram Mathematica, MATLAB, Minitab, NCSS, R, RATS, SageMath, SAS, SHAZAM, SPSS, Stata, Statgraphics, Statistica, SYSTAT, с акцентом на методы анализа данных.

2 МНОЖЕСТВЕННОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ЛИНЕЙНОГО РЕГРЕССИОННОГО УРАВНЕНИЯ

2.1 Математическое обеспечение решения задачи множественного оценивания параметров

По-прежнему будем рассматривать линейное регрессионное уравнение

$$y_k = a_0 + \sum_{i=1}^{m} \alpha_i x_{ki} + \varepsilon_k,\ k = \overline{1,n}, \tag{2.1}$$

где y_k – эндогенная (объясняемая, зависимая, выходная), а x_i – i-ая экзогенная (объясняющая, независимая, входная) переменная; α_i – i-ый, подлежащий оцениванию, параметр; ε – ошибки аппроксимации, k – номер наблюдения, n – число наблюдений (длина выборки).

Представим уравнение (2.1) в матричной форме:

$$y = X\alpha + \varepsilon, \tag{2.2}$$

где $y = (y_1, \ldots, y_n)^T$, $\alpha = (\alpha_0, \ldots, \alpha_m)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, X – $(n \times (m+1))$ матрица с компонентами x_{ki} первый столбец которой состоит из единиц.

Присутствие в уравнениях (2.1), (2.2) аппроксимационной составляющей может быть вызвано:

- неточностями в регистрации значений эндогенной и экзогенных переменных;
- неучетом каких-либо существенных факторов;
- влиянием помех;
- неточным (неудачным) выбором формы связи между переменными.

Вопросам оценивания неизвестных параметров регрессионных уравнений посвящена весьма обширная литература. Достаточно упомянуть фундаментальные монографии российских и зарубежных ученых [3–5, 9, 36, 37, 39, 57, 80, 81, 88], а также работы, посвященные некоторым частным вопросам оценивания [19 - 21, 79].

Отметим одно отмечаемое ранее важное обстоятельство. Как уже отмечалось, если следовать мнению известного российского специалиста в области анализа данных Айвазяна С. А., существует два подхода к интерпретации ε в (2.2) и, соответственно, два варианта формирования методических основ для оценивания вектора параметров α. В первом из них ε трактуется как случайная величина, распределенная по некоторому закону. В соответствии со вторым ε − это ошибки модельного описания реального объекта или процесса и только. Первый подход называется вероятностным, второй − аппроксимационным. Мы в данной работе придерживаемся второго подхода.

Общим для этих подходов является то, что и в том, и другом случае обработка данных осуществляется в соответствии с некоторым функционалом качества метода оценивания. А вот различие состоит в обосновании выбора такого функционала и в интерпретации полученных результатов.

В начале данной главы мы приведем ряд выкладок и положений, уже изложенных выше, с тем, чтобы привести их в соответствие с обозначениями, фигурирующими в уравнении (2.1).

Широкий класс методов оценивания параметров уравнения (2.2) связан с поиском уже упоминавшихся ранее так называемых L_ν-оценок посредством минимизации функций потерь вида [36]:

$$J_\nu(\alpha) = \sum_{k=1}^{n} \left|\varepsilon_k\right|^\nu. \tag{2.3}$$

Каждая из этих оценок характеризуется реакцией на так называемые выбросы, то есть наблюдения, несогласующиеся со всей выборкой в целом. При этом, чем больше значение ν, тем сильнее L_ν-оценка реагирует на выбросы. В регрессионном анализе методы оценивания, слабо реагирующие на выбросы, или вообще их игнорирующие, называют робастными [101].

Методом оценивания параметров уравнения (2.2), соответствующим $\nu = 2$, является описанный ранее, всем хорошо известный и наиболее популярный в регрессионном анализе метод наименьших квадратов (МНК), соответствующий метрике Эвклида, при $\nu = 1$ − это метод наименьших модулей (МНМ), соответствующий городскому расстоянию, при $\nu \to \infty$ − метод антиробастного оценивания (МАО), соответствующий расстоянию Чебышева. Отметим, что в соответствии с хорошо известным фактом имеет место соотношение:

$$J_\infty = \max_{k=\overline{1,n}} |\varepsilon_k|. \tag{2.4}$$

В упомянутых выше источниках описаны способы расчета вектора параметров α в соответствии с каждым из трех методов. При этом использование МНК приводит к аналитическому выражению:

$$\alpha = (X^T X)^{-1} X^T y.$$

Для отыскания вектора α по методу наименьших модулей можно использовать аппарат линейного программирования (ЛП) следующим образом, несколько отличавшемся от использованного в первой главе. Представим вектор ошибок ε в виде:

$$\varepsilon = u - \upsilon,$$

где u − положительная часть ε, а υ − отрицательная часть. Или, более строго,

$$u_k = \begin{cases} \varepsilon_k, & \varepsilon_k > 0, \\ 0, & \text{в пр. случ.} \end{cases} \qquad \upsilon_k = \begin{cases} -\varepsilon_k, & \varepsilon_k < 0, \\ 0, & \text{в пр. случ.} \end{cases}$$

Легко видеть, что имеют место равенства $|\varepsilon_k| = u_k + \upsilon_k$, $u_k \upsilon_k = 0$, $k = \overline{1,n}$. Поскольку среди компонент вектора α могут быть отрицательные, по аналогии с ε представим его в виде

$$\alpha = \alpha^1 - \alpha^2,$$

где α^1 и α^2 − положительные и отрицательные части α соответственно.

Представим уравнение (2.1) в виде

$$y_k = \sum_{i=1}^{m} (\alpha_i^1 - \alpha_i^2) x_{ki} + u_k - \upsilon_k, \quad k = \overline{1,n}. \tag{2.5}$$

Представим далее равенства (2.5) в матричной форме и их преобразуем:

$$y = X\alpha^1 - X\alpha^2 + u - \upsilon. \tag{2.6}$$

По определению все неизвестные переменные неотрицательны:

$$\alpha^1 \geq 0,\ \alpha^2 \geq 0,\ u \geq 0,\ \upsilon \geq 0. \qquad (2.7)$$

Целевая функция задачи ЛП имеет вид:

$$\sum_{k=1}^{n}(u_k + \upsilon_k) \to \min . \qquad (2.8)$$

Решение задачи ЛП (2.6) – (2.8) позволяет оценить параметры уравнения (2.2) в соответствии с МНМ. Она имеет $2((m+1)+n)$ переменных и n ограничений.

В [75] показано, что МАО-оценка также может быть определена с помощью решения задачи ЛП.

Действительно, поскольку в этом случае необходимо минимизировать функцию (2.4), а номер, на котором реализуется максимум модуля ошибки, заранее неизвестен, введем в рассмотрение новую неизвестную переменную r по правилу:

$$r = \max_{k=\overline{1,n}} |\varepsilon_k|.$$

Тогда естественным образом можно выписать систему линейных неравенств:

$$u_k + \upsilon_k - r \leq 0, \qquad k = \overline{1,n}. \qquad (2.9)$$

Целевая функция в задаче ЛП принимает вид:

$$r \to \min . \qquad (2.10)$$

МАО-оценка будет получена посредством решения задачи ЛП (2.6), (2.7), (2.9), (2.10). Эта задача по сравнению с (2.6) – (2.8) содержит дополнительные n ограничений-неравенств и еще одну неизвестную переменную.

МНМ- и МАО- оценки являются своего рода антиподами: первая вообще игнорирует выбросы, вторая к ним тяготеет.

Это обстоятельство наталкивает на мысль оценивания параметров уравнения (2.2) одновременно по двум критериям – $J_1(\alpha)$ и $J_\infty(\alpha)$, то есть по векторному критерию $J(\alpha) = (J_1(\alpha),\ \ J_\infty(\alpha))$. Это позволило бы максимально увеличить информативность процедуры оценивания, извлечь из выборки (X, y) всю заключающуюся в ней информацию при построении уравнения (2.2).

Прежде всего, заметим, что добавление к ограничениям (2.6), (2.7) ограничений (2.9) не изменит МНМ-оценки, если целевая функция по-прежнему будет иметь вид (2.8). Таким образом, задачи ЛП для определения МНМ- и МАО- оценок различаются лишь целевой функцией − соответственно (2.8) и (2.10).

Итак, задача состоит в минимизации векторного критерия $J(\alpha)$ при ограничениях (2.6), (2.7), (2.9). Она относится к классу задач многомерного линейного программирования (МЛП). Для удобства последующего изложения будем через $Z \subset R^{2(n+m+1)+1}$ обозначать множество векторов $z = (\alpha^1, \alpha^2, u, \upsilon, r)$, удовлетворяющих ограничениям (2.6), (2.7), (2.9).

Из теории принятия решений, занимающейся многокритериальными проблемами, известно, что классического решения задачи минимизации критерия J на симплексе Z, как правило, не существует, тем более, когда частные критерии полярны. То есть не существует элемента $z^* \in Z$ такого, что $J_1(z^*) \le J_1(z),\ J_\infty(z^*) \le J_\infty(z),\ \forall z \in Z$. Под решением такой многокритериальной задачи обычно понимают множество Парето $P \subseteq Z$, характеризующееся тем, что ни одно паретовское решение не может быть улучшено по какому-либо одному критерию без ухудшения значения хотя бы одного из оставшихся. Или, более строго,

$$P = \left\{ z \in Z \middle| (\forall z_1 \in Z) \quad z_1 \overline{P_1} z \right\},$$

где P_1 − отношение Парето, задаваемое по правилу

$$\begin{aligned}&(\forall z_1,\ z_2 \in Z)\ \ z_1 P_1 z_2 \Leftrightarrow \left[(J_1(z_1) \le J_1(z_2) \wedge (J_\infty(z_1) \le J_\infty(z_2) \right] \\ &\wedge \left[(\exists s = 1, \infty)\ \ (J_s(z_1) < J_s(z_2) \right].\end{aligned}$$

Таким образом, решением задачи оценивания параметров регрессии (2.2) по двум критериям $J_1(\alpha)$ и $J_\infty(\alpha)$ одновременно будет множество оценок. Назовем его L-множеством по аналогии с L_ν-оценками [76]. Существует фундаментальная работа американских математиков L. Ю и М. Зелены [108], где изложен так называемый многокритериальный симплекс-метод решения задач МЛП.

Будем в дальнейшем через P^* обозначать множество паретовских вершин многогранника Z, через $J(P)$ − образ множества P в критериальном пространстве, а через $S(a_1,....,a_l)$ − выпуклую оболочку векторов $a_1,....,a_l$.

Многокритериальный симплекс-метод основан на следующих фундаментальных результатах [108].

1. Множество P^* связно.
2. Если внутренняя точка грани (выпуклой комбинации вершин) Z является паретовской, недоминируемой, то вся грань паретовская.
3. Если внутренняя точка грани Z является непаретовской (доминируемой), то вся грань непаретовская.
4. Так называемая программа отсутствия мажорирования состоит в следующем.

Пусть $z^0 \in Z$. Сформируем задачу ЛП:

$$J^0 = d_1 + d_\infty \to \max_{\tilde{Z}},$$

$$\tilde{Z} = \{(z,d) \big| \; z \in Z, \; J_s(z) + d_s \le J_s(z^0), d_s \ge 0, \; s = 1,\infty\}.$$

Имеет место теорема [108]:

$$z^0 \in P \Leftrightarrow J^0 = 0.$$

Кроме того, в [108] изложены некоторые простые необходимые условия паретовости вершин Z.

В [108] описаны также два способа формирования множества P^*.

Первый из них имеет итерационный конструктивный характер.

Второй же предполагает применение приема последовательного свертывания критериев. Рассмотрим его более подробно.

Сформируем линейную свертку критериев J_1 и J_∞:

$$J_\gamma(z) = \gamma J_1(z) + (1-\gamma) J_\infty(z), \qquad \gamma \in (0,1). \tag{2.11}$$

Построим на интервале (0, 1) равномерную сеть:

$$0 < \gamma_1 < \gamma_2 ... < \gamma_l < 1.$$

Для каждого узла γ_i, $i = \overline{1,l}$ решим обычную со скалярной целевой функцией задачу ЛП:

$$\min_{z \in Z} J_\gamma(z).$$

В [108] доказано, что ее решением является паретовская вершина. При достаточно мелкой сети, таким образом, формируется все множество P^*. Необходимо иметь в виду одно важное обстоятельство − каждой паретовской вершине в критериальном пространстве в качестве прообраза может соответствовать грань многогранника Z, представляющая собой выпуклую оболочку паретовских вершин.

Упорядочим по возрастанию одной из компонент (например, первой) значения элементов множества $J(P^*) = (J^1, J^2, ..., J^q)$. Очевидно, что множество Парето в критериальном пространстве $J(P)$ является объединением ребер S, соединяющих соседние паретовские вершины симплекса $J(Z)$:

$$J(P) = \bigcup_{j=1}^{q-1} S(J^j, J^{j+1}). \tag{2.12}$$

При $q = 5$ $J(P)$ может иметь вид, представленный на рис. 2.1.

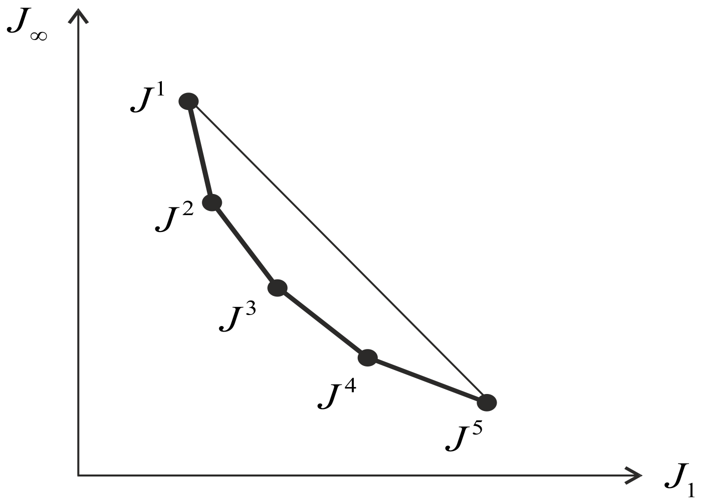


Рисунок 2.1. Множество $J(P)$ в критериальном пространстве

Прообразами ребер $S(J^j, J^{j+1})$ являются множества векторов $z \in Z$, удовлетворяющие ограничениям

$$\sum_{k=1}^{n}(u_k+\upsilon_k)=\lambda J_1^j+(1-\lambda)J_1^{j+1}, \tag{2.13}$$

$$r=\lambda J_\infty^j+(1-\lambda)J_\infty^{j+1}, \qquad j=\overline{1,q-1} \tag{2.14}$$

$$0\le\lambda\le1. \tag{2.15}$$

Обозначим через C^j множество векторов $z\in R^{2(m+1+n)+1}$, удовлетворяющих ограничениям (2.6), (2.7), (2.9), (2.13) – (2.15). Тогда множество P представимо в виде:

$$P=\bigcup_{j=1}^{q-1}C^j. \tag{2.16}$$

В качестве вектора параметров α регрессионного уравнения (2.2) можно принять любой вектор $\alpha=(z_1-z_{m+1},\quad z_2-z_{m+2},...,z_m-z_{2m})$ при $z\in P$.

Безусловно, с такой формой задания модели, в которой параметры определены неявным образом, работать трудно. Поэтому представляется целесообразным иметь какие-то конструктивные приемы, облегчающие эту работу. Рассмотрим некоторые из них.

В работе [77] описан способ точечной характеризации множества Парето, позволяющий оперировать не со всем множеством, а с неким его «полномочным представителем», который в какой-то степени отражает в себе свойства всего множества. Таким представителем может быть, например, центр тяжести J^* множества $J(P)$, характеризующий его конфигурацию. Он рассматривается как выпуклая комбинация паретовских вершин многогранника $J(P)$ с равными коэффициентами:

$$J^*=\frac{1}{g}\sum_{i=1}^{g}J^i. \tag{2.17}$$

Очевидно, что J^* не будет являться паретовской точкой многогранника $J(P)$. Определим точку $J^\#$, максимально улучшающую J^* по обоим критериям одновременно. Воспользуемся для этого программой отсутствия мажорирования и решим задачу

$$J(z^*)=d_1+d_\infty\to\max_{z\in\tilde{Z}},$$

$$\tilde{Z} = \{(z,d) \mid z \in Z, \sum_{k=1}^{n}(u_k + \upsilon_k) + d_1 \le J_1^*,$$
$$r + d_\infty \le J_\infty^*, d_1 \ge 0, \ d_\infty \ge 0\}$$

Решение этой задачи z^* и будет являться искомой точечной характеризацией множества P. Оно может также трактоваться как компромиссное решение задачи оценивания параметров уравнения (2.2). Заметим также, что будет справедливо равенство $J(z^*) = J^\#$.

Рассмотрим способ повышения «осязаемости» в восприятии множества P [10]. Это может быть сделано, в частности, посредством построения множества A

$$A = \{\alpha \in R^m \mid \alpha_i \in [\underline{\alpha_i}, \overline{\alpha_i}]\}, \tag{2.18}$$

которому гарантированно будут принадлежать вектора $\alpha = (\alpha_0, ..., \alpha_m)^T$ с компонентами

$$\alpha_i = z_i - z_{i+m}, \ i = \overline{0,m}. \tag{2.19}$$

А представляет собой параллелепипед в $m+1$-мерном пространстве, в который вписана проекция $\overline{P}$ множества P на R^m. Легко видеть, что справедливы равенства

$$\underline{\alpha}_i = \underline{\alpha_i^1} - \overline{\alpha_i^2}, \qquad \overline{\alpha}_i = \overline{\alpha_i^1} - \underline{\alpha_i^2}.$$

Здесь $\underline{\alpha_i^1}, \underline{\alpha_i^2}$ и $\overline{\alpha_i^1}, \overline{\alpha_i^2}$ – соответственно минимально и максимально возможные значения положительных и отрицательных частей компонент вектора оцениваемых параметров.

Для построения множества A необходимо для каждого параметра α_i решить $2(g-1)$ следующих задач ЛП:

$$\underline{\alpha}_i^j = \min_{z \in C^j}\left(\alpha_i^1 - \alpha_i^2\right), \tag{2.20}$$

$$\overline{\alpha}_i^j = \max_{z \in C^j}\left(\alpha_i^1 - \alpha_i^2\right), \tag{2.21}$$

$$j = \overline{1, g-1}.$$

Тогда $\underline{\alpha}_i$ и $\overline{\alpha}_i$ в представлении (2.18) отыщутся по формулам:

$$\underline{\alpha}_i = \min_{j=\overline{1,g-1}} \underline{\alpha}_i^j , \ \overline{\alpha}_i = \max_{j=\overline{1,g-1}} \overline{\alpha}_i^j .$$

Имея множество A, легко формировать вектора $\tilde{\alpha}$ для регрессии (2.2). Но, как правило, имеет место свойство

$$A \setminus \overline{P} \neq \varnothing ,$$

т. е. A содержит «лишние» параметры, являющиеся компонентами непаретовских векторов $\tilde{z} \in Z$.

Для того чтобы выявить такие вектора $\tilde{z} \in Z \setminus P$, необходимо всякий раз реализовывать программу отсутствия мажорирования. Она формируется следующим образом.

Найдем вектор $\tilde{\varepsilon} = y - X\tilde{\alpha}$. Далее легко выписать вектора $\tilde{u} - \tilde{\upsilon} = \tilde{\varepsilon}$, $\tilde{u} + \tilde{\upsilon} = |\tilde{\varepsilon}| = (|\varepsilon_1|,\ldots,|\varepsilon_n|)$, $\tilde{r} = \max_{k=\overline{1,n}} |\tilde{\varepsilon}_k|$.

Вектор z^0, фигурирующий в программе отсутствия мажорирования, примет вид:

$$z^0 = (\tilde{\alpha}^1, \quad \tilde{\alpha}^2, \quad \tilde{u}, \quad \tilde{\upsilon}, \quad \tilde{r}).$$

В случае, если z^0 окажется непаретовским, в качестве компонент вектора параметров α регрессии (2.2) следует принять разности (2.19) по отношению к вектору, который мажорирует z^0.

Вместо программы отсутствия мажорирования можно проверить выполнение условия, следующего из (2.13) – (2.15):

$$\frac{\sum(\tilde{u}_k + \tilde{v}_k) - J_1^{j+1}}{J_1^j - J_1^{j+1}} = \frac{r - r_\infty^{j+1}}{J_\infty^j - J_\infty^{j+1}} = \lambda \in [0;1]$$

для какого-либо $j = \overline{1, g-1}$.

Рассмотрим проблему прогнозирования значений эндогенной переменной y регрессии (2.2) с множественной оценкой ее параметров такой, что $\tilde{z} \in P$, или, что то же, $J(\tilde{z}) \in J(P)$.

Пусть заданы значения экзогенных переменных уравнения

$$x_i = \tilde{x}_i, \qquad i = \overline{1,m}.$$

Поскольку оценка параметров имеет множественный характер, естественно считать, что соответствующее прогнозное значение переменной y также будет принадлежать множеству – отрезку $[\underline{y}, \overline{y}]$. Ниже приведен способ расчета его границ.

Решим $2(q-1)$ задач ЛП:

$$\underline{y}^{j} = \min_{\tilde{z} \in C^{j}} \sum_{i=1}^{m} \tilde{\alpha}_i \tilde{x}_i,$$
$$\overline{y}^{j} = \max_{\tilde{z} \in C^{j}} \sum_{i=1}^{m} \tilde{\alpha}_i \tilde{x}_i,$$
$$j = \overline{1, q-1}.$$

Тогда получим

$$\underline{y} = \min_{j=\overline{1,\, q-1}} \underline{y}^{j},\ \ \overline{y} = \max_{j=\overline{1,\, q-1}} \overline{y}^{j},$$

где $\underline{y}$ – нижняя граница $\overline{y}$ – верхняя граница прогнозного интервала эндогенной переменной.

2.2 Выбор среды программирования и формулировка требований к ПК МОРМ

Очевидно, что описанный в предыдущем параграфе математический аппарат построения множественных оценок регрессионных моделей не рационально стремиться реализовывать вручную. Это связано с тем, что в процессе множественного оценивания регрессии требуется решать большое количество задач линейного программирования с различными целевыми функциями и ограничениями. Следовательно, возникает необходимость в разработке специализированного программного комплекса, позволяющего быстро и точно справляться со всеми тремя рассмотренными ранее приемами: точечной характеризацией множества Парето, построением множества А и прогнозированием. В данном параграфе приводится описание программного комплекса множественного оценивания регрессионных моделей (ПК МОРМ) [87]. Его разработка осуществлялась в порядке решения следующих задач.

1. Выбор среды программирования.
2. Формулировка требований к системе.
3. Разработка архитектуры, информационной схемы и общего алгоритма работы ПК МОРМ.
4. Разработка алгоритмов работы подсистем ПК МОРМ.
5. Написание программного кода.

Методика множественного оценивания регрессионных моделей предполагает решение большого количества задач линейного программирования. Понятно, что для их решения нет никакого смысла заниматься программированием, например, симплекс-метода, потому что для этого в настоящее время существует большое количество специализированных программ. Выбор был остановлен на бесплатном пакете для решения задач линейного программирования – LPSolve. Этот пакет достаточно эффективно справляется с задачами с большим количеством переменных и ограничений.

К сожалению, в пакете LPSolve нет встроенного языка программирования, поэтому разработку ПК МОРМ, в частности, его интерфейса, было принято реализовывать в среде программирования Delphi. При этом была поставлена и решена задача организации взаимодействия Delphi с пакетом LPSolve.

Сформулируем требования к ПК МОРМ.

В программном комплексе должен быть целиком реализован рассмотренный в предыдущей главе математический аппарат: точечная характеризация множества Парето, построение множества А и прогнозирование. По желанию пользователя комплекс должен выводить не только конечные, но и все промежуточные результаты вычислений. Должно быть предусмотрено графическое представление множества Парето в критериальном пространстве.

В программном комплексе должна быть предусмотрена система ввода/вывода данных. При этом должна быть предусмотрена возможность создания исходных данных вручную или с помощью их экспорта из текстового файла. Комплекс должен предусматривать изменение структуры данных, т. е. реагировать на заданное пользователем количество наблюдений и объясняющих переменных.

Интерфейс комплекса, в том числе его главное меню, должен быть дружественным, реализовывать преимущества графического и табличного отображения.

При работе с комплексом должен быть организован всесторонний контроль ошибок. Например, если статистические данные будут введены в систему с ошибкой, то комплекс должен сигнализировать об этом пользователю.

Составные части комплекса должны быть достаточно гибкими и настраиваемыми. Комплекс должен позволять подключать новые модули, расширяя и дополняя ранее созданные программные блоки.

2.3 Архитектура, информационная схема и общий алгоритм работы ПК МОРМ

В соответствии со сформулированными требованиями была сформирована архитектура ПК МОРМ, представленная на рис. 2.2. Она включает в себя семь взаимосвязанных блоков: «Блок управления», «Блок формирования исходных данных», «Блок формирования множества Парето», «Блок формирования множества А», «Блок прогнозирования», «Блок реализации программы отсутствия мажорирования», «Блок контроля».

Рассмотрим эти подсистемы подробнее. Главными в архитектуре на рис. 2.2 являются «Блок управления» и «Блок контроля».

«Блок управления» предназначен для организации работы и взаимодействия всех основных подсистем комплекса. Он производит запуск и остановку независимых блоков, прием и передачу необходимой информации между блоками, хранение значений различных переменных, необходимых для функционирования комплекса.

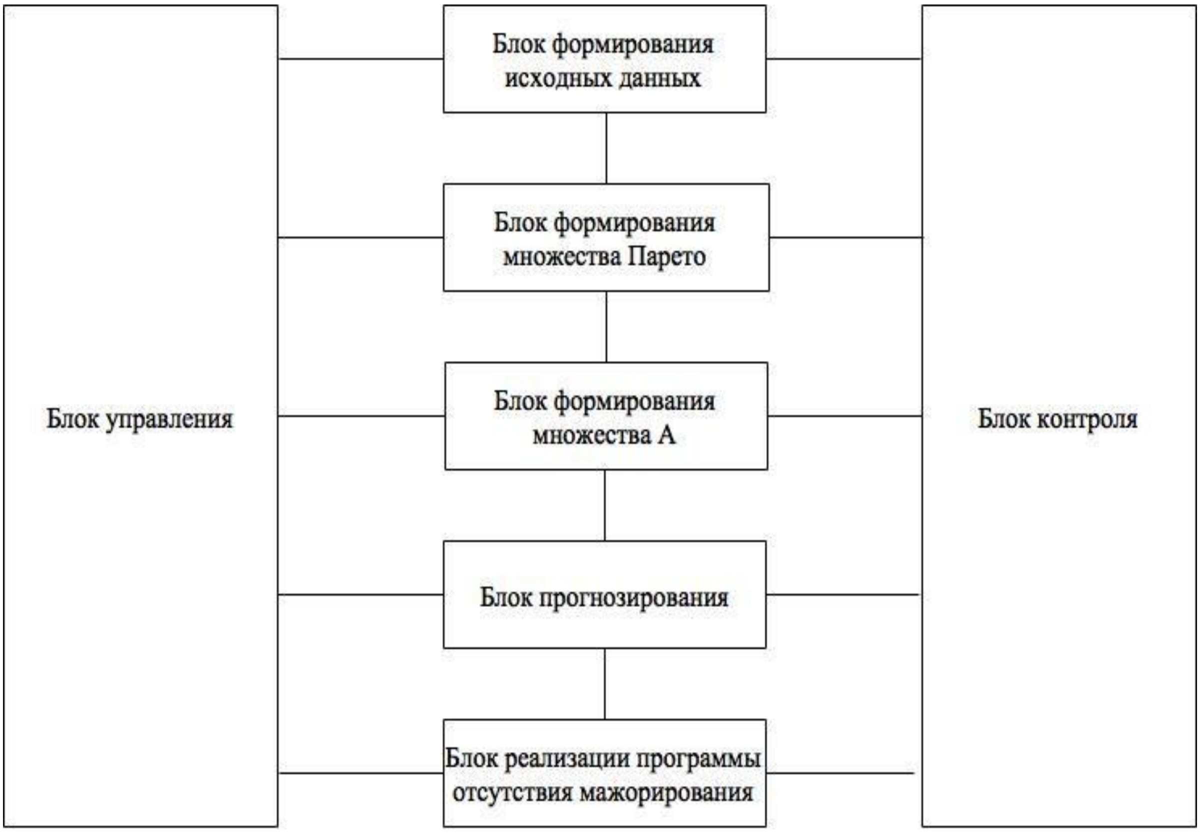


Рисунок 2.2. Архитектура ПК МОРМ

«Блок контроля» контролирует правильность работы программного комплекса. В случае возникновения каких-либо ошибок он сигнализирует об этом «Блоку управления», который принимает решение о прерывании процесса оценивания регрессионной модели, изменения режима работы или игнорирования ошибки.

«Блок формирования исходных данных» содержит процедуры для создания исходных данных в комплексе, либо для экспорта этих данных из текстового файла с расширением .txt. Имеются также процедуры изменения структуры данных, т. е. можно изменять количество наблюдений и число объясняющих переменных.

«Блок реализации программы отсутствия мажорирования» предназначен для проверки произвольного вектора параметров на «паретовость». Обращение к этому блоку может инициироваться «Блоком формирования множества Парето» и «Блоком формирования множества А».

«Блок формирования множества Парето» предназначен для формирования множества Парето посредством перебора узлов сети. В нем содержится раздел графического представления множества Парето в критериальном пространстве. В этом блоке также предусмотрено построение точечной характеризации множества Парето, поэтому он тесно связан с «Блоком реализации программы отсутствия мажорирования».

«Блок формирования множества А» предназначен для формирования множества А посредством решения соответствующих задач линейного программирования. В этом блоке предусмотрена проверка вектора на «паретовость», поэтому он тесно связан с «Блоком реализации программы отсутствия мажорирования».

В «Блоке прогнозирования» на основе сформированного множества А и прогнозных значений объясняющих переменных решаются соответствующие задачи линейного программирования, и определяется нижняя и верхняя границы прогнозного значения.

Информационная схема взаимодействия подсистем ПК МОРМ представлена на рис. 2.3.

В информационной схеме на рис. 2.3 главным блоком является «Блок управления». Пользователь непосредственно осуществляет работу именно с этим блоком. Ее главным «помощником» является «Блок контроля», который следит за правильностью работы других подсистем, и, в случае возникновения каких-либо ошибок, сигнализирует об этом «Блоку управления». С использованием «Блока формирования исходных данных» данные вводятся в систему либо вручную, либо с помощью импорта из текстового файла с расширением .txt.

После ввода данных «Блок контроля» проверяет правильность работы «Блок формирования исходных данных» и передает информацию «Блоку управления». Отметим, что выходные данные о работе программного комплекса также экспортируются в текстовый файл с расширением .txt.

После ввода данных пользователь, в первую очередь, должен использовать «Блок формирования множества Парето», потому что только благодаря полученной в нем информации возможна работа с «Блоком формирования множества А»,

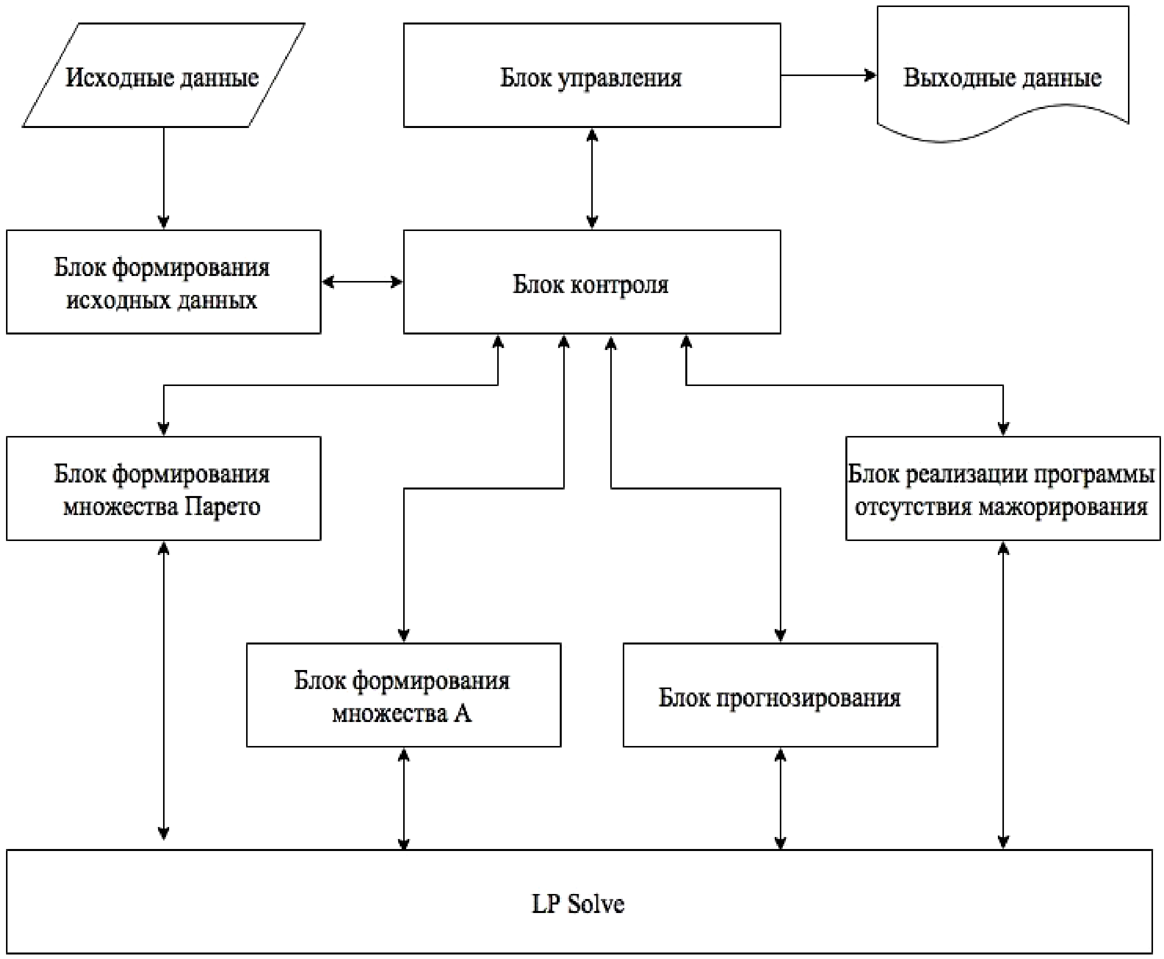


Рисунок 2.3. Информационная схема взаимодействия блоков ПК МОРМ

«Блоком прогнозирования» и «Блоком реализации программы отсутствия мажорирования». При обращении к «Блоку формирования множества Парето» в среде программирования Delphi осуществляется подключение к свободно размещенному в Интернете пакету LPSolve, затем осуществляется автоматическая формализация задачи линейного программирования, т. е. формируется целевая функция и система линейных ограничений задачи, и вся эта информация передается в пакет LPSolve. В пакете LPSolve находится решение сформулированной задачи, которое вновь передается «Блоку формирования множества Парето». Полученная информация расшифровывается, проверяется «Блоком контроля» и, если ошибки не обнаружены, полученное множество Парето для заданных параметров выводится на экран. Отметим, что информационное взаимодействие между пакетом LPSolve и тремя оставшимися блоками («Блоком формирования множества А», «Блоком прогнозирования» и «Блоком реализации программы отсутствия мажорирования») осуществляется аналогичным образом. Отметим также, что программный комплекс способен формировать необходимую выходную информацию даже в том случае, когда задача линейного программирования не имеет решения.

Общий алгоритм программного комплекса представлен на рис. 2.4. При запуске ПК МОРМ сразу осуществляется запуск двух главных блоков: «Блока управления» и «Блока контроля». Все остальные шаги осуществляются под их комплексным управлением. Перед тем, как приступить к математическому моделированию, необходимо ввести исходные статистические данные, запустив «Блок формирования исходных данных». Как видно на рис. 2.4, после работы с этим блоком осуществляется проверка корректности введенных данных. Если статистические данные сформированы правильно, то они сохраняются в системе, иначе пользователю требуется их скорректировать. Как было отмечено выше, после ввода данных, в первую очередь, необходимо воспользоваться «Блоком формирования множества Парето». Если этого не сделать, другие подсистемы ПК МОРМ будут не доступны. Если пользователь принял решение построить множество Парето, то запускается «Блок формирования множества Парето» и «Блок реализации программы отсутствия мажорирования», который предназначен для проверки вектора на «паретовость».

После работы этих блоков сформированное множество Парето сохраняется в системе и осуществляется переход на следующий этап.

Следующим этапом является либо построение множества А, либо прогнозирование. Если пользователь принял решение построить множество А, то происходит запуск «Блока формирования множества А», определяется множество А и результаты сохраняются в ПК МОРМ. При этом при желании можно проверить вектор на «паретовость». В этом случае запускается «Блок реализации программы отсутствия мажорирования», происходит проверка и результаты выводятся на экран. Если пользователь принял решение получить прогнозы, то запускается «Блок прогнозирования», благодаря решению соответствующих задач линейного программирования, находятся прогнозные значения и сохраняются в системе. Отметим, что на любом этапе пользователь может вернуться к любому предыдущему шагу алгоритма, либо завершить работу с программным комплексом.

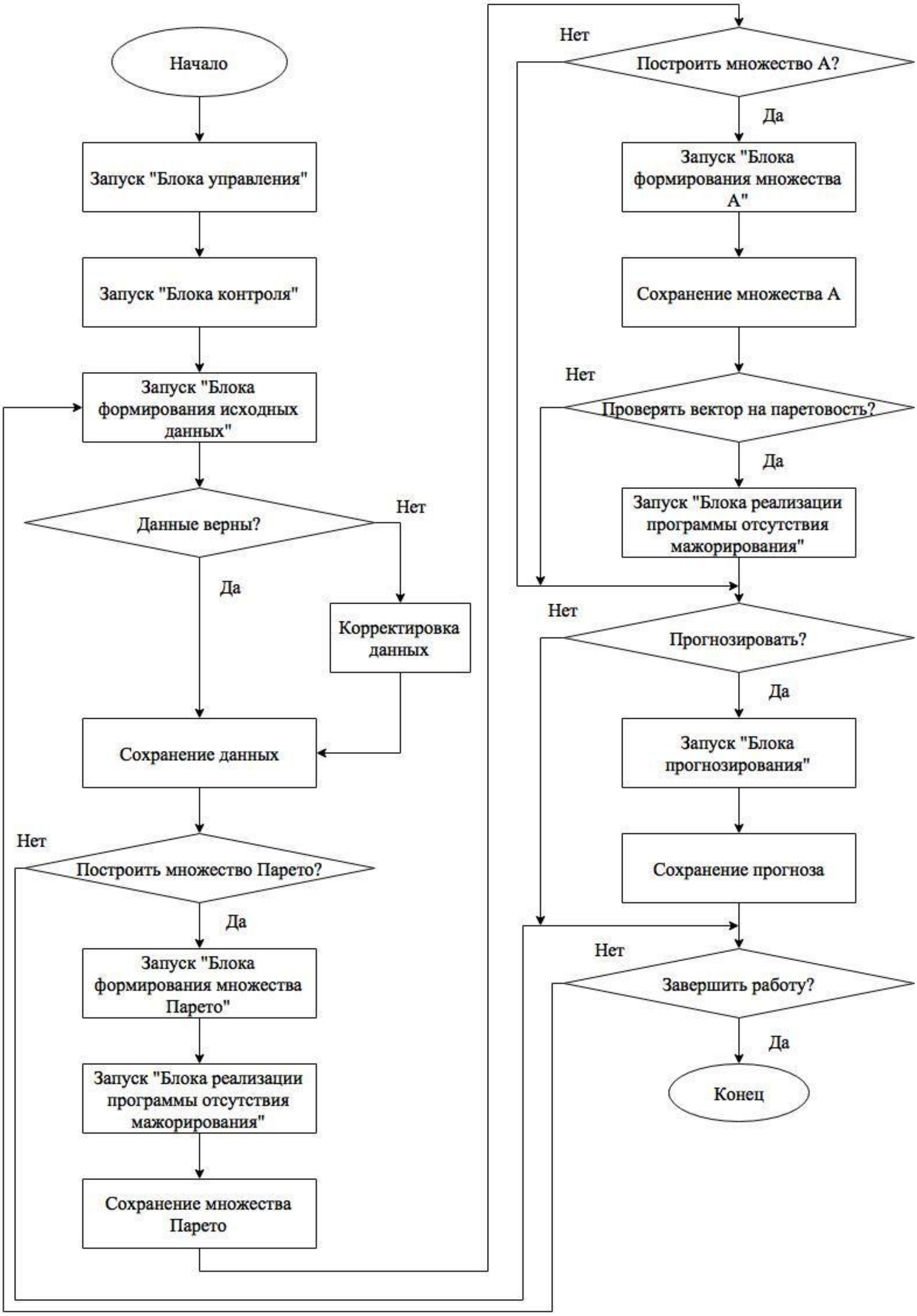


Рисунок 2.4. Общий алгоритм работы ПК МОРМ

2.4 Алгоритмы работы подсистем ПК МОРМ

Рассмотрим подробнее алгоритмы работы подсистем ПК МОРМ. Алгоритм работы «Блока формирования множества Парето» представлен на рис. 2.5. Для построения сети первоначально пользователь должен задать параметр $iter$ – количество ее узлов, т. е. количество точек, на которые будет разбит интервал $(0,1)$. По умолчанию параметр lam – текущая точка из интервала $(0,1)$, равен 0. Параметр $shag$ – шаг сети находится по формуле:

$$shag = \frac{1}{iter}. \tag{2.22}$$

После задания количества узлов сети начинается основной итерационный цикл «Блока формирования множества Парето». В цикле для каждой точки сети формируется соответствующая задача линейного программирования с целевой функцией:

$$lam \cdot J_1 + (1 - lam) \cdot J_\infty, \tag{2.23}$$

где J_1 – функция потерь для МНМ, а J_∞ – для МАО.

Задачи линейного программирования в Delphi формируются в специально организованной матрице, в которой по столбцам расположены переменные задачи, а по строкам – их ограничения. Затем на каждой итерации и для каждой точки сети в LPSolve решается задача линейного программирования, и все промежуточные результаты сохраняются в матрице *H*. Матрица *H* представляет собой матрицу, в которой каждая строка соответствует заданной точке сети, а по столбцам расположены найденные значения неизвестных переменных задачи линейного программирования. После чего по матрице *H* происходит формирование множества Парето и строится график.

Алгоритм работы «Блока прогнозирования» изображен на рис. 2.6. Данный алгоритм представляет собой двойной вложенный цикл. Внешний цикл по переменной i осуществляется с изменением от 1 до $kolPareto - 1$, где $kolPareto$ –

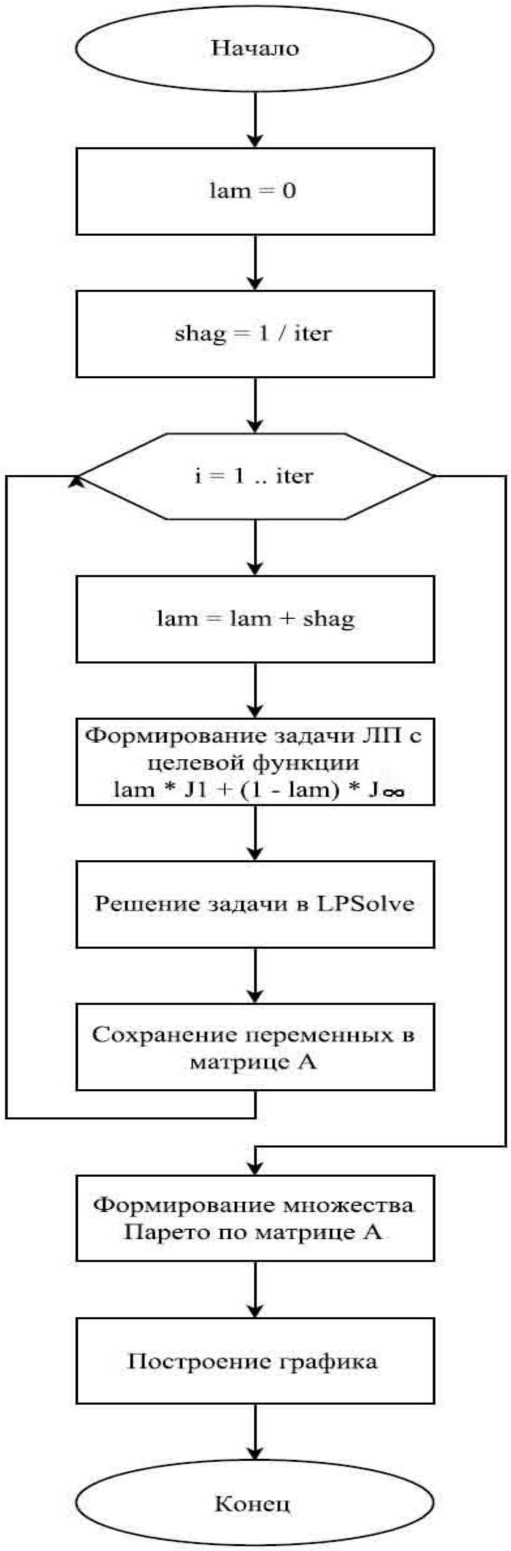


Рисунок 2.5. Алгоритм работы «Блока формирования множества Парето»

количество паретовских граней. Для каждой такой грани автоматически формируется матрица системы ограничений задачи линейного программирования. Затем включается внутренний цикл. Его главная задача заключается в том, чтобы сформировать две целевых функции: на максимум или на минимум. При этом тип задачи будет зависеть от параметра цикла k. Так, если $k=1$, то формируется целевая функция на минимум, а если $k=2$, то на максимум. После чего задача передается пакету LPSolve, который осуществляет поиск решения, и результаты сохраняются в матрице *H*. Структура матрицы *H* для случая модели парной регрессии представлена в таблице 2.1.

Таблица 2.1

Структура матрицы *H* в «Системе прогнозирования»

Ребро	Экстремум	Значение
1	мин	…
1	макс	…
2	мин	…
2	макс	…
3	мин	…
3	макс	…
4	мин	…
4	макс	…
…	…	…

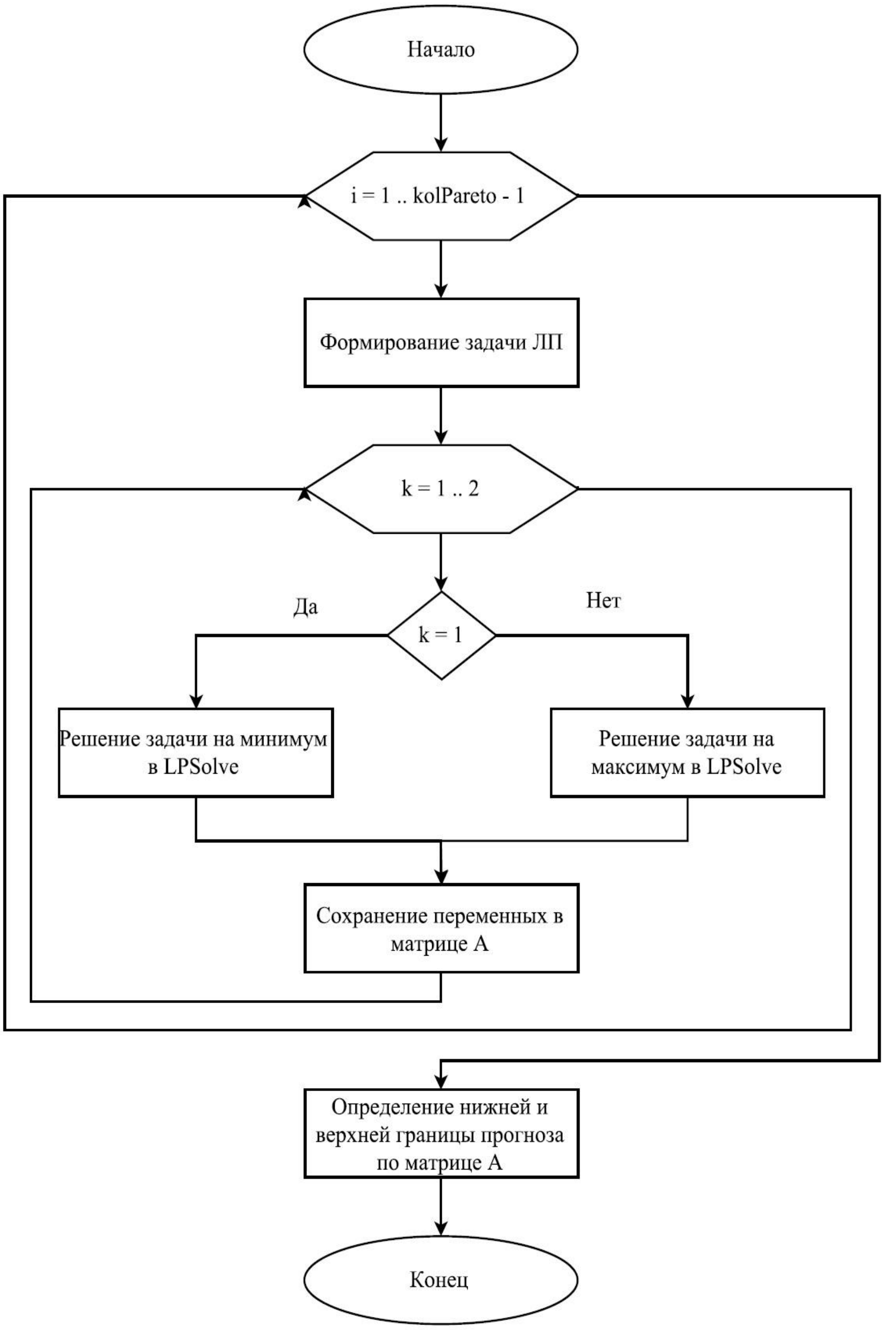


Рисунок 2.6. Алгоритм «Блока прогнозирования»

Общее число задач линейного программирования, решаемых в «Блоке прогнозирования», составляет $(kolPareto-1)\cdot k$. После того, как матрица H сформирована, по ней осуществляется определение верхней и нижней границы прогноза.

Алгоритм работы «Блока формирования множества H» проиллюстрирован на рис. 2.7.

Данный алгоритм представляет собой тройной вложенный цикл. Внешний цикл по переменной i осуществляется с изменением от 1 до $kolPareto-1$, где

$kolPareto$ – количество паретовских граней. Затем первый внутренний цикл по переменной j осуществляется с изменением от 1 до $perem$, где $perem$ – количество неизвестных параметров регрессионной модели. Для каждой такой грани и переменной автоматически формируется матрица системы ограничений задачи линейного программирования. Затем включается второй внутренний цикл. Его главная задача также заключается в том, чтобы сформировать две разных целевых функции: на максимум или на минимум. Если параметр цикла $k=1$, то формируется целевая функция на минимум, а если $k=2$, то на максимум. После чего, задача передается пакету LPSolve, который осуществляет поиск решения, и результаты сохраняются в матрице *H*. Структура матрицы *H* для случая модели парной регрессии представлена в таблице 2.2.

Таблица 2.2

Структура матрицы *H* в «Блоке формирования множества *H*»

Ребро	Параметр	Экстремум	Значение
1	a0	мин	…
1	a0	макс	…
1	a1	мин	…
1	a1	макс	…
2	a0	мин	…
2	a0	макс	…
2	a1	мин	…
2	a1	макс	…
…	…	…	…

Общее число задач линейного программирования, решаемых в «Системе формирования множества *H*», составляет $(kolPareto-1)\cdot perem\cdot k$. После того, как матрица *H* получена, по ней осуществляется формирование множества H.

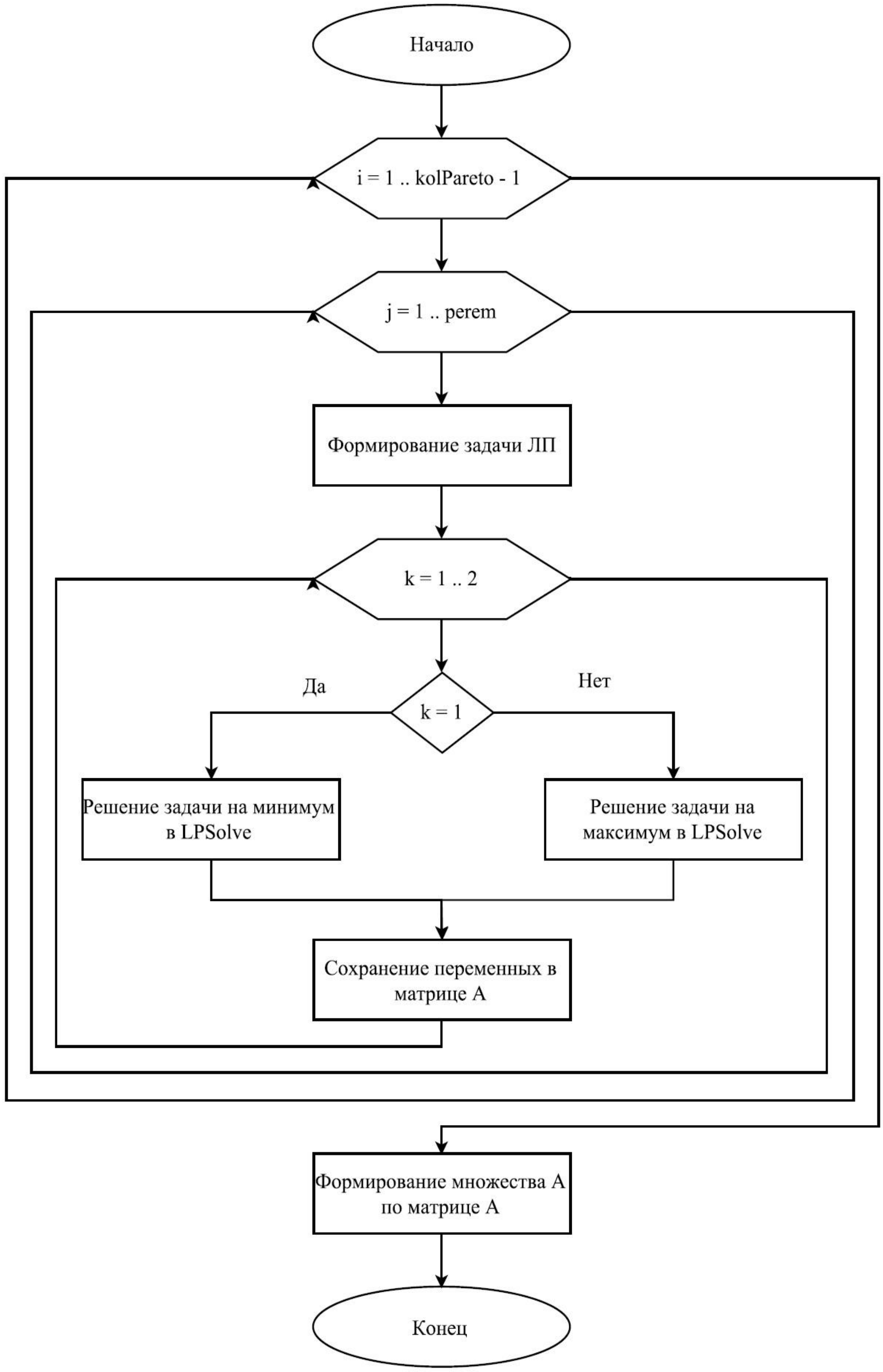


Рисунок 2.7. Алгоритм работы «Блока формирования множества *H*»

2.5 Особенности работы в ПК МОРМ

Рассмотрим особенности работы с ПК МОРМ на примере множественного оценивания линейной модели парной регрессии по статистическим данным, представленным в таблице 2.3.

Таблица 2.3

Пример статистических данных

y	5	3	9	8
x	2	4	7	5

Главное окно ПК МОРМ содержит четыре перекрывающих друг друга страницы (панели): «Исходные данные», «Основная задача», «Построение множества А», «Прогнозирование». После ввода данных из таблицы 2.3 в программный комплекс панель «Исходные данные» принимает вид, показанный на рис. 2.8.

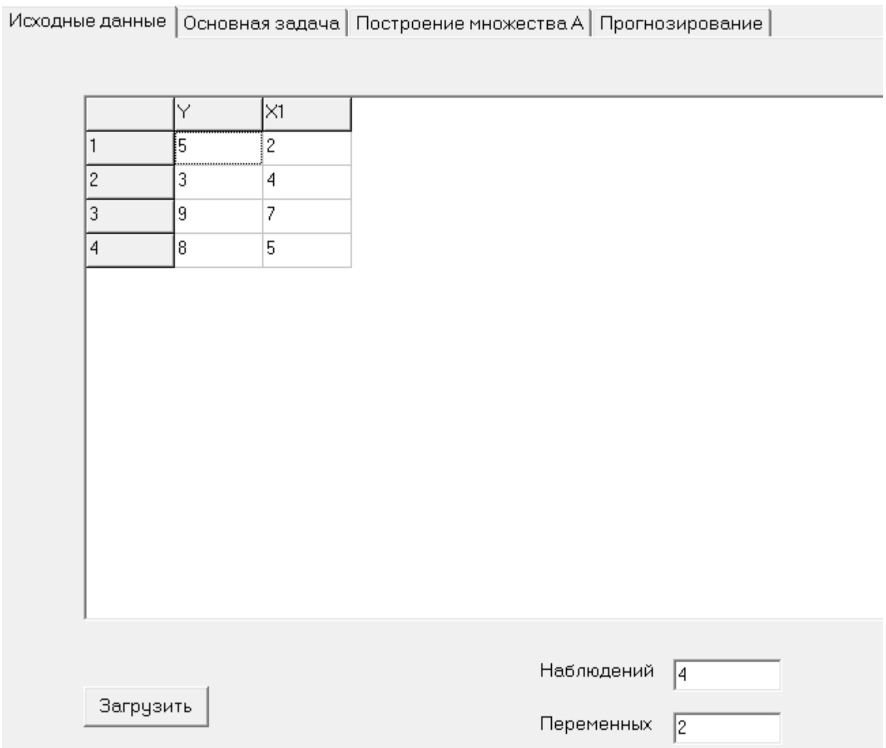


Рисунок 2.8. Панель «Исходные данные»

Кнопка «Загрузить» на панели «Исходные данные» предназначена для загрузки статистических данных из текстового файла с расширением .txt. При этом в текстовом файле не допускается именование переменных, первый столбец обязательно должен соответствовать значениям объясняемой переменной, столбы должны быть разделены клавишей Tab, а разделитель между целой и дробной частями вещественных чисел – «,». При загрузке данных переменным автоматически присваивается имя: зависимой переменной – «Y», независимым переменным «X1», «X2» и т. д. При желании пользователь может изменить любое значение, дважды нажав по нему левой клавишей мыши. Для изменения числа наблюдений или количества переменных в одноименных полях «Наблюдений» или «Переменных» необходимо ввести с клавиатуры требуемое значение.

Если данные введены правильно, то для построения множества Парето необходимо активировать панель «Основная задача», ввести число узлов сети, которое по умолчанию равно 100 и нажать кнопку «Решить». Результаты этих шагов представлены на рис. 2.9.

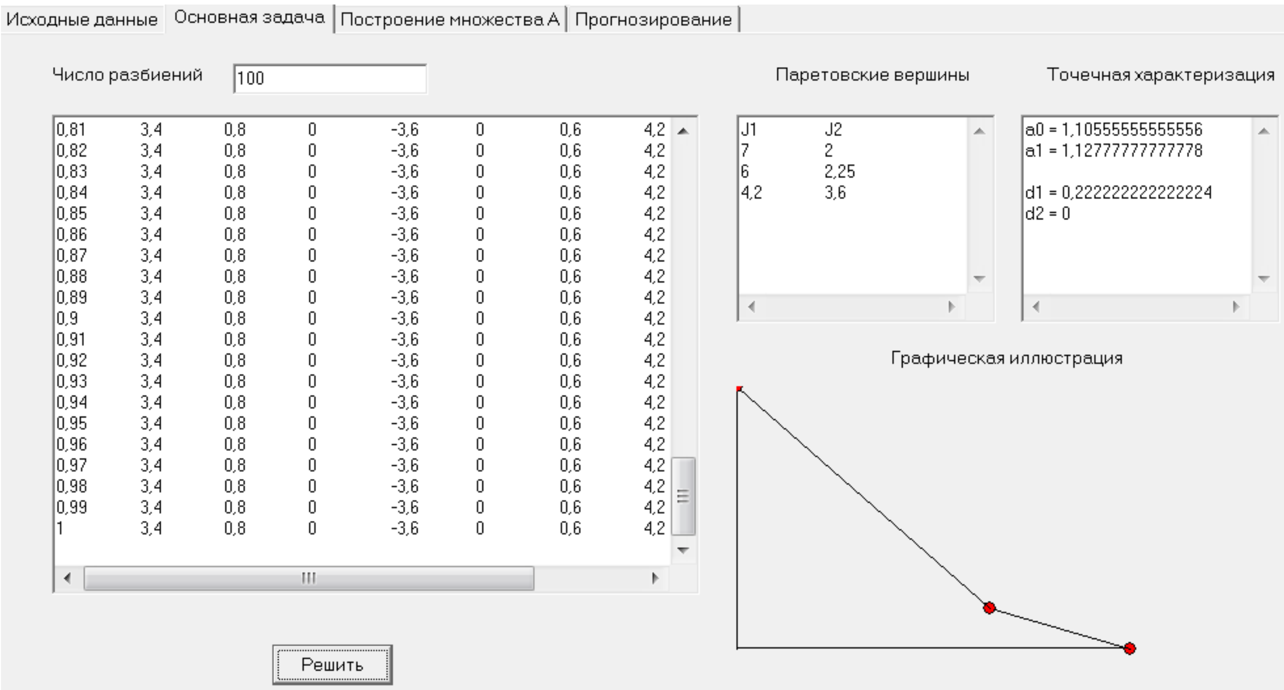


Рисунок 2.9. Панель «Основная задача»

Панель «Основная задача» условно можно разделить на 4 поля.

1. Поле «Подробные результаты», расположенное слева. В каждой строке этого поля содержатся результаты решения задачи ЛП для каждого числа λ: оценки параметров «a0», «a1», ошибки «e1», «e2», «e3», «e4», значения критериев «J1», «$J\infty$».

2. Поле «Паретовские вершины», содержащее информацию о количестве и координатах паретовских вершин.

3. Поле «Точечная характеризация», в котором содержатся оценки коэффициентов и значения вспомогательных переменных «d1» и «d2».

4. Поле «Графическая иллюстрация», содержащее представление множества Парето в критериальном пространстве.

Для определения точечной характеризации в ПК МОРМ автоматически была сформирована и решена следующая задача ЛП:

$\max : d_1 + d_\infty$;

$a_1 - a_2 + 2a_3 - 2a_4 + u_1 - v_1 = 5$,

$a_1 - a_2 + 4a_3 - 4a_4 + u_2 - v_2 = 3$,

$a_1 - a_2 + 7a_3 - 7a_4 + u_3 - v_3 = 9$,

$a_1 - a_2 + 5a_3 - 5a_4 + u_4 - v_4 = 8$,

$u_1 + v_1 - r \le 0$,

$u_2 + v_2 - r \le 0$,

$u_3 + v_3 - r \le 0$,

$u_4 + v_4 - r \le 0$,

$u_1 + v_1 + u_2 + v_2 + u_3 + v_3 + u_4 - v_4 + d_1 \le 5{,}733333$,

$r + d_2 \le 2{,}616667$.

При активации панели «Построение множества А» происходит автоматический запуск «Системы формирования множества А» и все результаты заносятся в соответствующие текстовые поля. Панель «Построение множества А» для нашего случая представлена на рис. 2.10.

Панель «Построение множества А» разделяется на 4 поля.

1. Поле «Подробные результаты», содержащее информацию о минимальных и максимальных значениях оценок параметров модели для каждого ребра.

2. Поле «Множество А» содержит информацию о том, в каких пределах изменяется оценка для каждого параметра.

3. Поле «Проверка вектора на паретовость» содержит таблицу, в которую необходимо внести координаты вектора, подлежащего проверке.

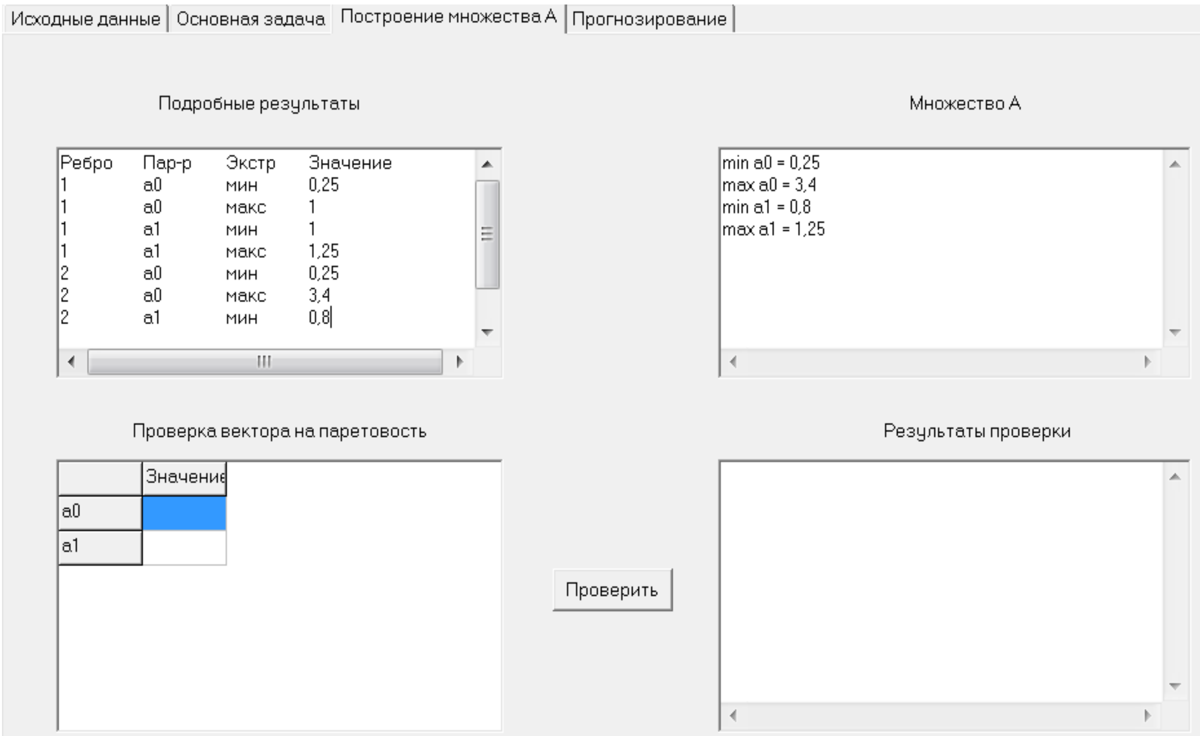


Рисунок 2.10. Панель «Построение множества А»

4. Поле «Результаты проверки» содержит значения вспомогательных переменных «d1» и «d2», а также вывод: либо «Вектор является паретовским», либо «Вектор не является паретовским».

В нашем случае для построения множества А было решено 8 задач ЛП (2 ребра, 2 переменных на максимум и 2 ребра, 2 переменных на минимум). Для нижней оценки свободного члена для первого ребра задача ЛП имеет вид:

$$\min : a_1 + a_2,$$

$$a_1 - a_2 + 2a_3 - 2a_4 + u_1 - v_1 = 5,$$

$$a_1 - a_2 + 4a_3 - 4a_4 + u_2 - v_2 = 3,$$

$$a_1 - a_2 + 7a_3 - 7a_4 + u_3 - v_3 = 9,$$

$$a_1 - a_2 + 5a_3 - 5a_4 + u_4 - v_4 = 8,$$

$$u_1 + v_1 - r \le 0,$$

$$u_2 + v_2 - r \le 0,$$

$u_3 + v_3 - r \le 0$,

$u_4 + v_4 - r \le 0$,

$u_1 + v_1 + u_2 + v_2 + u_3 + v_3 + u_4 + v_4 = 7\lambda + 6 - 6\lambda$,

$r = 2\lambda + 2{,}25 - 2{,}25\lambda$,

$\lambda \le 1$.

Для проверки на паретовость возьмем, например, вектор $\tilde{\alpha} = (2;1)$. Тогда задача ЛП программы отсутствия мажорирования имеет вид:

$\max : d_1 + d_\infty$,

$a_1 - a_2 + 2a_3 - 2a_4 + u_1 - v_1 = 5$,

$a_1 - a_2 + 4a_3 - 4a_4 + u_2 - v_2 = 3$,

$a_1 - a_2 + 7a_3 - 7a_4 + u_3 - v_3 = 9$,

$a_1 - a_2 + 5a_3 - 5a_4 + u_4 - v_4 = 8$,

$u_1 + v_1 - r \le 0$,

$u_2 + v_2 - r \le 0$,

$u_3 + v_3 - r \le 0$,

$u_4 + v_4 - r \le 0$,

$u_1 + v_1 + u_2 + v_2 + u_3 + v_3 + u_4 - v_4 + d_1 \le 5$,

$r + d_2 \le 3$.

Целевая функция этой задачи равна 0, следовательно, вектор $\tilde{\alpha} = (2;1)$ является паретовским.

Панель «Прогнозирование» содержит 2 поля: «Прогнозные значения», в которое вводятся прогнозные значения объясняющих переменных, и «Результаты», в котором после нажатия кнопки «Прогнозировать» появится информация о нижней и верхней границах прогноза по каждому ребру и итоговый интервальный прогноз.

Панель «Прогнозирование» с прогнозом для $\tilde{x} = 8$ представлена на рис. 2.11.

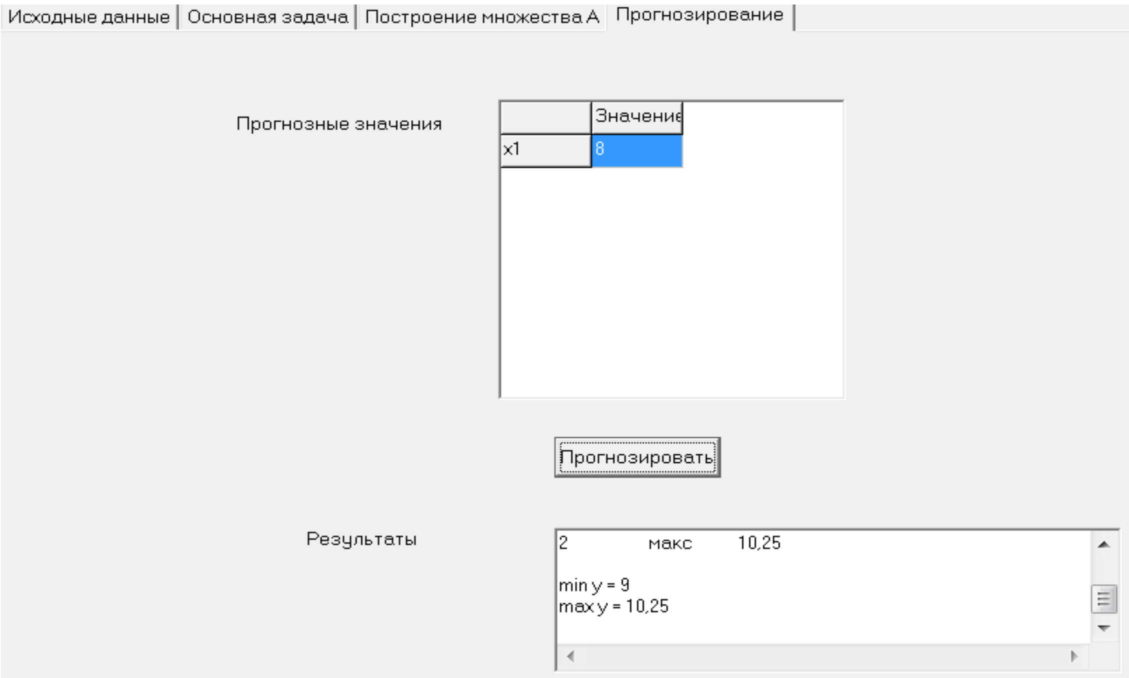


Рисунок 2.11. Панель «Прогнозирование»

Для построения прогноза решалось 4 задачи ЛП (2 ребра на максимум и 2 ребра на минимум). Задача ЛП для нижней границы первого ребра имеет вид:

$$\min : a_1 - a_2 + 8a_3 - 8a_4,$$

$$a_1 - a_2 + 2a_3 - 2a_4 + u_1 - v_1 = 5,$$

$$a_1 - a_2 + 4a_3 - 4a_4 + u_2 - v_2 = 3,$$

$$a_1 - a_2 + 7a_3 - 7a_4 + u_3 - v_3 = 9,$$

$$a_1 - a_2 + 5a_3 - 5a_4 + u_4 - v_4 = 8,$$

$$u_1 + v_1 - r \le 0,$$

$$u_2 + v_2 - r \le 0,$$

$$u_3 + v_3 - r \le 0,$$

$$u_4 + v_4 - r \le 0,$$

$$u_1 + v_1 + u_2 + v_2 + u_3 + v_3 + u_4 + v_4 = 7\lambda + 6 - 6\lambda,$$

$$r = 2\lambda + 2{,}25 - 2{,}25\lambda,$$

$$\lambda \le 1.$$

2.6 Выводы

В данной главе рассмотрено математическое и программное обеспечение множественного оценивания регрессионных моделей (ПК МОРМ). Описанная методика множественного оценивания предполагает решение достаточно большого числа громоздких задач линейного программирования. Поэтому было принято решение интерфейс программного комплекса разработать в среде программирования Delphi, а для решения задач линейного программирования использовать бесплатный пакет LPSolve. При этом технически не сложно организовать взаимодействие между Delphi и LPSolve.

Согласно сформулированным требованиям, разработана архитектура, информационная схема и общий алгоритм работы ПК МОРМ. Приведены алгоритмы работы основных подсистем комплекса, рассмотрены особенности работы в ПК МОРМ.

Перечислим основные возможности разработанного ПК МОРМ.

1. Создание исходных статистических данных вручную или путем их импорта из текстового файла. При этом в первом случае количество наблюдений и переменных задается вручную, а во втором – определяется автоматически.

2. Формирование множества Парето оценок параметров моделей линейной регрессии.

3. Определение точечной характеризации множества Парето, основанной на программе отсутствия мажорирования.

4. Построение множества Парето в критериальном пространстве, представляющим собой объединение ребер образа многогранника.

5. Формирование множества A возможных оценок параметров регрессии.

6. Проверка произвольно заданного вектора на «паретовость».

7. Получение интервальных прогнозов на основе построенной модели регрессии.

8. Формирование полной информации о результатах моделирования. Если система линейных ограничений окажется несовместной, то система выводит соответствующее сообщение. Если система совместна, то организуется вывод на экран следующей информации: оценки параметров модели, ошибки аппроксимации и т. д.

3 МОДЕЛИРОВАНИЕ ДИНАМИКИ ВАЛОВОГО РЕГИОНАЛЬНОГО ПРОДУКТА ИРКУТСКОЙ ОБЛАСТИ

3.1 Прогнозирование валового регионального продукта

Непосредственным результатом воздействия производительных сил на экономический потенциал в масштабе страны является валовой внутренний продукт (ВВП), а в масштабе региона – валовой региональный продукт (ВРП). Он является важнейшим показателем регионального развития, который обобщает результат всей экономической деятельности региона [11].

Являясь фактором экономического роста, ВРП позволяет дать полную структурную характеристику своих составляющих элементов, а также характеризует экономическое развитие общества в целом и перспективы его развития [102].

В России, с учетом ее площади и многоукладности экономики, проблема оценки уровня и динамики развития экономик регионов, а также проведения межрегиональных сопоставлений, стоит особенно остро [47].

В системе национальных счетов ВРП определяется следующим образом: «ВРП – это сумма вновь созданных стоимостей продукции и услуг, произведенных в отдельном регионе» и равен сумме валовой добавленной стоимости по отраслям и чистым налогам на продукты» [7].

Из-за методической сложности часто используют упрощенные методики для расчета данного показателя, которые снижают точность и аналитические возможности, что, в свою очередь, негативно влияет на качество управленческих решений, принимаемых на основе этих показателей.

Существует ряд проблем при расчете составляющих ВРП:

- Результаты видов деятельности, направленных на осуществление общенациональных задач (оборона, денежная система и т.д.);
- Результаты функционирования тех предприятий, деятельность кото-

рых выходит за пределы региона;

- Теневой сектор экономики;
- Различия в уровне цен на товары и услуги.
- Учет налогов предприятий, чья деятельность выходит за пределы региона;
- Распределение добавленной стоимости между регионами [7];
- Миграция (легальная и нелегальная) и ее последствия.

Однако, несмотря на некоторые успехи в этой сфере, в последние годы проблема учета перечисленных явлений продолжает оставаться актуальным [47].

Одной из главных задач региональной экономической политики последних лет является развитие конкурентно-способной инфраструктуры. Однако основывать управленческие решения на основе показателя ВРП по-прежнему остается достаточно сложно. В ВРП не включается часть результатов экономической деятельности, т. е. усилия по формированию благоприятного климата для развития предпринимательства, и инвестирования отражаются в ключевом макроэкономическом показателе не полностью. В этой ситуации любой разумный, хотя бы приближенный метод распределения лучше, чем его полное отсутствие.

При наличии стольких недостатков альтернативных показателей, которые могли бы заменить ВРП, в настоящее время нет. Поэтому объемы и темпы роста ВРП являются целевыми показателями при разработке стратегий, программ, прогнозов и формировании социально-экономической политики регионов.

В современной литературе проблемам прогнозирования макроэкономических показателей уделяется значительное внимание. В отечественной литературе проблемы макроэкономического моделирования рассматривались в трудах Гринберга Л., Канторовича Л., Макарова В., Яременко Ю, Андреева А. В., Зайцевой. Ю. С. Среди зарубежных работ по прикладным вопросам прогнозирования в экономике представлены исследования Дж. Бокса и Г Дженкивса, Ф Робертса, Х. Таха, Дж. Хапка и др. В частности, в работах Дж. Стока и М. Ватсона приводятся ограничения, определяющие достоверность расчетных моделей для малых выборок [56].

Методика прогноза ВРП основана на ежегодном сборе и анализе значительного объема первичной статистической информации в региональном разрезе. Это крайне сложная методическая, аналитическая и техническая задача, решаемая с разной степенью успешности по отдельным компонентам.

Но есть еще несовершенства в плане учета ВРП, а, значит, и составления прогноза.

Определение прогноза можно сформулировать следующим образом: «Прогноз – это научно обоснованное суждение о возможном состоянии объекта в будущем» [7].

Но непосредственно сам процесс построения прогнозов является весьма сложным. Выделяются несколько видов прогноза по периоду: краткосрочный (один год); среднесрочный (от 3 до 5 лет); долгосрочный (более 5 лет), а также три вида по типу: пессимистичный, умеренный и оптимистичный [7].

Прогностика взаимосвязана с такими дисциплинами, как экономическая теория, макроэкономика, государственное регулирование экономики, статистика, теория вероятностей, математическая статистика, эконометрика и т. д.

Процесс прогнозирования базируется сегодня на Федеральном законе от 28 июня 2014 г. № 172-ФЗ «О стратегическом планировании в Российской Федерации» (с изменениями и дополнениями)[91].

Кроме того, особое внимание прогнозу уделяется и в таком важном документе, как Бюджетный кодекс РФ. Так, в ст. 184 БК РФ устанавливается, что первым этапом формирования бюджета является разработка прогноза, а также выбор варианта развития экономики в тот или иной период [90].

Так, начиная с мая каждого года, Минэкономразвития РФ передает сценарные условия по бюджету субъектам Федерации. В течение года идет постоянное уточнение цифр. Ответственность при составлении прогноза огромна. В случае неточности показателей страдают те сферы экономики, которые напрямую зависят от бюджета.

Существуют два основных метода прогноза уровня ВРП – производственный и распределительный. Минэкономразвития РФ при прогнозировании ВРП рекомендует пользоваться только производственным методом. Распределительный метод, к сожалению, пока не нашел должного отражения в системе национальных счетов (СНС) и поэтому сложен для использования [7].

Справедлива следующая словесная формула: Валовой выпуск – Промежуточное потребление = валовая добавленная стоимость + чистые налоги.

Чтобы полностью спрогнозировать показатель выпуска, необходимо детализировать данные по отдельным составляющим, т. е. прогнозировать все показатели по базовым отраслям экономики. В СНС учитывается пять таких базовых отраслей: промышленность, сельское хозяйство, строительство, транспорт и связь, торговля и общественное питание.

Существуют различные способы обработки информации: метод дефлятирования, метод экстраполяции, метод моделирования, экономико-математический метод, метод экспертных оценок и др.

На практике используют, как правило, все методы одновременно. Так, метод дефлятирования применяется в тех случаях, когда точно известно, что физический объем по данному продукту не увеличится в следующем году. Метод экстраполяции применяют тогда, когда изменяется физический объем производства.

Расчет валовой добавленной стоимости во многом несовершенен, поскольку должен базироваться на показателе ВРП за предыдущий год. Однако эти данные регионы получают только через год, поскольку раньше органы статистики не готовы предоставить информацию. Т. е. прогноз показателя ВДС (валовой добавленной стоимости), например, за 2016 год, составляется на базе 2014 года. В процессе итоговых расчетов изменяются не только стоимостные показатели, но и индексы физического объема производства.

Если есть несовершенство эмпирического анализа, т. е. фактических данных, то эта ошибка автоматически транспортируется в прогнозные данные.

Что касается распределительного метода, то, несмотря на то, что он не рекомендован Минэкономразвития, на практике применяется довольно часто. Фактически это метод расчета по источникам дохода. Он основан на анализе долей, составляющих ВРП в динамике (фонд оплаты труда, прибыли, косвенных налогов, амортизации и т. д.). Главная причина нежелательности использования данного метода состоит в том, что последнюю статью ВРП «валовая прибыль и прочие смешанные доходы» очень важно просчитывать. Здесь отражаются теневые доходы, доходы от собственности, рентные платежи, дивиденды и т. д. Но поскольку пока не существует отработанной методики учета подобных доходов, то и вероятность получения достоверных данных весьма низка.

Прежде чем приступить к прогнозу, анализируются все факторы, влияющие на экономику региона и России в целом с учетом того, что регион есть часть национальной экономики [7].

Таблица 3.1

Учитываемые факторы при построении прогноза

Внешние	Внутренние	
	объективные	субъективные
Темп роста мировой экономики Темпы развития ведущих стран мира Динамика мировых цен на нефть Спрос и объемы экспорта российской нефти Интеграция России в ВТО Выплаты по внешнему долгу	Дифференциация доходов Состав производственных фондов и степень их изношенности Уровень капиталоемкости Степень развития инфраструктуры Демографическая ситуация в регионе	Темпы инфляции Тарифная политика Налоговая политика

Необходимо учитывать региональные риски: роль и доля наиболее значимых секторов экономики, динамика цен на ресурсы, добываемые на территории данного региона, прогноз добычи данных ресурсов, изменение демографических показателей, инфляционный рост, степень развитости инфраструктуры.

Основные потребители прогнозов – органы исполнительной власти, которые составляют годовые бюджеты разных уровней. На федеральном уровне прогноз социально-экономического развития на год и параметры развития на три года разрабатывает Министерство экономики и развития РФ, а проект федерального бюджета на будущий год и прогноз основных характеристик доходов и расходов федерального бюджета на три последующих года – Министерство финансов РФ. Работа по краткосрочному прогнозированию возложена на Центр экономической конъюнктуры при правительстве РФ, который осуществляет мониторинг состояния экономики и прогнозирует макроэкономическую ситуацию в стране. В регионе распределение прогнозных функций аналогично федеральным: Министерство экономики составляет прогноз социально-экономического развития, а Министерство финансов на его основе составляет бюджет на следующий год.

Прогнозирование ВРП, как правило, опирается на временные ряды, и выбор моделей осуществляется исходя из этого. Также для оценки динамики наиболее часто используются регрессионные модели, с помощью которых оцениваются динамика и факторы развития экономики. Такие модели дают возможность объединения всех накопленных знаний об основных пропорциях и законах развития экономики [7].

Закономерности развития экономических процессов могут быть выражены как одним регрессионным уравнением, так и группой. Среди большеразмерных моделей (содержащих два и более уравнения) различают простые системы и системы одновременных регрессионных уравнений (СОРУ). В простых системах уравнения не связаны между собой (рекурсивны) и используются, как правило, для описания взаимозависимости региональных переменных и их влияния на ВРП. Системы одновременных регрессионных уравнений применяются, как правило, в развернутых моделях. СОРУ могут состоять из тождеств, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы.

В отличие от простых систем, где возможен расчет параметров для каждого уравнения в отдельности, используя метод наименьших квадратов (МНК), параметры модели и эндогенные переменные в СОРУ определяют путем решения всей системы, что требует более сложных методов оценки параметров, таких как косвенный МНК или двухшаговый МНК.

Регрессионные модели имеют ряд недостатков. Во-первых, они применимы только для стационарных процессов, когда в системе не происходит качественных изменений, способных изменить коэффициенты уравнения, во-вторых, при построении эконометрических моделей, особенно систем регрессионных моделей, часто возникают проблемы информационного плана. Это связано, прежде всего, с необходимостью наличия представительных рядов динамики, а также с отсутствием таких рядов для некоторых показателей из-за произошедших структурных изменений в экономике и в системе государственной статистики.

Другой метод, применимый для прогнозирования ВРП – так называемая простая экстраполяция или продолжение имеющихся в настоящее время трендов. На практике простая экстраполяция сводится к аналитическому выравниванию, то есть к процедуре подбора наиболее подходящей функции. Помимо недостатков, присущих эконометрическим моделям, у простой экстраполяции есть свой недостаток – полученные таким методом прогнозы «хороши» в те периоды времени, когда изменения экономической ситуации носят инерционный характер, но в иных случаях совершенно бесполезны.

Подход более сложный, чем прогноз по тренду, представляет компонентный анализ, рассматривающий временной ряд прогнозируемого показателя, состоящего из нескольких компонент. При компонентном анализе данные, представленные в виде временных рядов, в общем случае, раскладываются на четыре компоненты: тренд, строго периодическая, нестрого периодическая и случайная компоненты. Полученные составные части анализируются и прогнозируются в отдельности. Прогноз исходного временного ряда получается суммированием прогнозов всех компонент.

Разбиение на другого рода компоненты исследуемого временного ряда предполагает модель, в которой стационарный временной ряд рассматривается как сочетание авторегрессионного процесса с моделью скользящего среднего. Стационарность ряда достигается учетом разностей порядка, заданного исследователем. Авторегрессионный процесс представляет собой линейную зависимость значения временного ряда от предыстории, то есть реализуется гипотеза о влиянии поведения процесса в прошлом на будущие его изменения [7].

Основной недостаток эконометрических и экстраполяционных моделей, а именно: неприменимость к неоднородным рядам, отсутствует в адаптивных моделях. Малый объем ретроспективной информации не позволяет применять классические методы прогнозирования временных рядов, так как они обладают невысокими точностью и достоверностью прогноза. Таким образом, возникает потребность в разработке новых методов прогнозирования, способных улучшить качество прогнозов и достоверно определять количественную оценку ВРП.

В основу адаптивных моделей положено предположение о том, что будущее в той или иной степени будет похоже на недавнее прошлое, чем на более ранние периоды. Реализуется это с помощью адаптируемых параметров, которые на основе рекуррентных формул корректируются, исходя из разницы фактических и прогнозных значений на последнем предпрогнозном этапе.

Ограниченность использования их в среднесрочном прогнозировании и невозможность использования в долгосрочном объясняется тем, что модель не рассматривает причинно-следственные связи, определяющие этот процесс (как в регрессионном анализе), и только частично описывает структуру процесса.

Лучший эффект дает комбинированное использование нескольких моделей для прогнозирования.

Наиболее развернуто и детально отображают сложившиеся пропорции воспроизводства и тенденции их изменения динамические и статические межотраслевые балансы (МОБ). Давая целостное представление о структуре и взаимосвязях отраслей регионального хозяйственного комплекса, МОБ может являться ядром системы моделей регионального прогнозирования и связывать показатели макро- и микроуровней региональной экономики.

Последнее время все большее внимание уделяется долгосрочным прогнозам (с горизонтом свыше 5 лет) в связи с необходимостью перехода к новым технологическим укладам и усилению действия инновационных факторов [51, 90]. При этом совершенно очевидно, что точно и однозначно предусмотреть характер изменения и степень влияния всех факторов в их взаимозависимости практически невозможно. Поэтому, чтобы сделать выводы о необходимости корректировки прогнозов, следует остановиться на методических подходах к их разработке, характерных для последнего периода, на национальном и региональных уровнях.

Возрастает актуальность повышения качества прогнозных исследований. Это требует углубленного изучения и разработки основных проблем, возникающих при прогнозировании. В странах с развитой рыночной моделью экономики прогнозирование и планирование являются важнейшим инструментом государственного регулирования. Нацелено применяя такой инструмент, можно добиться большого успеха в техническом прогрессе, повышении уровня жизни населения и других социально-экономических областях.

Недостаточное внимание к региональным проблемам не только отрицательно сказывается на эффективности экономики и уровне жизни населения, но и может при вести к усилению центробежных тенденции в государстве.

3.2 Анализ функциональной зависимости между переменными при моделировании ВРП Иркутской области

Важным макроэкономическим показателем развития любой области и Иркутской, в частности, является валовой региональный продукт. Поэтому весьма актуальной задачей является его моделирование. В работе предпринята попытка решения этой задачи с помощью регрессионного анализа. Зависимой переменной является валовой региональный продукт Иркутской области. К независимым факторам относятся: потребление электроэнергии, численность безработных, объем строительства жилых домов и оборот розничной торговли.

1. Одна из проблем модельного описания ВРП региона – так называемая теневая часть экономики. В настоящее время не существует стандартных, универсальных подходов, которые могли бы обеспечить успех подобного расчета в любой экономике, но за последние годы происходит заметное развитие соответствующей методологии. Большинство методов можно разделить на прямые и косвенные. Прямые методы позволяют получить детальную и высококачественную информацию. Они базируются на различного рода обследованиях (домашних хозяйств, рабочей силы и т.д.) и результатах налогового аудита. Множество косвен-

ных методов основывается на расхождениях в официальных статистических данных между производством и потреблением, т.е. при определении реального размера ВВП базируется на анализе потребления отдельных производственных ресурсов. В последние годы после выхода целого ряда статей специалистов Мирового банка и Международного валютного фонда (МВФ) чаще всего исследуется динамика энергопотребления. Преимущества данного метода заключаются в простоте измерения потребления данного ресурса, стабильности соотношения между объемом потребляемого ресурса и объемом производства, в уверенности, что теневой сектор не может утаить потребление электроэнергии [89]. В структуре валового регионального продукта Иркутской области за 2014 год производство и реализация электроэнергии составляет около 6% [98].

2. Безработица выступает как сложное социально-экономическое явление, когда часть экономически активного населения не занята в общественном производстве товаров и услуг. Одно из главных негативных последствий безработицы – нерабочее состояние трудоспособных граждан и, соответственно, невыпущенная продукция. Когда экономика не в состоянии удовлетворить потребности в рабочих местах для всех, кто хочет и может работать, происходит потеря потенциальной возможности производства товаров и услуг. В конечном итоге это приводит к снижению темпов экономического роста, снижению объемов валового регионального продукта. С макроэкономической точки зрения сокращение фактического уровня безработицы увеличивает объем ВРП региона, а вместе с ним и личный доход граждан, повышая благосостояние и уровень жизни населения. А. Оукен вывел зависимость, согласно которой превышение фактического уровня занятости на 1% по сравнению с естественным уровнем приводит к отставанию фактического ВВП по сравнению с потенциальным примерно на 2.5% («коэффициент Оукена»). Следовательно, логично предположить, что объем валового регионального продукта связан с безработицей [54].

3. В экономической литературе сохраняется дискуссия о роли жилищного строительства в экономической жизни регионов и стран, о том, могут ли являться жилищные инвестиции активатором экономического роста территории. Большин-

ство исследований подтверждает мультипликативный эффект жилищных инвестиций и их положительное воздействие на экономический рост. Строительство жилых домов является прямым следствием жилищных инвестиций.

Уровень развития жилищной сферы и её влияния на экономический рост страны и регионов является важным фактором формирования экономической и социальной политики государства. Развитие жилищного строительства увеличивает производимый валовой продукт, создает спрос на продукцию смежных отраслей промышленности, транспорта. Кроме того, развитие этой сферы влияет на мобильность рабочей силы, эффективность её использования, а также на уровень человеческого капитала в целом. Экономическая политика в развитых странах в значительной степени обращает внимание на состояние жилищного рынка и его строительства, используя различные виды косвенного воздействия на его развитие. Это внимание к жилищной сфере обусловлено множественностью эмпирических доказательств мультипликативного влияния развития жилищного рынка на экономической рост регионов.

Рост строительства жилья обеспечивается проведением либеральной экономической политики и снятием административных барьеров на земельном рынке, понижением ставок по ипотеке, льготным софинансированием и т д.

Существуют эмпирические доказательства того, что строительство жилых домов может быть, как двигателем экономического роста, так и результатом экономического роста. Этот показатель является важным, поскольку отражает состояние экономики, любой коллапс в жилищном строительстве может привести к большим колебаниям в экономике, угрожающим ее стабильности [48]. В структуре валового регионального продукта Иркутской области за 2014 год строительство составляет около 6% [98].

4. К важнейшим факторам, формирующим объем и уровень валового дохода, относятся оборот розничной торговли. Этот показатель равен выручке от продажи товаров населению для личного потребления или использования в домашнем хозяйстве за наличный расчет или оплаченных по кредитным карточкам, расчетным чекам банков и т. д. Не вызывает сомнения, что этот показатель наиболее

ярко отражает благосостояние граждан региона и валового регионального продукта. В структуре ВРП Иркутской области розничная торговля за 2014 год составляет около 10% [98].

Для построения регрессионной модели динамики валового регионального продукта (ВРП) Иркутской области были собраны статистические данные за период с 2005 по 2014 год по следующим переменным [98]:

y – валовой региональный продукт, млрд. руб.;

x_1 – потребление электроэнергии, млрд. кВт;

x_2 – численность безработных, тыс. чел.;

x_3 – строительство жилых домов, тыс. кв. м.;

x_4 – оборот розничной торговли, млрд. руб.

Значения этих переменных представлены в таблице 3.2.

Таблица 3.2

Динамика экономических показателей 2005–2014 гг.

Год	y	x_1	x_2	x_3	x_4
	млрд. руб	млрд. кВт	тыс. чел	тыс. кв. м.	млрд. руб
	ВРП Иркутской обл.	Потребл. электроэн.	Числен. безработн.	Строительство жилых домов	Оборот розничной торговли
2005	258,1	52,5	127,5	303,0	104,3
2006	330,8	53,6	108,3	331,0	128,0
2007	402,7	53,3	104,9	575,0	151,3
2008	438,9	55,1	109,9	585,0	192,1
2009	458,8	52,4	137,9	602,2	191,4
2010	546,1	54,3	127,3	629,5	197,3
2011	634,6	56,7	114,9	755,2	225,8
2012	738,0	58,0	97,8	871,4	250,0
2013	805,2	56,6	104,4	829,2	266,5
2014	907,4	56,3	109,7	716,9	285,9

По этим данным рассчитана матрица парных коэффициентов корреляции, представленная в таблице 3.3.

Таблица 3. 3.

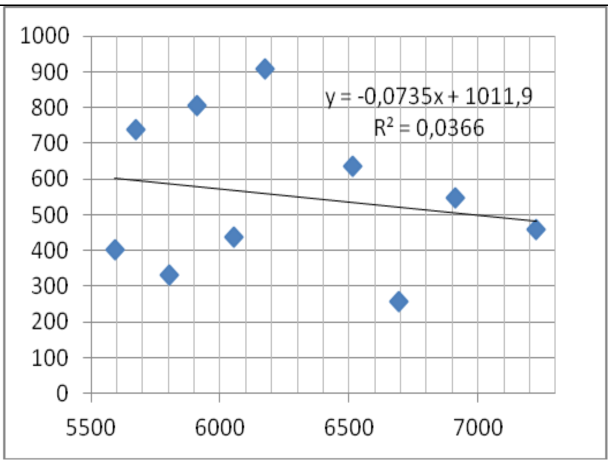
Корреляционная матрица

	y	x_1	x_2	x_3	x_4
y	1	0,8248	-0,4153	0,8647	0,9780
x_1		1	-0,6748	0,8277	0,8200
x_2			1	-0,4069	-0,3610
x_3				1	0,9088
x_4					1

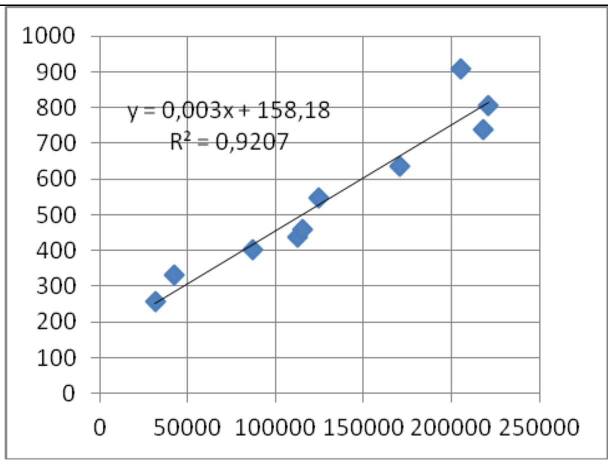
Между зависимой переменной y и каждым из объясняющих факторов x_1, x_3, x_4 имеется сильная положительная корреляция, а между y и x_2 – умеренная отрицательная, что вполне соответствует содержательному смыслу факторов. По корреляционной матрице также установлено, что некоторые из объясняющих переменных тесно коррелируют между собой.

Для более глубокого исследования зависимостей между переменными рассмотрим ряд построенных моделей, линейных по параметрам, но нелинейных по переменным, с целью выявить наиболее существенные закономерности между рассматриваемыми переменными.

1) $x = x_1 x_2$

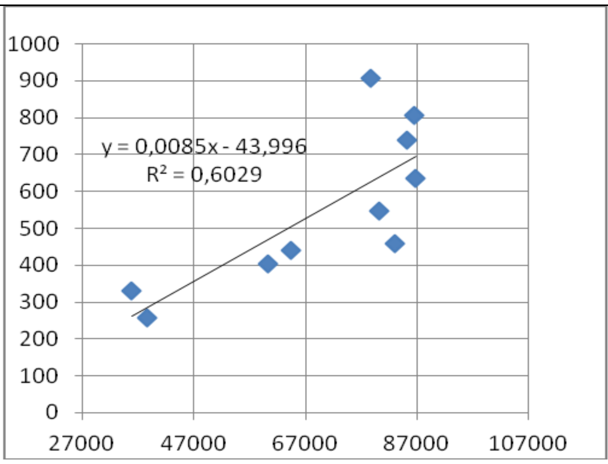


2) $x = x_3 x_4$

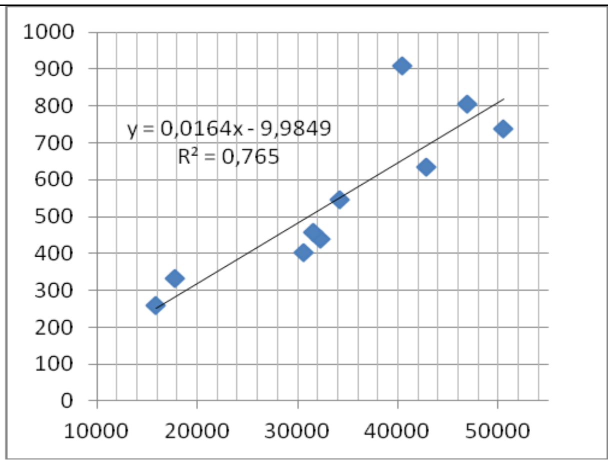


Далее на рисунке 3.1 представлены линейные регрессионные уравнения с одной переменной, где в качестве независимой переменной рассматриваются различные нелинейные сочетания переменных x_1, x_2, x_3, x_4, с целью выявления наиболее высокоинформативных зависимостей.

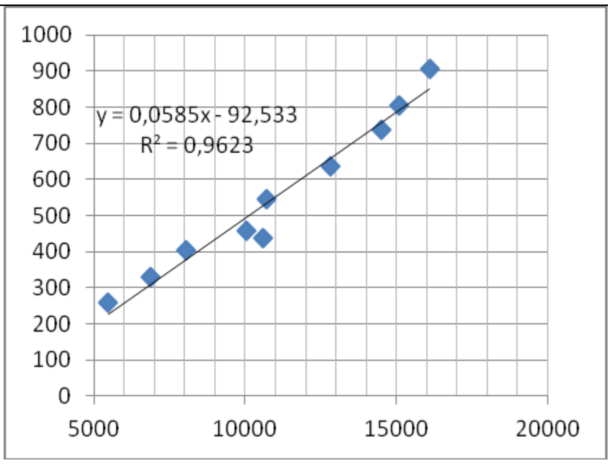
3) $x = x_2 x_3$



4) $x = x_1 x_3$



5) $x = x_1 x_4$



6) $x = x_2 x_4$

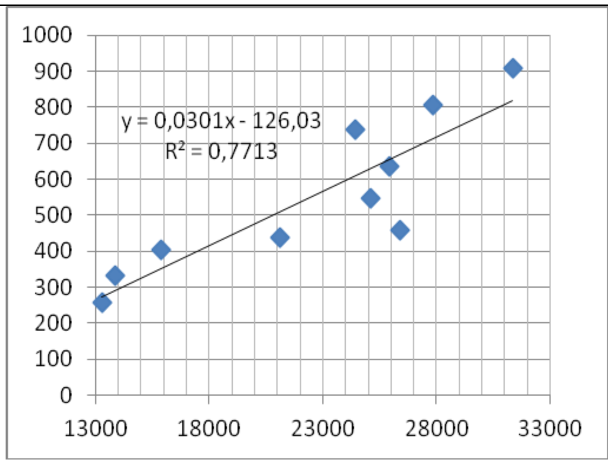
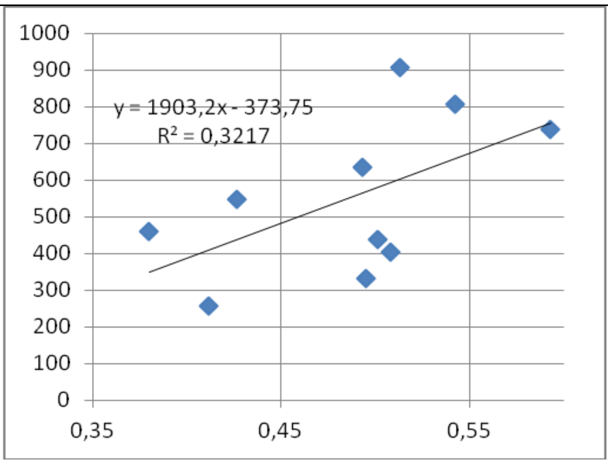
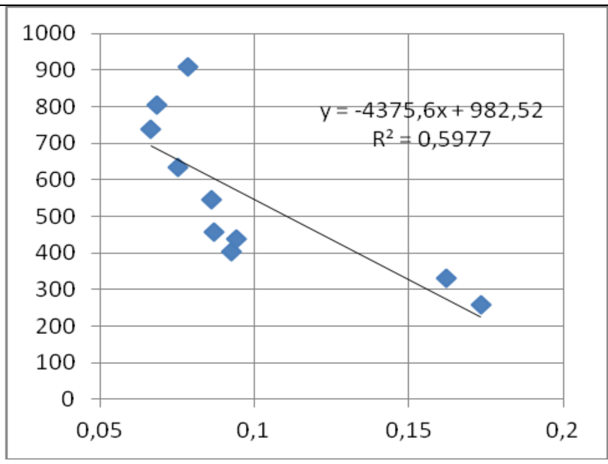


Рисунок 3.1. Комбинации переменных в виде произведений.

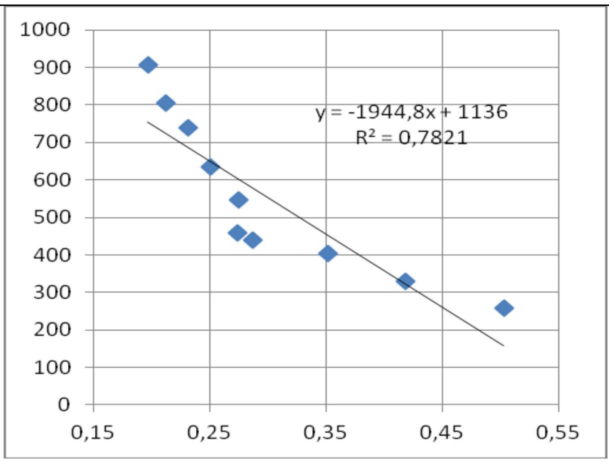
1) $x = (x_1 / x_2)$



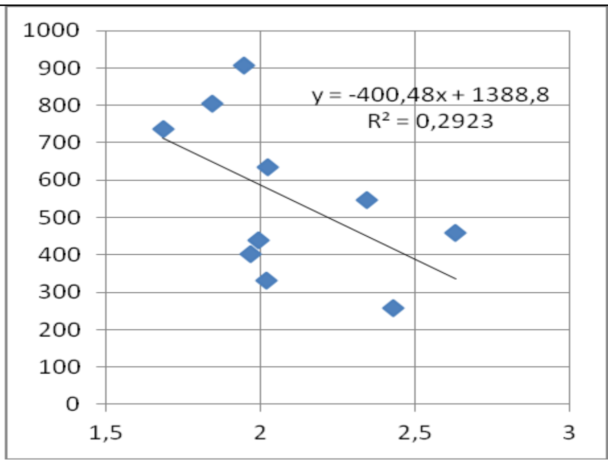
2) $x = (x_1 / x_3)$



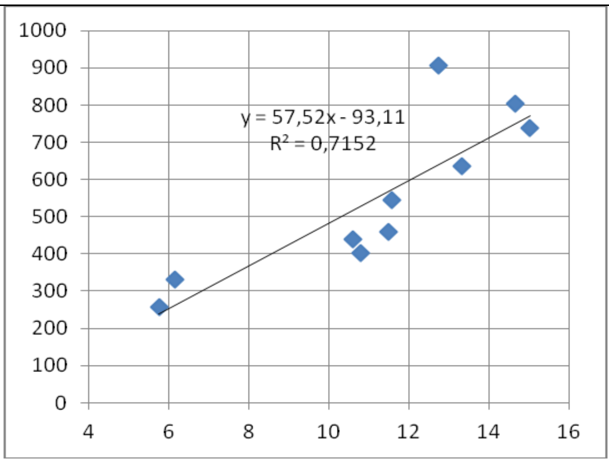
3) $x = (x_1/x_4)$



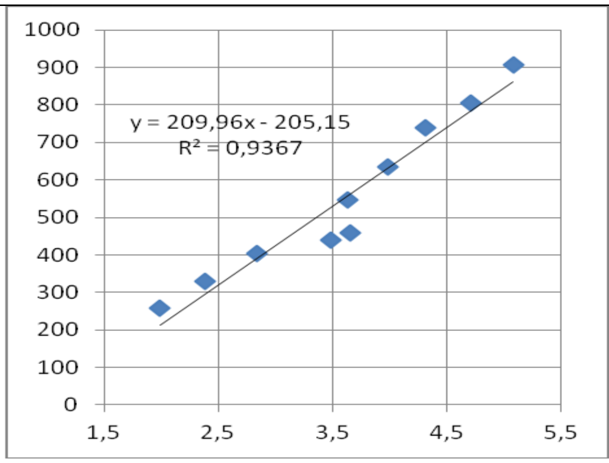
4) $x = (x_2/x_1)$



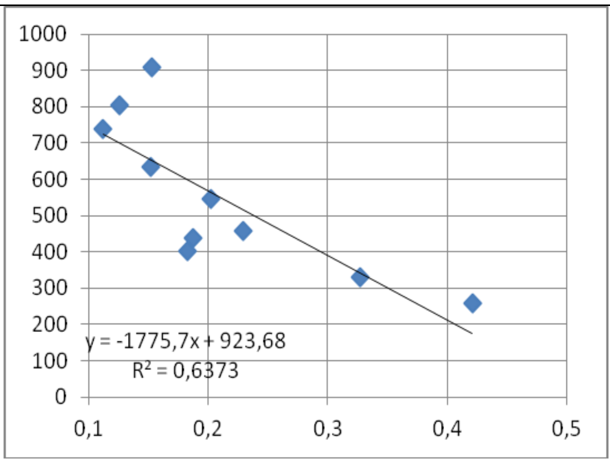
5) $x = (x_3/x_1)$



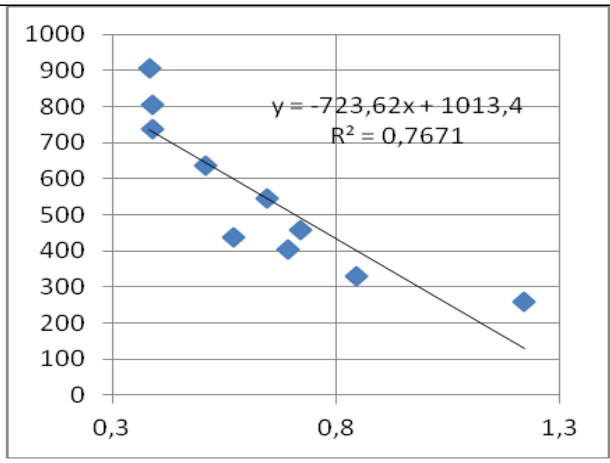
6) $x = (x_4/x_1)$



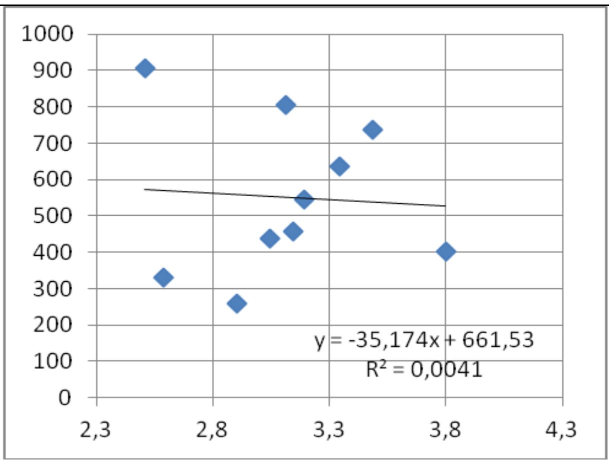
7) $x = (x_2/x_3)$



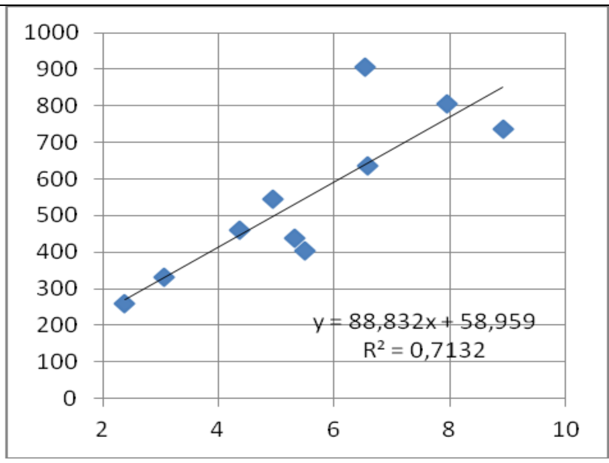
8) $x = (x_2/x_4)$



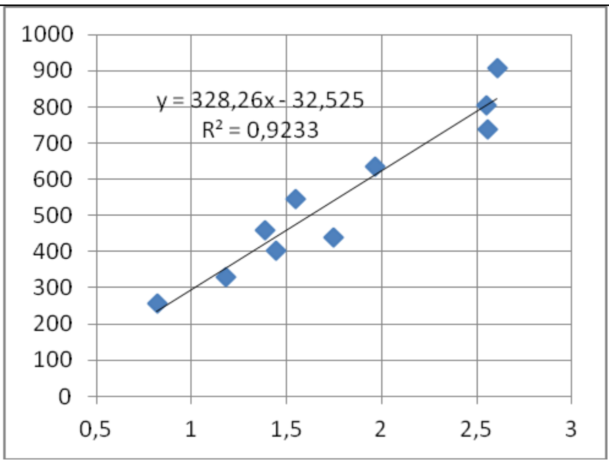
9) $x = (x_3/x_4)$



10) $x = (x_3/x_2)$



11) $x = (x_4/x_2)$



12) $x = (x_4/x_3)$

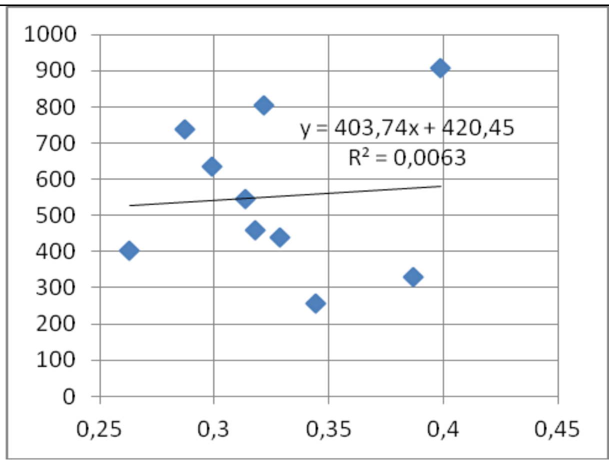
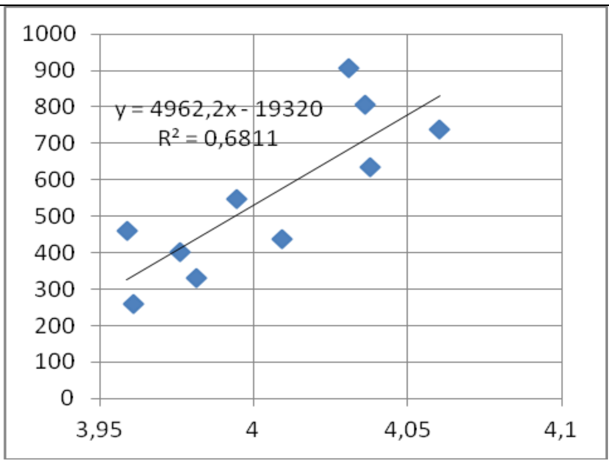
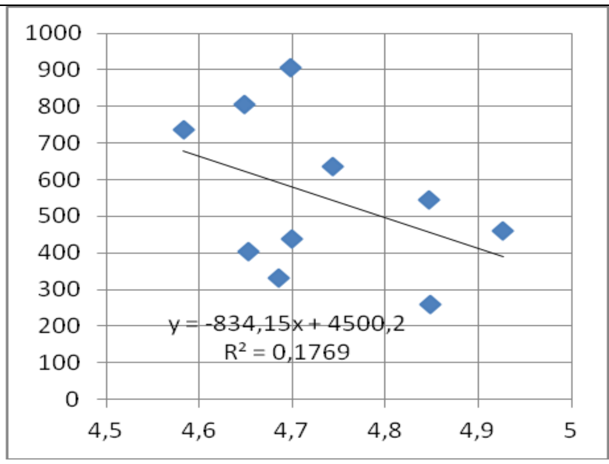


Рисунок 3.2. Комбинации переменных в виде частного.

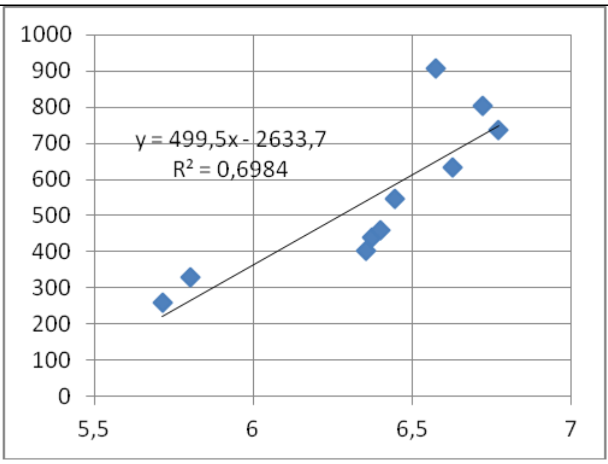
1) $x = \ln x_1$



2) $x = \ln x_2$



3) $x = \ln x_3$



4) $x = \ln x_4$

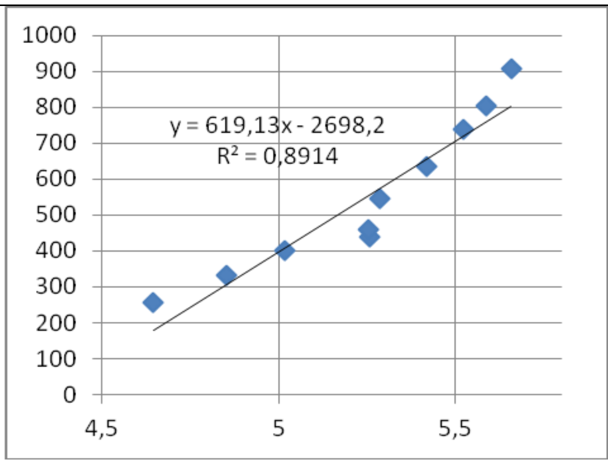
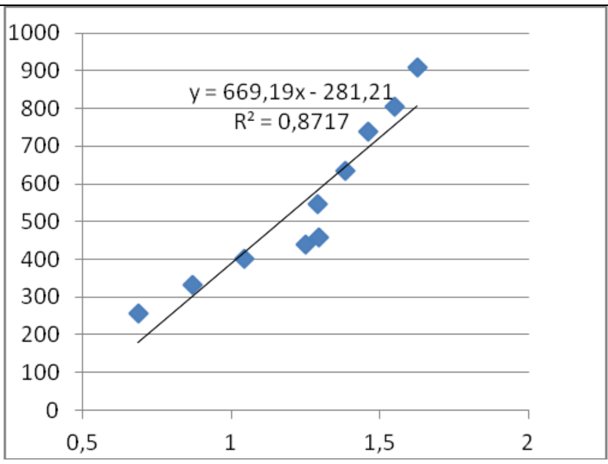
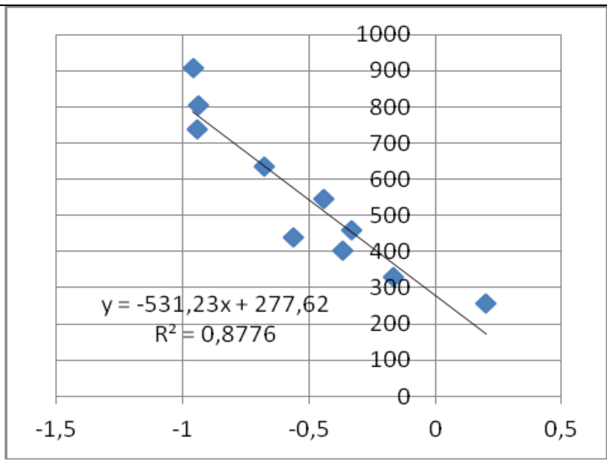


Рисунок 3.3. Нелинейная зависимость в виде логарифма.

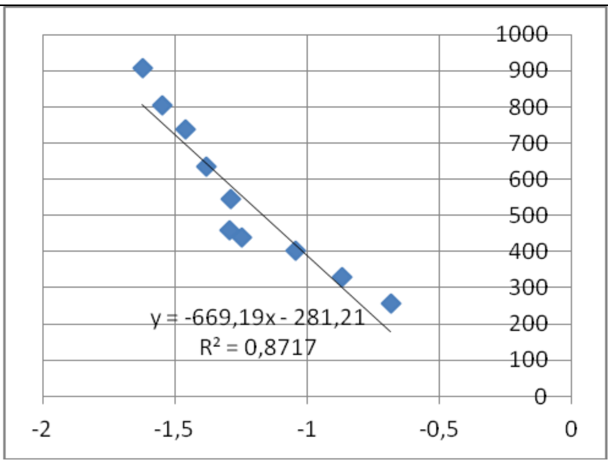
1) $x = \ln\left(x_4/x_1\right)$



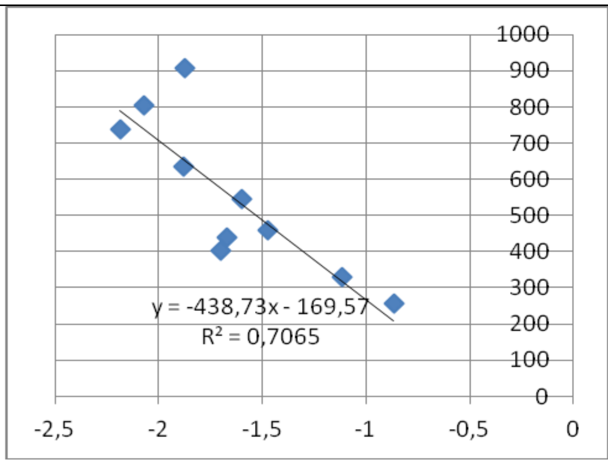
2) $x = \ln\left(x_2/x_4\right)$



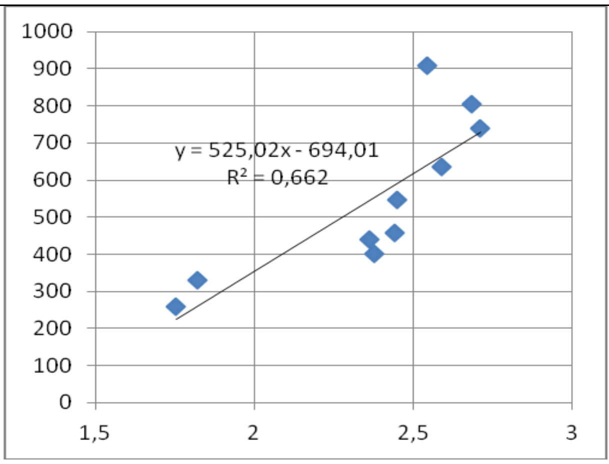
3) $x = \ln\left(x_1/x_4\right)$



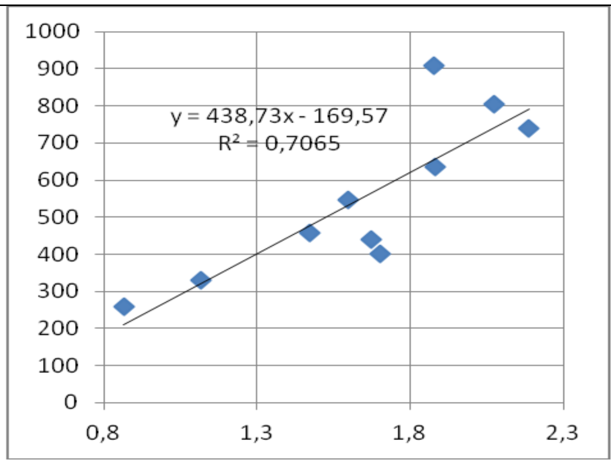
4) $x = \ln\left(x_2/x_3\right)$



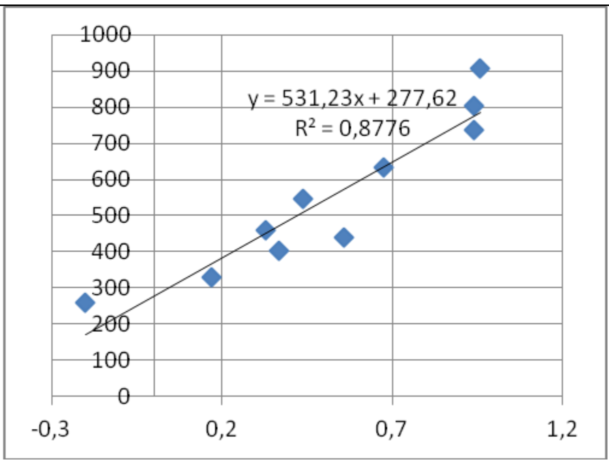
5) $x = \ln\left(x_3/x_1\right)$



6) $x = \ln\left(x_3/x_2\right)$



7) $x = \ln\left(x_2/x_4\right)$



8) $x = \ln\left(x_2/x_4\right)$

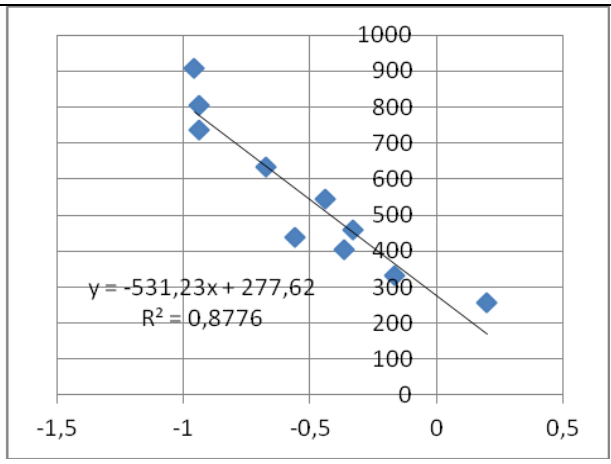


Рисунок 3.4. Комбинации переменных в виде логарифма частного.

Анализ рисунка 3.1 (1 – 6) показывает, что произведение переменных $x_1 \cdot x_2$ и y плохо связаны, чуть лучше обстоит с композицией переменных $x_2 \cdot x_3, x_1 \cdot x_3, x_2 \cdot x_4$ и y, и достаточно хорошая линейная зависимость между переменными $x_3\ x_4$ и y.

На рисунке 3.2 (1 – 12) представлен еще один вид нелинейной зависимости – частное. Анализ этих рисунков показал: следующее сочетание переменных $\dfrac{x_1}{x_2}, \dfrac{x_2}{x_1}, \dfrac{x_3}{x_4}, \dfrac{x_4}{x_3}$ и y дает слабую линейную связь, чуть лучше ситуация между

$\frac{x_1}{x_3}, \frac{x_3}{x_1}, \frac{x_2}{x_4}, \frac{x_4}{x_2}, \frac{x_3}{x_2}$ и y. Самая лучшая, судя по коэффициенту детерминации, линейная зависимость между $\frac{x_4}{x_1}$ с y.

На рисунке 3.3 (1 – 4) представлена зависимость виде логарифма от переменной. Лучшая закономерность между $\ln x_4$ и y. Менее выражена линейная зависимость между $\ln x_1, \ln x_3$ и y.

На рисунке 3.4(1 – 8) представлена зависимость виде логарифма от частного переменных. Лучшая закономерность между $\ln\frac{x_4}{x_2}$, $\ln\frac{x_2}{x_4}$ и y.

Построенные зависимости позволили определить какие элементарные функции и их преобразования следует использовать при организации конкурса моделей, а именно: произведение, частное и логарифм определенных переменных.

3.3 Выбор структурной спецификации регрессионной модели

Выбор структурной спецификации регрессии проведен с помощью реализации технологии «конкурса» моделей [12].

По корреляционной матрице было установлено, что некоторые из объясняющих переменных тесно коррелируют между собой. Следствием этого будет возникновение эффекта мультиколлинеарности при попытке построения, например, линейной модели множественной регрессии. Такая модель была оценена по методу наименьших квадратов (МНК) и имеет вид:

$$y = \underset{(-0,094)}{-128,092} + \underset{(0,135)}{3,140}x_1 - \underset{(-0,607)}{1,285}x_2 - \underset{(-0,886)}{0,210}x_3 + \underset{(5,093)}{3,939}x_4. \qquad (3.1)$$

Под коэффициентами этого уравнения расположены значения t-статистики. При уровне значимости 10 % значимым оказался только t-коэффициент переменной x_4 ($t_{кр}$ =2,01). Кроме того, знак коэффициента при переменной x_3 не удовлетворяет экономическому смыслу задачи. Как отмечено выше, причиной этого является эффект мультиколлинеарности. Для модели (3.1) коэффициент детерминации $R^2 = 0{,}966$, критерий Фишера $F = 35{,}86$, статистика Дарбина-Уотсона $DW = 1{,}85$. Эти значения говорят о высокой степени адекватности модели. Но применять полученное уравнение на практике можно лишь для прогнозирования, поскольку не все коэффициенты имеют адекватную экономическую интерпретацию.

Объединим переменные x_1, x_3, x_4, коэффициенты которых должны входить в модель со знаком «плюс», в один регрессор $x_1x_3x_4$. Оцененная модель y на регрессоры $x_1x_3x_4$ и x_2 имеет вид:

$$y = \underset{(0{,}461)}{120{,}997} + \underset{(7{,}942)}{5{,}137 \cdot 10^{-5}}\, x_1x_3x_4 + \underset{(0{,}218)}{0{,}451} x_2 . \tag{3.2}$$

$$R^2 = 0{,}917,\ F = 38{,}845,\ DW = 1{,}251.$$

Коэффициент корреляции между регрессорами равен –0,455. Это говорит об отсутствии мультиколлинеарности. Но неверным оказался знак коэффициента при переменной x_2, также этот коэффициент незначим. Поэтому была построена следующая однофакторная модель:

$$y = \underset{(4.320)}{222{,}770} + \underset{(7.366)}{0{,}00491} \frac{x_1x_3x_4}{x_2} . \tag{3.3}$$

$$R^2 = 0{,}871,\ F = 54{,}25,\ DW = 1{,}386.$$

Из-за того, что модель (3.3) состоит только из одного регрессора, то в ней автоматически отсутствует эффект мультиколлинеарности. Все коэффициенты этой модели значимы и удовлетворяют смыслу задачи. Однако по сравнению с моделями (3.1) и (3.2) снизились ее аппроксимационные качества. Поэтому было принято решение от уравнения (3.3) перейти к регрессии с двумя регрессорами. Для этого целесообразно организовывать технологию «конкурса» моделей [17,18, 19, 75] оценены параметры следующих шести моделей:

$$y = a_0 + a_1 \frac{x_1 x_3}{x_2} + a_2 x_4 + \varepsilon,$$

$$y = a_0 + a_1 \frac{x_1 x_4}{x_2} + a_2 x_3 + \varepsilon,$$

$$y = a_0 + a_1 \frac{x_3 x_4}{x_2} + a_2 x_1 + \varepsilon,$$

$$y = a_0 + a_1 \frac{x_1}{x_2} + a_2 x_3 x_4 + \varepsilon,$$

$$y = a_0 + a_1 \frac{x_3}{x_2} + a_2 x_1 x_4 + \varepsilon,$$

$$y = a_0 + a_1 \frac{x_4}{x_2} + a_2 x_1 x_3 + \varepsilon,$$

где a_0, a_1, a_2 – подлежащие оцениванию параметры, ε – ошибки аппроксимации.

Оцененные модели приведены ниже.

$$y = \underset{(-2,455)}{-147{,}022} + \underset{(0,103)}{0{,}028} \frac{x_1 x_3}{x_2} + \underset{(6,214)}{3{,}465} x_4. \tag{3.4}$$

Для модели (3.4) $R^2 = 0{,}956$, $F = 77$, $DW = 1{,}017$. Коэффициент при первом регрессоре (0.028) незначим. Коэффициент корреляции между регрессорами $r = 0{,}861$.

$$y = \underset{(0,1407)}{11{,}662} + \underset{(3,662)}{5{,}304} \frac{x_1 x_4}{x_2} + \underset{(0,093)}{0{,}027} x_3. \tag{3.5}$$

Для модели (3.5) $R^2 = 0{,}913$, $F = 36{,}96$, $DW = 1{,}590$. Свободный член (11,662) и коэффициент при втором регрессоре (0,027) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}900$.

$$y = \underset{(0,726)}{1268{,}99} + \underset{(3,602)}{0{,}342} \frac{x_3 x_4}{x_2} - \underset{(-0,609)}{20{,}537} x_1. \tag{3.6}$$

Для модели (3.6) $R^2 = 0{,}888$, $F = 27{,}73$, $DW = 1{,}551$. Свободный член (1268,99) и коэффициент при втором регрессоре (–20,537) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}912$.

$$y = \underset{(1,008)}{192{,}710} - \underset{(-0,187)}{83{,}786}\frac{x_1}{x_2} + \underset{(7,294)}{0{,}003}x_3x_4. \qquad (3.7)$$

Для модели (3.7) $R^2 = 0{,}921$, $F = 40{,}878$, $DW = 1{,}223$. Свободный член (192,71) и коэффициент при первом регрессоре (–83,786) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}607$.

$$y = -\underset{(-1,882)}{92{,}173} - \underset{(-0,651)}{10{,}392}\frac{x_3}{x_2} + \underset{(7,039)}{0{,}064}x_1x_4. \qquad (3.8)$$

Для модели (3.8) $R^2 = 0{,}964$, $F = 95{,}07$, статистика Дарбина-Уотсона $DW = 1{,}574$. Свободный член (–92,173) и коэффициент при первом регрессоре (−10,392) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}883$.

$$y = -\underset{(-0.402)}{28.350} + \underset{(3.816)}{343.875}\frac{x_4}{x_2} - \underset{(-0.188)}{0.00093}x_1x_3. \qquad (3.9)$$

Для модели (3.9) $R^2 = 0{,}923$, $F = 42{,}384$, $DW = 1{,}587$. Свободный член (−8,350) и коэффициент при втором регрессоре (–0,00093) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}918$.

Из моделей (3.4) – (3.9) была выбрана модель (3.8). Затем, используя нелинейное квадратичное преобразование для переменных модели (3.8), оценивались следующие спецификации:

$$y = a_0 + a_1\frac{x_3^2}{x_2} + a_2x_1x_4 + \varepsilon,$$

$$y = a_0 + a_1\frac{x_3}{x_2^2} + a_2x_1x_4 + \varepsilon,$$

$$y = a_0 + a_1\frac{x_3}{x_2} + a_2x_1^2x_4 + \varepsilon,$$

$$y = a_0 + a_1\frac{x_3}{x_2} + a_2x_1x_4^2 + \varepsilon.$$

Оцененные модели приведены ниже.

$$y = -\underset{(-1,841)}{109{,}984} - \underset{(-0,521)}{0{,}0075}\frac{x_3^2}{x_2} + \underset{(6,917)}{0{,}0626}x_1x_4. \qquad (3.10)$$

Для модели (3.10) $R^2 = 0{,}963$, $F = 93{,}05$, $DW = 1{,}497$. Свободный член (– 109,984) и коэффициент при первом регрессоре (–0,0075) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}88$.

$$y = \underset{(-1{,}886)}{-95{,}285} - \underset{(-0{,}366)}{451{,}715}\frac{x_3}{x_2^2} + \underset{(7{,}945)}{0{,}0608}x_1x_4. \tag{3.11}$$

Для модели (3.11) $R^2 = 0{,}963$, $F = 91{,}22$; $DW = 1{,}425$. Свободный член (–95,285) и коэффициент при первом регрессоре (–451,715) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}824$.

$$y = \underset{(-0{,}743)}{-33{,}661} - \underset{(-1{,}094)}{18{,}272}\frac{x_3}{x_2} + \underset{(7{,}147)}{0{,}0011}x_1^2x_4. \tag{3.12}$$

Для модели (3.12) $R^2 = 0{,}965$, $F = 97{,}75$, $DW = 2{,}189$. Свободный член (–33,661) и коэффициент при первом регрессоре (–18,272) не значимы. Коэффициент корреляции между регрессорами $r = 0{,}896$.

$$y = \underset{(5{,}639)}{178{,}999} + \underset{(0{,}071)}{0{,}689}\frac{x_3}{x_2} + \underset{(10{,}81)}{0{,}000154}x_1x_4^2. \tag{3.13}$$

Для модели (3.13) $R^2 = 0{,}984$, $F = 212{,}489$, $DW = 2{,}258$. Только коэффициент при первом регрессоре (0,071) не значим. Коэффициент корреляции между регрессорами $r = 0{,}849$.

Из моделей (3.10) – (3.13) лучшей оказалась модель (3.13). Затем, используя нелинейное преобразование $\ln x$ для переменных x_1, x_2, x_3 модели (3.8), оценивались следующие спецификации:

$$y = a_0 + a_1\frac{\ln x_3}{x_2} + a_2x_1x_4^2,$$

$$y = a_0 + a_1\frac{x_3}{\ln x_2} + a_2x_1x_4^2,$$

$$y = a_0 + a_1\frac{x_3}{x_2} + a_2\ln x_1x_4^2.$$

Оцененные модели приведены ниже.

$$y = \underset{(1,913)}{167,117} + \underset{(0,160)}{287,144}\frac{\ln x_3}{x_2} + \underset{(15,22)}{0,000154}x_1x_4^2. \qquad (3.14)$$

Для модели (3.14) $R^2 = 0,984$, $F = 213,12$, $DW = 2,262$. Только коэффициент при первом регрессоре (287,144) не значим. Коэффициент корреляции между регрессорами $r = 0,67$.

$$y = \underset{(4,445)}{177,427} + \underset{(0,096)}{0,0487}\frac{x_3}{\ln x_2} + \underset{(10,05)}{0,000154}x_1x_4^2. \qquad (3.15)$$

Для модели (3.15) $R^2 = 0,984$, $F = 212,618$, $DW = 2,248$. Только коэффициент при первом регрессоре (0,0487) не значим. Коэффициент корреляции между регрессорами $r = 0,871$.

$$y = \underset{(5,074)}{159,896} + \underset{(0,404)}{3,826}\frac{x_3}{x_2} + \underset{(10,69)}{0,00215}x_4^2\ln x_1. \qquad (3.16)$$

Для модели (3.16) $R^2 = 0,983$, $F = 207,85$, $DW = 2,013$. Только коэффициент при первом регрессоре (3.826) не значим. Коэффициент корреляции между регрессорами $r = 0,841$.

Из моделей (3.14) – (3.16) лучшей оказалась зависимость (3.15). Используя кубическое нелинейное преобразование для переменных x_1, x_3, x_4 модели (3.15), оценивались следующие спецификации:

$$y = a_0 + a_1\frac{x_3^3}{\ln x_2} + a_2x_1x_4^2,$$

$$y = a_0 + a_1\frac{x_3}{\ln x_2} + a_2x_1^3x_4^2,$$

$$y = a_0 + a_1\frac{x_3}{\ln x_2} + a_2x_1x_4^3.$$

Оцененные модели приведены ниже.

$$y = \underset{(8,767)}{180,797} + \underset{(0,023)}{1,009\cdot 10^{-8}}\frac{x_3^3}{\ln x_2} + \underset{(10,79)}{0,000155}x_1x_4^2. \qquad (3.17)$$

Для модели (3.17) $R^2 = 0{,}984$, $F = 212{,}347$, $DW = 2{,}268$. Только коэффициент при первом регрессоре ($1{,}009 \cdot 10^{-8}$) не значим. Коэффициент корреляции между регрессорами $r = 0{,}851$.

$$y = \underset{(4,348)}{212{,}873} - \underset{(-0,0714)}{0{,}0429} \frac{x_3}{\ln x_2} + \underset{(8,526)}{4{,}646 \cdot 10^{-8}}\, x_1^3 x_4^2 . \tag{3.18}$$

Для модели (3.18) $R^2 = 0{,}978$, $F = 156{,}011$, $DW = 2{,}306$. Только коэффициент при первом регрессоре (–0,0429) не значим. Коэффициент корреляции между регрессорами $r = 0{,}877$.

$$y = \underset{(5,141)}{203{,}795} + \underset{(2.,23)}{0{,}881} \frac{x_3}{\ln x_2} + \underset{(10,5)}{4{,}244 \cdot 10^{-7}}\, x_1 x_4^3 . \tag{3.19}$$

Для модели (3.19) $R^2 = 0{,}985$, $F = 231{,}36$, $DW = 1{,}964$. Все коэффициенты значимы. Коэффициент корреляции между регрессорами $r = 0{,}82$.

Из регрессий (3.17) – (3.19) лучшими свойствами обладает модель (3.19).

Таким образом, реализация «конкурса» моделей позволила выбрать структурную спецификацию модели валового регионального продукта Иркутской области. Это регрессия (3.19).

3.4 Множественное оценивание регрессионных параметров

В этом разделе рассматривается задача двухкритериального оценивания параметров линейного уравнения регрессии по методике, описанной в главе 2 моделирования валового регионального продукта Иркутской области [13].
Динамика рассматриваемых показателей представлена на рисунках 3.5 – 3.9.

Таблица 3.3

ВРП Иркутской области по годам (млрд руб)

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
258.1	330,8	402,7	438,9	458,8	546,1	634,6	738	805,2	907,4

Рисунок 3.5. Динамика ВРП Иркутской области

Таблица 3.4

Потребление электроэнергии по годам (млрд кВт)

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
52,5	53,6	53,3	55,1	52,4	54,3	56,7	58	56,6	56,3

Рисунок 3.6. Динамика потребления электроэнергии, млрд. кВт/ч

Таблица 3.5

Численность безработных по годам (тыс чел)

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
127,5	108,3	104,9	109,9	137,9	127,3	114,9	97,8	104,4	109,7

Рисунок. 3.7. Динамика численности безработных, тыс чел.

Таблица 3.6

Строительство жилых домов по годам (тыс кв м)

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
303	331	575	585	602,2	629,5	755,2	871,4	829,2	716,9

Рисунок. 3.8. Динамика строительства жилых домов, тыс. кв. м.

Таблица 3.7

Оборот розничной торговли по годам (млрд руб.)

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
104,3	128	151,3	192,1	191,4	197,3	225,8	250	266,5	285,9

Рисунок 3.9. Динамика оборота розничной торговли, млрд. руб.

Для реализации процедуры множественного оценивания регрессионных моделей в среде программирования Delphi был разработан специализированный программный комплекс [87]. Все представленные ниже результаты получены с его помощью по отношению к уравнению (3.19).

Множество Парето $J(P^*)$, состоящее из шести вершин, представлено в таблице 3.8.

Таблица 3.8

Паретовские вершины

J_1	J_∞
245,6	44,23
235,7	44,36
226,2	44,61
218,4	44,9
197,1	46,18
190,6	53,56

Множество Парето в критериальном пространстве представлено на рис. 3.10.

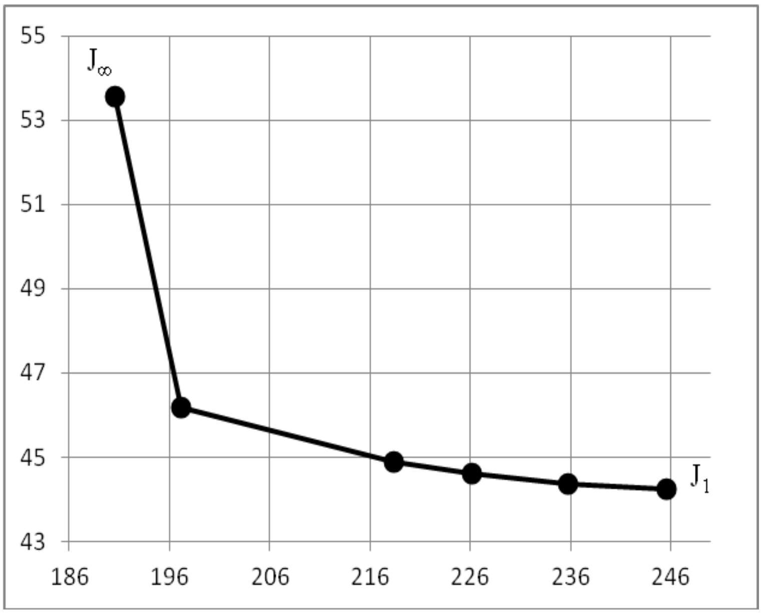


Рисунок 3.10. Множество Парето

1. Точечная характеризация множества P:

$$y = 239,057 + 4,382 \cdot 10^{-7} x_1 x_4^3 + 0,601 \frac{x_3}{\ln x_2}.$$

2. Множество A:

$$\alpha_0 \in [132,1;\ 251,1],$$

$$\alpha_1 \in \left[3,68 \cdot 10^{-7};\ 4,513 \cdot 10^{-7}\right],$$

$$\alpha_2 \in [0,4096;\ 1,665].$$

3. Для получения прогноза ВРП Иркутской области на 2015 год были использованы, известные на момент исследования значения объясняющих переменных: $x_1^* = 52,7$, $x_2^* = 103,1$, $x_3^* = 963,7$, $x_4^* = 290,8$ [98]. Интервальный прогноз имеет вид: $y \in [921,1;\ 962,6]$.

В марте 2017 года появлялись уточнения статистических данных по ВРП Иркутской области за 2014 год и фактическое значение ВРП за 2015 год, т.е появилась возможность проверить прогноз и фактическое значение ВРП. На рисунке 3.11 представлено сравнение фактического и прогнозного значения ВРП на 2015 год.

Рисунок 3.11. Сравнение фактического и прогнозного ВРП

Модель была пересчитана для скорректированных данных и получено следующее регрессионное уравнение:

$$y = \underset{(5.174)}{208{,}563} + \underset{(10.57)}{4{,}342 \cdot 10^{-7}}\, x_1 x_4^3 + \underset{(1.921)}{0{,}811}\frac{x_3}{\ln x_2}.$$

Для уровня значимости 10% все коэффициенты значимы. Коэффициент детерминации этой модели $R^2 = 0{,}985$, критерий Фишера $F = 227{,}2$. Статистика Дарбина-Уотсона $DW = 1{,}97$, что свидетельствует об отсутствии автокорреляции в ошибках модели.

Множество Парето $J(P^*)$, состоящее из шести вершин, представлено в таблице 3.9

Таблица 3.9

Паретовские вершины

J_1	J_∞
253,1	44,25
244,6	44,36
235,1	44,61
224,6	45
204,9	46,19
197,2	54,76

Множество Парето в критериальном пространстве представлено на рис. 3.12

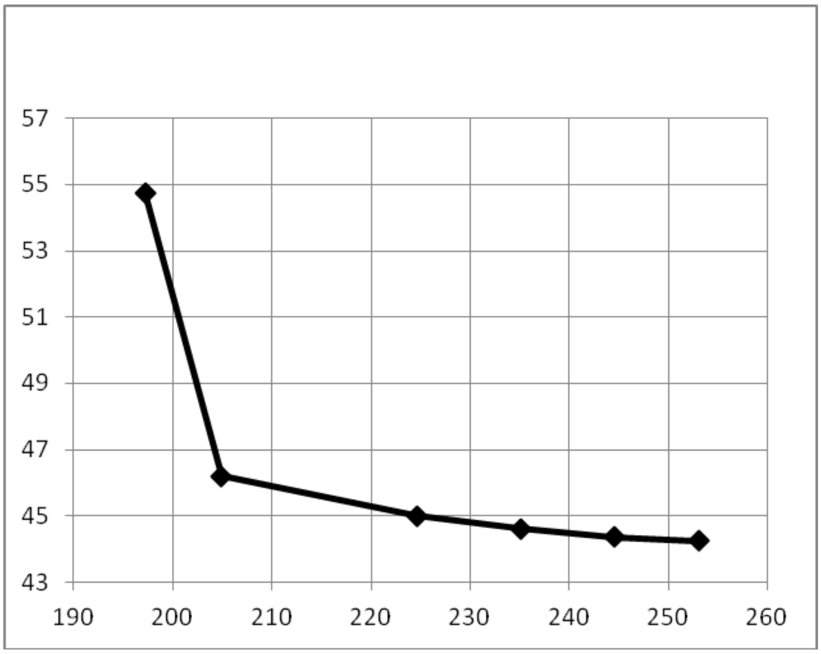


Рисунок 3.12 Множество Парето

1. Точечная характеризация множества P:

$$y = 241{,}933 + 4{,}497 \cdot 10^{-7} x_1 x_4^3 + 0{,}542 \frac{x_3}{\ln x_2}.$$

2. Множество A:

$$\alpha_0 \in [133{,}5;\ 256],$$

$$\alpha_1 \in \left[3{,}703 \cdot 10^{-7};\ 4{,}648 \cdot 10^{-7}\right],$$

$$\alpha_2 \in [0{,}3197;\ 1{,}647].$$

3. Для получения прогноза ВРП Иркутской области на 2015 год были использованы следующие значения объясняющих переменных: $x_1^* = 52{,}7$, $x_2^* = 103{,}1$, $x_3^* = 963{,}7$, $x_4^* = 290{,}8$. Значение объясняющей переменной – 967,7 млрд. рублей. Интервальный прогноз имеет вид: $y \in [926{,}7;\ 967{,}7]$.

При этом официальное значение ВРП за 2015 год составило 1013,5 млрд. рублей [98]. Абсолютная погрешность на 2015 год составляет 45,8 млрд. рублей. Средняя относительная ошибка аппроксимации составила 5%. Значения средней относительной ошибки аппроксимации, не превышающее 10%, свидетельствует о хорошем соответствии регрессии исходным данным [42].

Для сравнения построим прогноз по модели, полученной по МНК- методу со структурной спецификацией вида (3.19):

$$y = 203,7951 + 4,24 \cdot 10^{-7} x_1 x_4^3 + 0,88 \frac{x_3}{\ln x_2} \cdot$$

Прогноз ВРП Иркутской области на 2015 год по этой модели по тем же данным составил 776,775 млрд. рублей, и при этом отклоняется от истинного значения ВРП на 236,725 млрд. рублей. Интервальный прогноз имеет вид: $y \in [720,87; 832,68]$.

3.5 Выводы

Рассмотрена методология расчета валового регионального продукта, а также основные проблемы с этим сопряженные. Изложены основные методы прогнозирования данного показателя.

Проведен анализ функциональной зависимости между переменными при моделировании ВРП Иркутской области

С помощью технологии организации «конкурса» моделей выбрана структурная спецификация модели ВРП.

Проведена процедура множественного оценивания параметров выбранной модели регрессии, с помощью которой получен интервальный прогноз ВРП Иркутской области на 2015 год.

Из работы следует, что аппарат множественного оценивания параметров является весьма эффективным при моделировании сложных систем и более гибким (мягким) по сравнению с традиционным регрессионным анализом, в рамках которого возможно построение только так называемых точечных (жестких) оценок.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований в работе получены следующие результаты.

1. Выполнен критический анализ наиболее часто используемых в рамках регрессионного анализа методов численного оценивания параметров уравнений, главным образом сводящихся к минимизации функций потерь, приводящей к получению L_v-оценок.

2. Разработана алгоритмическая схема получения множества параметрических оценок Парето в двухкритериальной задаче идентификации на основе одновременного использования полярных по отношению к выбросам методов наименьших модулей и антиробастного оценивания.

3. Предложены способы конкретизации множества паретовских оценок, облегчающие оперирование им.

4. Разработан программный комплекс МОРМ множественного оценивания параметров линейного регрессионного уравнения и проведения прогнозных расчетов на основе полученных оценок.

5. Построена регрессионная модель динамики валового регионального продукта Иркутской области с множественной оценкой параметров и на ее основе проведены прогнозные расчеты на среднесрочную перспективу.

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян, С. А. Анализ качества и образа жизни населения: эконометрический подход / С. А. Айвазян. – М. : Наука, 2012. – 432 с.

2. Айвазян, С. А. Прикладная статистика и основы эконометрики. / С. А. Айвазян, В. С. Мхитарян. − М. : ЮНИТИ, 1998. – Т. 1. – 1022 с.

3. Айвазян, С. А. Прикладная статистика. Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 983. – 472 с.

4. Айвазян, С. А. Прикладная статистика. Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1985. – 488 с.

5. Айвазян, С. А. Прикладная статистика. Классификация и снижение размерности / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1989. – 607 с.

6. Акофф Рассел, Л. Планирование в больших экономических системах /Л. Акофф Рассел – М. : Советское радио, 1972. – 224 с.

7. Андреев, А. В. Региональная экономика: учебник для вузов. Стандарт третьего поколения / А. В. Андреев, Л. М. Борисова, Э. В. Плучевская. – Спб. : Питер, 2012. – 464 с.

8. Анализ и прогнозирование экономики региона. – М. : Наука, 1984. – 271 с.

9. Афифи, А. Статистический анализ: подход с использованием ЭВМ / А. Афифи, С. Эйзен. – М. : Мир, 1983. – 486 с.

10. Баенхаева, А. В. Множественное оценивание параметров линейного регрессионного уравнения / С. И. Носков // Современные технологии, системный анализ. Моделирование. – 2016. – № 3(51) – С.133 – 140.

11. Баенхаева, А. В. Прогнозирование валового регионального продукта // Экономика и бизнес. Теория и практика. – 2016.– Вып. 11.– С. 5 − 11.

12. Баенхаева, А. В. Выбор структурной спецификации регрессионной модели валового регионального продукта Иркутской области / М. П. Базилевский, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2016. – Вып. 16. – С. 30 − 37.

13. Баенхаева, А. В. Моделирование валового регионального продукта Иркутской области на основе применения методики множественного оценивания / М. П. Базилевский, С. И. Носков // Фундаментальные исследования. – 2016. – № 10 (часть 1). – С. 9 − 14.

14. Баенхаева, А.В. Один подход к исследованию эффективности вузов // Мы продолжаем традиции Российской статистики. Сборник докладов I Открытого российского статистического конгресса. Российская ассоциация статистиков; Федеральная служба государственной статистики РФ; Новосибирский государственный университет экономики и управления «НИНХ». – 2016. – С. 160-166.

15. Баенхаева, А.В. К вопросу о методике проведения мониторинга эффективности деятельности вузов / И.А. Слободняк, А.В. Баенхаева // Экономический анализ: теория и практика. – 2015. – №36(435). – С.50-60.

16. Баенхаева, А.В. Один подход к прогнозированию валового регионального продукта Иркутской области//Материалы научной конференции «Высокие технологии и инновации в науке» ГНИИ «Нацразвитие». Июль 2018: Сборник избранных статей.– СПб.: ГНИИ «Нацразвитие», 2018.– С. 178-185.

17. Базилевский, М. П. Технология организации конкурса регрессионных моделей/ М. П. Базилевский, С. И. Носков // Информационные технологии и проблемы математического моделирования сложных систем. – 2009. – Вып. 7. – С. 77 − 84.

18. Базилевский, М. П. Алгоритм построения линейно-мультипликативной регрессии / М. П. Базилевский, С. И. Носков // Современные технологии. Системный анализ. Моделирование. – 2011. – № 1(29). – С. 88 − 92.

19. Базилевский, М. П. Алгоритм формирования множества регрессионных моделей с помощью преобразования зависимой переменной / М. П. Базилевский,

С. И. Носков // Международный журнал прикладных и фундаментальных исследований. – 2011. – № 3. – С. 159 − 160.

20. Базилевский, М. П. Идентификация неизвестных параметров линейно-мультипликационной регрессии / М.П. Базилевский, С.И. Носков // Современные наукоемкие технологии. –2012. – № 3. – С. 14 – 18.

21. Базилевский, М. П. Методические и инструментальные средства построения некоторых типов регрессионных моделей / М. П. Базилевский, С. И. Носков // Системы. Методы. Технологии. – 2012. – № 13. – С. 81 – 86.

22. Банди, Б. Методы оптимизации. Вводный курс : пер. с англ. / Б. Банди. – М. : Радио и связь, 1988.– 128 с.

23. Бара, Ж.-Р. Основные понятие математической статистики / Ж. -Р. Бара. – М. : Мир, 1974. – 275 с.

24. Бестужев – Лада, И. В. Рабочая книга по прогнозированию/ И. В. Бестужев - Лада. – М.: Мысль, 1982. – 430 с.

25. Болдин, М. В. Знаковый статистический анализ линейных моделей / М. В. Болдин, Г. И. Симонова, Ю. Г. Тюрин. − М. : Наука Физматлит, 1997. – 208 с.

26. Боровков, А. А. Математическая статистика. Оценка параметров, проверка гипотез / А. А. Боровков. − М. : Наука, 1984. – 472 с.

27. Вилкас, Э. И. Решения: теория, информация, моделирование / Э. И. Вилкас, Е. З. Майминас. – М. : Радио и связь, 1981. – 328 с.

28. Власов, М. П. Моделирование экономических процессов / М. П. Власов, П. Д. Шимко. – Ростов на Дону : Феникс, 2005. – 409 с.

29. Вопросы анализа и процедуры принятия решений. – М. : Мир, 1976. – 229 с.

30. Вороненко, С. В. Нахождение множества Парето в линейной многокритериальной задаче / С. В. Вороненко, А. П. Селедкин // Дифференциальные уравнения и численные методы. – 1986. – С. 217 − 224.

31. Вучков, И. Прикладной линейный регрессионный анализ / И. Вучков, Л. Бояджилова, Е. Солаков. − М. : Финансы и статистика, 1987. – 239 с.

32. Гейци, И. И. Распределенные информационные системы территориального управления / И. И. Гейци, Г. И. Карпачев, Н. Г. Лавров. – Новосибирск : ВЦ СО АН СССР, 1989. – 248 с.

33. Головченко, В. Б. Об одном подходе к разработке подсистемы «Демография и социальное развитие» в составе АСУ «Город» / В. Б. Головченко, Б. Р. Жидиханов, С. И. Носков // Информатизация и моделирование территориальных социально-экономических объектов: сб. науч. тр. всесоюз. конф. – Новосибирск, 1990. – ч. 2. – С. 24 − 25.

34. Головченко, В. Б. О комплексе программ сравнения эконометрических моделей на примере некоторых регрессионных зависимостей / В. Б. Головченко, С. И. Носков // Пакеты прикладных программ. Функциональное наполнение. – Новосибирск : Наука, 1985. – С. 90 − 96.

35. Головченко, В. Б. Прогнозирование на основе дискретной динамической модели с использованием экспертной информации / В. Б. Головченко, С. И. Носков // Автоматика и телемеханика. – 1993. – №10. – С. 140 − 148.

36. Демиденко, Е. З. Линейная и нелинейная регрессии / Е. З. Демиденко. – М. : Финансы и статистика, 1981. – 302 с.

37. Джонстон, Дж. Эконометрические методы / Дж. Джонстон. – М. : Статистика, 1980. – 416 с.

38. Доугерти, К. Введение в эконометрику : учебник / К. Доугерти. – М. : ИНФРА–М, 2001. – 402 с.

39. Дрейпер, Н., Смит, Г. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М. : Финансы и статистика, 1981. – т. 1. – 366 с., т. 2. – 351 с.

40. Дрейпер, Н. Прикладной регрессионный анализ. / Н. Дрейпер, Н. Смит. − М. : Статистика, 1973. – 392 с.

41. Дубровский, С. А. Прикладной многомерный статистический анализ / С. А. Дубровский. − М. : Финансы и статистика, 1981. – 216 с.

42. Елисеева, И. И. Эконометрика : учебник / И. И. Елисеева [и др.]; под ред.И. И. Елисеевой. – М. : Юрайт, 2016. – 449 с.

43. Елисеева, И. И. Эконометрика : учебник для бакалавров / И. И. Елисеева [и др.]; под ред.И. И. Елисеевой. – М. : Проспект, 2014. – 228 с.

44. Емельянов, А. С. Промышленное производство: динамика, тенденции, модели / А. С. Емельянов. – Киев : Наукова Думка, 1980. – 428 с.

45. Емельянов, А. С. Эконометрия и прогнозирование / А. С. Емельянов. − М. : Экономика, 1985. – 208 с.

46. Еремин, И. И. Введение в теорию линейного и выпуклого программирования / И. И. Еремин, Н. Н. Астафьев. – М. : Наука, 1976. – 191 с.

47. Зайцева, Ю. С. Валовой региональный продукт, что и как мы измеряем [Электронный ресурс]: универсальные базы данных / Ю. С. Зайцева // ЭКО. Экономика и организация промышленного производства. – Новосибирск. – № 4. – 2012. – С. 86 − 103 – Режим доступа: http://www.ebiblioteka.ru / (дата обращения: 24.сен.2016).

48. Залкинд, Л. О. О взаимосвязи инвестиций в жилищное строительство и экономического роста /Л. О. Залкинд // Жилищные стратегии - 2014 – № 1.

49. Зорич, В. А. Математический анализ / В. А. Зорич. − М. : ФАЗИС, 1997. – 554 с.

50. Зуховицкий, С. И. Линейное и выпуклое программирование / С. И. Зуховицкий, Л. И. Авдеева – М. : Наука, 1964. – 348 с.

51. Иващенко, И. Ю. Сценарное прогнозирование и анализ рисков развития региональных систем : дис….кан. экон. наук: 08.00.05/ И. Ю. Иващенко; НИУ БелГУ. – Белгород, 2002. – 227 с.

52. Ивахненко, А. Г. Помехоустойчивость моделирования / А. Г. Ивахненко, В. С. Степашко. – Киев : Наукова думка, 1985. – 216 с.

53. Ивченко, Г. И. Математическая статистика : учеб. пособие / Г. И. Ивченко, Ю. И. Медведев. −М. : Высш. шк., 1984. – 248 с.

54. Илюхина, А. В. Безработица как фактор достижения потенциального ВВП [Электронный ресурс] III Общероссийская студенческая электронная научная конференция «Студенческий научный форум»15 – 20 февраля 2011 года– Режим доступа: https://www.rae.ru/forum2011/82/163 (дата обращения: 23.сен.2017).

55. Исследование операций. Модели и применение / под ред. Дж. Моудера и С. Элмаграби. – М. : Мир, 1981. – 677 с.

56. Карбаев, Д. С. Модели сценарного прогнозирования макроэкономических показателей региона в условиях малой выборки: дис…кан. экон. наук: 08.00.13 / Д. С. Карабаев; СГОУ. – Самара, 2010. – 174 с.

57. Кади, Дж. Количественные методы в экономике / Дж. Канди. – М.: Прогресс, 1977. – 297 с.

58. Корн, Г. Справочник по математике для научных работников и инженеров / Г. Корн, Т. Корн. – М. : Наука, 1973. – 832 с.

59. Корхин, А. С.Оценивание параметров регрессии как многокритериальная задача / А. С. Корхин // Автоматика и телемеханика. – 1992. – № 9. – С. 110 − 117.

60. Ковалев, С. В. Экономическая математика : учеб. пособие / С. В. Ковалев. – М. : КНОРУС, 2013. – 248 с.

61. Кремер, Н. Ш. Исследование операций в экономике : учеб. пособие для вузов / Н. Ш. Кремер, Б. А. Путко, И. М. Тришин, М. Н. Фридман; под ред. проф. Н. Ш. Кремера. – М. : Банки и биржи, ЮНИТИ, 1997. – 350 с.

62. Кукушкин, Д. К. Применение шкал эквивалентности для измерения уровня жизни / Д. К. Кукушкин // Научные труды: Институт народнохозяйственного прогнозирования РАН, 2003. – Т. 1. – С. 430 − 450.

63. Лакеев, А. В. Метод наименьших модулей для линейной регрессии. Число ненулевых ошибок аппроксимации // Современные технологии. Системный анализ. Моделирование. 2012. – № 2. – С. 48 – 50.

64. Лопатников, Л. И. Экономико-математический словарь: Словарь современной экономической науки. 5-е изд., перераб. и доп. − М. : Дело, 2003. – 520 с.

65. Макаров, Н. М. Теория выбора и принятия решений / Н. М. Макаров и др.– М. : Наука, 1982. – 392 с.

66. Математическая энциклопедия: в 6 т. / гл.ред. И. М. Виноградов. – М.: Советская энциклопедия, 1984. − Т.4: Ок – Сло. – 1216 с.

67. Магнус, Я. Р. Эконометрика. Начальный курс. / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 2007. − 504 с.

68. Мельников, Р. М. Эконометрика : учеб. пособие / Р. М. Мельников. – М. : Проспект, 2014. – 288 с.

69. Митропольский, А. К. Техника статистических вычислений. / А. К. Митропольский.– М. : Наука, 1971. – 576 с.

70. Моисеев, Н. Н. Введение в экономико-математические методы / Н. Н. Моисеев. − М. : Наука, 1984. – С. 5 − 10.

71. Моудера, Дж., Элмаграби, С. Исследование операций. Модели и применения. Том 2 / Пер. с англ., Под ред. Дж. Моудера, С. Элмаграби. – М. : Мир, 1981.– 684 с.

72. Мудров, В. И. Метод наименьших модулей / В. И. Мудров, В. Л. Кушко. – М. : Знание, 1971.– 61 с.

73. Мудров, В. И. Методы обработки информации: Квазиправдоподобные оценки / В. И. Мудров, В. Л. Кушко. – М. : Радио и связь, 1983. – 304 с.

74. Мхитарян, В. С. Эконометрика : учебник / В. С. Мхитарян и др. – М. : Проспект, 2009. – 354 с.

75. Носков, С. И. Технология объектов с нестабильным функционированием и неопределенностью в данных / С. И. Носков. – Иркутск : РИЦ ГП «Облинформпечать», 1996. – 321 с.

76. Носков, С. И. L–множество в многокритериальной задаче оценивания параметров регрессионных уравнений // Информационные технологии и проблемы математического моделирования сложных систем. – Иркутск, 2004. – № 1. – С. 64 − 71.

77. Носков, С. И. Точечная характеризация множества Парето в линейной многокритериальной задаче // Современные технологии. Системный анализ. Моделирование. – Иркутск, 2008. – № 17. – С. 99 − 102.

78. Носков, С. И., Базилевский М. П. Программный комплекс автоматизации процесса построения регрессионных моделей // Международный журнал прикладных и фундаментальных исследований. – Москва, 2010. – № 1. – С. 93 − 94.

79. Носков, С. И. Лоншаков Р. Л. К оцениванию параметров производственной функции с постоянными пропорциями // Успехи современного естествознания. – 2008. – № 8. С. 118 − 121.

80. Пирогов, Г. Г., Федоровский, Ю. П. Проблемы структурного оценивания и эконометрии / Г. Г. Пирогов, Ю. П. Федоровский. – М. : Статистика, 1979. – 327 с.

81. Поллард, Дж. Справочник по вычислительным методам статистики / Дж. Поллард. – М. : Финансы и статистика, 1985. – 344 с.

82. Просветов, Г. И. Эконометрика: задачи и решения: Учебно-практическое пособие. 5-е изд., доп / Г. И. Просветов. – М. : Альфа-Пресс, 2008. – 192 с.

83. Пугачев, В. С. Теория вероятностей и математическая статистика. / В. С. Пугачев. − М. : Наука, 1979. – 496 с.

84. Рао, С. Р. Линейные статистические методы и их применение / С. Р. Рао –М. : Наука, 1968.– 548 с.

85. Робастность в статистике. Поход на основе функций влияния: Пер. с англ. / Хампель Ф., Рончетти Э., Рауссеу П. – М. : Мир, 1989. – 512 с.

86. Самарский, А. А. Численные методы : учеб. пособие для вузов / А. А. Самарский, А. В. Гулин. − М. : Наука. Гл. ред. физ-мат. лит., 1989. – 432 с.

87. Свидетельство о государственной регистрации программ для ЭВМ № 2017611338 Программный комплекс множественного оценивания регрессионных моделей (ПК МОРМ) / А. В. Баенхаева, М. П. Базилевский, С. И. Носков (Россия); Правообладатель: Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Иркутский государственный университет путей сообщения» (ФГБОУ ВО ИрГУПС); заявка № 2016660373 06.10.2016; дата регистр. 1.02.17

88. Себер, Дж. Линейный регрессионный анализ / Дж. Себер. – М. : Мир, 1980. – 456 с.

89. Сенчагов, В. К. Экономическая безопасность России: Общий курс: Учебник / В.К. Сенчагов . 2-е изд. – М.: Дело, 2005. – 896 с.

90. Селин, В. С. Роль прогнозирования в формировании стратегии регио-

нального развития [Электронный ресурс]: научная электрон. библиотека «киберленинка» / Проблемы прогнозирования /УРАН ИНП РАН, М.– № 6. – 2009 – Режим доступа:http://cyberleninka.ru/article/n/rol-prognozirovaniya-v-formirovanii-strategii-regionalnogo-razvitiya#ixzz4L3QqySfP (дата обращения: 23.сен.2016).

91. Федеральном законе от 28 июня 2014 г. № 172-ФЗ «О стратегическом планировании в Российской Федерации» (с изменениями и дополнениями) [Электронный ресурс]. Дата обновления: 01.11.2016. Доступ из системы ГАРАНТ//ЭПС «Система ГАРАНТ»: ГАРАНТ-Максимум. ВСЯ Россия// НПП «ГАРАНТ-СЕРВИС-УНИВЕРСИТЕТ» Режим доступа: http://base.garant.ru/70684666/#ixzz4LQy3UyqA (дата обращения: 4. дек. 2016).

92. Соколов, А. П. Механизм экономического взаимодействия хозяйствующих субъектов и органов власти в регионе: дис…к. экон. наук: 08.00.05/ А. П. Соколов; ИРЭИ. – Москва, 2007. – 154 с.

93. Совершенствование методов и алгоритмов анализа сложных многофакторных объектов (этап 1) / А. Ю. Колесникова, Е. С. Морозова, В. С. Тимофеев, Е. А. Хайленко // Отчет о НИР / НГТУ. Каф. ТР. – № ГР 02201150649. Новосибирск, 2009. – 101 с.

94. Скороходов, А. В. Справочник по теории вероятностей и математической статистике / В. С. Королюк, Н. И. Портенко, А. В. Скороходов, А. Ф. Турбин. − М. : Наука, 1985. – 640 с.

95. Тимофеев, В. С. Адаптивное оценивание параметров регрессионных моделей с использованием обобщенного лямбда распределения / В. С. Тимофеев, Е. А. Хайленко // Доклады академии наук высшей школы РФ – Новосибирск : НГТУ, 2010. – № 2(15).– С. 25 − 36.

96. Трояновский, В. М. Математическое моделирование в менеджменте: учебное пособие. 2-е изд., испр. и доп. / В. М. Трояновский. − М. : РДЛ, 2000. – 256 с.

97. Фестер, Э. Методы корреляционного и регрессионного анализа / Э. Фестер, Б. Ренц. – М. : Финансы и статистика, 1983. – 293 с.

98. Федеральная служба государственной статистики [Электронный ресурс]: офиц. сайт – Режим доступа: http://www.gks.ru/ (дата обращения: 25.авг.2016).

99. Фишер, Р. А. Статистические методы для исследователей / Р. А. Фишер. − М. : Госстатиздат, 1958. – 268 с.

100. Хайленко, Е. А. Алгоритмы оценивания параметров регрессионных моделей и планирования эксперимента при наличии выбросов и неоднородности распределения ошибок: дис…к. тех. наук: 05.13.17 / Е. А. Хайленко; НГТУ, Новосибирск, 2013. – 185 с.

101. Хьюбер, П. Робастность в статистике / П. Хьюбер. − М. : Мир, 1984. – 303 с.

102. Шамуратов, Н. М. Моделирование и прогнозирование макроструктурных параметров экономики региона (на примере республики Башкортостан): дис…к. экон. наук: 08.00.13, 08.00.05/ Н. М. Шамуратов; БГУ. – Уфа, 2002. – 165 с.

103. Шапкин, А. С. Математические методы и модели исследования операций : учебник. – 4-е издание/ А. С. Шапкин, Н. П. Мазаева. – М. : Дашков и К, 2007. – 400 с.

104. Шимко, П. Д. Оптимальное управление экономическими системами: учеб. пособие / П. Д. Шимко. – СПб. : ИД «Бизнес-пресса», 2004. – 240 с.

105. Экономико-математический энциклопедический словарь / под ред. В. И. Данилов-Данильян. – М. : ИНФРА-М, 2003. – 688 с.

106. Ясин, Е. Г. Теория информации и экономические исследования / Е. Г. Ясин. − М. : Статистика, 1970. –112 с.

107. Wagner, H. M. Linear programming techniques for regression analysis //JASA. -1959. –V.54. -№ 285. – P. 114 – 121.

108. Yu, L., Zeleny, M. The set of all nondoinated solutions in linear cases and multicriteria simplex method // J. of Math. Anal. and Applic. – 1975. – v.49. – № 2. – P. 430 − 468.