

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**НЕЛИНЕЙНАЯ МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И
КОРРЕЛЯЦИЯ В ЭКОНОМИКЕ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 412 группы

направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Белоглазова Ильи Дмитриевича

Научный руководитель

доцент, к. ф.-м. н., доцент

М. Г. Плешаков

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Основное содержание работы	5
ЗАКЛЮЧЕНИЕ	11

ВВЕДЕНИЕ

Современная экономика требует эффективных методов анализа, позволяющих выявлять зависимости между различными экономическими показателями. Регрессионный анализ — один из ключевых инструментов в этом направлении. В настоящей работе рассматриваются как линейные, так и нелинейные модели регрессии, включая методы оценки параметров и анализа качества моделей.

Настоящая работа посвящена исследованию методов регрессионного анализа, как линейных, так и нелинейных, применяемых для анализа экономических данных. Основное внимание уделено теоретическим аспектам и практической реализации моделей множественной регрессии, а также сравнению их применимости и эффективности в экономических задачах.

Цель бакалаврской работы: исследовать особенности и различия линейных и нелинейных моделей множественной регрессии, а также их применение в экономике на примере реальных данных.

Задачи бакалаврской работы:

- Изучение теоретических основ множественной и нелинейной регрессии.
- Исследование методов оценки параметров моделей, включая метод наименьших квадратов и линеаризацию.
- Сравнение линейных и нелинейных моделей по степени пригодности к экономическим данным.
- Реализация и анализ моделей в программных средах Gretl и MS Excel.
- Моделирование сложных зависимостей с использованием языка программирования R, включая полиномиальные, сплайновые и обобщённые аддитивные модели.
- Оценка качества моделей на основе статистических критериев (коэффициенты корреляции и детерминации, t - и F -критерии, ошибка аппроксимации).
- Интерпретация результатов моделирования с учётом экономического смысла переменных.

Данная бакалаврская работа состоит из всего девяти разделов:

В первом разделе рассматриваются теоретические основы регрессии.

Даются определения ключевых понятий: парная и множественная регрессия, метод наименьших квадратов, коэффициенты корреляции и детерминации, гипотезы, гетероскедастичность и др. Представлены формулы множественной линейной регрессии (MLR), условия применимости модели, способы оценки параметров. Далее изучаются различные нелинейные модели — как линеаризуемые, так и внутренне нелинейные (полиномиальные, степенные, показательные, логистические функции). Подробно рассмотрены методы линеаризации и расчёта коэффициентов. Завершается раздел сравнением линейных и нелинейных моделей по форме, методам оценки, сложности интерпретации и области применения.

Во втором разделе проводится практический анализ с использованием прикладных программ. Рассматривается работа в средах Gretl и Microsoft Excel. В Gretl показаны примеры построения моделей, расчёт коэффициентов регрессии, анализ коллинеарности факторов, использование матричного и скалярного подходов. В Excel демонстрируется использование надстройки «Анализ данных»: построение модели, вывод коэффициентов, проверка значимости по t - и F -критериям. Приведён пример интерпретации модели в экономическом контексте. Также выполнено математическое решение задачи построения и анализа гиперболической, степенной и показательной регрессий с расчётом соответствующих статистик (коэффициенты корреляции, детерминации, F -критерий, ошибка аппроксимации).

В третьем разделе рассматривается применение языка R для построения и анализа регрессионных моделей. Используется набор данных о жилье в Бостоне. Исследуются различные подходы: линейная регрессия, полиномиальная, сплайновая регрессия и обобщённые аддитивные модели (GAM). Применяются пакеты `stats`, `ggplot2`, `splines`, `mgcv`. Даются примеры кода, визуализация результатов и сравнение моделей по критериям точности и гибкости. Подчёркиваются широкие возможности R в регрессионном анализе и визуализации.

1 Основное содержание работы

В теоретической части рассматриваются базовые понятия и методы регрессионного анализа. Последовательно излагаются определения ключевых терминов, математические формулы и проводится сравнение линейных и нелинейных моделей.

Даются определения основных понятий, используемых в регрессии и эконометрике:

- **Эконометрика** — наука, изучающая количественные взаимосвязи между экономическими показателями с применением математико-статистических методов.
- **Регрессия** — статистическая зависимость между зависимой переменной и одной или несколькими независимыми переменными.
- **Парная регрессия** — модель с одной зависимой и одной независимой переменной.
- **Множественная регрессия (MLR)** — модель, где зависимая переменная зависит от нескольких факторов.
- **Метод наименьших квадратов (МНК)** — метод оценки параметров модели путём минимизации суммы квадратов ошибок.
- **Коэффициент детерминации (R^2), коллинеарность, остатки, гетероскедастичность, нормальность ошибок** и другие важные термины также подробно объясняются.

Описываются принципы построения моделей на основе статистических данных. В экономике чаще встречаются корреляционные, а не функциональные зависимости, где влияние факторов может быть частичным, косвенным и случайным. Формулируется уравнение условного математического ожидания:

$$\mathbb{E}(Y|X) = f(X)$$

в парном и множественном виде.

Приводится общая формула множественной регрессии:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

Описываются параметры модели, её допущения (линейность, независимость, нормальность ошибок, отсутствие сильной корреляции между факторами) и способы оценки значимости регрессоров.

Подчёркивается роль MLR в прикладном анализе. Модель позволяет прогнозировать поведение экономических показателей при варьировании нескольких факторов. Отмечается важность критической оценки включаемых переменных и ограничений модели.

Подробно разбирается метод оценки параметров модели. Обосновывается переход от теоретического уравнения к эмпирическому:

$$\hat{y} = b_0 + b_1x$$

Выводятся формулы:

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \quad b_0 = \bar{y} - b_1\bar{x}$$

Поясняется принцип минимизации суммы квадратов отклонений между фактическими и расчётными значениями.

Излагаются основные виды моделей, выходящих за рамки линейной зависимости:

- **Линеаризуемые** (полиномиальная, гиперболическая, логарифмическая).
- **Нелинейные по параметрам** (степенная, показательная, логистическая).

Приводятся методы преобразования (например, логарифмирование), чтобы использовать МНК. Отдельно рассматривается коэффициент эластичности, его значение и способы расчёта для различных типов функций.

Представлено сравнение по следующим критериям:

- **Форма зависимости:** линейная модель — прямая, нелинейная — кривая.
- **Методы оценки:** аналитические против численных.
- **Интерпретация коэффициентов:** постоянный эффект против переменного.
- **Области применения:** простые и сложные реальные зависимости.

Делается вывод, что выбор модели должен зависеть от природы данных

и цели анализа.

В следующей части работы рассматривается применение современных прикладных программ для решения задач регрессионного анализа. Основное внимание уделено практической реализации моделей в средах Gretl и Microsoft Excel, а также сравнению методов оценки, проверки гипотез и интерпретации результатов.

Возможности прикладных программ. Описываются функциональные возможности программных средств, таких как Gretl и Excel:

- автоматизация расчётов регрессии и корреляции;
- вычисление t - и F -статистик, доверительных интервалов;
- анализ остатков: графики, нормальность, гетероскедастичность;
- визуализация результатов (графики регрессий, остатков, подгонки);
- построение моделей с несколькими регрессорами и трансформацией переменных;
- импорт/экспорт данных в различных форматах;
- поддержка нелинейной регрессии и моделирования временных рядов.

Указывается, что использование таких инструментов позволяет существенно ускорить процесс анализа и повысить точность и наглядность выводов.

Построение моделей в среде Gretl. Рассматривается практическое применение среды Gretl на примере данных 20 организаций торговли. Используются следующие переменные:

- Y — прибыль, тыс. руб.;
- X_1 — издержки обращения, тыс. руб.;
- X_2 — товарооборот, тыс. руб.

Пошагово показано:

- импорт данных из Excel;
- построение корреляционной матрицы и выявление коллинеарности факторов ($r_{x_1x_2} = 1$);
- построение парных моделей с каждым фактором отдельно;
- применение матричного и скалярного метода МНК;

— интерпретация модели:

$$Y = 0,538 + 1,800X_1 + \varepsilon, \quad R^2 = 0,999$$

Это уравнение показывает высокую точность и значимость влияния фактора X_1 .

Объясняется, что в случае сильной коллинеарности необходимо исключать один из взаимозависимых факторов из модели.

Реализация анализа в Microsoft Excel. Показано, как в Excel активировать надстройку «Анализ данных» и использовать инструмент «Регрессия». Представлен пример построения модели зависимости годового объёма производства от величины основных фондов:

$$\hat{y} = 17,593 + 0,242x, \quad R^2 = 0,853$$

Осуществляется:

- расчёт коэффициентов регрессии и их статистическая оценка (t -статистика, p -значения);
- построение доверительных интервалов;
- проверка значимости модели по F -критерию Фишера;
- экономическая интерпретация коэффициентов.

Отмечается, что Excel позволяет быстро выполнять анализ и наглядно представлять результаты, что делает его удобным инструментом в прикладной экономике.

Математическое решение задачи для нелинейных моделей. Анализируются данные по семи регионам РФ. Строятся три типа моделей:

1. гиперболическая: $y = a + \frac{b}{x}$;
2. степенная: $y = ax^b$;
3. показательная: $y = ab^x$.

Проводится линеаризация моделей с помощью преобразования переменных, после чего применяется метод наименьших квадратов. Для каждой модели рассчитываются:

- коэффициент корреляции;
- коэффициент детерминации (R^2);

- F -критерий Фишера;
- средняя ошибка аппроксимации:

$$A = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Результаты сравниваются, делается вывод о наилучшей модели для описания данных, с учётом статистической значимости и точности аппроксимации.

Также в работе уделено внимание практическому применению языка R для построения и сравнения различных типов регрессионных моделей. В качестве примера используется открытый набор данных о жилье в Бостоне, включающий переменные, характеризующие качество окружающей среды, уровень преступности, удалённость от центра и другие параметры, влияющие на стоимость жилья.

Описание данных и инструментов. Для анализа используется встроенный в пакет MASS набор данных `Boston`. Для построения моделей и визуализации задействованы следующие библиотеки:

- `stats` — базовые статистические функции и регрессии;
- `ggplot2` — графическое представление данных и моделей;
- `splines` — сплайновая регрессия;
- `mgcv` — обобщённые аддитивные модели (GAM).

Линейная регрессия. Строится базовая модель линейной регрессии, в которой цена жилья объясняется, например, уровнем преступности (`crim`) или долей бедных жителей района (`lstat`):

$$\text{medv} = \beta_0 + \beta_1 \cdot \text{lstat} + \varepsilon$$

Модель оценивается с помощью функции `lm()`, проверяется статистическая значимость коэффициентов и анализируются остатки.

Полиномиальная регрессия. Для описания возможной нелинейности зависимости стоимости жилья от фактора `lstat` применяется регрессия второй и третьей степени. Используется конструкция вида:

$$\text{medv} = \beta_0 + \beta_1 \cdot \text{lstat} + \beta_2 \cdot \text{lstat}^2 + \varepsilon$$

Проверяется улучшение качества модели с ростом степени, сравниваются значения R^2 и визуализируются результаты.

Сплайновая регрессия. Для гибкого моделирования нелинейной зависимости используется регрессия на основе сплайнов. Применяется функция `lm()` с включением `bs()` (базисные сплайны) из пакета `splines`. Это позволяет локально аппроксимировать сложные формы зависимости без переобучения модели.

Обобщённые аддитивные модели (GAM). Построение модели осуществляется с помощью функции `gam()` из пакета `mgcv`, что позволяет включать в модель нелинейные сглаженные функции переменных:

$$\text{medv} = \beta_0 + s(\text{lstat}) + \varepsilon$$

Здесь $s(\cdot)$ обозначает гладкую функцию, автоматически определяемую по данным. Модель оценивается, строится график зависимости, производится сравнение с линейной и полиномиальной моделями.

Сравнение моделей. Все построенные модели сравниваются по следующим критериям:

- визуальное совпадение с данными;
- значение коэффициента детерминации R^2 ;
- остаточные графики и признаки переобучения;
- гибкость и интерпретируемость.

На основании анализа делается вывод, что обобщённые аддитивные модели обладают наибольшей гибкостью и способны наиболее точно аппроксимировать реальные данные, особенно при наличии ярко выраженной нелинейности. Однако для их корректного применения требуется больше вычислительных ресурсов и статистической грамотности.

Отмечается, что язык R является подходящим инструментом для регрессионного анализа, обеспечивая широкие возможности по моделированию, оценке и визуализации экономических зависимостей.

ЗАКЛЮЧЕНИЕ

В результате проведённого исследования была всесторонне рассмотрена тема регрессионного анализа как с теоретической, так и с прикладной стороны. На первом этапе подробно изучены математические и статистические основы метода: дана классификация линейных и нелинейных моделей, изложены методы оценки параметров, в том числе метод наименьших квадратов, и рассмотрены критерии качества аппроксимации. Особое внимание уделено различиям между типами моделей и обоснованию выбора той или иной формы зависимости в зависимости от структуры данных.

На втором этапе проанализированы возможности программных средств Gretl и Microsoft Excel для построения и оценки регрессионных моделей. В среде Gretl реализованы матричный и скалярный подходы к построению моделей, проведён анализ коллинеарности и дана интерпретация коэффициентов. Работа в Excel показала простоту применения встроенных инструментов и удобство визуализации результатов. Также был выполнен математический расчёт параметров для нелинейных моделей, что подтвердило применимость аналитических подходов при отсутствии специализированного программного обеспечения.

Заключительный этап был посвящён моделированию в программной среде R. Используя широкий набор инструментов и пакетов, реализованы и сравнены модели линейной, полиномиальной, сплайновой регрессии и обобщённых аддитивных моделей (GAM). Проведённая работа продемонстрировала преимущества R в плане гибкости настройки, точности построения и глубины визуального анализа, что особенно важно при работе с реальными, сложными и нелинейными данными.

Таким образом, проделанная работа показала, что сочетание строгой теоретической базы с практикой программной реализации и математического анализа позволяет эффективно решать широкий спектр прикладных задач. Такой подход особенно ценен для применения в экономике, где модели часто требуют как высокой точности, так и интерпретируемости.