В. И. МЕНЬЩИКОВА

ЭКОНОМЕТРИКА

Тамбов
Издательский центр ФГБОУ ВО «ТГТУ»
2024

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Тамбовский государственный технический университет»

В. И. МЕНЬЩИКОВА

ЭКОНОМЕТРИКА

Утверждено Ученым советом университета
в качестве учебного пособия для студентов 2 и 3 курсов
направления подготовки 38.01.03 «Экономика»
очной и очно-заочной форм обучения

Учебное электронное издание

Тамбов
Издательский центр ФГБОУ ВО «ТГТУ»
2024

УДК 336
ББК 65.26
М51

Рецензенты:

Кандидат экономических наук, доцент,
доцент кафедры «Коммерция и бизнес-информатика» ФГБОУ ВО «ТГТУ»
Н. В. Дюженкова

Кандидат экономических наук, доцент, заведующий кафедрой
экономики и менеджмента ФГБОУ ВО «РАНХиГС»
при Президенте Российской Федерации
Т. Н. Харламова

Меньщикова, В. И.

М51 Эконометрика [Электронный ресурс] : учебное пособие / В. И. Меньщикова. – Тамбов : Издательский центр ФГБОУ ВО «ТГТУ», 2024. – 1 электрон. опт. диск (CD-ROM). – Системные требования : ПК не ниже класса Pentium II ; CD-ROM-дисковод ; 2,4 Mb ; RAM ; Windows 95/98/XP ; мышь. – Загл. с экрана.

ISBN 978-5-8265-2846-4

Представлены основы эконометрики, начиная от базовых понятий и заканчивая более сложными методами анализа данных. Представлены основы построения и оценки регрессионных моделей, анализа временных рядов и др. Особое внимание уделено практическим аспектам применения эконометрических методов в реальных исследованиях.

Предназначено для студентов 2 и 3 курсов направления подготовки 38.01.03 «Экономика» очной и очно-заочной форм обучения.

УДК 336
ББК 65.26

ISBN 978-5-8265-2846-4

ВВЕДЕНИЕ

Эконометрика – это «наука о применении статистических методов для анализа экономических данных и построения моделей, позволяющих лучше понять экономические процессы и явления. Она объединяет в себе элементы экономики, статистики и математики, предоставляя инструменты для количественного исследования сложных взаимосвязей между различными экономическими переменными».[1]

В данной работе рассмотрены основы эконометрики, начиная от базовых понятий и заканчивая более сложными методами анализа данных. Мы изучим, как строить и оценивать регрессионные модели, анализировать временные ряды, проводить тесты гипотез и многое другое. Особое внимание будет уделено практическим аспектам применения эконометрических методов в реальных исследованиях.

Книга предназначена для студентов экономических специальностей, а также для всех тех, кто интересуется применением математических и статистических инструментов в экономике. Независимо от уровня подготовки, здесь можно найти полезную информацию, которая поможет глубже понять принципы работы эконометрии и научиться применять ее методы на практике.

[1] Скляров, Ю. С. Эконометрика. Краткий курс : учебное пособие / Ю. С. Скляров. – 2-е изд., испр. – СПб. : ГУАП, 2007. – 140 с.

1. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Сегодня эконометрика занимает важное место в системе экономического образования, так как «объединяет множество теоретических выводов, подходов, методов и моделей, а также статистических и математических инструментов, необходимых для измерения социально-экономических процессов и явлений».[2]

Считается, что классическое определение эконометрики дано основателем этой науки Р. Фриш в 1926 г. Фриш Р. писал, что «*эконометрика* – это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек – статистика, экономическая теория и математика – необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это единство всех трех составляющих. И это единство образует эконометрику».[3]

Один из российских ученых, занимающихся эконометрическими измерениями, – Ю. С. Скляров дал несколько упрощенное по своей конструкции определение эконометрики. Он писал, что «*эконометрика* – это наука, которая изучает количественное выражение взаимосвязей экономических объектов и процессов, используя и разрабатывая для этой цели специфические математико-статистические методы».[4]

Основная цель эконометрики заключается в создании и предложении исследователям эффективного инструмента для прогнозирования поведения экономического объекта в разнообразных ситуациях. Это предполагает использование количественных методов и моделей для анализа исторических данных, что, в свою очередь, позволяет выявлять устойчивые закономерности и тенденции в экономических процессах.

Эконометрика предоставляет возможность не только понимать, как объект вел себя в прошлом, но и предсказывать его поведение в будущем при различных условиях. Это особенно важно в условиях нестабильной экономики, где

[2] Скляров, Ю. С. Эконометрика. Краткий курс : учебное пособие / Ю. С. Скляров. – 2-е изд., испр. – СПб. : ГУАП, 2007. – 140 с.

[3] Frisch R. Editorial. Econometrica. – 1933. – № 1.

[4] Скляров, Ю. С. Эконометрика. Краткий курс : учебное пособие / Ю. С. Скляров. – 2-е изд., испр. – СПб. : ГУАП, 2007. – 140 с.

изменения могут происходить стремительно и непредсказуемо. Например, наличие точных прогнозов может помочь компании адаптировать свою стратегию, минимизировать риски и максимизировать прибыль.

На базе полученных прогнозов можно решать широкий спектр практических задач. Прогнозирование может служить основой для оптимального управления экономическими объектами, будь то предприятия, отрасли или даже целые государства. Это включает в себя принятие решений о ресурсах, необходимых для достижения заявленных целей, управлении запасами, планировании бюджета и других аспектах охвата управления.

Выбор стратегии поведения на рынке также является важным направлением, где эконометрика играет ключевую роль. С использованием эконометрических данных исследователи могут определить, какая стратегия будет наилучшей для достижения конкурентных преимуществ. Это может включать в себя анализ ценовой эластичности спроса, оценку влияния маркетинговых кампаний или изучение реакции потребителей на изменения в характеристиках продукции.

К тому же, эконометрика помогает в выведении оптимальных решений на основе комплексного анализа данных. Это может подразумевать не только выбор правильной стратегии, но и помощь в понижении издержек и улучшении качества. Предоставляя богатую информацию и статистическую обоснованность для принимаемых решений, эконометрика позволяет организовать и оптимизировать процессы внутри объекта, адаптируя его к изменяющимся условиям внешней среды.

Таким образом, сочетая теорию и практические навыки, эконометрика становится мощным инструментом для исследователей и практиков, которые стремятся использовать данные для принятия обоснованных и эффективных решений, что в конечном итоге ведет к улучшению общей эффективности и результативности экономических объектов.

Основная задача эконометрики заключается в обнаружении связей между количественными показателями экономических объектов для создания математических моделей, способных предсказывать (приближенно рассчитывать) ненаблюдаемые характеристики этих объектов на основе известных или заданных значений других характеристик. Как отмечает нобелевский лауреат Л. Р. Клейн, «основная задача эконометрики – наполнить эмпирическим содержанием априорные экономические рассуждения».[5] Иными словами, именно эконометрика переводит на количественный уровень основные результаты общей экономической теории, носящих качественный характер.

[5] Klein L. R. Economic Theory and Econometrics, 1985.

В целом, эконометрика решает множество задач, связанных с количественным и качественным анализом экономических взаимосвязей, не только предоставляя инструменты для экономических измерений, но и методологию оценки параметров экономических моделей. Это позволяет использовать эконометрические модели для прогнозирования экономических явлений и процессов как в масштабах экономики в целом, так и на уровне отдельных предприятий.

Базовые понятия эконометрики включают в себя ключевые термины, такие как «объект», «переменная» и «модель». Эти понятия служат основой для построения и применения эконометрических исследований, обеспечивая понимание и интерпретацию экономических явлений.

Экономический **объект** представляет собой любую хозяйствующую единицу, которая участвует в производственной, потребительской или другой экономической деятельности. Это может быть как индивидуальное предприятие, так и целый сектор экономики или даже государство. Объект в эконометрике рассматривается как источник данных, которые помогут исследователям и аналитикам выявить закономерности и взаимосвязи в экономических процессах.

Переменная – это количественная характеристика объекта, обладающая способностью принимать различные значения в процессе хозяйственной деятельности. Переменные могут быть независимыми (которые влияют на другие) и зависимыми (которые находятся под воздействием других переменных). Например, переменными могут быть показатели дохода, расходы, объем производства и уровень безработицы. Анализ переменных позволяет выявить тенденции, зависимости и корреляции в экономических данных, тем самым углубляя понимание взаимодействия между различными экономическими факторами.

Модель в эконометрике представляет собой математически выраженную связь между переменными объекта. Эта модель может быть представлена в разных форматах, включая набор графиков или таблиц, а также в виде системы математических уравнений и неравенств. Моделирование позволяет формализовать изучение взаимосвязей и эффективно анализировать данные, связывая воедино все переменные объекта. Модели могут быть как простыми, так и сложными, включая линейные и нелинейные зависимости, что позволяет проводить более детальный анализ.

В основе эконометрики лежат несколько ключевых областей знаний, осуществляющих поддержку и структуру для анализа данных.

– Экономическая теория предоставляет концептуальные рамки и принципы, позволяющие формулировать гипотезы и предсказания о поведении экономических агентов. Эти теоретические основоположения служат основой для построения эконометрических моделей и анализа.

– Социально-экономическая статистика основывается на сборе, обработке и интерпретации данных, связанных с экономикой и обществом. Она предоставляет необходимую информацию для анализа и помогает верифицировать модели, созданные на основе экономической теории.

– Теория вероятностей и математическая статистика формируют статистическую основу для анализа данных, интерпретации результатов и оценки надежности моделей. Эти методы применяются для обработки случайных данных, что позволяет учитывать неопределенности и возможности разных сценариев в экономических исследованиях.

Таким образом, понимание этих базовых понятий и основ эконометрики позволяет более глубоко анализировать экономические процессы, разрабатывать прогнозы и принимать обоснованные решения на основе полученных данных. Эконометрика как наука становится важным инструментом в исследованиях, финансах и управлении, помогая выявлять и устранять проблемы, возникающие в современных экономических системах.

Схематично проведение эконометрического исследования можно представить следующим образом (рис. 1).

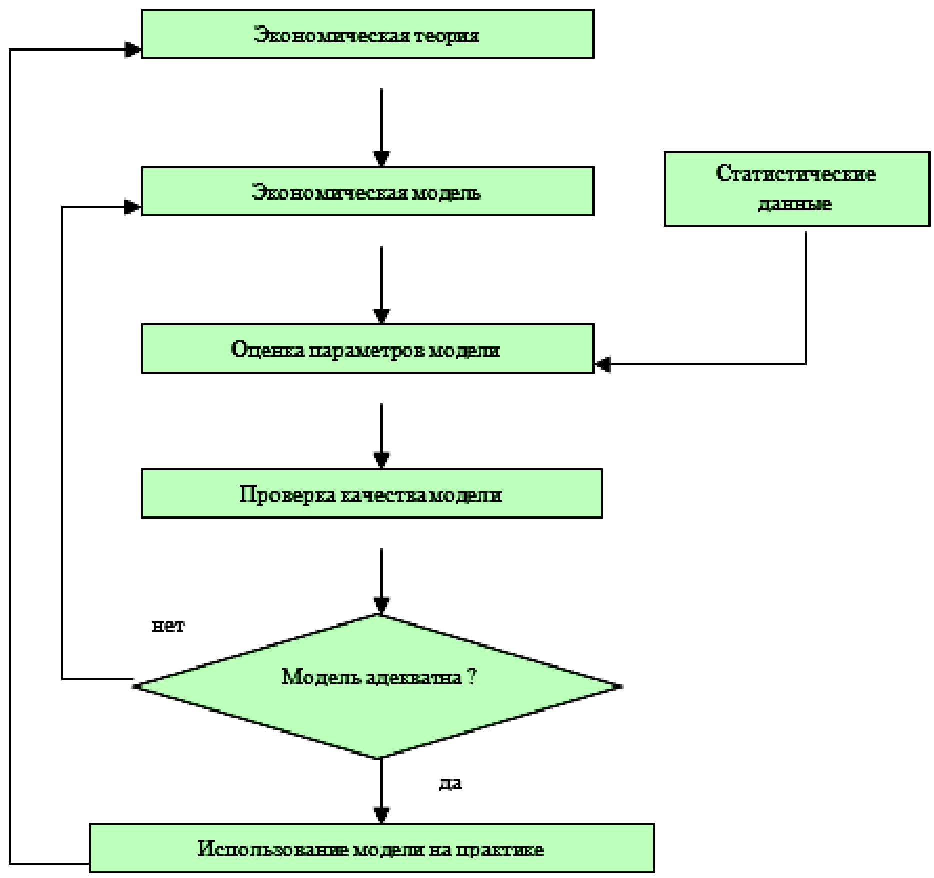


Рис. 1. Схема проведения эконометрического исследования

Последовательность этапов эконометрического исследования включает несколько ключевых шагов, каждый из которых важен для получения достоверных результатов. Вот примерная последовательность этих этапов[6]:

1. Постановка задачи и формулировка гипотезы

На этом этапе исследователь определяет цель своего исследования и формулирует гипотезу, которую он хочет проверить. Например, целью может быть изучение влияния образования на уровень доходов населения, а гипотеза может звучать так: «Уровень образования положительно влияет на доходы».

2. Сбор данных

Следующий шаг – сбор необходимых данных. Это могут быть данные из различных источников: официальные статистические отчеты, опросы, эксперименты и т.д. Важно убедиться, что собранные данные являются надежными и репрезентативными.

3. Предварительный анализ данных

После сбора данных проводится их предварительный анализ. На этом этапе проверяется качество данных, выявляются аномалии и пропуски, рассчитываются базовые статистические показатели (среднее значение, дисперсия, корреляция и др.). Также может проводиться визуализация данных для лучшего понимания их структуры.

4. Выбор модели

На основе предварительного анализа данных выбирается подходящая модель для описания исследуемого процесса. Чаще всего используются линейные регрессионные модели, но могут применяться и другие типы моделей (логит-, пробит-модели, модели временных рядов и т.п.).

5. Оценка параметров модели

После выбора модели необходимо оценить ее параметры. «Для этого обычно используется метод наименьших квадратов (МНК), который позволяет найти такие значения параметров, при которых сумма квадратов отклонений наблюдаемых значений от предсказанных моделью минимальна»[7].

6. Проверка адекватности модели

Проверка адекватности модели включает в себя оценку ее способности объяснять вариацию зависимой переменной. Для оценки качества модели используются различные критерии, такие как коэффициент детерминации (R^2), F-тест, t-тесты для коэффициентов модели и др.

---

[6] Мхитарян, В. С. Эконометрика : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин // Цифровой образовательный ресурс IPR SMART : [сайт]. – М. : Евразийский открытый институт, 2012. – 224 с. – URL : https://www.iprbookshop.ru/11125.html

[7] Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.

7. Анализ остатков

Анализ остатков (разницы между наблюдаемыми значениями и предсказанными моделью) помогает выявить возможные проблемы с моделью, такие как гетероскедастичность, автокорреляция или наличие выбросов. Если обнаружены значительные отклонения, возможно, потребуется корректировка модели.

8. Интерпретация результатов

На последнем этапе интерпретируются результаты моделирования. Исследователь делает выводы относительно того, подтверждена ли исходная гипотеза, насколько точно модель описывает реальные данные и какие факторы оказывают наибольшее влияние на исследуемую переменную.

Эта последовательность является общей схемой эконометрического исследования. В зависимости от специфики задачи и используемых данных некоторые этапы могут быть изменены или дополнены другими процедурами.

«Методы корреляционного и регрессионного анализа являются основными инструментами математической статистики, применяемыми для разработки эконометрических моделей. Целью **корреляционного анализа** является проверка наличия и значимости линейной связи между переменными, при этом они не делятся на зависимые и независимые. Для ответа на этот вопрос используются коэффициенты корреляции».[8]

Регрессионный анализ заключается «в представлении исследуемой зависимости в форме аналитического уравнения, где предварительно выделяются зависимая и независимые (объясняющие) переменные».[9]

Регрессионный анализ предназначен для решения ряда задач:

– определить переменные, которые влияют на поведение других величин, и возможность их использования в качестве объясняющих переменных;

– определить математически формулы зависимости между переменными, а также экономическую интерпретацию включаемых в модель коэффициентов;

– построить уравнение регрессии, которое будет отражать форму, вид и экономический смысл зависимостей между переменными;

– проверить свойства построенного уравнения регрессии: типичность, значимость, существенность.

---

[8] Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.

[9] Там же.

ТЕСТ

1. В чем состоит основное отличие эконометрики от экономической статистики согласно Р. Фришу?

а) эконометрика включает в себя статистику;

б) эконометрика основана на математиках;

в) эконометрика использует только качественные данные;

г) они идентичны.

2. Какова основная цель эконометрики?

а) изучение качественных характеристик;

б) прогнозирование поведения экономического объекта;

в) разработка экономической теории;

г) сбор статистических данных.

3. Эконометрика – это:

а) связная наука о статистике;

б) наука о математических моделях;

в) наука о количественных взаимосвязях;

г) научный метод статистических измерений.

4. Какой метод обычно используется для оценки параметров модели в эконометрике?

а) метод максимального правдоподобия;

б) метод корреляции;

в) метод линейной регрессии;

г) метод наименьших квадратов.

5. Какой элемент эконометрики подразумевает количественную характеристику объекта?

а) модель;

б) объект;

в) переменная;

г) гипотеза.

6. Какой этап не является частью последовательности эконометрического исследования?

а) создание таблицы;

б) сбор данных;

в) анализ остатков;

г) выбор модели.

7. Какой параметр используется для оценки качества модели в эконометрике?

а) стандартное отклонение;

б) корреляция;

в) среднее значение;

г) коэффициент детерминации.

8. Что изучает эконометрика в контексте количественного анализа?

а) только математику;

б) взаимосвязи между экономическими объектами;

в) только экономическую теорию;

г) только статистику.

9. Какой шаг следует после сбора данных в эконометрическом исследовании?

а) оценка параметров модели;

б) выбор модели;

в) предварительный анализ данных;

г) постановка задачи.

10. Какова основная задача эконометрики?

а) обнаружение связей между количественными показателями;

б) создание графиков и таблиц;

в) сбор статистических данных;

г) разработка экономических теорий.

11. Что такое экономический объект в контексте эконометрики?

а) финансовый инструмент;

б) качественная характеристика процесса;

в) хозяйствующая единица;

г) демографическая группа.

12. Какое из следующих понятий является базовым в эконометрике?

а) модель;

б) рынок;

в) статистика;

г) маржинальность.

13. Что такое «переменная» в эконометрике?

а) количество товаров;

б) количество, принимающее различные значения;

в) качественная характеристика объекта;

г) хозяйствующая единица.

14. Как можно представить модель в эконометрике?

а) только в текстовом формате;

б) только в виде таблиц;

в) только в виде графиков;

г) в виде системы математических уравнений и неравенств.

15. Что именно связывает модель в эконометрике?

а) теоретические предположения;

б) демографические данные;

в) финансовые и административные показатели;

г) переменные объекта.

2. ПРОСТАЯ ЛИНЕЙНАЯ МОДЕЛЬ РЕГРЕССИИ

Ключевой задачей эконометрики является исследование объективно существующих связей между явлениями. Эти связи, учитывая их разнообразие, классифицируют по различным критериям.

По роли признаков в изучении взаимосвязей принято их подразделять на два класса:

Факторные признаки (факторы) – это те признаки, которые вызывают изменения других признаков, связанных с ними. **Результативные признаки (результат)** – такие признаки, которые изменяются под влиянием факторных признаков.[10]

В эконометрических и статистических исследованиях выделяют две ключевых формы взаимосвязи между факторным и результативным признаком:[11]

1) **функциональная (полная) связь:** каждому значению факторного признака соответствует одно и только одно значение результативного признака:[12]

$$y = f(x),$$

где y – зависимая переменная (результативный признак); x – независимая, объясняющая переменная (факторный признак).

2) **стохастическая (неполная) связь:** одному значению зависимой переменной может соответствовать несколько возможных значений независимой переменной[13]:

$$y = \acute{y}_x + \varepsilon,$$

где y – фактическое значение результативного признака; $\acute{y}_x$ – теоретическое значение результативного признака, найденное исходя из соответствующей математической функции связи y и x, т.е. из уравнения регрессии; ε – случайная величина, характеризующая отклонение результативного признака от теоретического, найденного по уравнению регрессии.

---

[10] Новиков, А. И. Эконометрика: учебное пособие / А. И. Новиков. – 2-е изд. – М. : Дашков и К, 2019. – 224 с.

[11] Там же.

[12] Там же.

[13] Там же.

Случайная величина (ε) характеризует в построенной модели влияние факторов, которые мы могли не включить в модель, а также различных ошибок и специфику измерений переменных. Случайная величина (ε) – это возмущение[14].

Связи между признаками могут различаться *по направлению и аналитической форме*.

Связи могут быть классифицированы по направлению на прямые и обратные, а также на положительные и отрицательные.

Прямая связь означает, что при изменении значений фактора соответствующим образом изменяются и значения результата. К примеру, увеличение уровня автоматизации труда способствует росту рентабельности в промышленном производстве.

В отличие от этого, **обратная связь** наблюдается, когда изменение фактора вызывает противоположное изменение результата. Например, рост фондоотдачи чаще всего приводит к снижению себестоимости продукции.

Что касается *аналитической формы*, связи могут быть линейными и нелинейными.

Линейная связь наблюдается, когда зависимость между признаками приблизительно выражается уравнением прямой линии.

Нелинейная, или криволинейная связь, возникает, когда эта зависимость лучше описывается уравнением кривой (например, параболы, гиперболы, степенного или экспоненциального закона). Поэтому от правильной спецификации модели зависит надежность построенных моделей.

При построении эконометрических моделей мы должны учитывать возможные ошибки. Принято выделять ошибки спецификации, выборки, измерения.

«*Ошибки спецификации* включают не только неверный выбор математической функции для описания зависимости, но и игнорирование существенных факторов в уравнении регрессии, например, использование парной регрессии вместо множественной».[15] Ключевыми примерами здесь служат практически все показатели, характеризующие уровень жизни населения. Например, при построении модели влияния факторов на величину спроса изначальное предположение о том, что спрос зависит от цены, подтверждается, но будут и другие факторы, очень сильно влияющие на величину спроса – доход на душу населения. Ошибки спецификации можно уменьшить, изменив модель.

[14] Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков. – 2-е изд. – М. : Дашков и К, 2019. – 224 с.

[15] Там же.

Ошибки выборки связаны с неоднородностью данных в исходной статистической совокупности. Исследователи зачастую для анализа взаимосвязей между признаками используют именно выборочные данные, которые отбирали изначально, полагаясь на свою интуицию и опыт. Здесь важно помнить о том, что чем неоднороднее совокупность, тем бессмысленнее уравнение регрессии. «Для улучшения результатов исследования обычно исключаются наблюдения с аномальными значениями признаков, что позволяет рассматривать результаты регрессии как выборочные характеристики».[16] Особенно внимательными надо быть при построении моделей временных рядов, где выборка представлена хронологическими датами. Ошибки выборки можно снизить, увеличив объем данных.

Ошибки измерения вызывают наиболее сложные проблемы для практического использования эконометрических моделей. Этот тип ошибок значительно усложняет оценку взаимосвязей между признаками.[17] Однако, ошибки измерения нивелировать полностью невозможно. но через спецификацию модели можно минимизировать данный тип ошибок.

Спецификация модели – это выбор соответствующей функциональной формы и переменных для отражения связей между факторными и результативными признаками.

Например, для выбора вида математической функции $y = f(x)$ в парной регрессии используются такие методы, как: графический; аналитический метод; экспериментальный.

Зависимости $\hat{y} = f(x)$ соответствует некоторая кривая на плоскости и по форме облака наблюдений (поля корреляции) можно определить вид регрессионной функции (рис. 2 – 5).

Когда «результативный и факторный признаки увеличиваются равномерно, примерно в арифметической прогрессии, это говорит о наличии линейной связи между ними, а при обратной связи – гиперболической. Если результативный признак растет в арифметической прогрессии, а факторный – значительно быстрее, применяются параболическая или степенная функции».[18]

Линейная регрессия широко применяется в эконометрике благодаря ясному экономическому толкованию ее параметров.

$$y = a + bx + \varepsilon.$$

---

[16] Орлов, А. И. Эконометрика : учебное пособие / А. И. Орлов // Электронно-библиотечная система IPR BOOKS: [сайт]. – 3-е изд. – М., Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. – 676 с. – URL : https://www. iprbookshop.ru/89481.html

[17] Тихомиров, Н. П. Эконометрика : учебник / Н. П. Тихомиров, Е. Ю. Дорохина. – 2-е изд., стереотип. – М. : Изд-во «Экзамен», 2007. – 512 с.

[18] Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков. – 2-е изд. – М. : Дашков и К, 2019. – 224 с.

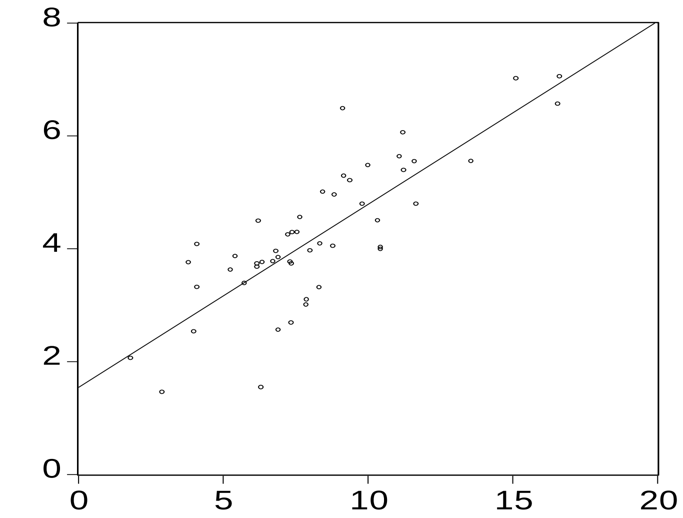


Рис. 2. Поле корреляции и тип функциональной зависимости
(прямая $y = a + bx + \varepsilon$)

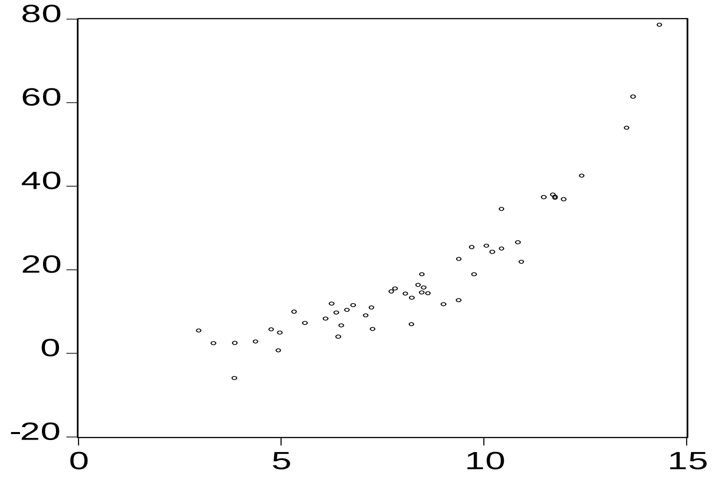


Рис. 3. Поле корреляции и тип функциональной зависимости
(степенная $y = x^n\varepsilon$)

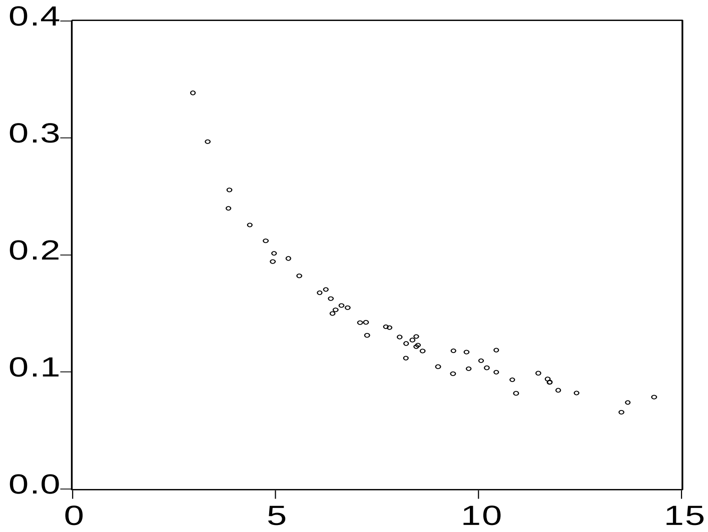


Рис. 4. Поле корреляции и тип функциональной зависимости
(гиперболическая $y = b + a/x + \varepsilon$)

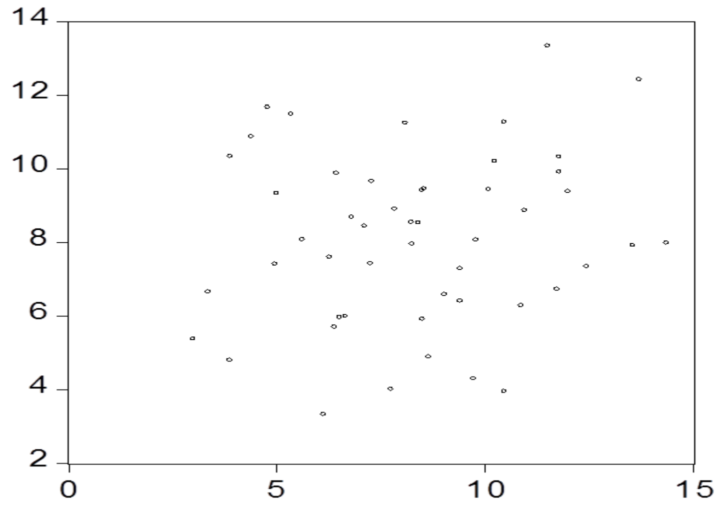


Рис. 5. Поле корреляции при отсутствии зависимости между *x* и *y*

Чтобы формализовать задачу, рассмотрим разность между расчетными (теоретическими) и наблюдаемыми значениями *y*:

$$\varepsilon = y - \acute{y}.$$

Лучшей считается та зависимость, при которой сумма квадратов отклонений минимальна, т.е.:

$$S = \sum(y - \acute{y})^2 \to \min \quad \text{или} \quad \sum\varepsilon^2 \to \min;$$

$$S = \sum(y - \acute{y})^2 = \sum(y - a - bx)^2.$$

Традиционно для оценки параметров линейной регрессии используют **метод наименьших квадратов** (МНК).[19] Его суть заключается в том, что за счет минимизации суммы квадратов отклонений фактических значений зависимой переменной *y* от теоретически рассчитанных значений $\acute{y}$ можно найти значения параметров *a* и *b*.

Для определения минимума функции следует рассчитать частные производные по каждому из параметров *a* и *b* и приравнять их к нулю:

$$S'_a = -2\sum y + 2na + 2b\sum x = 0;$$

$$S'_b = -2\sum yx + 2a\sum x + 2b\sum x^2 = 0.$$

Из этого следует, что система нормальных уравнений примет вид:

$$\begin{cases} na + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum yx. \end{cases}$$

[19] Яковлева, А. В. Эконометрика : учебное пособие / А. В. Яковлева // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – Саратов : Научная книга, 2019. – 223 с. – URL : https://www.iprbookshop.ru/81090.html

При решении данной системы уравнений получаем следующие формулы для нахождения параметров a и b:

$$a = \frac{\sum y}{n} - \frac{b\sum x}{n} = \bar{y} - bx;$$

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\operatorname{cov}(x,\ y)}{\sigma_x^2},$$

где $\operatorname{cov}(x, y)$ – ковариация признаков; σ_x^2 – дисперсия признака x.

Значение параметра b отражает среднее изменение зависимой переменной при изменении независимого признака на одну единицу. Значение параметра b – это *коэффициент регрессии*.

Например, уравнение затрат предприятия $y = 4000 + 12x$ можно интерпретировать следующим образом. Если объем производства вырастет на одну единицу, то затраты вырастут на 12 единиц.

Параметр a в уравнении не всегда подлежит буквальному объяснению, как правило можно объяснить знак при данном параметре. Если $a > 0$, это указывает на то, что относительное изменение зависимой переменной происходит медленнее по сравнению с изменением независимого фактора. В этом случае коэффициент вариации (отношение стандартного отклонения к среднему) для факторного признака будет выше, чем для результативного показателя. Это также подразумевает, что изменения в факторе оказывают менее выраженное влияние на результат, чем само изменение фактора.

Корреляция представляет собой «статистическую зависимость между случайными величинами, которая не носит строгого функционального характера. Это значит, что изменение одной из случайных величин приводит к изменению ожидаемого значения другой».[20]

Каждое уравнение регрессии дополняется показателем силы связи. При линейной регрессии вычисляется *линейный коэффициент корреляции*:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = b\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x,\ y)}{\sigma_x \cdot \sigma_y}$$

[20] Скляров, Ю. С. Эконометрика. Краткий курс : учебное пособие / Ю. С. Скляров. – 2-е изд., испр. – СПб. : ГУАП, 2007. – 140 с.

или

$$r = \frac{\sum yx - \frac{\sum y \sum x}{n}}{\sqrt{\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right]}}.$$

Значение линейного коэффициента корреляции варьируется следующим образом:

$$-1 \le r \le 1.$$

Знаки коэффициентов корреляции и регрессии совпадают. Трактовку значений коэффициента корреляции представим в табл. 1.

Для определения *тесноты связи* между факторным и результативным признаками принято пользоваться шкалой Чеддока (табл. 2).

1. Оценка линейного коэффициента корреляции[21]

Значение линейного коэффициента связи	Характер связи	Интерпретация связи
$r = 0$	Отсутствует	–
$0 < r < 1$	Прямая	С увеличением x увеличивает y
$-1 < r < 0$	Обратная	С увеличением x уменьшается y и наоборот
$r = 1$	Функциональная	Каждому значению факторного признака строго соответствует одно значение результативного

2. Шкала Чеддока[22]

Показатели тесноты связи	0,1…0,3	0,3…0,5	0,5…0,7	0,7…0,9	0,9…0,99
Характеристика силы связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

[21] Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков. – 2-е изд. – М. : Дашков и К, 2019. – 224 с.

[22] Скляров, Ю. С. Эконометрика. Краткий курс: учебное пособие / Ю. С. Скляров. – 2-е изд., испр. – СПб. : ГУАП, 2007. – 140 с.

Для оценки точности подбора линейной функции вычисляется квадрат линейного коэффициента корреляции r_{yx}^2, который называется коэффициентом детерминации. Этот коэффициент показывает, какая доля дисперсии зависимой переменной y объясняется регрессионной моделью в рамках общей вариации. Например, если $r_{yx}^2 = 0{,}988$, это означает, что уравнение регрессии объясняет 98,8% дисперсии зависимого признака, в то время как 1,2% приходятся на влияние других факторов (остаточная дисперсия). Более высокий коэффициент детерминации свидетельствует о большей надежности полученного уравнения регрессии.

Предпосылки регрессионного анализа. Теорема Гаусса–Маркова. Как было отмечено ранее, важно качество построения эконометрических моделей. Для достижения наилучших результатов применения методики регрессионного анализа на основе МНК необходимо выполнение теоремы Гаусса–Маркова. Речь идет о единовременном соблюдении четырех условий: «математическое ожидание случайной ошибки для любого наблюдения должно равняться нулю, т.е.: $\sigma_\varepsilon = 0$; дисперсия случайной ошибки должна оставаться постоянной для всех наблюдений; случайные ошибки должны быть статистически независимыми (некоррелированными) друг от друга; объясняющая переменная x должна быть фиксированной, т.е. не случайной величиной».[23]

Когда соблюдаются условия Гаусса–Маркова, модель можно классифицировать как *классическую нормальную линейную регрессионную модель*. В дополнение к этому, часто делается предположение о том, что случайная ошибка распределена нормально. Одним из значимых моментов является то, что в данном контексте условие о некоррелированности случайных ошибок также фактически является эквивалентом их независимости. Это важное различие следует учитывать при работе с данными.

Теперь давайте подробнее остановимся на каждом из этих условий, поскольку они играют ключевую роль в построении надежной регрессионной модели.

Первое условие касается среднего значения случайной ошибки, которое должно равняться нулю. Это предполагает, что в данных отсутствует систематическое смещение, т.е. ошибки не имеют предвзятости в сторону увеличения или уменьшения значений. Важно отметить, что если в регрессионное уравнение включен свободный член, это условие автоматически соблюдается.

[23] Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.

Свободный член служит корректировочным элементом, который позволяет модели компенсировать любые систематические ошибки, что делает ее более точной и сбалансированной.

Следующее условие предполагает наличие постоянной дисперсии ошибок, что известно как гомоскедастичность. Это означает, что разброс ошибок остается постоянным – не меняется в зависимости от значений независимых переменных. Дисперсия случайной ошибки отражает потенциальные вариации ошибки до того, как был осуществлен отбор выборки. Эта величина, как правило, неизвестна, и одной из главных задач регрессионного анализа является ее оценка. Существуют различные методики для анализа и оценки дисперсии ошибок, которые помогают выявить, насколько значительными могут быть эти вариации.

Когда постоянное значение дисперсии не наблюдается, возникает явление, известное как *гетероскедастичность*. В этом случае эффективность оценок коэффициентов регрессии значительно снижается, даже если они все еще остаются несмещенными. То есть, несмотря на то, что оценки правильны в среднем, они могут иметь большую вариативность, что ведет к ухудшению предсказательной способности модели.

В практике существует множество методов для диагностики и смягчения последствий гетероскедастичности. К таким методам относятся визуализирующие тесты, такие как графики остатков, а также статистические тесты, например, тест Бройша–Пагана или тест Уайта. Если после диагностики гетероскедастичности обнаруживается, необходимо принимать меры по ее устранению, такие как использование методов весов, которые помогают корректировать оценки и вернуть модели нормальные параметры.

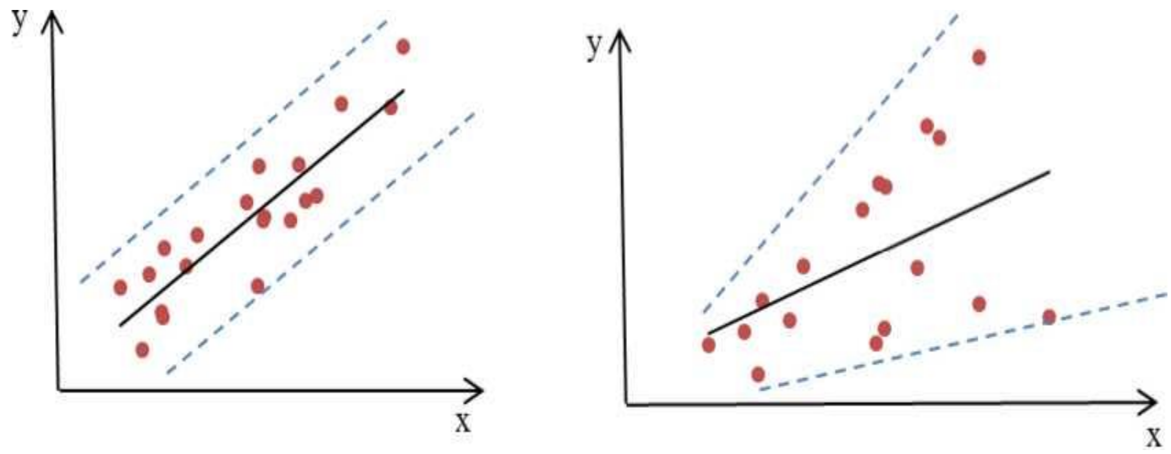
Таким образом, соблюдение условий Гаусса–Маркова, включая среднее значение ошибок и постоянную дисперсию, играет критически важную роль в построении и интерпретации линейной регрессионной модели. Это не только делает модели более надежными, но и обеспечивает правильные выводы, основанные на полученных данных.

Характерные диаграммы рассеяния гомоскедастичности и гетероскедастичности показаны на рис. 6, *а* и *б* соответственно.

Условие независимости случайных ошибок в регрессионной модели обозначает отсутствие корреляции между ошибками, возникающими при различных наблюдениях. Это значит, что каждая ошибка должна быть независимой от других, что является ключевым аспектом для обеспечения точности и надежности получаемых оценок. Однако важно отметить, что это требование нередко оказывается нарушенным в анализе данных временных рядов. В таких данных, особенно когда рассматриваются временные последовательности, вероятно появление автокорреляции остатков.

Автокорреляция означает, что ошибки в одном периоде могут быть связаны с ошибками в предыдущих периодах. Это создает проблемы, так как отрицательно сказывается на эффективности оценок коэффициентов модели. При наличии автокорреляции значения остаются несмещенными, но их вариация увеличивается, что приводит к менее точным интервалам доверия и снижению качества предсказаний. Пример того, как выглядит автокорреляция в данных, можно визуализировать с помощью графиков, аналогичных тем, которые могут быть представлены на рис. 7.

Чтобы корректно анализировать данные и минимизировать влияние автокорреляции, существует ряд методов для идентификации и устранения этой проблемы. К ним относятся использование тестов на автокорреляцию, таких как тест Дарбина–Уотсона, а также корректировка модели, например, с помощью методов, учитывающих временные лаги.



а)　　　　　　　　*б*)

Рис. 6. Диаграммы рассеяния гомоскедастичности и гетероскедастичности

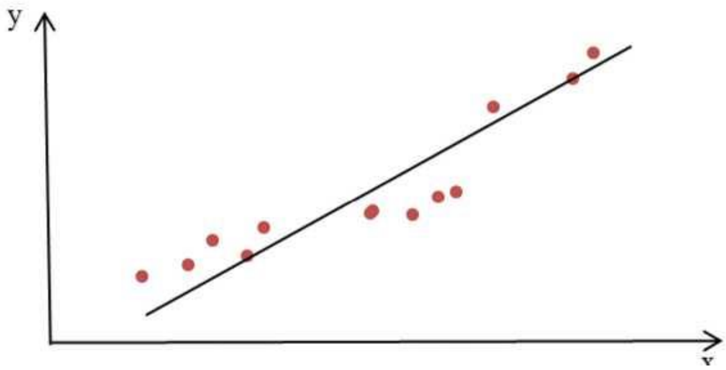


Рис. 7. Автокорреляция остатков

Другим важным условием, необходимым для корректного применения регрессионного анализа, является то, что объясняющая переменная должна быть детерминированной, т.е. неслучайной. Это условие необходимо для предотвращения получения смещенных и неэффективных оценок коэффициентов модели. В случае, если это условие нарушается, например, из-за ошибок в измерении объясняющих переменных, результат анализа может оказаться недостаточно надежным.

Нарушение этого условия может происходить и за счет использования запаздывающих переменных, которые зависят от значений, предшествующих моменту времени, когда осуществляется анализ. Это может ввести дополнительную неопределенность в модель, так как запаздывающие переменные могут быть подвержены тем же самым проблемам измерения и автокорреляции, что и ошибки. В результате, если объясняющие переменные не являются детерминированными, это может привести к ошибочным выводам и неправильным интерпретациям результатов эконометрического анализа.

Таким образом, соблюдение этих условий является критически важным для повышения качества и надежности регрессионных моделей, что, в свою очередь, улучшает интерпретируемость и практическую применимость полученных результатов.

Еще раз акцентируем внимание на том, что только одновременное выполнение четырех условий может гарантировать качественные и надежные результаты регрессионного анализа.

Оценка существенности параметров линейной регрессии и корреляции. Для оценки существенности параметров уравнения регрессии используют несколько подходов.

1. Оценка существенности параметров уравнения регрессии с помощью *F*-критерия Фишера.

Для проверки существенности параметров уравнения с помощью F-критерия Фишера вначале «выдвигается нулевая гипотеза, что коэффициент регрессии равен нулю, следовательно, факторный признак не оказывает влияния на результативный показатель».[24]

Фактически мы проверяем практическую значимость моделей, построенных с помощью корреляционно-регрессионного анализа. Это осуществляется путем определения тесноты связи между признаками x и y. Для этого использу-

[24] Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – М. : Дашков и К, 2019. – 224 с. – URL : https://www.iprbookshop.ru/85184.html

ет метод дисперсионного анализа. Иными словами, рассчитываем показатели общей, факторной и остаточной дисперсий.

Общая дисперсия результативного признака σ_y^2 отображает общее влияние всех факторов:

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}.$$

Факторная дисперсия результативного признака $\sigma_{y_x}^2$ отражает вариацию ***y*** только от воздействия изучаемого фактора *x*:

$$\sigma_{y_x}^2 = \frac{\sum (\hat{y}_{xi} - \bar{y})^2}{n}.$$

Данная формула характеризует отклонение выровненных значений y_x от их общей средней величины $\bar{y}$.

Остаточная дисперсия σ_ξ^2 отражающая вариацию результативного признака *y* от всех прочих факторов, кроме *x*:

$$\sigma_\xi^2 = \frac{\sum (y - \hat{y}_{xi})^2}{n};$$

$$\sigma_{\text{общ}}^2 = \sigma_{\text{факторная}}^2 + \sigma_{\text{остаточная}}^2.$$

Следует отметить, что любая сумма квадратов отклонений связана с числом свободы независимого варьирования признака, т.е. с числом степеней свободы. А число степеней свободы, в свою очередь, связано с числом единиц совокупности n и с числом определяемых по ней констант (табл. 3). При этом n – это число наблюдений, а m – число параметров при переменной *x*.

3. Схема дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum (y - \bar{y})^2$	$n-1$	$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}$
Факторная	$\sum (\hat{y}_{xi} - \bar{y})^2$	m	$\sigma_{yx}^2 = \frac{\sum (\hat{y}_{xi} - \bar{y})^2}{m}$
Остаточная	$\sum (y - \hat{y}_{xi})^2$	$n-m-1$	$\sigma_\varepsilon^2 = \frac{\sum (y - \hat{y}_{xi})^2}{n-m-1}$

Существует известная взаимосвязь между количеством степеней свободы, которые соответствуют общей, факторной и остаточной суммам квадратов. Эта взаимосвязь определяет количество степеней свободы для различных типов дисперсий.

В контексте линейной регрессии для остаточной дисперсии (суммы квадратов) количество степеней свободы составляет $n - 2$. Что касается общей дисперсии, то число степеней свободы определяется на основании числа наблюдений. Однако следует учесть, что при расчетах используется среднее значение, которое вычисляется по выборке, в результате чего теряется одна степень свободы. Таким образом, число степеней свободы для общей дисперсии равно $n - 1$.

Сравнение факторной и остаточной дисперсии с учетом одной степени свободы позволяет вычислить F-критерий Фишера, который выражается через формулу:

$$F = \frac{\sigma^2_{y_x}}{\sigma^2_{\varepsilon}}.$$

Для оценки значимости параметров уравнения регрессии с использованием F-критерия Фишера необходимо сопоставить фактическое значение F-критерия с табличным значением $F_{\text{табл}}$ (α, k_1, k_2) при указанном уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$.

Это может быть выполнено с помощью таблиц критических значений F-отношений, составленных английским статистиком Снедекором, для различных уровней значимости нулевой гипотезы и разных значений степеней свободы. В свою очередь, это является проверкой нулевой гипотезы: если она верна, то имеет место равенство между факторной и остаточной дисперсиями. Для того чтобы опровергнуть нулевую гипотезу, необходимо, чтобы факторная дисперсия была значительно больше остаточной.

Главное правило сравнения: достоверным является рассчитанное значение F-отношения, если оно превышает табличное значение. В этом случае имеет место быть значимость связи между признаками x и y.

Для оценки значимости коэффициента линейной корреляции r F-критерий рассчитывается по определенной формуле:

$$F = \frac{r^2}{1 - r^2}(n - 2).$$

2. Оценка значимости уравнения регрессии с помощью *F*-критерия Фишера.

При линейной регрессии значимость коэффициентов регрессии оценивается с использованием *t*-критерия Стьюдента. В этом случае формулируется гипотеза о случайном характере коэффициентов, т.е. о том, что их отклонение от нуля не является значимым. Затем вычисляются фактические значения данного критерия $t_{\text{факт}}$ для каждого из коэффициентов регрессии, т.е.

$$t_b = \frac{b}{m_b}, \qquad t_a = \frac{a}{m_a},$$

где m_b и m_a – *стандартные ошибки* параметров линейной регрессии определяются по формулам:

$$m_b = \sqrt{\frac{\sum (y-\hat{y})^2}{(n-2)\sum (x-\bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x \sqrt{n}},$$

$$m_a = \sqrt{\frac{\sum (y-\hat{y})^2}{(n-2)} \frac{\sum x^2}{n\sum (x-\bar{x})^2}} = \frac{S_{\text{ост}}}{n\sigma_x}\sqrt{\sum x^2}.$$

$t_{\text{табл}}(1-\alpha;\, k)$ – максимально возможное значение критерия Стьюдента под влиянием случайных факторов при данной степени свободы $k = n - 2$ и уровне значимости α находится из таблицы критерия Стьюдента (табл. 2 приложения).

Если $t_{\text{табл}} < |t_{\text{факт}}|$, то гипотеза о несущественности коэффициента регрессии отклоняется с уровнем значимости α т.е. коэффициент (*a* или *b*) не случайно отличается от нуля и сформировался под влиянием систематически действующего фактора *x*.

Если $t_{\text{табл}} > |t_{\text{факт}}|$, то гипотеза не отклоняется и признается случайная природа формирования параметра.

Значимость линейного коэффициента корреляции также проверяется с помощью t-критерия Стьюдента, т.е.

$$t_r = \frac{r_{xy}}{m_r} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}; \qquad m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}}.$$

Гипотеза о несущественности коэффициента корреляции отклоняется с уровнем значимости α если $t_{\text{табл}} < |t_r|$.

Следует отметить, что для случая с линейной парной регрессии проверки гипотез о значимости коэффициента b и коэффициента корреляции r_{xy} равносильны проверке гипотезы о существенности уравнения регрессии в целом, т.е. $t_b = t_r = \sqrt{F}$.

Для расчета доверительного интервала определяют *предельную ошибку* для каждого показателя, т.е. $\Delta_a = t_{\text{табл}} m_a$, $\Delta_b = t_{\text{табл}} m_b$.

Доверительные интервалы для коэффициентов линейной регрессии:

$$a - \Delta_a < \gamma_a < a + \Delta_a, \qquad b - \Delta_b < \gamma_b < b + \Delta_b.$$

Если ноль находится внутри доверительного интервала, т.е. нижняя граница интервала отрицательна, а верхняя положительна, то считается, что оцениваемый параметр равен нулю. Это объясняется тем, что он не может одновременно принимать и положительные, и отрицательные значения.

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии $\hat{y} = a + bx$ соответствующего прогнозного значения x_p. Затем вычисляется *средняя стандартная ошибка прогноза*:

$$m_{y_p} = \sigma_{\text{ост}} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{nVar(x)}}, \quad \text{где } \sigma_{\text{ост}} = \sqrt{\frac{\sum (y - \hat{y})^2}{(n-2)}},$$

и строится *доверительный интервал прогноза*

$$y_p - t_{\text{табл}} m_{y_p} < \gamma_{y_p} < y_p + t_{\text{табл}} m_{y_p}.$$

Интервал может быть достаточно широк за счет малого объема наблюдений.

Проверка адекватности всей модели осуществляется также с помощью расчета величины средней ошибки аппроксимации (ξ).

Значение средней ошибки аппроксимации, определяемой по формуле:

$$\bar{\varepsilon} = \frac{1}{n} \sum \frac{|y_i - \hat{y}_x|}{y_i} \cdot 100.$$

Ее величина не должна превышать 10…15%.

При оценке соответствия уравнения регрессии реальному процессу могут возникнуть следующие варианты:

1. Модель, согласно критериям Стьюдента (t-критерий) и Фишера (F-критерий), оказывается адекватной, и все коэффициенты регрессии являются

значимыми. Эту модель можно использовать для принятия решений и прогнозирования.

2. Модель соответствует критериям Стьюдента (t-критерий) и Фишера (F-критерий), однако некоторые коэффициенты регрессии незначимы. Поэтому модель не подходит для прогнозирования, но для принятия ограниченного числа решений вполне приемлема.

3. Модель согласно критериям Стьюдента (t-критерий) и Фишера (F-критерий) оказывается адекватной, но все коэффициенты регрессии незначимы. В такой ситуации модель признается полностью неадекватной, и на ее основе нельзя принимать решения или делать прогнозы.

Для того, чтобы понять, на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%, в эконометрическом анализе применяют *частные коэффициенты эластичности*, которые можно рассчитать по формуле:

$$Эx_i = a_i \frac{\bar{x}_i}{\bar{y}},$$

где $\bar{x}_i$ – среднее значение соответствующего факторного признака; $\bar{y}$ – среднее значение результативного признака; a_i – коэффициент регрессии при соответствующем факторном признаке.

Рассмотрим пример выявления зависимости затрат на ремонт оборудования (y) от срока его службы (x) (рис. 8).

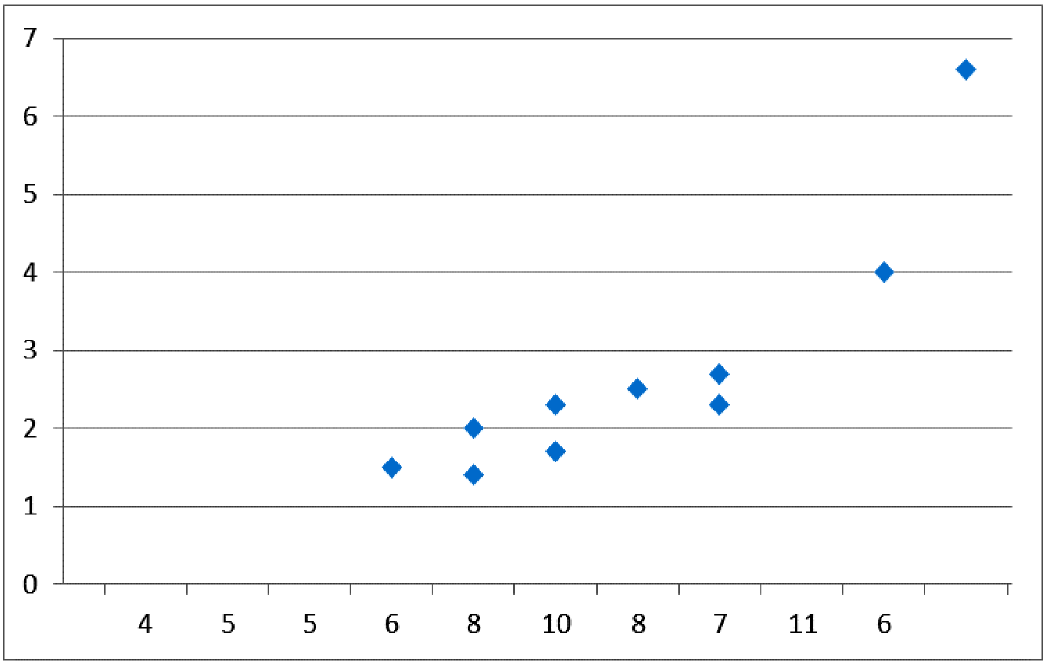


Рис. 8. Поле корреляции зависимости затрат на ремонт оборудования (y) от срока его службы (x)

Анализ рис. 8 показывает наличие близкой к прямолинейной зависимости, так как точки расположены практически по прямой линии.

Прямолинейная форма зависимости y от x описывается уравнением прямой:

$$\bar{y}_x = a + bx.$$

Параметры уравнения прямой найдем с помощью метода наименьших квадратов. Исходные данные и расчетные показатели представлены в табл. 4.

$$\begin{cases} na + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy; \end{cases}$$

$$\begin{cases} 10a + 70b = 27; \\ 70a + 536b = 217. \end{cases}$$

Параметр α характеризует то, каково в среднем влияние на результативный признак тех факторов, которые не были учтены и включены в модель.

Параметр b выступает в качестве индикатора, отражающего среднее изменение значения результативного признака при изменении факторного на одну единицу измерения. По сути это *коэффициент регрессии*.

4. Вспомогательная расчетная таблица

№ п/п	Затраты на ремонт (тыс. руб.) Y	Срок службы оборудования (лет) X	x^2	xy	y^2	$\hat{y}$	$(y-\hat{y}_x)$	$(y-\hat{y}_x)^2$	$(x-\bar{x})^2$	$\lvert y-\hat{y}_x \rvert / y_i$	$(y-\bar{y})^2$
1.	1,5	4	16	6	2,25	0,875	0,625	0,391	9	0,417	1,44
2.	2	5	25	10	4	1,484	0,516	0,266	4	0,258	0,49
3.	1,4	5	25	7	1,96	1,484	–0,084	0,007	4	0,060	1,69
4.	2,3	6	36	13,8	5,29	2,093	0,207	0,043	1	0,090	0,16
5.	2,7	8	64	21,6	7,29	3,311	–0,611	0,373	1	0,226	1,97E-31
6.	4	10	100	40	16	4,529	–0,529	0,280	9	0,132	1,69
7.	2,3	8	64	18,4	5,29	3,311	–1,011	1,022	1	0,440	0,16
8.	2,5	7	49	17,5	6,25	2,702	–0,202	0,041	0	0,081	0,04
9.	6,6	11	121	72,6	43,56	5,138	1,462	2,137	16	0,222	15,21
10.	1,7	6	36	10,2	2,89	2,093	–0,393	0,154	1	0,231	1
Сумма	27	70	536	217,1	94,78	27,02	–0,02	4,715	46	2,156	21,88
Средняя	2,7	7	53,6	21,71	9,478	2,702	–0,002	0,471	4,6	0,216	2,188

Для того, чтобы рассчитать параметры системы уравнений, как правило, применяют метод определителей.

Расчеты осуществляются в несколько этапов. Сначала систему записывают в виде матрицы:

$$\begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x^2 \end{vmatrix} = \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix}.$$

Для нашего примера решение этой матрицы представляется следующим образом:

$$\Delta = \begin{pmatrix} n, \sum x \\ \sum x, \sum x^2 \end{pmatrix} = n\sum x^2 - \sum x \sum x = 460.$$

Чтобы найти определители Δa и Δb, элементы первого (a) и второго (b) столбцов заменяют свободными членами. В итоге получаем следующее:

$$\Delta a = \begin{pmatrix} \sum y, \sum x \\ \sum xy, \sum x^2 \end{pmatrix} = \sum y \sum x^2 - \sum xy \sum x = -718;$$

$$\Delta b = \begin{pmatrix} n, \sum y \\ \sum x, \sum xy \end{pmatrix} = n\sum xy - \sum x \sum y = 280;$$

$$a = \frac{\Delta a}{\Delta} = \frac{\sum y \sum x^2 - \sum xy \sum x}{n\sum x^2 - \sum x \sum x} = -1{,}576;$$

$$b = \frac{\Delta b}{\Delta} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x} = 0{,}61.$$

Решая систему уравнений, получаем уравнение прямой:

$$\overline{y_x} = a + bx.$$

Отсюда: $y = -1{,}576 + 0{,}61x$.

Расчеты можно ускорить с помощью Excel. В *Мастере функций* среди *Статистических* выбираем функцию **ЛИНЕЙН** и заполняем ее аргументы (рис. 9).

Алгоритм заполнения следующий. В зависимости от того, есть ли свободный член в уравнении заполняем параметр *Конст.* Это либо 1, либо 0. Аналогично заполняем параметр *Статистика.* Если нужно выводить дополнительную информацию по регрессионному анализу, то ставим 1.

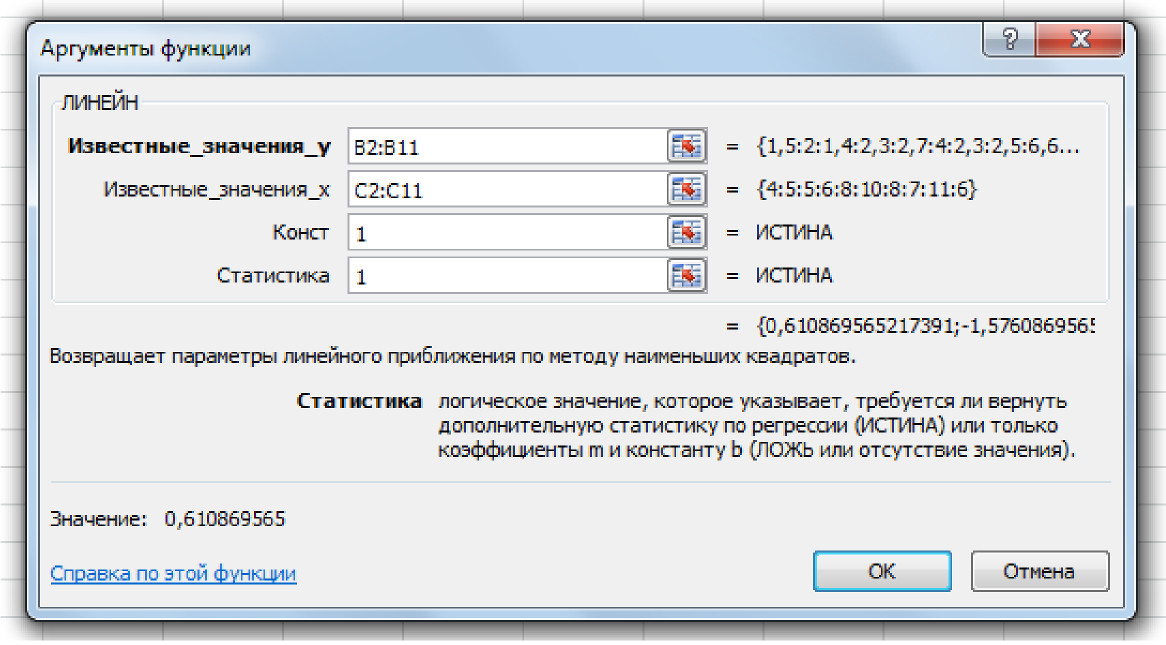


Рис. 9. Расчет коэффициента *b* с помощью функции ЛИНЕЙН

В Excel *линию тренда* можно добавить к диаграмме с областями гистограммы или к графику. Для этого необходимо сначала создать точечную диаграмму и затем добавить в нее линию тренда (рис. 10 и 11).

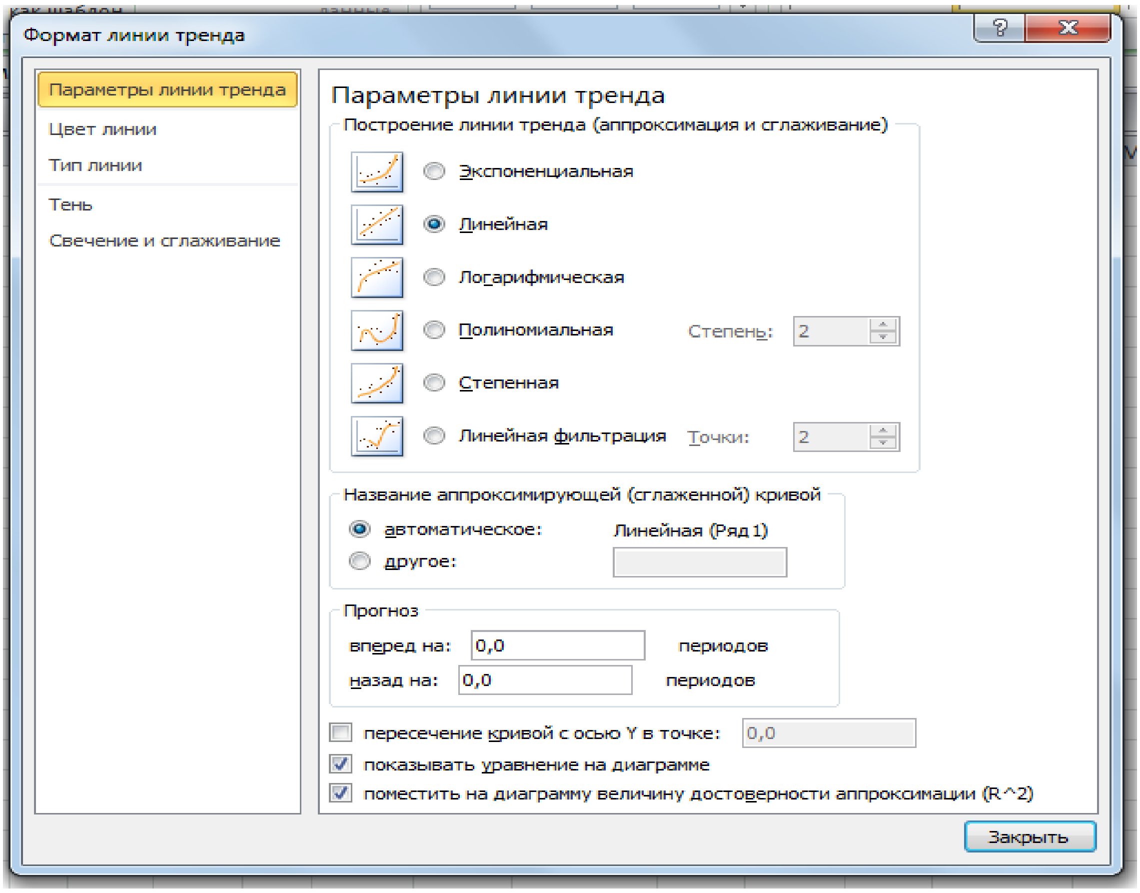


Рис. 10. Добавление линии тренда в точечную диаграмму

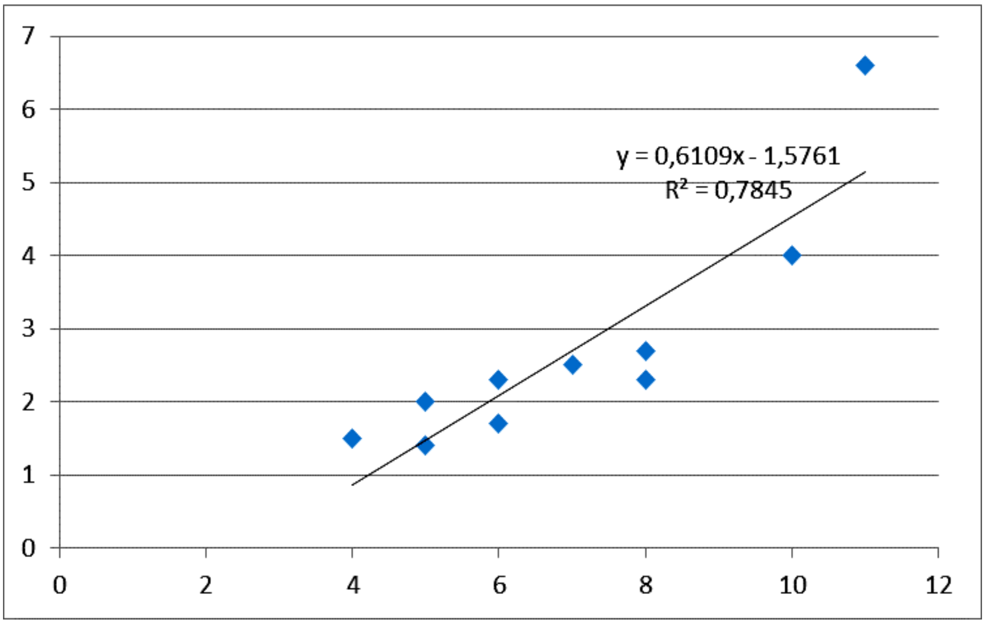


Рис. 11. Теоретическое уравнение линейной регрессии

Таким образом, мы получили, что в уравнении есть свободный член, что значит наличие в модели неучтенных факторов. А значение коэффициента регрессии b = 0,61 означает, что при увеличении срока службы оборудования на 1 год затраты на ремонт растут на 0,61 тыс. р.

Проверка параметров регрессионного уравнения на типичность показала, что:

$$t_a = \frac{a}{m_a}, \quad t_a = \frac{-1{,}576}{0{,}829} = -1{,}901; \quad t_b = \frac{b}{m_b}, \quad t_b = \frac{0{,}61}{0{,}113} = 5{,}397.$$

t-критерий Стьюдента находим либо по специальным таблицам при α = 0,05 и k = 10, либо с помощью встроенной статистической функции **СТЬЮДЕНТ.ОБР.2Х** (рис. 12).

Таким образом, $t_{\text{табл}}$ = 2,228.

Сравнения фактическое и табличное значения t-критерия получаем, что значение t_a меньше $t_{\text{табл}}$, а значение t_b больше $t_{\text{табл}}$. Это значит, что параметр a является нетипичным, а параметр b – типичным.

Проверку практической значимости построенной модели найдем с помощью показателя r:

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x \cdot \sigma_y} = \frac{21{,}71 - 7\cdot 2{,}7}{2{,}145\cdot 1{,}479} = 0{,}886;$$

$$\sigma_y = \sqrt{\frac{\sum(y-\bar{y})^2}{n}} = \sqrt{\frac{21{,}88}{10}} = \sqrt{2{,}188} = 1{,}479.$$

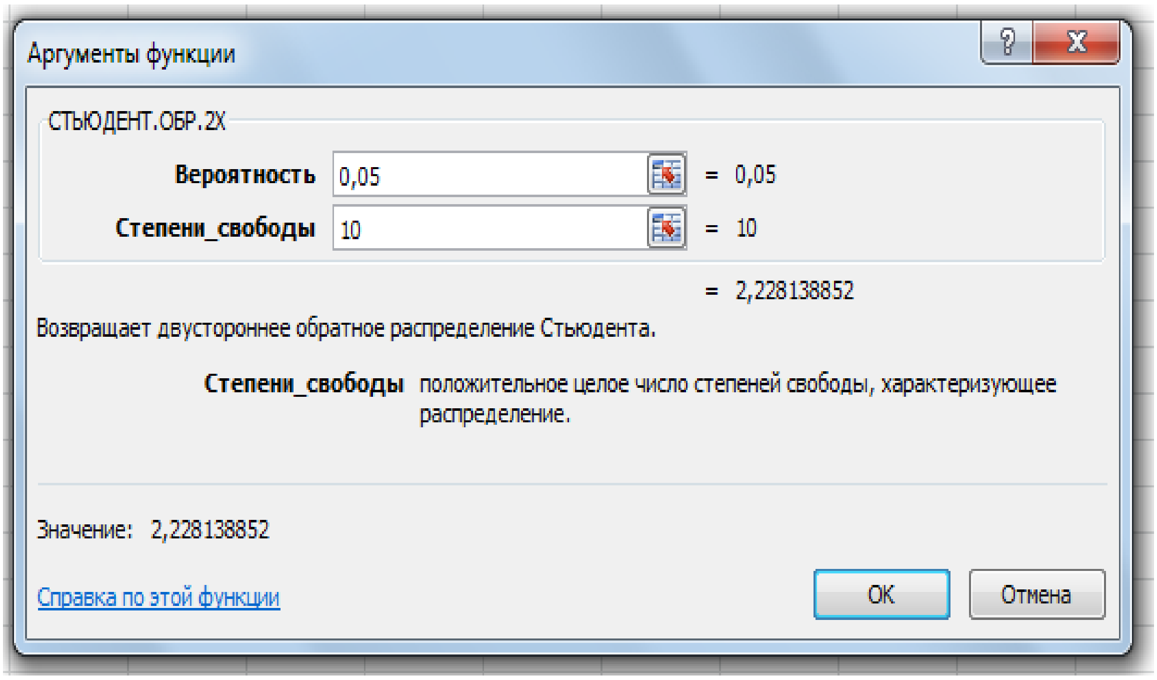


Рис. 12. *t*-критерий Стьюдента (функция СТЬЮДЕНТ.ОБР.2Х)

Сравнение значения r со шкалой Чеддока позволяет сделать вывод о высокой степени связи между сроком службы оборудования и затратами на ремонт, так как величина этой связи приблизительно 0,9.

Индекс детерминации равен:

$$r^2 = 0{,}886^2 = 0{,}785.$$

Что свидетельствует о том, что изменением факторного признака объясняется 78,5% общей вариации.

Оценка значимости коэффициента корреляции осуществляется по F-критерию. Фактическое значение этого критерия t_r определяется по формуле

$$F_{\text{факт}} = \frac{r^2}{1-r^2}(n-2) = \frac{0{,}785}{1-0{,}785}(10-2) = 29{,}21.$$

На уровне значимости 0,05 $F_{\text{табл}} = 5{,}31$ определяем по таблице F-критерия Фишера либо с помощью встроенной статистической функции **F.ОБР.ПХ** (рис. 13).

Так как $F_{\text{табл}} < F_{\text{факт}}$, $F_{\text{табл}} = 5{,}31 < F_{\text{факт}} = 29{,}21$ уравнение регрессии значимо при $\alpha = 0{,}05$.

Среднюю ошибку аппроксимации определим по формуле:

$$\bar{\varepsilon} = \frac{1}{n}\sum\frac{|y_i - \hat{y}_x|}{y_i}\cdot 100, \quad \bar{\varepsilon} = \frac{1}{10}2{,}156\cdot 100 = 21{,}56\%.$$

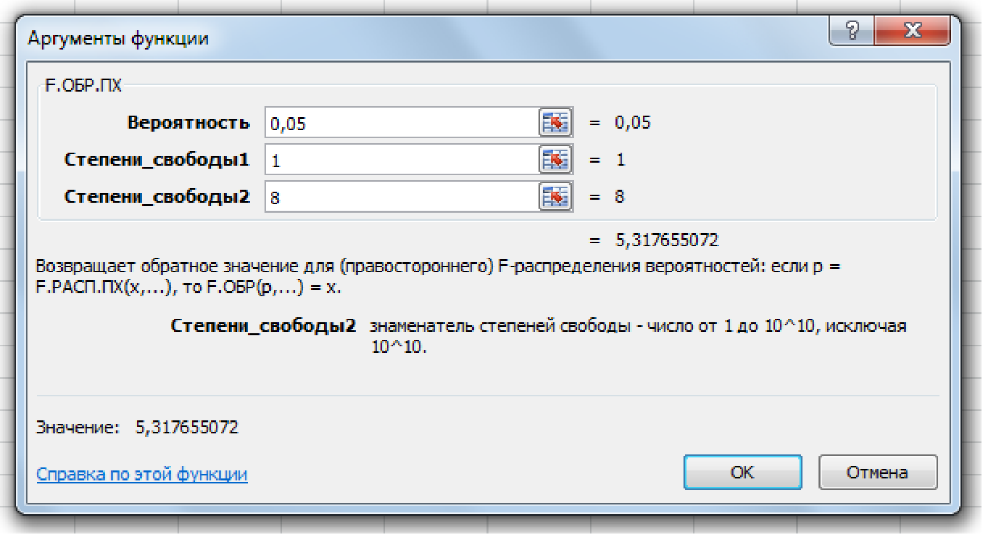


Рис. 13. Диалоговое окно функции FРАСПОБР

Средняя ошибка аппроксимации составляет 21,56%, что недопустимо велико.

Определим коэффициент эластичности по формуле:

$$Э_{x_i} = a_i \frac{\bar{x}_i}{\bar{y}_i} = 0{,}61 \frac{7}{2{,}7} = 1{,}579.$$

Коэффициент эластичности показывает, что при росте срока службы оборудования на 1% затраты на ремонт оборудования вырастут на 1,579%.

Таким образом, построенная нами модель зависимости y от x:

$$y = -1{,}576 + 0{,}61x.$$

В целом модель, проверенная по F-критерию, является адекватной, однако не все коэффициенты регрессии оказываются значимыми (например, коэффициент a отличается от ожидаемого значения). Эта модель может применяться на практике, но она не подходит для построения прогнозов.

В учебных целях приведем пример расчета точечного и интервального прогнозов размера затрат на ремонт оборудования при сроке его эксплуатации 5 лет.

Подставим в полученное уравнение регрессии значение x = 5, получим точечный прогноз:

$$У = -1{,}576 + 0{,}61 \cdot 5 = 1{,}474 \text{ тыс. руб.}$$

Определим среднюю ошибку прогнозируемого индивидуального значения

$$m = \sigma_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_{пр} - \bar{x})^2}{n\sigma_x^2}} = 0{,}687 \sqrt{1 + \frac{1}{10} + \frac{(5-7)^2}{10 \cdot 4{,}6}} 0{,}749.$$

Определим предельную ошибку при уровне значимости 0,05 ($t_{табл}$ = 2,2), $\Delta = 0{,}749 \cdot 2{,}2 = 1{,}647$

$$У_{пр} - \Delta \le У_{пр} \le У_{пр} + \Delta;$$

$$1{,}474 - 1{,}647 = -0{,}173;$$

$$1{,}474 + 1{,}647 = 3{,}121.$$

Так как затраты на ремонт не могут быть отрицательными, то доверительный интервал будет находится в пределах от 0 до 3,121 тыс. руб. при сроке службы оборудования 5 лет.

К экономической интерпретации коэффициента *a* следует относиться с известной долей осторожности, сообразуясь со здравым смыслом, поскольку выборочные данные находятся достаточно далеко от нуля. В ряде случаев ограничиваются интерпретацией коэффициента при объясняющей переменной.

ТЕСТ

1. Что показывает коэффициент регрессии *b* в уравнении $y = a + bx$?

а) сумму квадратов отклонений;

б) корреляцию между признаками;

в) изменение результативного показателя при изменении фактора на единицу;

г) функцию зависимости.

2. Что такое ε в контексте линейной регрессии?

а) гиперболическая зависимость;

б) ошибка измерения;

в) разность между расчетными и наблюдаемыми значениями;

г) коэффициент регрессии.

3. Какой метод используется для оценки параметров линейной регрессии?

а) метод наименьших модулей;

б) графический метод;

в) метод максимального правдоподобия;

г) метод наименьших квадратов.

4. Какой диапазон значений принимает линейный коэффициент корреляции *r*?

а) 0 до 1;

б) –2 до 2;

в) –100 до 100;

г) –1 до 1.

5. Как интерпретируется значение коэффициента детерминации r_{yx}^2?

а) сумма квадратов отклонений;

б) число наблюдений;

в) корреляция между переменными;

г) доля дисперсии результативного показателя, объясняемая регрессией.

6. Какова форма уравнения линейной регрессии?

а) $y = a + bx + \varepsilon$;

б) $y = a - bx + \varepsilon$;

в) $y = ax + b + \varepsilon$;

г) $y = abx + \varepsilon$.

7. Что такое корреляция?

а) сумма квадратов отклонений;

б) отклонение от уравнения регрессии;

в) зависимость между случайными величинами;

г) старое название регрессии.

8. Какова цель вычисления частных производных в методе наименьших квадратов?

а) вывести уравнение связи;

б) определить максимум функции;

в) определить градиент;

г) найти точки минимума функции.

9. Что такое ковариация?

а) гипотеза о зависимости;

б) средний квадрат отклонений;

в) статистическая мера зависимости между двумя признаками;

г) среднее значение двух признаков.

10. Что означает значение *r* = 0 в контексте линейной регрессии?

а) функциональная зависимость;

б) прямая связь между признаками;

в) обратная связь;

г) отсутствие зависимости.

11. Какое из следующих условий относится к Теореме Гаусса–Маркова?

а) математическое ожидание случайной ошибки равно нулю;

б) дисперсия случайной ошибки переменная;

в) случайные ошибки зависимы;

г) объясняющая переменная случайная.

12. Что означает гомоскедастичность в контексте регрессионного анализа?

а) независимость случайных ошибок;

б) случайные ошибки коррелированы;

в) постоянство дисперсии случайной ошибки;

г) систематическое смещение ошибок.

13. Какое условие описывает отсутствие систематического смещения случайной ошибки?

а) независимость ошибок;

б) дисперсия постоянная;

в) математическое ожидание ошибок равно нулю;

г) объясняющая переменная случайная.

14. Что происходит, если условия гомоскедастичности не выполнены?

а) оценки коэффициентов становятся несмещенными;

б) оценки коэффициентов становятся неэффективными;

в) все оценки становятся нереальными;

г) оценки коэффициентов становятся систематически смещенными.

15. Что из следующего можно сказать о случайной ошибке в регрессионной модели?

а) она должна быть равна 1;

б) она должна быть фиксированной величиной;

в) ее математическое ожидание должно быть равно нулю;

г) ее дисперсия должна быть разной для разных наблюдений.

16. Какова роль объясняющей переменной *X* в регрессионной модели?

а) она должна быть неслучайной величиной;

б) она должна быть статистически значимой;

в) она должна иметь случайный характер;

г) она должна быть постоянной величиной.

17. Что необходимо сделать, если обнаружена автокорреляция ошибок?

а) игнорировать ошибки;

б) снизить уровень значимости;

в) применить методы устранения автокорреляции;

г) увеличить размер выборки.

18. Какое предположение о распределении случайной ошибки обычно делается в регрессионном анализе?

а) оно должно быть равномерным;

б) оно должно быть нормально распределено;

в) к ситуации распределения неважно;

г) оно должно быть экспоненциальным.

3. НЕЛИНЕЙНАЯ ПАРНАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Нелинейная регрессия является важным инструментом статистического анализа и моделирования, позволяя исследовать зависимости между переменными, которые не могут быть адекватно описаны с помощью линейных моделей. В контексте нелинейной регрессии существует два основных типа: первая относится к ситуациям, когда относительная нелинейность наблюдается по включенным объясняющим переменным, но сохраняется линейность по оцениваемым параметрам, а вторая представляет собой случаи, когда наблюдается нелинейность и по параметрам регрессии.

1. Нелинейность относительно включенных объясняющих переменных, но линейность по оцениваемым параметрам.

В этом типе нелинейной регрессии функциональная форма модели включает в себя нелинейные преобразования объясняющих переменных, но при этом уравнение модели остается линейным по отношению к параметрам. Такой подход часто используется в практических задачах, где необходимо моделировать сложные зависимости без применения полностью нелинейных регрессионных техник.

Типичным примером такой модели является использование полиномов (многочленов), где зависимость между объясняющей переменной x и зависимой переменной $\boldsymbol{y}$ описывается многочленом. Несмотря на присутствие нелинейных членов (x^2, x^3, ..., x^k), эта модель остается линейной по параметрам, что упрощает их оценку и статистическую интерпретацию.

2. Регрессия нелинейная по оцениваемым параметрам.

Второй тип нелинейной регрессии включает в себя модели, в которых как функциональная форма, так и сами параметры являются нелинейными. Это может происходить в ситуациях, когда связь между зависимой переменной и объясняющими переменными не может быть адекватно выражена даже с помощью полинома. Такие модели требуют более сложных методов оценки, таких как метод максимального правдоподобия (ML), и могут значительно усложнять анализ.

Классическим примером являются модели, представленные экспоненциальными или логарифмическими функциями, включая модели роста или убыли. Здесь параметры оцениваются в рамках нелинейной модели. Методы оценки таких моделей часто требуют итеративных алгоритмов, таких как метод Ньютона–Рафсона или градиентного спуска.

Основные различия между этими двумя типами заключаются в следующем:

1. Структура модели: первый тип сохраняет линейность в оценивании параметров, что упрощает статистический анализ и интерпретацию; второй тип требует использования более сложных методов оценки и обработки результатов.

2. Методы оценки: в первом типе чаще можно применять метод наименьших квадратов, в то время как во втором необходимо использовать методы, такие как максимальное правдоподобие или другие численные методы.

3. Применение: первый тип чаще используется в случаях, когда есть основания ожидать не линейную зависимость, но линейную зависимость параметров. Это может включать модели с полиномиальными терминами; второй тип применяется, когда предположения о линейности параметров не оправдываются, и наблюдаются более сложные зависимые структуры.

Понимание различий между этими двумя типами нелинейной регрессии позволяет более эффективно применять их в анализе данных и построении предсказательных моделей. Выбор между этими подходами зависит от природы данных, исследуемых зависимостей и цели исследования.

Рассмотрим подробнее несколько таких функций.

Парабола второй степени $y_x = a + bx + cx^2$ приводится к линейному виду с помощью замены: $x = x_1$, $x^2 = x_2$.

В результате приходим к двухфакторному уравнению $y_x = a + bx_1 + cx_2$, оценка параметров которого при помощи МНК приводит к системе следующих нормальных уравнений:

$$\begin{cases} an + b\sum x_1 + c\sum x_2 = \sum y; \\ a\sum x_1 + b\sum x_1^2 + c\sum x_1x_2 = \sum x_1y; \\ a\sum x_2 + b\sum x_1x_2 + c\sum x_2^2 = \sum x_2y. \end{cases}$$

А после обратной замены переменных получим:

$$\begin{cases} an + b\sum x + c\sum x^2 = \sum y; \\ a\sum x + b\sum x^2 + c\sum x^3 = \sum xy; \\ a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum x^2y. \end{cases}$$

Парабола второй степени часто используется в тех случаях, когда в пределах определенного диапазона значений фактора происходит изменение характера взаимосвязи между исследуемыми признаками: прямая зависимость сменяется обратной, либо наоборот.

Широкое применение степенной функции объясняется тем, что параметр b в этой функции обладает ясным экономическим смыслом – он представляет собой коэффициент эластичности. (Коэффициент эластичности демонстрирует, на сколько процентов в среднем изменится результат при изменении фактора на 1%.) Формула для определения коэффициента эластичности выглядит следующим образом:

$$Э = f'(x)\frac{x}{y}.$$

Так как для остальных функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора x, то обычно рассчитывается средний коэффициент эластичности:

$$Э = f'(x)\frac{\bar{x}}{\bar{y}}.$$

Приведем формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии (табл. 5).

5. Формулы для расчета коэффициентов эластичности

Вид функции	Первая производная y	Средний коэффициент эластичности, Э
$y = a + bx$	b	$\frac{b\bar{x}}{a+b\bar{x}}$
$y = a + bx + cx^2$	$b + 2cx$	$\frac{(b+2c\bar{x})\bar{x}}{a+b\bar{x}+c\bar{x}^2}$
$y = a + b/x$	$-b/x^2$	$-\frac{b}{a\bar{x}+b}$
$y = ax^b$	abx^{b-1}	b
$y = ab^x$	$a\ln b b^x$	$\bar{x}\ln b$
$y = a + b\ln x$	b/x	$\frac{b}{a+b\ln\bar{x}}$
$y = a/(1 + be^{-cx})$	$abce^{-cx}/(1 + be^{-cx})^2$	$\frac{bc\bar{x}}{b+e^{c\bar{x}}}$
$y = 1/(a + bx)$	$b/(a + bx)^2$	$-\frac{b\bar{x}}{a+b\bar{x}}$

Однако существуют обстоятельства, при которых вычисление коэффициента эластичности становится неуместным или затруднительным, особенно в контексте нелинейной регрессии: это возникает при присутствии в данных разрывов; когда зависимость между переменными является явно нелинейной; в случаях, когда одна из переменных имеет строго ограниченные значения; а также ситуация, где зависимость представлена функцией минимальных или максимальных порогов.

В любом случае, аналогично линейной функции необходим расчет коэффициента корреляции:

$$R_{xy} = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}}.$$

Чем ближе индекс корреляции к единице, тем сильнее связь между рассматриваемыми признаками, и тем более надежным является уравнение регрессии.

Индекс детерминации также аналогично линейной функции. Для оценки значимости уравнения регрессии в целом применяем F-критерия Фишера:

$$F = \frac{R_{xy}^2}{1 - R_{xy}^2} \frac{n - m - 1}{m},$$

где R_{xy} – индекс детерминации; n – число наблюдений; m – число параметров при переменной x.

Если значение статистики Фишера (обычно обозначаемое как F) значимо, это указывает на то, что модель объясняет значительное количество вариации зависимой переменной. Фактическое значение F-критерия сопоставляется с табличным значением при заданном уровне значимости α и соответствующем числе степеней свободы: $k_2 = n - m - 1$ (для остаточной суммы квадратов) и $k_1 = m$ (для факторной суммы квадратов).

Качество нелинейного уравнения регрессии также можно оценить на основе средней ошибки аппроксимации. Этот критерий показывает, насколько хорошо модель предсказывает исходные данные. Он рассчитывается как среднее значение абсолютных ошибок между предсказанными и фактическими значениями.

Пример. Используя данные предыдущего примера (табл. 4) проверим наличие тесноты связи и регрессии с помощью парных нелинейных моделей (табл. 6).

6. Вспомогательная таблица для расчета параметров уравнения параболы второй степени

	x	y	x^2	x^3	xy	x^4	x^2y	$\hat{y}_{xi}$	$y-\hat{y}_{xi}$	$(y-\hat{y}_{xi})^2$	$\|y-\hat{y}_x\|/y_i$
1	4	1,5	16	64	6	256	24	1,814	–0,314	0,099	0,209
2	5	2	25	125	10	625	50	1,665	0,335	0,112	0,167
3	5	1,4	25	125	7	625	35	1,665	–0,265	0,070	0,189
4	6	2,3	36	216	13,8	1296	82,8	1,768	0,532	0,283	0,231
5	8	2,7	64	512	21,6	4096	172,8	2,727	–0,027	0,001	0,010
6	10	4	100	1000	40	10 000	400	4,692	–0,692	0,479	0,173
7	8	2,3	64	512	18,4	4096	147,2	2,727	–0,427	0,182	0,186
8	7	2,5	49	343	17,5	2401	122,5	2,122	0,378	0,143	0,151
9	11	6,6	121	1331	72,6	14 641	798,6	6,052	0,548	0,300	0,083
10	6	1,7	36	216	10,2	1296	61,2	1,768	–0,068	0,005	0,040
$\sum$	70	27	536	4444	217,1	39 332	1894,1	27		1,675	1,441
Ср.	7	2,7	53,6	444	21,71	3933,2	189,41	2,7		0,167	0,144

Составим систему уравнений для нахождения параметров уравнения *параболы второй степени*:

$$y_x = a + bx + cx^2.$$

$$\begin{cases} an + b\sum x + c\sum x^2 = \sum y; \\ a\sum x + b\sum x^2 + c\sum x^3 = \sum xy; \\ a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum x^2 y. \end{cases}$$

$$\begin{cases} 10a + 70b + 536c = 27; \\ 70a + 536b + 4444c = 217{,}1; \\ 536a + 4444b + 39\,332c = 1894{,}1. \end{cases}$$

Найдем общий определитель матрицы

$$\Delta_{\text{общ}} = \begin{vmatrix} 10 & 70 & 536 \\ 70 & 536 & 4444 \\ 536 & 4444 & 39\,332 \end{vmatrix} = 88\,464.$$

Найдем определитель матрицы a

$$\Delta_a = \begin{vmatrix} 27 & 70 & 536 \\ 217,1 & 536 & 4444 \\ 1894,1 & 4444 & 39\,332 \end{vmatrix} = 435\,628,8.$$

Найдем определитель матрицы b

$$\Delta_b = \begin{vmatrix} 10 & 27 & 536 \\ 70 & 217,1 & 4444 \\ 536 & 1894,1 & 39\,332 \end{vmatrix} = -113\,274.$$

Найдем определитель матрицы c

$$\Delta_c = \begin{vmatrix} 10 & 70 & 27 \\ 70 & 536 & 217,1 \\ 536 & 4444 & 1894,1 \end{vmatrix} = 11\,122.$$

Определим параметры уравнения параболы второго порядка:

$$a = \frac{\Delta_a}{\Delta} = \frac{435\,628,8}{88\,464} = 4,924;$$

$$b = \frac{\Delta_b}{\Delta} = \frac{-113\,274}{88\,464} = -1,28;$$

$$c = \frac{\Delta_c}{\Delta} = \frac{11\,122}{88\,464} = 0,126.$$

Таким образом, уравнение параболы примет вид:

$$y = 4,924 - 1,28x + 0,126x^2.$$

Далее определим индекс корреляции:

$$R = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}} = \sqrt{1 - \frac{0,167}{2,188}} = 0,961.$$

Согласно шкалы Чеддока теснота связи весьма высокая, а коэффициент детерминации составит соответственно $R^2 = 0,924$.

Коэффициент детерминации используется для проверки существенности в целом уравнения регрессии по F-критерию Фишера

$$F=\frac{R_{xy}^2}{1-R_{xy}^2}\frac{n-m-1}{m}=\frac{0{,}924}{1-0{,}924}\frac{10-2-1}{2}=42{,}55.$$

Сравним расчетное значение F-критерия Фишера с табличным $F_{\text{кр}}$ = 4,74. Так как расчетное значение F-критерия Фишера больше табличного, то величина уравнения является существенной, следовательно, его можно использовать в практических расчетах.

Определим среднюю ошибку аппроксимации:

$$\xi=\frac{1}{n}\sum\frac{|y_i-\bar{y}_x|}{y_i}\cdot 100=\frac{1}{10}\cdot 1{,}44\cdot 100=14{,}4\%.$$

Ошибка аппроксимации не превышает 15%, что свидетельствует о надежности построенной модели регрессии.

Парабола второго порядка для нашего примера будет выглядеть следующим образом (рис. 14).

Проведем расчет точечного и интервального прогнозов размера затрат на ремонт оборудования при сроке его эксплуатации 5 лет.

Подставим в полученное уравнение регрессии значение x = 5, получим точечный прогноз:

$$y=4{,}924-1{,}28\cdot 5+0{,}126\cdot 5^2=1{,}647 \text{ тыс. руб.}$$

Определим среднюю ошибку прогнозируемого индивидуального значения

$$m=\sigma_\varepsilon\sqrt{1+\frac{1}{n}+\frac{(x_{\text{пр}}-\bar{x})^2}{n\sigma_x^2}}=0{,}409\sqrt{1+\frac{1}{10}+\frac{(5-7)^2}{10\cdot 4{,}6}}=0{,}445.$$

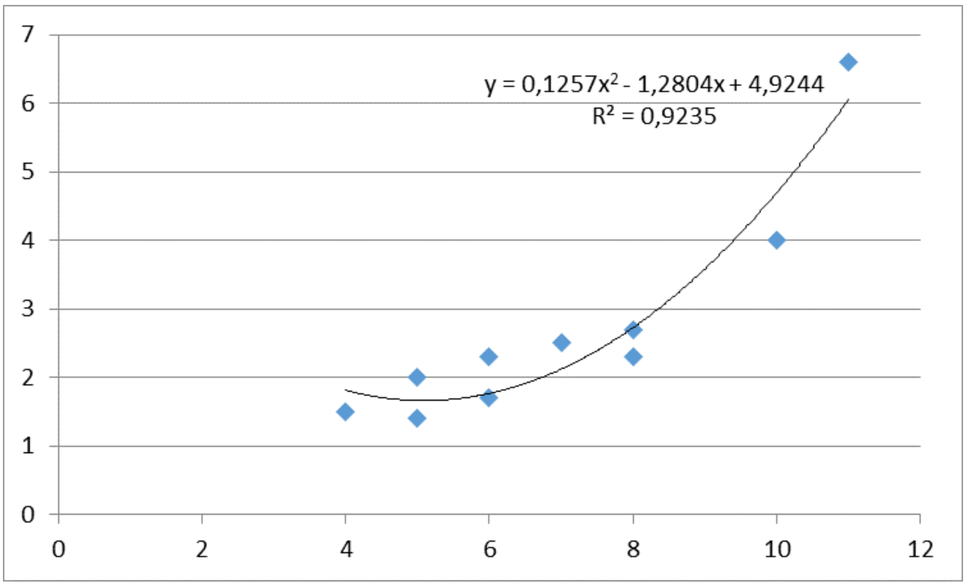


Рис. 14. Теоретическое распределение параболы второго порядка

Определим предельную ошибку при уровне значимости 0,05 ($t_{табл}$ = 2,2), $\Delta = 0{,}445 \cdot 2{,}2 = 0{,}98$

$$У_{пр} - \Delta \le У_{пр} \le У_{пр} + \Delta;$$

$$1{,}674 - 0{,}98 = 0{,}694;$$

$$1{,}674 + 0{,}98 = 2{,}654.$$

Таким образом при сроке службы оборудования 5 лет затраты на их ремонт будут находится в пределах от 0,694 до 2,654 тыс. руб. при сроке службы оборудования 5 лет.

Рассмотрим решение поставленной задачи с помощью экспоненциальной зависимости (табл. 7).

Экспоненциальная регрессия: $\hat{y} = e^{a+bx}$.

Линеаризующее преобразование: $x' = x; \quad y' = \ln y$.

$$b = \frac{n\sum x \ln y - \sum x \sum \ln y}{n\sum x^2 - \left(\sum x\right)^2}. \qquad b = \frac{10 \cdot 70{,}6 - 70 \cdot 8{,}814}{10 \cdot 536 - 70^2} = 0{,}193;$$

$$a = \frac{1}{n}\sum \ln y - \frac{1}{n} b \sum x; \qquad a = \frac{1}{10} \cdot 8{,}814 - \frac{1}{10} \cdot 0{,}193 \cdot 70 = -0{,}472.$$

Получаем уравнение:

$$y = e^{-0{,}472 + 0{,}193x}.$$

7. Вспомогательная таблица
для расчета параметров уравнения экспоненциальной регрессии

	x	y	ln$y$	$x$ln$y$	x^2	$\hat{y}_{xi}$	$y - \hat{y}_{xi}$	$(y - \hat{y}_{xi})^2$	$\mid y - \hat{y}_x \mid / y_i$
1	4	1,5	0,405	1,622	16	1,352	0,148	0,022	0,099
2	5	2	0,693	3,466	25	1,640	0,359	0,129	0,180
3	5	1,4	0,336	1,682	25	1,640	–0,240	0,057	0,171
4	6	2,3	0,833	4,998	36	1,999	0,310	0,096	0,135
5	8	2,7	0,993	7,946	64	2,929	–0,229	0,053	0,085
6	10	4	1,386	13,863	100	4,312	–0,312	0,098	0,078
7	8	2,3	0,833	6,663	64	2,929	–0,629	0,396	0,274
8	7	2,5	0,916	6,414	49	2,414	0,086	0,007	0,034
9	11	6,6	1,887	20,758	121	5,232	1,368	1,871	0,207
10	6	1,7	0,531	3,184	36	1,999	–0,289	0,084	0,171
$\sum$	70	27	8,814	70,595	536	26,43	0,570	2,814	1,434
Ср.	7	2,7	0,881	7,06	53,6	2,64	0,057	0,281	0,143

Далее определим индекс корреляции

$$R = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}} = \sqrt{1 - \frac{0{,}281}{2{,}188}} = 0{,}933.$$

Согласно шкалы Чеддока теснота связи весьма высокая, а коэффициент детерминации составит соответственно $R^2 = 0{,}871$.

Коэффициент детерминации используется для проверки существенности в целом уравнения регрессии по F-критерию Фишера

$$F = \frac{R_{xy}^2}{1 - R_{xy}^2} \frac{n - m - 1}{m} = \frac{0{,}871}{1 - 0{,}871} \frac{10 - 2 - 1}{2} = 23{,}63.$$

Сравним расчетное значение F-критерия Фишера с табличным $F_{кр} = 4{,}74$. Так как расчетное больше табличного величина уравнения является существенной, следовательно, его можно использовать в практических расчетах. Ошибка аппроксимации не превышает 15%, что свидетельствует о надежности построенной модели регрессии.

Определим среднюю ошибку аппроксимации:

$$\xi = \frac{1}{n} \sum \frac{|y_i - \bar{y}_x|}{y_i} \cdot 100 = \frac{1}{10} \cdot 1{,}43 \cdot 100 = 14{,}3\%.$$

Теоретическое распределение экспоненциальной зависимости для нашего примера графически будет выглядеть следующим образом (рис. 15).

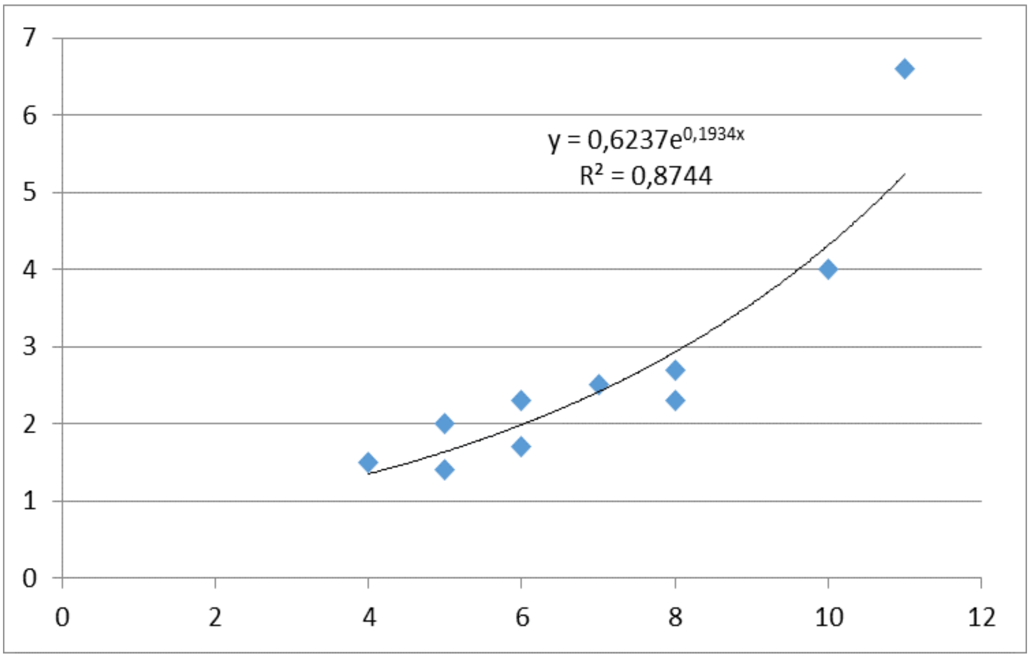


Рис. 15. Теоретическое распределение экспоненциальной зависимости

Проведем расчет точечного и интервального прогнозов размера затрат на ремонт оборудования при сроке его эксплуатации 5 лет.

Подставим в полученное уравнение регрессии значение $x = 5$, получим точечный прогноз:

Получаем уравнение

$$y = e^{-0,472+0,193x} = e^{-0,472+0,193\cdot 5} = 2,929.$$

Определим среднюю ошибку прогнозируемого индивидуального значения

$$m = \sigma_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{n\sigma_x^2}} = 0,281\sqrt{1 + \frac{1}{10} + \frac{(5-7)^2}{10 \cdot 4,6}} = 0,306.$$

Определим предельную ошибку при уровне значимости 0,05 ($t_{\text{табл}} = 2,2$), $\Delta = 0,306 \cdot 2,2 = 0,673$

$$У_{\text{пр}} - \Delta \le У_{\text{пр}} \le У_{\text{пр}} + \Delta;$$

$$2,929 - 0,673 = 2,256;$$

$$2,929 + 0,673 = 3,602.$$

Таким образом при сроке службы оборудования 5 лет затраты на их ремонт будут находится в пределах от 2,256 до 3,602 тыс. руб. при сроке службы оборудования 5 лет.

Сравним построенные модели по индексу детерминации и средней ошибке аппроксимации (табл. 8).

Наиболее адекватно описывает зависимость между сроком службы оборудования и затратами на ремонт модели параболы второго порядка и экспоненциальная регрессия.

Аналогичным образом решаются задачи по нахождению параметров других не линейных зависимостей.

8. Результаты сравнения различных моделей зависимости

Модель	Коэффициент детерминации, *R*	Средняя ошибка аппроксимации, %
Линейная	0,785	21,56
Парабола второго порядка	0,961	14,4
Экспоненциальная	0,933	14,3

Равносторонняя гипербола $y_x = a + b/x$ обычно находит свое применение в различных аспектах экономической теории для описания зависимостей между двумя переменными. Самые яркие примеры – кривая Филлипса и Э. Энгеля.

Кривая Филлипса иллюстрирует обратную связь между уровнем инфляции и уровнем безработицы. В этом контексте равносторонняя гипербола может служить моделью компромисса между данными макроэкономическими показателями. Значения параметров a и b позволяют описывать различные экономические ситуации, в которых снижение безработицы может сопровождаться увеличением инфляции и наоборот. Кривые Э. Энгеля отражают зависимость между доходами потребителя и его расходами по различным категориям товаров. В данном случае гипербола может быть использована для моделирования влияния изменений в доходах на уровень потребления. Например, определенная кривая может иллюстрировать, что с ростом дохода потребление одной категории товаров увеличивается, в то время как на другую – уменьшается, что также можно выразить через гиперболическое уравнение.

Гипербола может быть переведена в линейное уравнение благодаря математическим преобразованиям. Для этого можно выразить одну переменную через другую. Например, уравнение $y_x = a + b/x$ можно переписать как:

$$y = (a + b/x)/x = (a/x) + b/x^2.$$

Умножив обе стороны на x, мы получим линейную зависимость между переменными.

Ситуация в области нелинейных регрессионных моделей отличается по своей природе от линейных. Нелинейные модели можно классифицировать на два основных типа, в зависимости от характерных особенностей оценки их параметров. Первый тип включает в себя нелинейные модели, которые в рамках своей структуры представляют собой внутренне линейные системы. Эти модели обладают свойством, благодаря которому их можно преобразовать в линейный вид с помощью различных математических операций, таких как логарифмирование. Это преобразование позволяет упростить их анализ и интерпретацию. Второй тип охватывает нелинейные модели, которые остаются внутренне нелинейными, даже после применения преобразований. Эти модели не поддаются преобразованию в линейную форму, что делает их анализ более сложным и требует особого подхода для их оценки и интерпретации.

Таким образом, различие между двумя типами нелинейных регрессионных моделей существенно влияет на подходы к их исследованию и применение в реальных задачах.

При статистическом анализе нелинейной регрессии возможно применение уравнения регрессии *показательной функции*:

$$\hat{y} = ab^x.$$

Линеаризующее преобразование: $x' = x;\quad y' = \ln y.$

$$\ln a = \frac{1}{n}\sum \ln y - \frac{1}{n}\ln b\sum x; \qquad \ln b = \frac{n\sum x\ln y - \sum x\sum \ln y}{n\sum x^2 - \left(\sum x\right)^2}.$$

Далее проверим наличие зависимости с помощью степенной функции. Для определения параметров *степенной функции* $y = ax^b$ МНК степенную функцию необходимо привести к линейному виду также путем логарифмирования. В результате логарифмирования получим уравнение вида:

$$\ln y = \ln a + b\ln x;$$

$$Y = \ln y;\ A = \ln a;\ X = \ln x;$$

$$\begin{cases} An + b\sum X = \sum Y; \\ A\sum X + b\sum X^2 = \sum XY, \end{cases}$$

а затем потенцированием находим искомое уравнение.

Линеаризующее преобразование: $x' = \ln x;\quad y' = \ln y.$

$$b = \frac{n\sum(\ln x\cdot\ln y) - \sum \ln x\sum \ln y}{n\sum(\ln x)^2 - \left(\sum \ln x\right)^2}. \qquad \ln a = \frac{1}{n}\sum \ln y - \frac{1}{n}b\sum x.$$

ТЕСТ

1. Какие из перечисленных функций представляют собой регрессии, нелинейные относительно объясняющих переменных, но линейные по оцениваемым параметрам?

а) степенная функция ($y_x = ax^b$);

б) показательная функция ($y_x = ab^x$);

в) равносторонняя гипербола ($y_x = a + b/x$);

г) экспоненциальная функция ($y_x = e^{(a + bx)}$).

2. Что представляет собой коэффициент эластичности в степенной функции $y_x = ax^b$?

а) среднее значение зависимой переменной;

б) параметр a;

в) параметр b;

г) среднее значение независимой переменной.

3. В каких случаях расчет коэффициента эластичности не имеет смысла?

а) когда зависимая переменная принимает отрицательные значения;

б) когда независимая переменная принимает нулевое значение;

в) когда бессмысленно определение изменения в процентах для рассматриваемых признаков;

г) когда коэффициент корреляции близок к нулю.

4. Что представляет собой индекс корреляции (R_{xy}) в нелинейной регрессии?

а) среднюю ошибку аппроксимации;

б) показатель тесноты связи между переменными;

в) число наблюдений;

г) число параметров в уравнении.

5. Как используется индекс детерминации (R_{xy}^2) в проверке существенности уравнения регрессии?

а) для расчета средней ошибки аппроксимации;

б) для расчета коэффициента эластичности;

в) в F-критерии Фишера;

г) для определения знака корреляции.

6. Что показывает средняя ошибка аппроксимации в нелинейной регрессии?

а) тесноту связи между переменными;

б) точность аппроксимации эмпирических данных уравнением регрессии;

в) существенность уравнения регрессии;

г) коэффициент эластичности.

4. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

Отбор факторов при построении уравнения множественной регрессии. Отбор факторов при построении уравнения множественной регрессии – это важный этап анализа, который влияет на достоверность получаемой модели и ее прогностическую способность. В этом процессе исследователь сталкивается с рядом вызовов, так как общественные явления формируются под воздействием множества факторов, взаимодействия между которыми могут быть сложными и нелинейными. Давайте рассмотрим процесс отбора факторов более подробно.

1. *Понимание многофакторной модели*.

Многофакторная регрессия анализирует влияние нескольких независимых переменных на одну зависимую переменную. Модель может быть записана в виде:

$$y_i = f(x_1, x_2, \ldots, x_n) + \varepsilon_i.$$

Это выражение представляет собой *классическую нормальную модель множественной регрессии*.

2. *Этапы отбора факторов*.

Формулирование гипотезы. Перед началом анализа необходимо определить исходные гипотезы. Это важно, чтобы знать, какие факторы потенциально могут влиять на зависимую переменную. Обычно исследователь основывает свои гипотезы на теоретических рамках, предыдущих исследованиях или практическом опыте.

Сбор данных. Для многофакторного регрессионного анализа требуется собрать достаточно данных по всем рассматриваемым переменным. Важно, чтобы выборка была репрезентативной и охватывала все необходимые аспекты.

Первичный анализ данных. На этом этапе исследуются данные, чтобы выявить корреляции между переменными и провести предварительный анализ. Здесь могут использоваться такие методы, как: корреляционный анализ, чтобы оценить связь между зависимой и независимыми переменными; визуализация данных (например, построение диаграмм рассеяния), чтобы увидеть явные тренды и зависимости.

Проверка коллинеарности. Перед отбором факторов важно проверить наличие мультиколлинеарности – ситуации, когда две или более независимые переменные сильно коррелируют друг с другом. Это может привести к искажению оценок коэффициентов. Для диагностики мультиколлинеарности часто используют показатели, такие как: коэффициент вариации инфляции (VIF), корреляционные матрицы.

Если коллинеарность обнаружена, можно рассмотреть следующие варианты: Исключение одной из коррелирующих переменных; Объединение коррелирующих переменных в одну за счет создания индексов или составных переменных.

Выбор критериев отбора. Существует несколько методов, которые можно использовать для выбора факторов:

1. Метод наименьших квадратов: Начальный вариант, где исследуются все факторы, и отбираются наименее значимые по значению p. Если p-значение фактора превышает заданный уровень значимости (например, 0,05), он удаляется из модели.

2. Пошаговый (stepwise) отбор: Это автоматизированный процесс, при котором фактор поочередно добавляется или убирается из модели на основе определенных критериев (например, AIC, BIC, F-статистика).

3. Регрессия с LASSO и Ridge: Эти методы помогают одновременно уменьшить количество факторов и избежать переобучения модели, особенно в условиях высокой размерности данных. Они вводят штрафы за большие значения коэффициентов, что позволяет «забирать» некоторые из них, если они незначащие.

Проверка качества модели. После завершения этапа отбора факторов важно оценить качество полученной модели, используя следующие параметры:

– R^2 и скорректированный R^2. Они показывают, какую долю вариации зависимой переменной объясняет модель. Скорректированный R^2 будет более точным, если количество независимых переменных велико.

– средняя квадратическая ошибка или другие критерии оценки точности предсказаний;

– тесты на нормальность остатков, их гомоскедастичность и независимость.

Интерпретация результатов. На заключительном этапе интерпретируются полученные коэффициенты. Важно не только выяснить статистическую значимость факторов, но и понять их экономический или социологический смысл. Коэффициенты показывают величину и направление изменения зависимой переменной при изменении независимой на единицу.

Отбор факторов в множественной регрессии требует систематического и тщательного подхода. Ошибки на этом этапе могут привести к неверным выводам о характере и важности влияния факторов. Успешный отбор факторов не только повышает точность модели, но и углубляет понимание исследуемого явления, что особенно важно в социальных науках. Таким образом, многофакторный корреляционно-регрессионный анализ становится мощным инструментом для оценки сложных взаимосвязей между держащими влияние переменными.

«Отдельного внимания заслуживает обоснование необходимости включения определенных факторов в модель. Поэтому процесс отбора факторов обычно состоит из двух этапов: вначале они выбираются на основании поставленных задач, а затем, опираясь на матрицу корреляций, проводится анализ статистик для параметров регрессии».[25]

«*Коэффициенты интеркорреляции* – это важный инструмент в множественной регрессии, который помогает выявлять взаимосвязи между независимыми переменными. Это критически важно для создания качественной регрессионной модели, поскольку наличие коллинеарных факторов может искажать результаты и усложнять их интерпретацию. Корреляция между независимыми переменными показывает, насколько изменение одной переменной связано с изменением другой. Если корреляция между двумя переменными высока (например, $r_{x_i x_j} \geq 0{,}7$), это указывает на линейную зависимость между ними. В такой ситуации эти переменные могут дублировать друг друга в модели, что может привести к избыточной информации и усложнить интерпретацию».[26]

Наличие коллинеарных переменных может привести к нескольким проблемам:

– Искажение оценок: при наличии коллинеарных факторов оценка коэффициентов регрессии становится менее устойчивой, что может привести к ошибочным выводам о влиянии независимых переменных на зависимую.

– Усложнение интерпретации: интерпретировать коэффициенты регрессии становится труднее, так как взаимосвязанные переменные могут затенять влияние друг друга.

Чтобы избежать проблем с коллинеарностью, рекомендуется внедрить следующие методы.

Выявление коллинеарных факторов: первоначально необходимо вычислить коэффициенты корреляции для всех пар независимых переменных. Если обнаружены пары с $r_{x_i - x_j} \geq 0{,}7$, стоит рассмотреть их дальнейшее использование.

[25] Буравлев, А. И. Эконометрика : учебное пособие / А. И. Буравлев // Электронно-библиотечная система IPR BOOKS: [сайт]. – 4-е изд. – М. : Лаборатория знаний, 2021. – 165 с. – URL : https://www.iprbookshop.ru/109431.html

[26] Там же.

Удаление дублирующих переменных: важно решить, какой из коллинеарных факторов оставить в модели. Предпочтение лучше отдать тому фактору, который достаточно связан с зависимой переменной и имеет минимальную зависимость от других независимых. Это помогает сохранить информативность модели и избежать избыточности.

Оценочный метод: для выбора, какой из коллинеарных факторов оставить, можно использовать различные подходы. Например, шаговая регрессия, при добавлении или удалении переменных.

Обеспечение взаимной независимости независимых факторов является основополагающим принципом множественной регрессии. Качественный анализ коэффициентов интеркорреляции и грамотное исключение дублирующих факторов помогут создать более надежные и легко интерпретируемые модели. Это, в свою очередь, повышает точность прогнозов и углубляет понимание процессов, влияющих на зависимую переменную.

Пусть, например, при изучении зависимости $y = f(x_1, x_2, x_3)$ матрица парных коэффициентов корреляции оказалась следующей (табл. 9).

Очевидно, что факторы x_1 и x_2дублируют друг друга. В анализ целесообразно включить фактор x_2, а не x_1, хотя корреляция x_2 с результатом y слабее, чем корреляция фактора x_1 с y.

$$r_{yx_2} = 0{,}7 < r_{yx_1} = 0{,}8,$$

но зато значительно слабее межфакторная корреляция

$$r_{x_2x_3} = 0{,}2 < r_{x_1x_3} = 0{,}5.$$

Поэтому в данном случае в уравнение множественной регрессии включаются факторы x_2, x_3.

9. Матрица парных коэффициентов корреляции

	y	x_1	x_2	x_3
y	1	0,8	0,7	0,6
x_1		1	0,8	0,5
x_2			1	0,2
x_3				1

Парные коэффициенты корреляции способны обнаружить лишь очевидные случаи коллинеарности факторов. Более серьезные трудности при применении методов множественной регрессии возникают в условиях мультиколлинеарности, когда несколько факторов связаны между собой через линейную зависимость, что ведет к их взаимозависимому воздействию.

Мультиколлинеарность означает, что определенные факторы могут систематически влиять друг на друга. Это сокращает независимость вариаций исходных данных, усложняя оценку отдельного вклада каждого фактора.[27]

Включение мультиколлинеарных факторов в модель нежелательно по ряду причин:

1. Параметры множественной регрессии становятся труднее интерпретировать как чистые показатели воздействия, потому что факторы коррелируют, и значения этих параметров утрачивают свое экономическое содержание.

2. Оценка параметров становится ненадежной, сопровождается большими стандартными ошибками и может меняться вместе с увеличением объема данных (по величине и даже по знаку), делая модель непригодной для анализа и прогнозирования.

Выбор факторов для включения в регрессию представляет собой одну из ключевых стадий практического использования регрессионных методов.

Построение уравнения множественной регрессии позволяет исследовать взаимосвязь между зависимой переменной и несколькими независимыми факторами. Существуют различные методики для выборки значимых переменных и формирования модели. Рассмотрим три популярных метода: метод исключения, метод включения и шаговую регрессию.

Метод исключения заключается в поэтапном удалении факторов из первоначального набора переменных. Это позволяет определить, какие переменные оказывают наименьшее влияние на зависимую переменную и могут быть исключены из модели без значительной потери ее качества. Процесс включает следующие шаги:

– Инициализация: на начальном этапе рассматривается полный набор факторов.

– Оценка значимости: на каждом этапе проводится анализ значимости каждого фактора через критерии, такие как *p*-значение. Факторы, у которых *p*-значение выше заданного уровня значимости (например, 0,05), считаются незначительными.

[27] Буравлев, А. И. Эконометрика : учебное пособие / А. И. Буравлев // Электронно-библиотечная система IPR BOOKS : [сайт]. – 4-е изд. – М. : Лаборатория знаний, 2021. – 165 с. – URL : https://www.iprbookshop.ru/109431.html

– Удаление переменных: на основании оценки значимости исключаются наименее значимые переменные.

– Проверка модели: после каждого удаления модель пересматривается для проверки ее качества, таких как *R*-квадрат.

Данный метод позволяет создать более простую и точную модель, но требует осторожности, чтобы не исключить важные переменные.

Метод включения, в отличие от метода исключения, подразумевает добавление факторов в модель по одному. Начинается он с простейшей модели, состоящей лишь из одной переменной, и поэтапно добавляются новые переменные. Процесс включает следующие шаги:

– Изначальная модель: начинается с простой модели с одной зависимой переменной.

– Добавление переменных: на каждом шаге поочередно добавляются факторы из первоначального набора. Каждое добавление анализируется на значимость.

– Оценка значимости: каждое новое добавление проверяется с использованием коэффициентов корреляции, *p*-значений и других статистических критериев.

– Финальная модель: процесс продолжается до тех пор, пока не будут добавлены все значимые факторы или не будет достигнуто заданное количество факторов в модели.

Метод включения полезен, когда индивидуальные факторы могут иметь разные уровни влияния на зависимую переменную, что позволяет найти оптимальное количество переменных.

Шаговая регрессия (combines метод включения и метод исключения), позволяя одновременно добавлять и удалять переменные. Этот подход наиболее эффективен для создания модели с оптимальным числом факторов и высоким качеством. Процесс включает следующие шаги:

– Начальная модель: начинается с пустой модели или модели с одной переменной.

– Шаги добавления и удаления: на каждом шаге алгоритм рассматривает возможность добавления или удаления переменных, основываясь на критериях значимости. Если добавление нового фактора делает модель лучше (например, увеличивает *R*-квадрат), то этот фактор включается. Если какой-либо фактор становится незначительным после добавления новых, он исключается.

– Циклический процесс: процесс продолжается в циклическом режиме, где на каждом этапе происходит оценка значимости текущих факторов, пока не будет достигнута оптимальная конфигурация.

Шаговая регрессия позволяет эффективно находить баланс между количеством переменных и качеством модели, делая ее более гибкой и адаптируемой к данным.

Использование этих методов позволяет построить более точные и интерпретируемые модели множественной регрессии, что является важным этапом в статистическом анализе и прогнозировании.

«При отборе факторов также рекомендуется учитывать следующий принцип: количество включаемых факторов обычно должно быть примерно в 6–7 раз меньше общего числа наблюдений, использованных для построения регрессии. Нарушение этого соотношения может привести к недостаточному количеству степеней свободы для остаточной дисперсии. В результате параметры регрессионного уравнения могут оказаться статистически незначимыми, а значение *F*-критерия будет ниже критических значений».[28]

Свойства оценок на основе МНК. Ввиду четкой интерпретации параметров наиболее широко используется линейная функция. В линейной множественной регрессии

$$y_x = a + a_l x_l + a_2 x_2 + \ldots + a_m x_m$$

коэффициенты при переменных ***x*** в модели множественной линейной регрессии называются *коэффициентами «чистой» регрессии*. Эти коэффициенты отражают среднее изменение зависимой переменной в результате изменения соответствующей независимой переменной на одну единицу, при условии, что все остальные независимые переменные остаются неизменными и находятся на среднем уровне. Это помогает понять индивидуальное влияние каждого фактора на результирующую переменную.

Для расчета параметров множественной линейной регрессии используется система линейных уравнений. Чаще всего применяется метод наименьших квадратов (МНК), который минимизирует сумму квадратов отклонений фактических значений зависимой переменной от предсказанных значений.

Для двухфакторного уравнения, представленного в форме:

$$y_x = a + a_l x_l + a_2 x_2 + \varepsilon.$$

Коэффициенты a_1 и a_2 можно вычислить с помощью решения уравнений методом наименьших квадратов. Эти вычисления могут быть представлены в матричной форме, что позволяет удобно работать с большими наборами данных.

[28] Мхитарян, В. С. Эконометрика : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин // Цифровой образовательный ресурс IPR SMART : [сайт]. – М. : Евразийский открытый институт, 2012. – 224 с. – URL : https://www.iprbookshop.ru/11125.html

Метод наименьших квадратов также применим для уравнения множественной регрессии в стандартизированном масштабе. Стандартизация позволяет привести все переменные к общему масштабу, что облегчает интерпретацию коэффициентов и сравнение влияния разных факторов. Обычно стандартизация включает преобразование каждой переменной так, чтобы она имела среднее значение 0 и стандартное отклонение 1. Таким образом, все переменные становятся сопоставимыми, что особенно полезно, когда факторы имеют различные единицы измерения или различный разброс значений.

Предположим, что у нас есть два фактора: количество часов обучения (x_1) и количество полученных консультаций (x_2), и мы хотим исследовать их влияние на итоговую оценку студентов (y).

При использовании множественной регрессии мы можем получить коэффициенты a_1 и a_2. Если a_1 равен 2, это означает, что при увеличении количества часов обучения на 1, результат (итоговая оценка) увеличивается в среднем на 2 балла, при условии, что количество консультаций остается неизменным. Если a_2 равен 3, это значит, что добавление одной консультации также ведет к увеличению результата на 3 балла, при условии, что часы обучения остаются неизменными.

Коэффициенты «чистой» регрессии являются ключевым элементом множественной регрессионной модели. Они предоставляют ценную информацию о влиянии каждой независимой переменной на зависимую и помогают выявить важные зависимости в данных. Использование метода наименьших квадратов и стандартизированных масштабов делает анализ более точным и интерпретируемым, что, в свою очередь, улучшает качество предсказаний и позволяет принимать более обоснованные решения на основе полученных результатов.

Метод наименьших квадратов применим и к уравнению множественной регрессии в стандартизированном масштабе:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \cdots + \beta_m t_{x_m} + \varepsilon,$$

где $t_y, t_{x_1}, ..., t_{x_m}$ – стандартизированные переменные; β – стандартизированные коэффициенты регрессии

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}},$$

для которых среднее значение равно нулю: $\bar{t}_y = \bar{t}_{x_i} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{x_i} = 1$.

Стандартизованные коэффициенты регрессии играют важную роль в анализе результатов множественной линейной регрессии, так как они дают возможность понять, на сколько единиц в среднем изменится зависимая переменная (результат), если соответствующий независимый фактор x_i увеличится на одну единицу. Это происходит при условии, что остальные факторы сохраняют свои средние значения. Анализируя стандартизованные коэффициенты, можно получить более четкое представление о том, как каждое изменение в факторах непосредственно влияет на зависимую переменную.

Поскольку все переменные в модели регрессии центрированы и нормированы, стандартизованные коэффициенты регрессии β_i становятся сопоставимыми между собой. Это значит, что значения, которые принимают эти коэффициенты, отражают относительное влияние различных независимых факторов на итоговый результат. Например, если у вас есть несколько факторов, влияющих на результат, вы сможете легко оценить, какой из них оказывает наибольшее воздействие, просто сравнив их стандартизованные коэффициенты.

Такое сопоставление позволяет ранжировать факторы по степени их влияния на результат. Если один из коэффициентов значительно больше, чем другие, это может свидетельствовать о том, что именно этот фактор является наиболее значимым в контексте данного исследования. В свою очередь, это направление внимания на наиболее влиятельные факторы может быть полезным для принятия решений и дальнейшего анализа.

Основное преимущество стандартизованных коэффициентов регрессии заключается в том, что в отличие от коэффициентов «чистой» регрессии, которые могут быть трудными для интерпретации из-за различных единиц измерения или масштабов переменных, стандартизованные коэффициенты позволяют проводить всесторонний анализ и делать выводы о влиянии переменных. Например, если один коэффициент равен 0,5, а другой – 0,2, это дает четкое понимание того, что первый фактор оказывает большее влияние на зависимую переменную, чем второй.

Стандартизованные коэффициенты также помогают избежать проблемы мультиколлинеарности, когда два или более независимых фактора могут быть сильно позитивно или негативно коррелированы друг с другом. В таких случаях сложно определить, какой именно фактор влияет на результат. Стандартизация упрощает этот процесс, так как каждый коэффициент отражает влияние фактора относительно других факторов, а не только в контексте его исходного масштаба.[29]

[29] Агаларов, З. С. Эконометрика : учебник / З. С. Агаларов, А. И. Орлов // Электронно-библиотечная система IPR BOOKS : [сайт]. – М. : Дашков и К, 2021. – 380 с. – URL : https://www.iprbookshop.ru/107834.html

Таким образом, стандартизованные коэффициенты регрессии представляют собой мощный инструмент в статистическом анализе, позволяющий не только выявить относительное влияние факторов, но и упрощать интерпретацию результатов. Благодаря возможности сравнения и ранжирования факторов, исследователи могут принимать более обоснованные решения и улучшать качество моделей на основе полученных данных.

Частные уравнения регрессии играют ключевую роль в анализе влияния отдельных факторов на зависимую переменную, позволяя исследовать, как именно изменяется результат в ответ на изменения определенного фактора. Они отражают влияние одного конкретного фактора на итоговый результат при условии, что остальные факторы остаются на постоянном уровне. Такой подход позволяет сосредоточиться на изучаемом факторе, не отвлекаясь на воздействие других переменных, которые могут повлиять на результаты исследования.

Важно отметить, что при построении частных уравнений регрессии влияние других факторов, которые также могут оказывать воздействие на результат, учитывается в свободном члене уравнения множественной регрессии. Свободный член представляет собой среднее значение зависимой переменной, когда все факторы находятся на нулевом уровне. Таким образом, этот элемент уравнения служит своего рода корректировкой, обеспечивая возможность отдельного анализа влияния интересующего фактора без искажения данных, вызванного другими переменными.

На основе частных уравнений регрессии $(\hat{y}_{x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m})$ можно также определять *частные коэффициенты эластичности*, которые помогают измерить, каким образом изменение в отдельном факторе приведет к изменению в результате:

$$Э_i = b_i \frac{x_i}{\hat{y}_{x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m}}.$$

Частный коэффициент эластичности показывает, на сколько процентов изменится зависимая переменная при изменении определенного независимого фактора на один процент, при этом предполагается, что все остальные факторы остаются неизменными. Это дает исследователям и практикам мощный инструмент для количественной оценки влияния отдельных переменных.

Частные коэффициенты эластичности и самим частные уравнения регрессии являются важными для понимания взаимосвязей между переменными в различных контекстах, таких как экономика, социология, медицина и многих других областях. Например, в экономике они могут помочь понять, как изменение цен на определенный товар влияет на спрос на него, при этом не принимая во внимание влияние других факторов, таких как доход потребителей. С помощью анализа частных уравнений регрессии исследователи могут проводить более детальные и точные оценки, что в дальнейшем позволяет грамотно интерпретировать данные и принимать обоснованные решения на основе полученных результатов.

Таким образом, частные уравнения регрессии и коэффициенты эластичности являются неотъемлемой частью современного статистического анализа, способствуя более глубокому пониманию динамики отдельных факторов в контексте исследования. Это позволяет не только выявлять закономерности, но и использовать их для разработки эффективных стратегий и прогнозов, что делает их незаменимыми инструментами в работе исследователей и аналитиков.

Средние коэффициенты эластичности представляют собой важный инструмент анализа и позволяют демонстрировать, на сколько процентов в среднем изменится езультат при увеличении соответствующего фактора на 1%. Это количественное измерение помогает исследователям и аналитикам лучше понять, насколько значимым является влияние конкретного фактора на зависимую переменную. Формула для вычисления среднего коэффициента эластичности выглядит следующим образом:

$$\overline{Э}_i = b_i \frac{\bar{x}_i}{\bar{y}}.$$

Такой подход дает возможность оценить чувствительность результата к изменениям в исследуемом факторе и позволяет в дальнейшем использовать полученные данные для построения прогнозов.

Важно отметить, что средние коэффициенты эластичности можно сравнивать друг с другом, что позволяет проводить анализ и определять, какие факторы оказывают наибольшее влияние на результат. Это дает возможность ранжировать факторы по силе их воздействия. Например, если один коэффициент эластичности составляет 0,8, а другой – 0,5, можно однозначно констатировать, что первый фактор оказывает более значительное влияние на результирующую переменную, чем второй. Такое ранжирование помогает в практических приложениях, где необходимо принять обоснованные решения на основе данных. В бизнесе, например, понимание влияния различных факторов на продажи может позволить сосредоточить маркетинговые усилия на наиболее критичных областях, повышая общую эффективность.

Анализ средних коэффициентов эластичности и их сравнение играют важную роль в принятии стратегических решений не только в бизнесе, но и в государственной политики, социальных исследованиях, экологии и других областях. Это позволяет оценивать рентабельность инвестиций, формировать бюджет, направлять ресурсы на наиболее перспективные направления и заниматься оптимизацией различных процессов.

Таким образом, средние коэффициенты эластичности не только служат инструментом для описания и понимания взаимосвязей между переменными, но и становятся основой для принятия более осознанных и взвешенных решений по различным вопросам, что в конечном итоге приводит к более эффективному использованию ресурсов и достижению поставленных целей.

Проверка существенности факторов и показатели качества регрессии. Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – коэффициента детерминации. Показатель множественной корреляции демонстрирует уровень связи между выбранным набором факторов и изучаемым показателем, а также позволяет оценить совокупное влияние факторов на результат. Независимо от характера этой связи, показатель множественной корреляции можно вычислить как индекс множественной корреляции:

$$R_{yx_1 \ldots x_m} = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}}.$$

Индекс множественной корреляции варьируется в диапазоне от 0 до 1. Чем ближе его значение к 1, тем теснее связь между результативным показателем и всеми включенными в анализ факторами. При этом значение индекса множественной корреляции должно быть не меньше максимального парного индекса корреляции.

Когда факторы корректно включены в регрессионную модель, индекс множественной корреляции существенно отличается от парного индекса корреляции. Но если в уравнение регрессии вводятся факторы, оказывающие незначительное влияние, индекс множественной корреляции может практически совпасть с парным индексом корреляции (различие будет заметным лишь в третьем-четвертом знаке после запятой). Это дает возможность сделать вывод, что сопоставление индексов множественной и парной корреляции помогает оценить необходимость включения определенного фактора в уравнение регрессии.

Для линейной зависимости признаков формула индекса множественной корреляции может быть представлена следующим образом:

$$R_{yx_1 \dots x_m} = \sqrt{\sum \beta_i r_{yx_i}},$$

где β – стандартизованные коэффициенты регрессии; r_{yx_i} – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии именуется ***линейным коэффициентом множественной корреляции или совокупным коэффициентом корреляции***.

В анализе показателей множественной корреляции учитывается остаточная дисперсия, которая имеет систематическую ошибку, склоняющуюся к занижению. Эта ошибка становится более значительной, по мере увеличения количества параметров в регрессионном уравнении при фиксированном объеме наблюдений n. Если количество параметров x_j равно m и близко к числу наблюдений, остаточная дисперсия будет стремиться к нулю, и коэффициент (индекс) корреляции может принять значение, близкое к единице, даже при слабой зависимости факторов от результата.[30]

Чтобы избежать возможного искажения оценки связи, применяется скорректированный индекс (коэффициент) множественной корреляции. Формула данного скорректированного индекса множественной детерминации выглядит следующим образом:

$$\hat{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-m-1}.$$

С увеличением величины m усиливается разница между коэффициентом множественной детерминации и скорректированным индексом множественной детерминации.

Как было упомянуто ранее, «факторы, участвующие в множественной линейной регрессии, можно ранжировать с использованием стандартизированных коэффициентов регрессии (β-коэффициентов). Этой же цели можно достичь с помощью частных коэффициентов корреляции (для линейных зависимостей)».[31] Кроме того, частные показатели корреляции широко применяют-

---

[30] Мхитарян, В. С. Эконометрика : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин // Цифровой образовательный ресурс IPR SMART : [сайт]. – М. : Евразийский открытый институт, 2012. – 224 с. – URL : https://www.iprbookshop.ru/11125.html

[31] Яковлева, А. В. Эконометрика : учебное пособие / А. В. Яковлева // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – Саратов : Научная книга, 2019. – 223 с. – URL : https://www.iprbookshop.ru/81090.html

ся при выборе факторов: обоснование включения того или иного фактора в модель можно подтвердить значением показателя частной корреляции.

Частные коэффициенты корреляции демонстрируют силу связи между результатом и соответствующим фактором после исключения воздействия других факторов, включенных в регрессионное уравнение.[32]

Показатели частной корреляции представляют собой отношение уменьшения остаточной дисперсии в результате добавления нового фактора к остаточной дисперсии, существовавшей до его включения в модель.

Частная корреляция позволяет оценить взаимосвязь между двумя переменными при фиксированном влиянии других переменных. Это особенно полезно, когда необходимо определить, насколько сильно одна переменная влияет на другую, исключая влияние других факторов.

Например, если вычисляется частный коэффициент корреляции между двумя переменными при фиксированном влиянии третьей, это означает, что мы определяем количественную меру линейной зависимости между ними, учитывающую удаление влияния этой третьей переменной.

Таким образом, показатели частной корреляции позволяют более точно и глубоко анализировать данные, выявляя скрытые взаимосвязи между переменными и исключая влияние посторонних факторов. Это делает их важным инструментом для проведения научных исследований и анализа данных в различных областях, таких как экономика, социология, психология и другие.

Частные коэффициенты корреляции, отражающие воздействие одного фактора на зависимую переменную при постоянном уровне остальных факторов, могут быть вычислены:

$$r_{yx_i|x_1 \ldots x_{i-1}x_{i+1}\ldots x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1\ldots x_m}}{1 - R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}}$$

или по рекуррентной формуле:

$$r_{yx_i|x_1 \ldots x_{i-1}x_{i+1}\ldots x_m} = \frac{r_{yx_i|x_1\ldots x_{m-1}} - r_{yx_m|x_1\ldots x_{m-1}} r_{x_i x_m|x_1\ldots x_{m-1}}}{\sqrt{(1 - r^2_{yx_m|x_1\ldots x_{m-1}})(1 - r^2_{x_i x_m|x_1\ldots x_{m-1}})}}.$$

Для двухфакторного уравнения:

$$r_{yx_1|x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \quad r_{yx_2|x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}$$

[32] Тихомиров, Н. П. Эконометрика : учебник / Н. П. Тихомиров, Е. Ю. Дорохина. – 2-е изд., стереотип. – М. : Изд-во «Экзамен», 2007. – 512 с.

или

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}};$$

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}};$$

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{yx_2})}},$$

где r – соответствующие парные коэффициенты корреляции.

$$r_{yx_1} = \frac{\overline{x_1 y} - \bar{x}_1 \bar{y}}{\sigma_y \sigma_{x_1}};$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \bar{x}_2 \bar{y}}{\sigma_y \sigma_{x_2}};$$

$$r_{x_1x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}}.$$

«Частные коэффициенты корреляции, рассчитанные с использованием рекуррентной формулы, варьируются от –1 до +1, в то время как при применении формул, основанных на множественных коэффициентах детерминации, их диапазон составляет от 0 до 1. Сравнение этих показателей позволяет упорядочить факторы по степени их влияния на результат. Частные коэффициенты корреляции отражают степень связи каждого фактора с результатом в наиболее чистом виде».[33] Если рассмотреть стандартизированное уравнение регрессии:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \cdots + \beta_m t_{x_m} + \varepsilon.$$

Следует, что $\beta_1 > \beta_2 > \beta_3$, т.е. по силе влияния на результат порядок факторов таков: x_1, x_2, x_3, то этот же порядок факторов определяется и по соотношению частных коэффициентов корреляции:

$$r_{yx_1x_2x_3} > r_{yx_2x_1x_3} > r_{yx_3x_1x_2}.$$

В эконометрике частные коэффициенты корреляции, как правило, не имеют самостоятельной ценности и применяются на стадии разработки

[33] Тихомиров, Н. П. Эконометрика : учебник / Н. П. Тихомиров, Е. Ю. Дорохина. – 2-е изд., стереотип. – М. : Изд-во «Экзамен», 2007. – 512 с.

модели. Сначала строится уравнение регрессии, включающее в себя все факторы. Затем вычисляется матрица частных коэффициентов корреляции. Далее из модели исключается фактор с наименьшим и незначительным показателем частной корреляции по *t*-критерию Стьюдента. После этого создается новое уравнение регрессии. Этот процесс повторяется до тех пор, пока все частные коэффициенты корреляции не станут существенно отличаться от нуля. Если исключается незначимый фактор, то множественные коэффициенты детерминации на двух последовательных этапах формирования регрессионной модели практически не изменяются, $R_{m+1}^2 \approx R_m^2$, где m – число факторов.

Из указанных выше формул частных коэффициентов корреляции видно, что эти показатели связаны с совокупным коэффициентом корреляции. Зная частные коэффициенты корреляции (в том числе первого, второго и более высокого порядка), можно вычислить совокупный коэффициент корреляции с помощью соответствующей формулы:

$$R_{yx_1 \ldots x_m} = \sqrt{1-(1-r_{yx_1}^2)(1-r_{yx_2x_1}^2)(1-r_{yx_3x_1x_2}^2)\ldots(1-r_{yx_mx_1x_2\ldots x_{m-1}}^2)}.$$

В частности, для двухфакторного уравнения формула примет вид:

$$R_{yx_1 \ldots x_m} = \sqrt{1-(1-r_{yx_1}^2)(1-r_{yx_2x_1}^2)}.$$

При полном соответствии результативного признака с рассматриваемыми факторами коэффициент их совокупного влияния равен единице. Это означает, что все изменения результативного признака полностью объясняются изменениями факторов. Из этого значения вычитается доля остаточной вариации результативного признака ($1 - r^2$), которая объясняется последовательно включаемыми в анализ факторами. В итоге подкоренное выражение отражает общее воздействие всех исследуемых факторов на результативный признак.

Качество созданной модели в целом оценивается с помощью *коэффициента* (*индекса*) *множественной детерминации*, который вычисляется как квадрат индекса множественной корреляции. Этот индекс фиксирует долю объясненной вариации результативного признака из-за факторов, использованных в регрессии. Влияние других факторов, не учтенных в модели, оценивается отдельно.

Если количество параметров близко к числу наблюдений, то коэффициент множественной корреляции может приблизиться к единице, даже если связь факторов с результатом слабая. Это может привести к переоценке значимости модели и ее способности предсказывать будущие значения результативного признака. Поэтому важно учитывать этот факт при интерпретации результатов регрессионного анализа. В таких случаях рекомендуется использовать дополнительные методы проверки качества модели, такие как проверка гипотез о значимости коэффициентов регрессии, анализ остатков и т.д. Эти методы позволяют оценить адекватность модели реальным данным и ее способность давать надежные прогнозы.

Чтобы минимизировать возможность перекоса результатов и избежать чрезмерного преувеличения силы связи, специалисты в области статистики применяют *скорректированный индекс множественной корреляции*. Этот индекс не только учитывает число наблюдений, но и корректирует значение коэффициента множественной корреляции в зависимости от количества параметров, включенных в модель. Таким образом, скорректированный индекс позволяет получить более адекватную и надежную оценку связи между факторами и результирующей переменной, принимая во внимание число степеней свободы. Это позволяет исследователям лучше понять истинную силу взаимодействий и принять более обоснованные решения на основе анализа данных. *Скорректированный индекс множественной корреляции* учитывает число степеней свободы:

$$\hat{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-m-1}.$$

Чем больше величина m, тем сильнее различия $\hat{R}^2$ и R^2.

Значимость частных коэффициентов корреляции проверяется так же, как и в случае с парными коэффициентами корреляции. Главное отличие заключается в том, что число степеней свободы следует определять как разницу между количеством наблюдений и количеством исследуемых факторов, включая одну дополнительную степень свободы для константы модели:

$$k = n - m - 2.$$

Значимость уравнения множественной регрессии играет ключевую роль в анализе и интерпретации результатов модели, и для ее оценки используется F-критерий Фишера. Как было отмечено ранее, важность этого критерия заключается в том, что он позволяет выяснить, есть ли статистически значимое влияние хотя бы одной из независимых переменных на результативную переменную. В ходе использования F-критерия Фишера, сначала вычисляется значение F-статистики. Это значение затем сравнивается с критическим значением F, которое извлекается из статистических таблиц с учетом заданного уровня значимости (например, 0,05) и соответствующих степеней свободы. Если вычисленное значение F-статистики оказывается выше критического значения, то нулевая гипотеза отвергается. Это означает, что по меньшей мере одна из независимых переменных действительно оказывает значительное влияние на зависимую переменную, что подтверждает значимость модели в целом.

Если полученное значение F-статистики превосходит критическое значение из таблицы, то нулевую гипотезу отвергают, и делается вывод о значимости модели:

$$F_{\text{факт}} = \frac{R^2}{1-R^2}\frac{n-m-1}{m}.$$

где n – число наблюдений; m – число параметров при переменной x; R^2 – коэффициент (индекс) множественной детерминации.

Таким образом, использование F-критерия Фишера в контексте множественной регрессии позволяет исследователям и аналитикам делать обоснованные выводы о эффективности разработанной модели и о том, насколько хорошо она описывает анализируемые данные.

Оценивается не только общая значимость регрессионной модели, но и вклад дополнительных факторов, введенных в нее. Это важно, поскольку не всякий фактор, добавляемый в модель, обязательно существенно увеличит долю объясненной вариации зависимой переменной. Также стоит учесть, что в модели может присутствовать несколько факторов, которые могут вводиться в разном порядке. Из-за наличия корреляции между факторами значимость одного и того же фактора может варьироваться в зависимости от порядка его добавления в модель. Мерой для оценки включения фактора в модель служит частный F-критерий, т.е. Fx_i.

Частный F-критерий основан на сравнении увеличения факторной дисперсии, вызванного добавлением дополнительного фактора, с остаточной дисперсией на одну степень свободы для всей регрессионной модели. В общем виде для фактора x. частный F-критерий определится как

$$F_{x_i} = \frac{R^2_{yx_1\ldots x_i\ldots x_m} - R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}{1-R^2_{yx_1\ldots x_i\ldots x_m}}\frac{n-m-1}{1}.$$

$R^2_{yx_1\ldots x_i\ldots x_m}$ – коэффициент множественной детерминации для модели с полным набором факторов;

$R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}$ – коэффициент множественной детерминации для той же модели, но без включения в модель фактора x_i; n – число наблюдений; m – число параметров в модели (без свободного члена).

Фактическое значение частного F-критерия сопоставляется с табличным при заданном уровне значимости α и соответствующим числом степеней свободы: 1 и (n – m – 1). Если фактическое значение F_x превышает табличное значение $F_{\text{табл}}(\alpha, k_1, k_2)$, это свидетельствует о том, что дополнение модели фактором x_i оправдано с точки зрения статистики, и коэффициент чистой регрессии b_i для фактора x_i будет считаться статистически значимым. В противном случае, если F_x оказывается меньше табличного значения, то включение фактора x_i в модель не приводит к значительному увеличению объясненной вариации признака y, что делает его внесение в модель нецелесообразным; коэффициент регрессии для этого фактора в таком случае будет статистически незначимым.

Для двухфакторного уравнения частные F-критерии имеют вид:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}}(n-3); \quad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}}(n-3).$$

С использованием частного F-критерия можно оценить значимость всех коэффициентов регрессии при предположении, что каждый соответствующий фактор x вводился в уравнение множественной регрессии в последнюю очередь. Частный F-критерий позволяет оценить значимость коэффициентов чистой регрессии. Зная значение F, можно также рассчитать t-критерий для коэффициента регрессии при i-м факторе, t_{b_i}, а именно:

$$t_{b_i} = \sqrt{F_{x_i}}.$$

Оценка значимости коэффициентов чистой регрессии может быть выполнена с помощью t-критерия Стьюдента и без необходимости вычисления частных F-критериев. В этом случае, аналогично парной регрессии, используется формула:

$$t_{b_i} = \frac{b_i}{m_{b_i}},$$

где b_i – коэффициент чистой регрессии при факторе x; m_b – средняя квадратическая (стандартная) ошибка коэффициента регрессии b_i.

Для уравнения множественной регрессии средняя квадратическая ошибка коэффициента регрессии может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1 \dots x_m}}}{\sigma_{x_i} \sqrt{1 - R^2_{x_1 x_2 \dots x_m}}} \frac{1}{\sqrt{n - m - 1}},$$

где σ_y – среднее квадратическое отклонение для признака y; σ_{x_i} – среднее квадратическое отклонение для признака x_i; $R^2_{yx_1 \dots x_m}$ – коэффициент детерминации для уравнения множественной регрессии; $R^2_{x_1 x_2 \dots x_m}$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии; $(n - m - 1)$ – число степеней свободы для остаточной суммы квадратов отклонений.

Таким образом, для применения этой формулы требуется наличие матрицы межфакторной корреляции и расчет соответствующих коэффициентов детерминации $R^2_{x_1 x_2 \dots x_m}$. Например, для уравнения $\hat{y} = a_0 + bx_1 + cx_2 + dx_3$ оценка значимости коэффициентов регрессии b, c, d подразумевает вычисление трех межфакторных коэффициентов детерминации $R^2_{x_1 x_2 x_3}$, $R^2_{x_2 x_1 x_3}$, $R^2_{x_3 x_1 x_2}$.

Взаимоотношения между показателями частного коэффициента корреляции, частным F-критерием и t-критерием Стьюдента для коэффициентов чистой регрессии имеют значительное значение в процессе отбора факторов, влияющих на целевую переменную. Эти связи позволяют более эффективно и точно анализировать данные, что особенно важно в контексте построения регрессионных моделей.

Исключение факторов при формировании уравнения регрессии методом исключения представляет собой подход, который реализуется на основе анализа частных коэффициентов корреляции. Этот метод подразумевает удаление фактора, который имеет наименьшее незначимое значение коэффициента корреляции на каждом этапе, что способствует оптимизации модели и повышению ее предсказательной способности.

Дополнительно, для более тщательной оценки значимости переменных могут быть использованы критические значения t-критерия для выявления силы взаимосвязей, а также частный F-критерий, который служит важным инструментом для проверки гипотез о значимости моделей.

Частный F-критерий активно применяется не только в процессе построения модели методом исключения переменных, но также и в методах включения и шагового регрессионного анализа, что позволяет гибко подходить к построению статистических моделей и улучшать их качество. Таким образом, комплексное использование этих критериев способствует созданию надежных и точных моделей, способных адекватно отражать реальные зависимости в анализируемых данных. «При включении нового фактора в регрессионную модель обычно ожидают увеличение коэффициента детерминации и уменьшение остаточной дисперсии. Если этого не происходит, значит, новый фактор не улучшает модель и скорее всего является лишним. Наличие лишних факторов не только не снижает остаточную дисперсию и не повышает коэффициент детерминации, но также может привести к статистической незначимости параметров регрессии по *t*-критерию Стьюдента».[34]

При построении уравнения множественной регрессии может возникнуть проблема мультиколлинеарности между факторами. Две переменные считаются коллинеарными, когда они находятся в линейной зависимости друг от друга, что означает, что их корреляционный коэффициент r_{ij} близок к 1. В таких ситуациях факторы могут дублировать информацию, и целесообразно исключить один из них из анализа. При этом рекомендуется выбрать тот фактор, который, несмотря на сильную связь с зависимой переменной, имеет меньшую зависимость от других факторов.

Для оценки уровня мультиколлинеарности факторов можно использовать определитель матрицы парных коэффициентов корреляции между ними. Чем ближе этот определитель к нулю, тем больше выражена мультиколлинеарность и менее надежными становятся результаты множественной регрессии. В то же время, если определитель близок к единице, это указывает на слабую мультиколлинеарность.

Для использования метода наименьших квадратов (МНК) необходимо, чтобы дисперсия остатков оставалась гомоскедастичной. Это означает, что для каждого значения фактора остатки имеют равную дисперсию. Если это условие нарушается, то наблюдается гетероскедастичность, что негативно сказывается на достоверности регрессионного анализа (рис. 16). При нарушении гомоскедастичности выполняются неравенства $\sigma^2_{\varepsilon_i} \neq \sigma^2_{\varepsilon_j} \neq \sigma^2,\ \ j \neq i.$

Для проверки выборки на гетероскедастичность можно использовать метод Гольдфельда–Квандта (при малом объеме выборки) или критерий Бартлетта (при большом объеме выборки).[35]

---

[34] Мхитарян В.С. Эконометрика : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин // Цифровой образовательный ресурс IPR SMART : [сайт]. – М. : Евразийский открытый институт, 2012. – 224 с. – URL : https://www.iprbookshop.ru/11125.html

[35] Яковлева, А. В. Эконометрика : учебное пособие / А. В. Яковлева // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – Саратов : Научная книга, 2019. – 223 с. – URL : https://www.iprbookshop.ru/81090.html

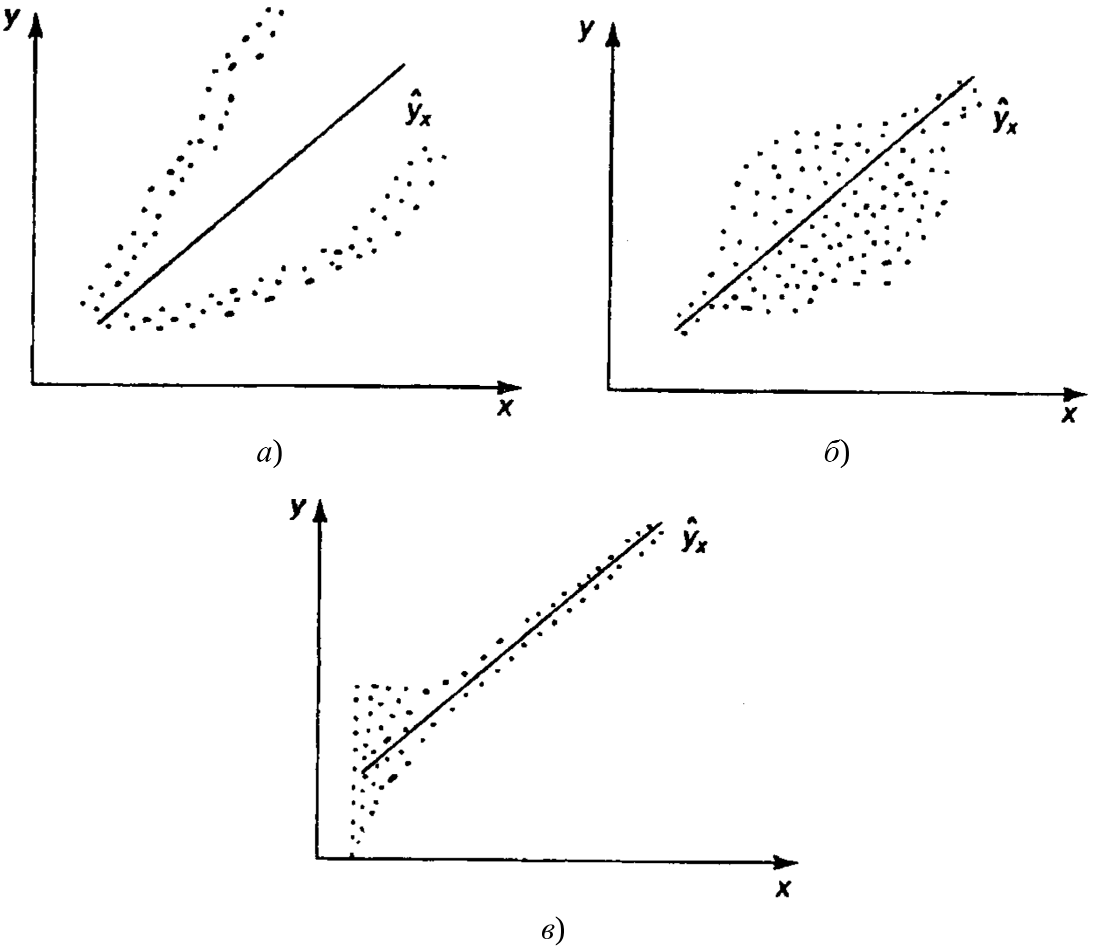


Рис. 16. Примеры гетероскедастичности

На представленном рисунке можно увидеть несколько значительных примеров гетероскедастичности, каждый из которых свидетельствует о различиях в остаточной дисперсии в зависимости от значений независимых переменных:

а) в первом примере наблюдается явное увеличение остаточной дисперсии по мере роста значений независимой переменной. Это говорит о том, что рассеяние остатков существенно возрастает с увеличением параметров, что может указывать на недостающие переменные или неправильную спецификацию модели;

б) второй пример демонстрирует ситуацию, где наибольшая дисперсия остатков фиксируется при средних значениях переменной, тогда как при крайних значениях, как высоких, так и низких, дисперсия значительно снижается. Такая ситуация говорит о наличии нестабильности в данных и может возникать в результате внезапных изменений верифицируемых факторов;

в) третий пример иллюстрирует сценарий, когда максимальная дисперсия остатков проявляется при низких значениях независимой переменной, а по мере ее увеличения эта дисперсия остается более-менее стабильной. Этот случай может свидетельствовать о том, что при небольших значениях переменной модель сталкивается с большими вариациями, создавая трудности для точного прогнозирования.

Тест Гольдфельда–Квандта является важным инструментом для проверки гипотезы о гомоскедастичности в анализе регрессионных моделей и включает в себя несколько ключевых этапов:

1. Упорядочивание данных: на самом первом этапе необходимо отсортировать все наблюдения по возрастанию или убыванию независимой переменной, для которой существует предположение о наличии гетероскедастичности. Такой порядок позволяет лучше увидеть паттерны в распределении данных.

2. Исключение центральных наблюдений: следующий шаг подразумевает исключение средних наблюдений, соблюдая условие $(n - C):2 > m$, где m – это количество оцениваемых параметров. В эмпирических исследованиях, связанных с однофакторным уравнением регрессии, рекомендуется, чтобы при $n = 30$ использовать $C = 8$, а при $n = 60$ – $C = 16$. Это помогает избежать искажения результатов за счет влияния средних значений.

3. Деление выборки: после этого необходимо разделить всю выборку на две группы: одну с малыми значениями фактора и другую – с большими значениями. На этом шаге важно определить уравнения регрессии для каждой из этих групп, что позволит детально проанализировать различия между ними.

4. Расчет остаточной суммы квадратов: в заключении необходимо вычислить остаточную сумму квадратов для каждой из групп и найти их отношение. Если нулевая гипотеза о гомоскедастичности верна, это отношение будет соответствовать F-критерию с заданными степенями свободы для каждой остатковой суммы квадратов. Чем выше величина отношения, тем значительнее нарушение предпосылки о равенстве дисперсий остаточных значений, что может требовать внимания к корректировке модели.

Обобщенный метод наименьших квадратов (ОМНК). При условии, что возникает нарушение гомоскедастичности, а также наблюдается автокорреляция ошибок, традиционный метод наименьших квадратов (OLS), который используется для оценки параметров регрессионной модели, может дать искаженные результаты. В таких случаях экспертами рекомендуется использовать *обобщенный метод наименьших квадратов* (GLS). Этот метод позволяет более эффективно обрабатывать данные, особенно когда данная информация перестает соответствовать предпосылкам классической линейной регрессии.

Главное преимущество метода GLS заключается в том, что он позволяет производить оценки, которые не только остаются несмещенными, но и обладают меньшей выборочной дисперсией по сравнению с оценками, полученными с помощью традиционных OLS. Таким образом, применение GLS становится важным шагом для повышения точности и надежности оценки параметров модели.

Теперь давайте более подробно рассмотрим, как обобщенный метод наименьших квадратов (ОМНК) может применяться для устранения гетероскедастичности. В этом контексте мы будем предполагать, что среднее значение остатка, как и в традиционной модели, по-прежнему равно нулю. Однако необходимо учитывать, что дисперсия остаточных величин уже не является постоянной, а начинает варьироваться в зависимости от значений независимой переменной.

Это варьирование дисперсии остаточных величин может быть пропорционально некоторой константе *K*. Таким образом, в процессе расчета оценок при условии гетероскедастичности будет учитывать пропорциональность между дисперсией и значениями переменной, что в итоге позволит получить более надежные и точные оценки.

Таким образом, использование обобщенного метода наименьших квадратов в условиях нарушения гомоскедастичности и наличия автокорреляции становится необходимым инструментом, который позволяет не только улучшить качество анализа регрессионной модели, но и добиться более справедливых результатов, что в конечном итоге повышает общую достоверность исследуемых данных

$$\sigma^2_{\varepsilon_i} = \sigma^2 K_i,$$

где $\sigma^2_{\varepsilon_i}$ – дисперсия ошибки при конкретном *i*-м значении фактора; σ^2 – постоянная дисперсия ошибки при соблюдении предпосылки о гомоскедастичности остатков; *K* – коэффициент пропорциональности, меняющийся с изменением величины фактора, что и обусловливает неоднородность дисперсии.

При этом предполагается, что σ^2 неизвестна, а в отношении величин *K* выдвигаются определенные гипотезы, характеризующие структуру гетероскедастичности.

В общем виде для уравнения $y = a + bx_f + \varepsilon_i$ при $\sigma^2_{\varepsilon_i} = \sigma^2 K_i,$ модель примет вид:

$$y = a = bx_i + \sqrt{K_i}\varepsilon_i.$$

В ней остаточные величины гетероскедастичны. Предполагая в них отсутствие автокорреляции, можно перейти к уравнению с гомоскедастичными остатками, поделив все переменные, зафиксированные в ходе i-го наблюдения, на $\sqrt{K_i}$. Тогда дисперсия остатков будет величиной постоянной, т.е. $\sigma^2_{\varepsilon_i} = \sigma^2$.

Иными словами, от регрессии y по x мы перейдем к регрессии на новых переменных: $\frac{y}{\sqrt{K}}$ и $\frac{x}{\sqrt{K}}$. Уравнение регрессии примет вид:

$$\frac{y_i}{\sqrt{K_i}} = \frac{a}{\sqrt{K_i}} + b\frac{x_i}{\sqrt{K_i}} + \varepsilon_i.$$

По отношению к обычной регрессии уравнение с новыми, преобразованными переменными представляет собой взвешенную регрессию, в которой переменные y и x взяты с весами $\frac{1}{\sqrt{K}}$.

Оценка параметров обновленного уравнения с трансформированными переменными приводит к использованию взвешенного метода наименьших квадратов, в котором требуется минимизировать сумму квадратов отклонений следующим образом:

$$S(a, b) = \sum_{i=1}^{n} \frac{1}{K_i}(y_i - a - bx_i)^2.$$

Соответственно получим следующую систему нормальных уравнений:

$$\begin{cases} \sum \frac{y}{K} = a\sum \frac{1}{K} + b\sum \frac{x}{K}; \\ \sum \frac{yx}{K} = a\sum \frac{x}{K} + b\sum \frac{x^2}{K}. \end{cases}$$

Если преобразованные переменные x и y взять в отклонениях от средних уровней, то коэффициент регрессии b можно определить как

$$b = \frac{\sum \frac{1}{K} xy}{\sum \frac{1}{K} x^2}.$$

При обычном применении метода наименьших квадратов к уравнению линейной регрессии для переменных в отклонениях от средних уровней коэффициент регрессии b определяется по формуле:

$$b = \frac{\sum xy}{\sum x^2}.$$

Как видно, при применении обобщенного метода наименьших квадратов для исправления гетероскедастичности коэффициент регрессии b является взвешенной величиной по сравнению с обычным методом наименьших квадратов, где вес равен $1/K$. Подобный подход может использоваться не только для парной регрессии, но и для множественной. Предположим, что рассматривается модель следующего типа.

$$y = a + b_1x_1 + b_2x_2 + \varepsilon,$$

для которой дисперсия остаточных величин оказалась пропорциональна $K_i^2 K_i$ представляет собой коэффициент пропорциональности, принимающий различные значения для соответствующих i значений факторов x_1 и x_2. Ввиду того, что

$$\sigma_{\varepsilon_i}^2 = \sigma^2 K_i^2.$$

рассматриваемая модель примет вид

$$y = a + b_1x_1 + b_2x_2 + K\varepsilon,$$

где ошибки гетероскедастичны.

Чтобы получить уравнение с гомоскедастичными остатками ε_i, мы перейдем к новым преобразованным переменным, разделив все члены исходного уравнения на коэффициент пропорциональности K. Уравнение с преобразованными переменными будет выглядеть следующим образом:

$$\frac{y_i}{K_i} = \frac{a}{K_i} + b_1\frac{x_1}{K_i} + b_2\frac{x_2}{K_i} + \varepsilon_i.$$

Это уравнение не содержит свободного члена. Вместе с тем, найдя переменные в новом преобразованном виде и применяя обычный МНК к ним, получим иную спецификацию модели:

$$\frac{y_i}{K_i} = A + \frac{a}{K_i} + b_1\frac{x_1}{K_i} + b_2\frac{x_2}{K_i} + \varepsilon_i.$$

Параметры такой модели зависят от концепции, принятой для коэффициента пропорциональности K_i. В эконометрических исследованиях довольно часто выдвигается гипотеза, что остатки ε_i пропорциональны значениям фактора. Так, если в уравнении $y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + \varepsilon$ предположить, что $e = \varepsilon x_i$, т.е. $K = x_1$ и $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i$, то обобщенный МНК предполагает оценку параметров следующего трансформированного уравнения:

$$\frac{y}{x_1} = b_i + b_2\frac{x_2}{x_1} + \ldots + b_m\frac{x_m}{x_1} + \varepsilon.$$

Применение обобщенного метода наименьших квадратов в данной ситуации приводит к тому, что наблюдения с меньшими значениями преобразованных переменных x/K оказывают большее влияние на оценку параметров регрессии по сравнению с исходными переменными. Важно отметить, что новые преобразованные переменные приобретают другое экономическое значение, и интерпретация регрессии по ним отличается от интерпретации регрессии по исходным данным.

ТЕСТ

1. Какой из перечисленных факторов НЕ является обязательным требованием к факторам, включаемым в уравнение множественной регрессии?

а) количественная измеримость;

б) отсутствие сильной корреляции между факторами;

в) наличие значимого влияния на результативный показатель, подтвержденное предварительным анализом;

г) линейная зависимость от результативного показателя;

д) объяснение значительной части вариации зависимой переменной.

2. К чему может привести включение в модель множественной регрессии высококоррелированных факторов?

а) к увеличению показателя детерминации r^2;

б) к уменьшению остаточной дисперсии;

в) к повышению точности прогноза;

г) к неустойчивости и ненадежности оценок коэффициентов регрессии;

д) к упрощению интерпретации результатов.

3. Какой показатель используется для оценки доли объясненной вариации результативного признака в модели множественной регрессии?

а) коэффициент корреляции Пирсона;

б) коэффициент детерминации r^2;

в) частный коэффициент корреляции;

г) коэффициент вариации;

д) стандартная ошибка регрессии.

4. Что характеризуют стандартизованные коэффициенты регрессии (β_i)?

а) среднее изменение результата, когда соответствующий фактор изменяется на одну единицу в исходных данных;

б) среднее изменение результата при изменении соответствующего фактора на одну стандартную единицу отклонения;

в) процентное изменение результата при увеличении соответствующего фактора на 1%;

г) абсолютное изменение результата при увеличении соответствующего фактора на одну единицу;

д) связь между результативным признаком и соответствующим фактором.

5. Какой метод используется для оценки параметров уравнения множественной линейной регрессии?

а) метод наименьших модулей;

б) метод максимального правдоподобия;

в) метод наименьших квадратов;

г) метод моментов;

д) метод средних.

6. Что такое мультиколлинеарность в контексте множественной регрессии?

а) сильная корреляция между результативным признаком и одним из факторов;

б) слабая корреляция между результативным признаком и всеми факторами;

в) линейная зависимость между двумя или более объясняющими переменными;

г) нелинейная зависимость между результативным признаком и факторами;

д) отсутствие связи между факторами и результативным признаком.

7. Как интерпретируются коэффициенты «чистой» регрессии?

а) как процентное изменение результата при изменении соответствующего фактора на 1%;

б) как среднее изменение результата при изменении соответствующего фактора на единицу, при неизменном среднем уровне других факторов;

в) как корреляция между результативным признаком и соответствующим фактором;

г) как абсолютное изменение результата при изменении всех факторов на единицу;

д) как стандартное отклонение результата при изменении соответствующего фактора.

8. Какой метод отбора факторов предполагает последовательное добавление факторов в модель?

а) метод исключения;

б) метод включения;

в) шаговый регрессионный анализ;

г) метод главных компонент;

д) метод наименьших квадратов.

9. Что показывает скорректированный коэффициент детерминации (R^2)?

а) долю объясненной вариации результативного признака, скорректированную на число факторов и наблюдений;

б) корреляцию между результативным признаком и всеми факторами;

в) среднюю ошибку прогноза;

г) силу связи между двумя переменными;

д) долю необъясненной вариации результативного признака.

10. Как проверяется значимость уравнения множественной регрессии в целом?

а) с помощью t-критерия Стьюдента;

б) с помощью критерия хи-квадрат;

в) с помощью F-критерия Фишера;

г) с помощью критерия Колмогорова–Смирнова;

д) с помощью критерия знаков.

11. Что такое частный коэффициент корреляции?

а) корреляция между результативным признаком и одним фактором, без учета других факторов;

б) корреляция между двумя факторами, без учета результативного признака;

в) корреляция между результативным признаком и одним фактором, с учетом других факторов;

г) теснота связи между всеми факторами и результативным признаком;

д) среднее значение корреляции между всеми переменными.

12. В чем заключается проблема гетероскедастичности?

а) неравенство дисперсий остатков для разных значений независимых переменных;

б) нелинейность зависимости между переменными;

в) мультиколлинеарность факторов;

г) автокорреляция остатков;

д) нарушение условия нормальности остатков.

13. Какой метод используется для проверки на гетероскедастичность при большом объеме выборки?

а) метод Гольдфельда–Квандта;

б) критерий Дарбина–Уотсона;

в) критерий Бреуша–Пагана;

г) критерий Бартлетта;

д) критерий ранговой корреляции Спирмена;

14. Какой показатель используется для оценки значимости дополнительно включенного фактора в модель?

а) коэффициент детерминации R^2;

б) частный F-критерий;

в) частный коэффициент корреляции;

г) t-критерий Стьюдента для коэффициента регрессии;

д) коэффициент множественной корреляции R.

15. При каком условии дополнительное включение фактора в модель считается статистически оправданным?

а) уменьшение остаточной дисперсии и увеличение коэффициента детерминации;

б) увеличение остаточной дисперсии и уменьшение коэффициента детерминации;

в) увеличение остаточной дисперсии и увеличение коэффициента детерминации;

г) уменьшение остаточной дисперсии и уменьшение коэффициента детерминации;

д) отсутствие изменений в остаточной дисперсии и коэффициенте детерминации.

16. Что характеризует показатель множественной корреляции *R*?

а) долю объясненной вариации результативного признака;

б) тесноту связи между результативным признаком и набором факторов;

в) среднюю ошибку прогноза;

г) значимость уравнения регрессии в целом;

д) влияние одного фактора на результативный признак при фиксированных остальных.

5. РЕГРЕССИОННЫЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ (ФИКТИВНЫЕ ПЕРЕМЕННЫЕ)

На сегодняшний день в большинстве исследовательских работ, которые сосредоточены на изучении различных экономических процессов, в качестве факторов, оказывающих влияние на результаты, чаще всего рассматриваются экономические переменные. Эти переменные представляют собой количественные показатели, которые принимают конкретные числовые значения и находятся в определенном диапазоне. Благодаря этому исследователи получают возможность количественно измерять, анализировать и оценивать влияние различных факторов на экономические результаты и процессы.

Тем не менее, в практической деятельности часто возникает необходимость учитывать также переменные качественного характера. Такие переменные могут иметь два или более уровня и не всегда могут быть сведены к простым числовым значениям. Например, факторы, связанные с качеством обслуживания клиентов или уровнем удовлетворенности, могут существенно влиять на экономические результаты, однако их сложно оценить только с помощью количественных параметров. Поэтому важно учитывать как количественные, так и качественные аспекты для более точного анализа экономических процессов и получения более глубоких выводов.

К качественным переменным можно отнести такие атрибутивные характеристики, как профессия, пол, уровень образования, климатические условия, а также принадлежность к конкретному региону. Эти характеристики играют важную роль в анализе и могут оказывать значительное влияние на экономическое поведение индивидов и организаций, однако они не могут быть легко представлены в количественной форме. Например, профессия может оказывать различное влияние на уровень дохода, а пол может влиять на доступ к определенным ресурсам или возможностям.

Чтобы интегрировать такие качественные переменные в регрессионную модель, им присваиваются цифровые, т.е. *числовые метки*. Этот процесс фактически представляет собой преобразование качественных переменных в количественные, что значительно упрощает их использование в математических расчетах и моделировании. Присвоение числовых значений позволяет количественно учитывать эффект качественных параметров, превращая их в весьма важные инструменты анализа. В области эконометрики такие преобразованные переменные получили название *фиктивных* (или dummy) переменных. Эти фиктивные переменные могут принимать только два значения – 0 или 1, что фактически обозначает присутствие или отсутствие определенного качественного признака. Такое преобразование является необходимым этапом для корректного включения качественных факторов в аналитические модели, позволяя исследователям создавать более сложные и адекватные регрессионные модели, отражающие реальное состояние дел в экономике.

В итоге, использование фиктивных переменных становится важным инструментом для глубинного анализа и повышения точности эконометрических исследований, так как позволяет не забывать о важном влиянии, которое могут оказывать качественные характеристики на экономические показатели и тенденции.

Давайте рассмотрим пример использования фиктивных переменных в функции спроса. Предположим, мы исследуем линейную зависимость потребления кофе от его цены среди людей различного пола. В общем виде уравнение регрессии для этой совокупности будет выглядеть следующим образом:

$$y = a + bx + \varepsilon,$$

где y – количество потребляемого кофе; x – цена.

Аналогичные уравнения могут быть найдены отдельно для лиц мужского пола:

$$y_1 = a_1 + b_1 x_1 + \varepsilon_1,$$

и женского пола:

$$y_2 = a_2 + b_2 x_2 + \varepsilon_2.$$

Различия в потреблении кофе проявятся в различии средних y_1 и y_2. Вместе с тем сила влияния x на у может быть одинаковой, т.е. $b \approx b_1 \approx b_2$.

В этом случае возможно построение общего уравнения регрессии с включением в него фактора «пол» в виде фиктивной переменной. Объединяя уравнения y_1 и y_2 и, вводя фиктивные переменные, можно прийти к следующему выражению:

$$y = a_1 z_1 + a_2 z_2 + bx + \varepsilon,$$

где z_1 и z_2 – фиктивные переменные, принимающие значения:

$$z_1 = \begin{cases} 1 - \text{мужской пол;} \\ 0 - \text{женский пол;} \end{cases} \qquad z_2 = \begin{cases} 0 - \text{мужской пол;} \\ 1 - \text{женский пол.} \end{cases}$$

В общем уравнении регрессии зависимая переменная y рассматривается как функция не только цены x, но и пола (z_1, z_2). Переменная z рассматривается как дихотомическая переменная, принимающая всего два значения: 1 и 0. При этом когда z_1 – 1, то z_2 – 0, и наоборот.

Для лиц мужского пола, когда $z_1 = 1$ и $z_2 = 0$, объединенное уравнение регрессии составит: $\hat{y} = a_1 + bx.$

Для лиц женского пола, когда $z_1 = 0$ и $z_2 = 1$: $\hat{y} = a_2 + bx.$ Иными словами, различия в потреблении для лиц мужского и женского пола вызваны различиями свободных членов уравнения регрессии: $a_1 \neq a_2$. Параметр b является общим для всей совокупности лиц, как для мужчин, так и для женщин.

Однако при введении двух фиктивных переменных z_1 и z_2 в модель $y = a_1 z_1 + a_2 z_2 + bx + \varepsilon$ применение МНК для оценивания параметров a_1 и a_2 приведет к вырожденной матрице исходных данных, а, следовательно, и к невозможности получения их оценок. Объясняется это тем, что при использовании МНК в данном уравнении появляется свободный член, т.е. уравнение примет вид

$$y = A + a_1 z_1 + a_2 z_2 + bx + \varepsilon.$$

Предполагая при параметре A независимую переменную, равную 1, имеем следующую матрицу исходных данных:

$$\begin{bmatrix} 1 & 1 & 0 & x_1 \\ 1 & 1 & 0 & x_2 \\ 1 & 0 & 1 & x_3 \\ 1 & 1 & 0 & x_4 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & x_n \end{bmatrix}.$$

В приведенной матрице есть линейная зависимость между первым, вторым и третьим столбцами: первый столбец представляет собой сумму второго и третьего. Это указывает на то, что матрица исходных факторов является вырожденной. Одним из решений этой проблемы может быть переход к системе уравнений.

$$y = A + A_1 z_1 + bx + \varepsilon$$

или

$$y = A + A_2 z_2 + bx + \varepsilon,$$

т.е. каждое уравнение включает только одну фиктивную переменную z_1 или z_2.

Предположим, что определено уравнение

$$y = A + A_1 z_1 + bx + \varepsilon,$$

где z принимает значения 1 для мужчин и 0 для женщин.

Теоретические значения размера потребления кофе для мужчин будут получены из уравнения

$$\hat{y} = A + A_1 + bx.$$

Для женщин соответствующие значения получим из уравнения

$$\hat{y} = A + bx.$$

Сопоставляя полученные результаты, мы можем наблюдать, что различия в уровне потребления среди мужчин и женщин заключаются в различии свободных членов, которые присутствуют в упоминаемых уравнениях: для женщин этот свободный член обозначен как A, в то время как для мужчин он представлен в виде выражения $A + A_1$. Это различие свидетельствует о том, что факторы, влияющие на потребление, могут существенно варьироваться в зависимости от пола.

Когда речь заходит о качественных факторах, которые имеют всего лишь два состояния, мы можем легко представить их в виде бинарной переменной. В этом случае одно из состояний будет соответствовать значению 1, а другое – 0, что позволяет нам просто и эффективно включать такие переменные в статистические модели. Однако ситуация усложняется, когда у качественного фактора имеется больше двух градаций или классов. В таких случаях для кодирования этих градаций в модели необходимо прибегнуть к использованию фиктивных переменных, что является распространенной практикой в подобной аналитике.

Важно отметить, что количество вводимых фиктивных переменных должно составлять на одну единицу меньше, чем общее количество градаций, существующих для данного качественного фактора. Это правило необходимо для предотвращения линейной зависимости между введенными переменными, что может негативно отразиться на адекватности модели. В случае, если бы мы решили ввести фиктивные переменные для каждой градации, это привело бы к вырожденности матрицы, что, в свою очередь, затруднило бы или сделало невозможным проведение оценки параметров нашей модели.

Таким образом, используя фиктивные переменные, мы способны эффективно моделировать качественные факторы, что не только сохраняет практическую применимость нашего анализа, но и обеспечивает статистическую корректность оценки всей модели в целом. Это, в свою очередь, позволяет нам более точно интерпретировать полученные результаты и делать обоснованные выводы на основе анализа данных.

Пример. Чтобы проанализировать зависимость цены двухкомнатной квартиры от ее полезной площади, мы можем воспользоваться регрессионным анализом. В модель можно включить как количественную переменную (полезная площадь квартиры), так и качественные переменные (тип дома). Качественные переменные вводятся в модель с помощью фиктивных переменных.

Допустим, у нас есть три типа домов: «хрущевки», панельные и кирпичные. Мы создадим две фиктивные переменные:

– $D_1 = 1$, если дом панельный, иначе 0;

– $D_2 = 1$, если дом кирпичный, иначе 0.

Таким образом, если обе фиктивные переменные равны нулю ($D_1 = 0$, $D_2 = 0$), то это означает, что дом относится к типу «хрущевка» (это базовый уровень).

Теперь наша модель зависимости цены квартиры (P) от полезной площади (A) и типа дома может выглядеть следующим образом:

$$P = \beta_0 + \beta_1 A + \gamma_1 D_1 + \gamma_2 D_2 + \varepsilon,$$

где β_0 – константа (средняя цена квартиры в хрущевке); β_1 – коэффициент, показывающий изменение цены на квартиру при изменении полезной площади на единицу; γ_1, γ_2 – коэффициенты, показывающие разницу в цене между панельными/кирпичными домами и хрущевками соответственно; ε – случайная ошибка.

Эту модель можно оценить с использованием метода наименьших квадратов или других методов регрессии. После оценки модели можно интерпретировать полученные коэффициенты:

– коэффициент β_1 покажет, насколько изменяется цена квартиры при изменении ее площади на 1 м^2;

– коэффициенты γ_1 и γ_2 покажут, насколько дороже или дешевле квартиры в панельных и кирпичных домах по сравнению с квартирами в хрущевках.

Этот подход поможет учесть влияние разных типов жилья на стоимость квартир и выявить потенциальные закономерности.

ТЕСТ

1. Что представляют собой фиктивные переменные в эконометрике?

а) качественные переменные;

б) датированные переменные;

в) статистические данные;

г) числовые переменные.

2. Какой формой является общее уравнение регрессии с включением фиктивной переменной «пол»?

а) $y = a_1z_1 + a_2z_2 + bx$;

б) $y = a + bx$;

в) $y = a + a_1z_1 + bx$;

г) $y = a + bx$.

3. Что произойдет при введении двух фиктивных переменных z_1 и z_2 в модель без исключения одной из них?

а) коэффициенты будут оценены корректно;

б) модель станет более точной;

в) матрица станет вырожденной;

г) увеличится количество данных.

4. Почему важно исключать одну фиктивную переменную из каждой группы?

а) чтобы избежать линейной зависимости;

б) для упрощения модели;

в) чтобы увеличить число переменных;

г) для повышения точности модели.

5. Какой подход помогает учесть влияние разных типов жилья на стоимость квартир?

а) регрессионный анализ;

б) метод проб и ошибок;

в) использование средних значений;

г) изолирование переменных.

6. МОДЕЛИРОВАНИЕ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

При создании эконометрических моделей используются два основных типа данных, каждый из которых имеет свои особенности и предназначение:

1. **Пространственные данные** – это информация, которая характеризует группу различных объектов в определенный момент времени. Эти данные позволяют экономистам и исследователям проводить анализ, сравнивая разные объекты или группы объектов в рамках одного временного среза. Такие модели полезны для изучения различий между регионами, отраслями или населением.

2. **Данные временных рядов** – в отличие от предыдущего типа, этот вид данных относится к одному объекту и собирается за определенные временные промежутки. Модели временных рядов используются для анализа изменения показателя во времени, что позволяет изучать его динамику и делать предсказания.

Временной ряд представляет собой последовательность значений определенного экономического показателя, зарегистрированных через равные интервалы времени. Каждый элемент такого ряда состоит из трех основных компонентов.

1. Тренд – это общее долгосрочное направление изменения показателя, которое может быть восходящим или нисходящим. Тренд отражает влияние совокупности факторов на динамику показателя на протяжении времени и позволяет исследователям видеть, как изменения в экономике или на других уровнях влияют на данный показатель в долгосрочной перспективе.

2. Цикл – представляет собой колебания, которые происходят в временном ряду на протяжении более коротких периодов по сравнению с трендом. Циклические изменения связаны, как правило, с экономическими условиями, такими как рецессии или подъемы, которые имеют временные колебания.

3. Случайный фактор – это компонента, которая не поддается прогнозированию и проявляется во временных рядах как неожиданные изменения, такие как природные катастрофы, политические события или другие случайные факторы, влияющие на показатель.

Теперь давайте рассмотрим влияние каждого из этих компонентов на временной ряд более подробно. Наличие тренда в большинстве временных рядов экономических индикаторов свидетельствует об общем долгосрочном воздействии множества факторов на изменение изучаемого показателя. Несмотря на то, что каждый из этих факторов может оказывать самостоятельное влияние, их совокупность формирует общую тенденцию, которую можно наблюдать в динамике ряда.

Восходящий тренд может свидетельствовать о устойчивом росте экономики, увеличении производительности и улучшении жизненного уровня населения, тогда как нисходящий тренд указывает на возможные экономические проблемы или снижение темпов роста. Изучение этих трендов с помощью различных эконометрических методов позволяет делать выводы о будущем развитии и определять меры для коррекции негативных тенденций (рис. 17, 18).

Таким образом, понимание структуры временных рядов и компонентов, которые на них влияют, является важным для успешного применения эконометрических моделей, что позволяет более точно анализировать и предсказывать экономические явления.

Изучаемый экономический показатель может подвергаться воздействию циклических колебаний, которые обычно возникают в результате воздействия различных рыночных и сезонных факторов. Эти колебания могут существенно влиять на анализ и интерпретацию данных, поэтому их необходимо учитывать при построении эконометрических моделей.

Иногда циклические колебания имеют ярко выраженный сезонный характер. Это означает, что экономическая активность в определенных отраслях напрямую зависит от времени года. Например, в аграрном секторе цены на сельскохозяйственную продукцию, как правило, достичь своего пика летом, когда происходит активный сбор урожая. Напротив, зимой эти цены могут снижаться из-за уменьшения предложения свежих продуктов и увеличения затрат на транспортировку.

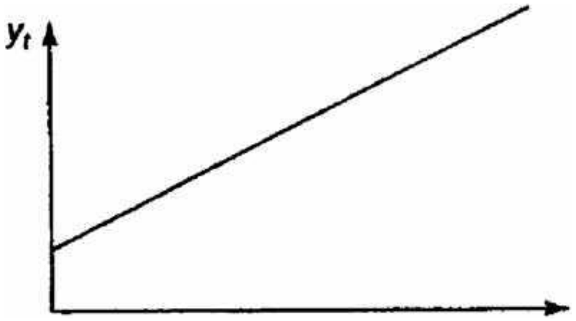


Рис. 17. Возрастающая тенденция

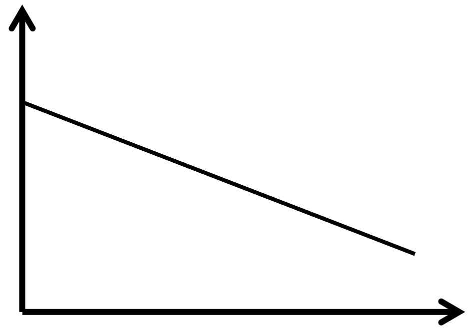


Рис. 18. Убывающая тенденция

Также стоит отметить, что в курортных городах уровень безработицы может значительно колебаться с сезоном: зимой, когда поток туристов уменьшается, количество рабочих мест может сокращаться, приводя к увеличению безработицы, тогда как летом, в сезон отпусков, занятость возрастает, и уровень безработицы снижается. Эти сезонные колебания играют важную роль в планировании экономической стратегии и прогнозировании будущих тенденций для различных отраслей.

Если у исследователей есть доступ к большому объему данных за длительный период, это открывает возможности для более глубокого анализа. В таком случае можно выявить циклы, которые связаны с общими изменениями рыночной конъюнктуры и экономическими условиями, а не только ограничиваться выявлением сезонных закономерностей. Например, они могут изучать, как в определенные годы происходят циклы boom-bust (подъемов и спадов) в отчетах о валовом внутреннем продукте, ценах на товары и другие ключевые индикаторы.

На таких больших временных рядах исследователи могут наблюдать, как различные экономические факторы взаимодействуют друг с другом, усиливая или ослабляя циклические колебания в период кризисов или роста. Это позволяет лучше понять динамику экономических процессов и заранее предсказывать возможные изменения в экономической среде.

На рисунке 19 представлен пример временного ряда, наглядно иллюстрирующего данные, состоящие исключительно из сезонной компоненты. Этот пример демонстрирует, как четко выделяются периодические колебания показателя в зависимости от сезонов, что полезно для дальнейшего анализа и построения более точных эконометрических моделей.

Таким образом, исследование циклических колебаний и их сезонной составляющей является важным аспектом эконометрического анализа, так как оно способствует более глубокому пониманию динамики экономических показателей и позволяет повышать точность прогнозов.

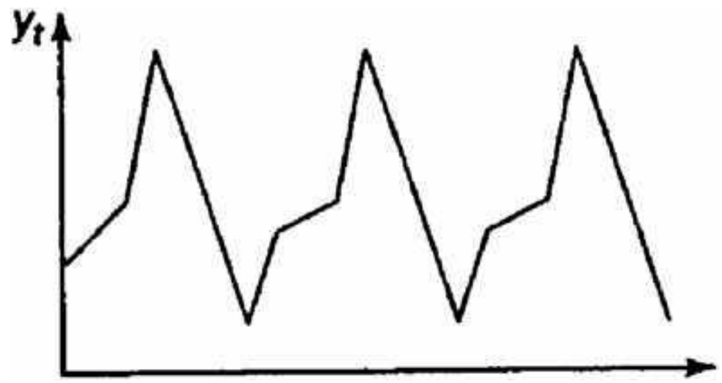


Рис. 19. Сезонные колебания

Существуют также временные ряды, в которых нет ни тенденции, ни цикличности. В таких рядах каждый последующий уровень формируется как сумма среднего значения и определенной случайной величины (как положительной, так и отрицательной).[36] На рисунке 20 представлен пример ряда, состоящего исключительно из случайной компоненты.

В большинстве ситуаций реальный уровень временного ряда можно интерпретировать как результат сложения или умножения трех ключевых компонент: трендовой, циклической и случайной. Это значит, что временной ряд, который мы наблюдаем, формируется под воздействием этих трех факторов, каждый из которых играет свою уникальную роль.

Важно отметить, что реальные данные редко соответствуют какой-либо из этих моделей в чистом виде. На практике большинство временных рядов содержат все три компонента одновременно, и их взаимодействие может быть достаточно сложным. Тренд представляет собой общий направленный поток данных, который может наблюдаться в течение длительного периода времени, тогда как сезонные колебания отражают краткосрочные изменения, связанные с определенными периодами года. Случайные флуктуации, в свою очередь, являются непредсказуемыми и могут возникать из-за различных факторов, таких как экономические события, природные катастрофы или технологические изменения.

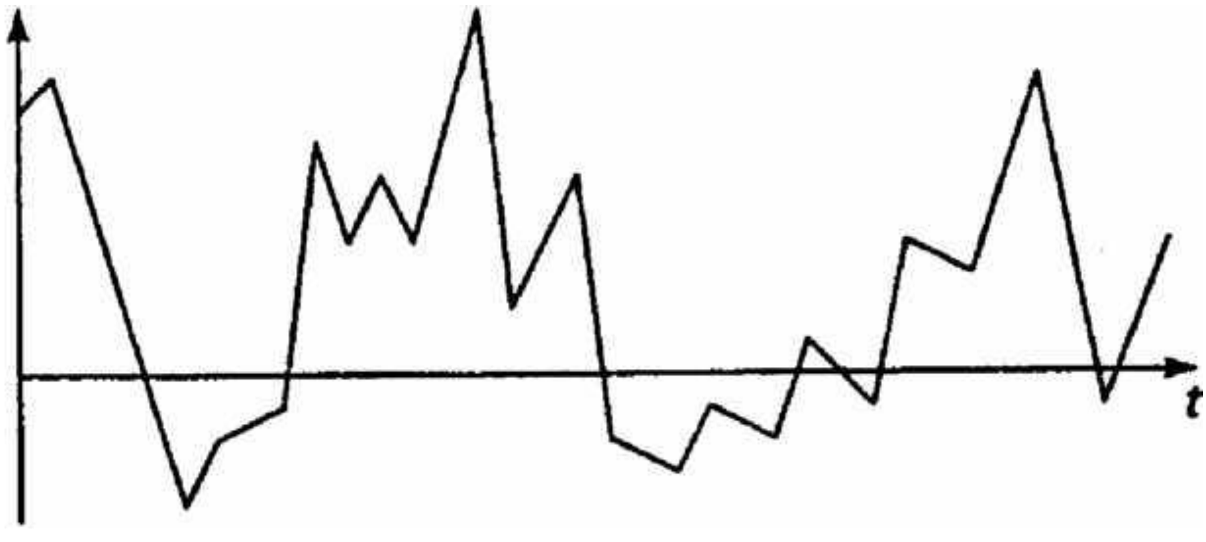


Рис. 20. Динамика, содержащая случайную компоненту

[36] Яковлева, А. В. Эконометрика : учебное пособие / А. В. Яковлева // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – Саратов : Научная книга, 2019. – 223 с. – URL : https://www.iprbookshop.ru/81090.html

При построении моделей, представляющих временной ряд, выделяют два основных подхода: аддитивные и мультипликативные модели. Модель, которая рассматривает временной ряд как сумму этих компонент, называется **аддитивной моделью**. Это означает, что каждая из компонент, включая тренд, сезонные колебания и случайные флуктуации, добавляется к общей величине ряда. С другой стороны, *мультипликативная модель* временного ряда предполагает, что эти три компоненты взаимодействуют через произведение. То есть, изменение одной компонент влияет на остальные, что приводит к большему или меньшему общему значению ряда. Этот подход часто используется в тех случаях, когда изменения в уровне одного из факторов могут значительно повлиять на общую величину.

Основная цель эконометрического анализа отдельного временного ряда заключается в обнаружении, выявлении и количественном описании каждой из упомянутых компонент, что включает в себя не только тренды, сезонные колебания, но и случайные флуктуации, которые могут оказывать влияние на общий динамический процесс. Углубленный анализ этих компонентов жизненно важен для более точного предсказания будущих значений ряда. Понимание этих свойств, в свою очередь, позволяет более эффективно анализировать ожидаемые изменения в данных и корректировочные действия, которые могут быть предприняты на основании предсказательных результатов.[37]

Более того, результаты такого анализа могут быть использованы не только для предсказания, но и для разработки моделей взаимосвязи между несколькими временными рядами. Это многообразие подходов имеет огромное значение, поскольку позволяет строить более сложные и точные экономические прогнозы, учитывающие взаимное влияние различных экономических индикаторов друг на друга. Например, анализ взаимосвязей между ценами на товары и уровнем спроса может дать бесценную информацию для принятия обоснованных управленческих решений.

Построение как аддитивных, так и мультипликативных моделей требует тщательных расчетов значений для каждого уровня ряда. Каждая из этих моделей имеет свои особенности, которые необходимо учитывать при анализе данных. Например, в аддитивной модели компоненты простым сложением предоставляют результат, тогда как в мультипликативной модели предполагается, что компоненты взаимодействуют друг с другом по умножению. Это значит, что правильный выбор типа модели может значительно повлиять на конечные результаты анализа.

---

[37] Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.

Прежде чем приступить к разработке модели, необходимо провести предварительную проверку исходных данных на несколько ключевых критериев. Эта предварительная проверка может включать в себя анализ на наличие выбросов, проверку на стационарность, а также оценку периодичности данных. Такие шаги позволят избежать ошибок и сделать анализ более надежным, что в конечном итоге приведет к созданию эконометрических моделей, которые обладают высокой точностью и могут быть эффективно использованы для принятия стратегических решений. Правильный подход на этом этапе – залог успешного анализа и его практического применения в дальнейшем.

Прежде всего, данные должны быть сопоставимыми, что предполагает использование единой методики сбора или расчета данных. Далее, необходимо убедиться в однородности данных, что означает отсутствие случайных выбросов, способных исказить результаты анализа. Устойчивость данных также является важным аспектом: наличие закономерностей в изменениях уровней ряда подтверждает, что данные поддаются анализу и вполне предсказуемы. И, наконец, данные должны быть достаточными для анализа; число наблюдений должно в 7-10 раз превышать количество параметров модели. Это важно для обеспечения статистической значимости и надежности предложенной модели.

Автокорреляция уровней временного ряда. Автокорреляция уровней временного ряда возникает в тех случаях, когда в этих данных выражены тренды и циклические колебания. Это явление наблюдается тогда, когда последовательные уровни временного ряда зависят друг от друга, что означает, что значение в одном моменте времени может подсказать о значении в другом. Такую корреляционную связь между последовательными уровнями временного ряда принято называть *автокорреляцией*. Это свойство отражает внутренние зависимости данных, благодаря которым можно выявить закономерности и предсказать будущие значения ряда на основе уже существующих. Например, если значения временного ряда имеют устойчивый восходящий или нисходящий тренд, то будущие уровни, вероятно, будут находиться в том же направлении.

Автокорреляцию можно количественно оценить с помощью линейного коэффициента корреляции, который является одним из основных инструментов в статистическом анализе временных рядов. Этот коэффициент не только позволяет измерять степень связи между уровнями исходного временного ряда и уровнями того же ряда, но уже смещенными на определенное количество временных интервалов, но и предоставляет довольно глубокое понимание структуры данных.

Фактически, автокорреляция отражает, насколько текущее значение ряда зависит от его предыдущих значений. Это очень полезная информация для аналитиков и экономистов, поскольку она помогает выявить паттерны и тренды в данных. Например, если мы рассматриваем временной ряд с ежедневными данными, мы можем исследовать, как значения одного дня связаны с значениями, наблюдаемыми на следующий, а также через несколько дней. Это может быть особенно полезно в таких областях, как финансовый анализ, где предсказание рыночных трендов часто основывается на анализе предыдущих значений.

Коэффициент автокорреляции дает возможность аргументированно оценить, насколько сильно текущие значения могут предсказывать будущие, и это может значительно упростить процесс построения прогнозных моделей. При этом необходимо учитывать, что автокорреляция может варьироваться в зависимости от сезонности и других циклических факторов, которые могут влиять на данные. Поэтому важно разрабатывать обоснованные модели, учитывающие все эти нюансы, что, в свою очередь, требует глубокой аналитической работы.

Кроме того, используя коэффициент автокорреляции, аналитики могут оценивать, когда временной ряд имеет значительную зависимость от своих собственных предыдущих значений, что может указать на наличие постоянных трендов или циклических колебаний. Это может быть дальнейшим индикатором для принятия более обоснованных управленческих решений и оптимизации бизнес-процессов. В конечном счете, исследование автокорреляции временных рядов является неотъемлемой частью эконометрического анализа и позволяет более четко и точно понять динамику данных, что в свою очередь опирается на надежные статистические методы и приемы.

Коэффициент корреляции может принимать значения от –1 до 1. Значение 1 указывает на полную положительную связь, –1 – на полную отрицательную связь, а 0 указывает на отсутствие связи. Значения, близкие к 1 или –1, свидетельствуют о сильной автокорреляции, что в свою очередь может быть признаком наличия формальных трендов или сезонных колебаний в данных.

Выявление и анализ автокорреляции крайне важно для построения надежных прогнозных моделей, так как эта информация помогает лучше понять структуру временного ряда и сделать более обоснованные предсказания о его будущих изменениях. Более того, понимание автокорреляции может также играть ключевую роль в оценке сроков и периодичности, поможет в планировании ресурсов и оценке рисков в различных сценариях, особенно в экономическом и финансовом анализе.

Формула для вычисления коэффициента автокорреляции выглядит следующим образом:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}},$$

где $\bar{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t$, $\bar{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}$.

Аналогично определяются коэффициенты автокорреляции второго и более высоких порядков. Например, коэффициент автокорреляции второго порядка:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}},$$

где $\bar{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t$, $\bar{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}$.

Количество периодов, применяемых для вычисления коэффициента автокорреляции, называется лагом. *Лаг* представляет собой важную характеристику временных рядов, поскольку он напрямую влияет на количество пар значений, используемых в расчете коэффициента. С увеличением лага число пар значений, задействованных в расчете, уменьшается, что следует учитывать при проведении анализа, так как это может существенно повлиять на конечные результаты. Рекомендуется следовать общепринятому правилу, согласно которому максимальный лаг не должен превышать $n/4$, где n – это общее количество наблюдений. Это правило призвано обеспечить статистическую надежность коэффициентов автокорреляции и повысить вероятность того, что выводы, сделанные на основе этих коэффициентов, будут корректными и достоверными.

Основные свойства коэффициента автокорреляции[38]:

1. **Метод расчета:** коэффициент автокорреляции вычисляется по принципу линейного коэффициента корреляции и, следовательно, отражает только

---

[38] Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.

линейную взаимосвязь между текущими и предшествующими уровнями ряда. Исходя из значения этого коэффициента, можно судить о наличии линейной (или близкой к ней) тенденции в данных. Это важно, поскольку наличие линейной зависимости может указывать на предсказуемость ряда. Однако в случае временных рядов, обладающих ярко выраженной нелинейной тенденцией (например, параболической или экспоненциальной), коэффициент автокорреляции может стремиться к нулю, что может вводить в заблуждение исследователей.

2. **Интерпретация знака:** по знаку коэффициента автокорреляции невозможно однозначно определить, растет или падает тенденция ряда. Важно учитывать, что во многих временных рядах экономических данных может наблюдаться положительная автокорреляция, но при этом общая тенденция может быть нисходящей. Такая ситуация свидетельствует о том, что данные могут иметь сложную структуру и требует более глубокого анализа для понимания происходящих процессов.

Последовательность коэффициентов автокорреляции первого, второго и последующих порядков называется автокорреляционной функцией временного ряда. График этой функции, показывающий зависимость от величины лага (порядка коэффициента автокорреляции), называется *коррелограммой*. Коррелограмма представляет собой важный инструмент в визуализации и анализе автокорреляции временных рядов, так как помогает исследователям сразу увидеть, на каких лагах наблюдается наибольшая корреляция.[39]

Анализ автокорреляционной функции и коррелограммы является важным инструментом в исследовании временных рядов. Он помогает определить задержку или лаг, при котором автокорреляция достигает своего максимального значения. Это максимальное значение лога указывает на период, в котором связь между текущими и прошлыми уровнями ряда наиболее сильно выражена. Такой анализ позволяет исследователю выявить глубинные зависимости в данных и понять, какие факторы могут влиять на изменения временного ряда. Понимание структуры и динамики временных рядов является ключевым моментом для успешного прогнозирования и принятия решений в сфере экономики, финансирования и других областях.

Такой анализ не только позволяет определить наилучший лаг для оценки автокорреляции, но и раскрывает структуру временного ряда.

[39] Агаларов, З. С. Эконометрика : учебник / З. С. Агаларов, А. И. Орлов // Электронно-библиотечная система IPR BOOKS : [сайт]. – М. : Дашков и К, 2021. – 380 с. – URL : https://www.iprbookshop.ru/107834.html

Если максимальный коэффициент автокорреляции принадлежит первому порядку, это указывает на наличие только одной тенденции в исследуемом ряде. Это означает, что изменения во времени происходят последовательно, и значения, полученные в одном временном интервале, имеют тенденцию смещаться в одном и том же направлении в следующем. Такой вид взаимодействия может быть полезен для понимания общего направления данных, например, в случае роста или падения экономических показателей.

Если самый высокий коэффициент автокорреляции приходится на порядок t, это свидетельствует о наличии циклических колебаний с периодом в t периодов. Данный случай может указывать на то, что временной ряд демонстрирует повторяющиеся тенденции на определенных интервалах, что может быть связано с сезонными изменениями, экономическими циклами или другими предсказуемыми влияниями. Для практического применения это означает, что можно прогнозировать будущие значения ряда, основываясь на таких циклических паттернах.

Если ни одному из коэффициентов автокорреляции не удается достичь значимого уровня, это может привести к двум основным выводам о структуре временного ряда. Во-первых, это может свидетельствовать о том, что в ряде отсутствуют как тенденции, так и циклические колебания. Во-вторых, это также может указывать на наличие сильной нелинейной тенденции, которая требует дальнейшего анализа для выявления скрытых закономерностей. В таких случаях может потребоваться применение более сложных статистических методов или моделей, таких как регрессионный анализ или временные модели, учитывающие сложные взаимосвязи.

Таким образом, использование коэффициента автокорреляции уровней и автокорреляционной функции является важным этапом для обнаружения наличия или отсутствия трендов и циклических (сезонных) компонентов во временном ряде. Это позволяет формировать более точные прогнозы и принимать обоснованные решения на основе анализа данных. В конечном счете, понимание структуры временного ряда помогает не только в научных исследованиях, но и в практических применениях, таких как финансовый анализ, прогнозирование спроса и управление производственными процессами.

Моделирование тенденции временного ряда. Моделирование тенденции временного ряда является ключевой задачей классического анализа экономических временных рядов. Это важный процесс, поскольку выявление и корректное представление долгосрочных тенденций позволяет экономистам, аналитикам и другим специалистам лучше понимать динамику изменений во времени и делать более точные прогнозы. Решение этой задачи обычно начинается с тщательной проверки наличия тенденции в данных и формирования предположений о характере долгосрочного тренда. На данном этапе важно не только подтвердить наличие тренда, но и оценить его направление, силу и стабильность.

После этого, на основе сделанных выводов, строится модель тенденции, рассматриваемая как функция времени. Моделирование позволяет детализировать и формализовать наблюдения, что, в свою очередь, открывает возможности для дальнейшего анализа и прогнозирования. Разработанная модель может служить основой для принятия решений и построения стратегий в различных областях, будь то бизнес, экономика или социальные науки.

Метод сравнения средних. Метод сравнения средних применим для выявления монотонной тенденции.

Временной ряд разбивается на две примерно равные части y_1, y_2, ..., y и $y_n + 1$, $y_n + 2$, ..., $y_n + n$ с количеством уровней n_1 и n_2 и для каждой части вычисляются средние (y_1, y_2) и выборочные дисперсии (s_1, s_2) соответственно.

Далее рассчитывается значение критерия Стьюдента по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{S_1^2}{n_1}} + \sqrt{\frac{S_2^2}{n_2}}},$$

если предполагается, что значения дисперсий на этих участках не равны между собой, т.е. $\sigma_1^2 \neq \sigma_2^2$ и по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{S^2}\sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

где S^2 – общая выборочная дисперсия ряда, если предполагается, что дисперсии одинаковы: $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Расчет генеральной дисперсии показал (табл. 3), что $\sigma_1^2 = 15 > \sigma_2^2 = 9{,}83$.

При этом выборочные дисперсии равны $S_1^2 = 18{,}01$ и $S_2^2 = 11{,}79$, тогда

$$\tau = \frac{|35{,}82 - 53{,}786|}{\sqrt{\frac{18{,}01}{6}} + \sqrt{\frac{11{,}79}{6}}} = 5{,}733.$$

Нулевая гипотеза о равенстве средних (об отсутствии тенденции) отвергается, если выполняется условие:

$$\tau > t_{\alpha,m}, \quad \tau > t_{1-\alpha,m},$$

где $t_{1-\alpha,m}$ – табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы ($m = n_1 + n_2 - 2$).

$$5{,}733 > 2{,}23.$$

Таким образом, метод средних указывает на наличие тенденции в приведенных данных.

Метод Фостера–Стюарта считается более универсальным и обеспечивает более достоверные результаты. Каждому уровню ряда y, начиная со второго, ставятся в соответствие два значения p и $\$q$ согласно следующим правилам:

$p_i = 1$, если уровень y_i меньше всех предыдущих уровней;

$p_i = 0$, в противном случае;

$q_i = 1$, если уровень y_i больше всех предыдущих уровней;

$q_i = 0$, в противном случае.

Тогда t-критерий определяется по формуле

$$t_p = \frac{\sum_{i=2}^{n}(p_i - q_i)}{2\sum_{i=2}^{n}\frac{1}{i}}.$$

Гипотеза об отсутствии тенденции отвергается, если выполняется условие

$$t_p > t_{1-\alpha,n-1},$$

где $t_{\alpha,n-1}$ – табличное значение t-критерия Стьюдента при уровне значимости α при числе степеней свободы ($n - 1$).

Метод укрупнения интервалов применяется, чтобы выявить тренд в рядах динамики. Этот метод основан на укрупнении периодов времени, к которым относятся уровни ряда. Например, ряд ежемесячного выпуска продукции заменяется рядом ежеквартального и т.д. После укрупнения интервалов с месячных данных на квартальные можно сразу же увидеть основную тенденцию – увеличение производства скота и птицы на убой в живом весе из квартала в квартал.

Метод скользящей средней. Сглаживание динамического ряда с помощью скользящей средней заключается в расчете среднего значения из определенного количества начальных уровней ряда, затем – среднего значения из такого же количества уровней, начиная со второго, потом – начиная с третьего и так далее. Таким образом, при расчетах средние значения как бы «скользят» вдоль ряда от его начала к концу, каждый раз отбрасывая один уровень в начале и добавляя один следующий. Именно поэтому этот метод получил название «скользящая средняя». Каждое звено скользящей средней представляет собой среднее значение за соответствующий период, которое относится к середине выбранного промежутка.

Используя данные за первые три месяца, рассчитываются трехмесячные суммы, а затем находится среднее значение. Интервал сглаживания можно выбирать и четным числом (четыре, шесть и т.д.). Вычисление скользящей средней по четному количеству членов усложняется тем, что среднее значение может относиться к середине между двумя датами. Чтобы устранить этот сдвиг, применяется процедура центрирования, т.е. нахождение средней из средних для привязки полученного уровня к конкретной дате. При центрировании также необходимо рассчитывать скользящие суммы, скользящие средние по этим суммам и средние из средних. Взяв данные за первые четыре месяца, исчисляем четырехчленные суммы, а затем среднюю.

Применение в анализе рядов динамики методов укрупнения интервалов и скользящей средней позволяет выявить тренд для его описания, но получение обобщенной статистической оценки тренда достигается методом аналитического выравнивания.

Аналитическое выравнивание – это процесс выявления основной тенденции развития изучаемого явления во времени. Развитие представляется исследователю как зависящее преимущественно от течения времени. В результате выравнивания временного ряда получается наиболее общее, суммарное отражение во времени результата действия всех причинных факторов. Отклонения конкретных уровней ряда от уровней, соответствующих общей тенденции, объясняются действием случайных или циклических факторов.

В результате приходят к трендовой модели

$$У_t = f(t) + \varepsilon_t,$$

где $f(t)$ – уровень, определяемый тенденцией развития; e_t – случайное и циклическое отклонение от тенденции.

Целью аналитического выравнивания динамического ряда является нахождение аналитической или графической зависимости $f(t)$. На практике для имеющегося временного ряда определяют вид и параметры функции $f(t)$, а затем анализируют отклонения от выявленной тенденции. Вид функции $f(t)$ выбирают таким образом, чтобы он обеспечивал осмысленное объяснение изучаемого процесса.

Чаще всего при выравнивании используются следующие зависимости:

– линейная регрессия $\hat{y}_t = a + bt$;

– полиномы разных степеней $\hat{y}_t = a + b_1 t + b_2 t^2 + \cdots + b_m t^m$;

– гипербола $\hat{y}_t = a + \dfrac{b}{t}$;

– степенная $\hat{y}_t = at^b$;

– показательная $\hat{y}_t = ab^t$;

– экспоненциальная $\hat{y}_t = e^{a+bt}$.

Параметры функций оцениваются методом наименьших квадратов (МНК), который минимизирует сумму квадратов отклонений фактических уровней от выровненных. Для нелинейных трендов предварительно выполняется стандартная процедура их линеаризации.

Есть несколько способов определения типа тенденции. Среди наиболее популярных методов – **качественный анализ изучаемого процесса и построение**, а также визуальный анализ графика зависимости уровней ряда от времени.

Определение вида тенденции на основе качественного анализа. Социально-экономические процессы, в зависимости от характера их протекания, можно разделить на три класса (см. рис. 21).

Процессы с монотонным характером развития и отсутствием пределов роста (рис. 21, *а*). Эти условия справедливы для поведения многих экономических показателей, например, для большинства натуральных показателей промышленного производства. В этом случае для моделирования тенденции могут использоваться функции: линейная $\hat{y}_t = a + bt$, полиномы разных степеней $\hat{y}_t = a + b_1 t + b_2 t^2 + \cdots + b_m t^m$, степенная $\hat{y}_t = at^b$, экспоненциальная $\hat{y}_t = e^{a+bt}$.

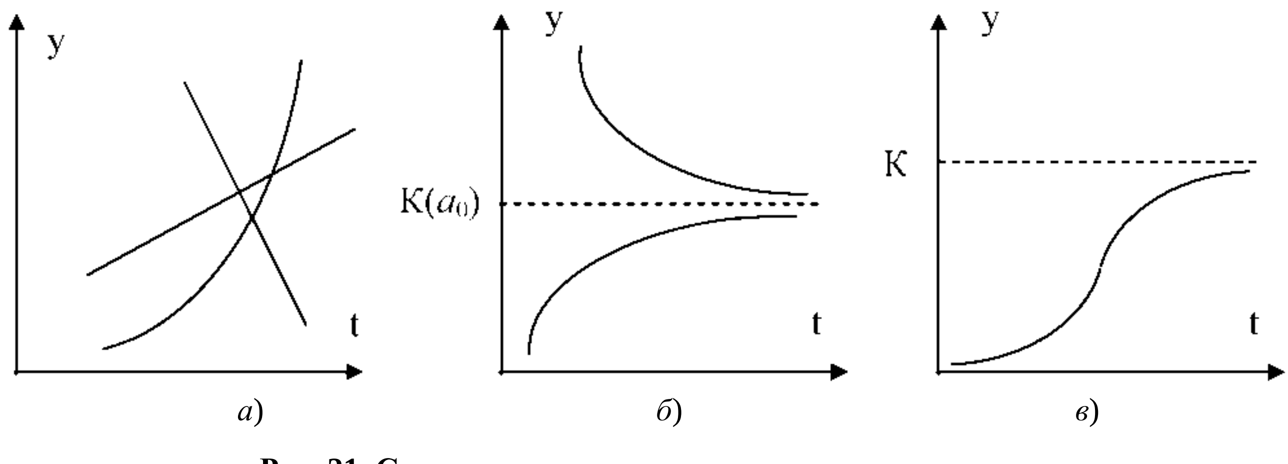


Рис. 21. Схемы протекания динамических процессов:
а – I класс; *б* – II класс; *в* – III класс

Процессы, которые достигают предела роста (спада) в течение исследуемого периода, известные как процессы с «насыщением» (рис. 21, *б*). Развитие таких процессов происходит под влиянием некоторых сдерживающих факторов, сила воздействия которых усиливается по мере достижения определенного уровня. Подобные процессы часто встречаются в демографических исследованиях, при анализе потребительских нужд (рассчитанных на душу населения), при оценке эффективности использования ресурсов и т.п. Примеры показателей, для которых могут существовать пределы роста, включают, например, среднедушевое потребление определенных видов продуктов питания, затраты удобрений на единицу площади и другие подобные индикаторы. В этом случае для моделирования тенденции используются гиперболическая функция $\hat{y}_t = a + \frac{b}{t}$, либо модифицированная экспонента $y_t = K + a_0 a_1^t$ с параметром a_1, удовлетворяющим условию $0 < a_1 < 1$.

В случае гиперболы параметр a_0 равен пределу роста, к которому значение уровня процесса приближается (при росте t) снизу в случае $a_i < 0$, либо сверху при $a_1 > 0$ (рис. 21, *б*).

В случае модифицированной экспоненты параметр K равен пределу роста, к которому значение уровня процесса приближается (при росте t) снизу в случае $a_0 < 0$, либо сверху при $a_0 > 0$ (рис. 21, *б*).

При решении экономических задач часто можно определить значение предела роста исходя из свойств прогнозируемого процесса (например, коэффициент использования оборудования не может превышать 1 и т.п.). Иногда значение предела роста задается экспертным путем.

Процессы, именуемые S-образными (рис. 21, *в*), представляют собой два последовательных процесса лавинного характера (где прирост зависит от достигнутого уровня): первый характеризуется ускоренным развитием, второй – замедленным. Такие процессы часто встречаются в демографических исследованиях, в страховании, при прогнозировании научно-технического прогресса, а также при оценке спроса на новую продукцию.

К S-образным процессам можно отнести процесс развития новой отрасли (нового производства). Сначала производство развивается крайне медленно, поскольку технологии еще недостаточно развиты, производственные расходы остаются высокими, а рыночный спрос на продукт пока невелик, что замедляет темпы роста. Впоследствии, благодаря совершенствованию технологий, переходу к массовому производству и расширению рынка сбыта, темпы роста начинают ускоряться. Позже наступает этап насыщения рынка, рост производства постепенно замедляется и в конечном счете почти останавливается. Производство стабилизируется на определенном уровне. Стоит отметить, что выявленные закономерности развития следует обобщать с известной долей осторожности, и делать это только для сравнительно коротких периодов, так как выявленная динамика производства может нарушаться под влиянием внешних факторов, таких как технологическая революция в данной отрасли или изменения, связанные с ней.

Для моделирования тенденции *S*-образных процессов следует использовать либо логистическую функцию $y_t = \dfrac{K}{1 + a_0 e^{-a_1 t}}$ (с параметром $a_1 < 1$), либо кривую Гомперца $y_t = K a_0 a_1^t$ с параметрами, удовлетворяющими условиям $0 < a_0, a_1 < 1$. Предел роста в обоих случаях равен параметру K.

Выбор вида тенденции на основе анализа показателей динамики временного ряда. Линейная зависимость выбирается в тех случаях, когда в исходном временном ряду наблюдаются более или менее постоянные абсолютные цепные приросты, которые не демонстрируют ни увеличения, ни уменьшения.

Параболическая зависимость используется, если абсолютные цепные приросты показывают определенную динамику, но абсолютные цепные приросты абсолютных цепных приростов (разницы второго порядка) не проявляют никаких изменений.

Экспоненциальные зависимости применяются, если в исходном временном ряду наблюдается устойчивый относительный рост (постоянство цепных темпов роста, темпов прироста, коэффициентов роста) или, при отсутствии такого постоянства, стабильность в изменении показателей относительного роста (цепных темпов роста, цепных темпов роста, цепных коэффициентов роста).

Использование коэффициентов автокорреляции уровней ряда. Тип тенденции можно определить, сравнив коэффициенты автокорреляции первого порядка, рассчитанные по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его смежные уровни y_t и y_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка для исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, экспоненту, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем аналогичный коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем больше различаются значения этих коэффициентов.

Высокие значения коэффициентов автокорреляции первого, второго и третьего порядков свидетельствуют о том, что ряд содержит тенденцию. Приблизительно равные значения коэффициентов автокорреляции по уровням этого ряда и по логарифмам уровней позволяют сделать следующий вывод: если ряд содержит нелинейную тенденцию, то она выражена в неявной форме. Поэтому для моделирования его тенденции в равной мере целесообразно использовать и линейную, и нелинейную функции. Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

Параметры линейного и экспоненциального трендов имеют самую простую экономическую интерпретацию.

Параметры линейного тренда можно трактовать следующим образом: a – начальный уровень временного ряда в момент времени $t = 0$; b – средний за период абсолютный прирост уровней ряда.

Параметры экспоненциального тренда имеют следующую интерпретацию. Параметр a – это начальный уровень временного ряда в момент времени $t = 0$. Величина e^b – это средний за единицу времени коэффициент роста уровней ряда.

Как и в линейной модели, расчетные значения уровней ряда по экспоненциальному тренду можно получить путем подстановки в уравнение тренда значений $t = 1, 2, ..., n$, а также в соответствии с интерпретацией параметров экспоненциального тренда: каждый последующий уровень ряда является произведением предыдущего уровня на соответствующий коэффициент роста.

При наличии неясной нелинейной тенденции следует дополнить перечисленные выше методы выбора оптимального уравнения тренда качественным анализом динамики изучаемого показателя, чтобы избежать ошибок спецификации при выборе типа тренда. Качественный анализ предполагает исследование возможных изменений в динамике временного ряда, таких как появление поворотных точек или изменения темпов прироста (ускорения темпов прироста) начиная с какого-то момента (периода) времени под влиянием различных факторов. Если уравнение тренда было выбрано неправильно, особенно при больших значениях t, результаты анализа и прогнозирования динамики временного ряда с использованием выбранного уравнения окажутся ненадежными из-за ошибки спецификации. Пример возможной ошибки спецификации показан на рис. 22.

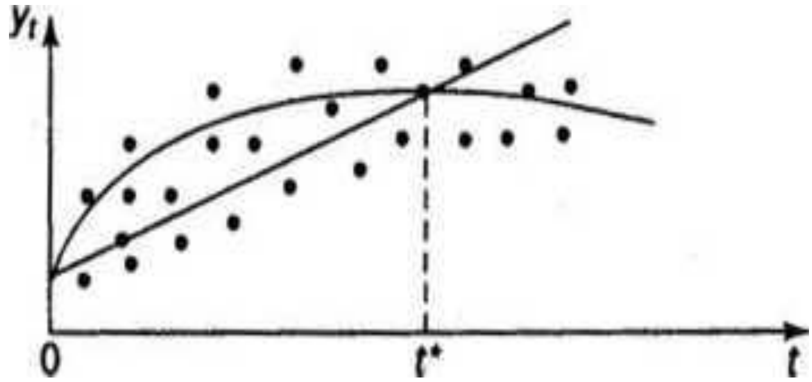


Рис. 22. Ошибка спецификации при выборе уравнения тренда

Если считать, что наилучшая форма тренда – парабола второго порядка, хотя на самом деле имеет место линейная тенденция, то при больших значениях t парабола и линейная функция будут по-разному описывать тенденцию в уровнях ряда. При $t > t^*$ парабола второго порядка будет показывать убывающую тенденцию в уровнях ряда y_t, тогда как линейная функция будет описывать возрастающую тенденцию.

Моделирование периодических колебаний. Самый простой способ моделирования сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и создание аддитивной или мультипликативной модели временного ряда.

Общая структура аддитивной модели выглядит так:

$$Y = T + S + E.$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = TSE.$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (T), сезонной (S) и случайной (E) компонент.

Решение о выборе между аддитивной и мультипликативной моделями принимается на основе анализа структуры сезонных колебаний. Если амплитуда колебаний примерно одинакова, строится аддитивная модель временного ряда, в которой значения сезонной компоненты рассматриваются как постоянные для различных циклов. Если амплитуда сезонных колебаний увеличивается или уменьшается, создается мультипликативная модель временного ряда, которая устанавливает зависимость уровней ряда от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S и E для каждого уровня ряда.

Процесс построения модели следующий:

Шаг 1. Выравнивание исходного ряда методом скользящей средней.

Шаг 2. Расчет значений сезонной компоненты S

Шаг 3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных в аддитивной $(T+E)$ или в мультипликативной $(T \cdot E)$ модели.

Шаг 4. Аналитическое выравнивание уровней $(T+E)$ или $(T \cdot E)$ и расчет значений T с использованием полученного уравнения тренда.

Шаг 5. Расчет полученных по модели значений $(T+S)$ или $(T \cdot S)$.

Шаг 6. Расчет абсолютных и(или) относительных ошибок.

Моделирование сезонных колебаний с помощью фиктивных переменных. Рассмотрим метод моделирования временного ряда, содержащего сезонные колебания, основанный на включении в модель фиктивных переменных. «Количество фиктивных переменных принимается равным числу наблюдений в пределах одного цикла колебаний без единицы».[40] Например, при моделировании поквартальных данных необходимо ввести три дополнительные переменные:

$$z_1 = \begin{cases} 1, \text{ весна;} \\ 0, \text{ невесна;} \end{cases} \quad z_2 = \begin{cases} 1, \text{ лето;} \\ 0, \text{ нелето;} \end{cases} \quad z_3 = \begin{cases} 1, \text{ осень;} \\ 0, \text{ неосень.} \end{cases}$$

Зиме в этом случае соответствуют нулевые значения всех фиктивных переменных. Уравнение регрессии с учетом фиктивных переменных принимает вид

$$y = a + bt + c_1z_1 + c_2z_2 + c_3z_3 + s.$$

Коэффициенты с характеризуют отклонение уровней первых трех сезонов по отношению к последнему. Поэтому модель с фиктивными переменными может рассматриваться как частный случай аддитивной модели временного ряда.

Моделирование сезонных колебаний с помощью гармонического анализа. Согласно гармоническому анализу, временной ряд представляется как

[40] Агаларов, З. С. Эконометрика : учебник / З. С. Агаларов, А. И. Орлов // Электронно-библиотечная система IPR BOOKS: [сайт]. – М. : Дашков и К, 2021. – 380 с. – URL : https://www.iprbookshop.ru/107834.html

совокупность гармонических колебательных процессов. Для каждой точки этого ряда справедливо выражение:[41]

$$y_t = f(t) + \sum_{k=1} \left(a_k \cos\left(kt \frac{2\pi}{n} \right) + b_k \sin\left(kt \frac{2\pi}{n} \right) \right),$$

где t = 1, 2, 3, …, n).

Здесь y_t – фактический уровень ряда в момент (интервал) времени t; $f(t)$ – выравненный уровень ряда в тот же момент времени, a_k, b_k – параметры колебательного процесса (гармоники) с номером k, в совокупности оценивающие размах (амплитуду) отклонения от общей тенденции и сдвиг колебаний относительно начальной точки.

Это уравнение представляет собой ряд Фурье, где время (t) выражается в радиальной мере или в градусах (табл. 10).

Общее число колебательных процессов, которые можно выделить для ряда, состоящего из n уровней, равно $n/2$. Обычно ограничиваются меньшим числом наиболее важных гармоник. Параметры гармоники с номером k определяются по формулам:

$$\varepsilon_t = y_t - f(t);$$

$$a_k = \frac{2}{n} \sum_{t=1}^{n} \varepsilon_t \cos\left(kt \frac{2\pi}{n} \right), \ k = 1, 2, ..., \frac{n}{k} - 1;$$

$$b_k = \frac{2}{n} \sum_{t=1}^{n} \varepsilon_t \sin\left(kt \frac{2\pi}{n} \right), \ k = 1, 2, ..., \frac{n}{k} - 1;$$

$$a_{n/2} = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \cos(\pi t), \ b_{n/2} = 0.$$

Этот метод хорошо подходит для аналитического выражения сезонных колебаний, имеющих синусоидальную форму.

10. Значение *t* для ряда Фурье

Месяцы (t)	1	2	3	4	5	6	7	8	9	10	11	12
Радиальная мера	0	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$	π	$\frac{7\pi}{6}$	$\frac{4\pi}{3}$	$\frac{3\pi}{2}$	$\frac{5\pi}{3}$	$\frac{11\pi}{6}$
Градусы	0	30	60	90	120	150	180	210	240	270	300	320
Уровни (y)	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}

[41] Яковлев, В. П. Эконометрика : учебник для бакалавров / В. П. Яковлев // Электронно-библиотечная система IPR BOOKS : [сайт]. – М. : Дашков и К, 2019. – 384 с. – URL : https://www.iprbookshop.ru/85674.html

Обычно при выравнивании по ряду Фурье рассчитывают не более четырех гармоник и затем уже определяют, с каким числом гармоник наилучшим образом отражается периодичность изменения уровней ряда.

Например, при $k = 1$ уравнение Фурье будет иметь вид:

$$y_t = a_0 + a_1 \cos\left(t\frac{2\pi}{2}\right) + b_1 \sin\left(t\frac{2\pi}{2}\right),$$

где

$$a_0 = \bar{y}\, a_1 = \frac{2}{n}\sum_{t=1}^{n} y_t \cos\left(t\frac{2\pi}{n}\right) b_1 = \frac{2}{n}\sum_{t=1}^{n} y_t \sin\left(t\frac{2\pi}{n}\right)$$

или

$$y_t = a_0 + a_1\text{cost} + b\text{sin}t.$$

При $k = 2$ соответственно

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t;$$

$$a_0 = \frac{\sum y}{n};\ a_2 = \frac{2\sum y \cos 2t}{n};\ b_2 = \frac{2\sum y \sin 2t}{n}.$$

Параметры уравнения выравненных уровней, определяемых рядом Фурье, находят по способу наименьших квадратов (табл. 11):

$$y_t = 41{,}82 - 2{,}913\text{cost} - 7{,}076\text{sin}.$$

11. Расчет параметров гармоники ряда Фурье

Месяцы	Производство скота и птицы на убой в живом весе Y (тыс. т)	t	cos$t$	sin$t$	y cos$t$	y sin$t$	yt
Январь	31,4	0	1	0	31,40	0,00	38,9
Февраль	31,3	π/6	0,87	0,50	27,11	15,64	35,8
Март	35,2	π/3	0,50	0,87	17,62	30,47	34,2
Апрель	35,6	π/2	0,00	1,00	0,03	35,60	34,7
Май	41,7	2π/3	–0,50	0,87	–20,81	36,14	37,1
Июнь	39,7	5π/6	–0,87	0,50	–34,35	19,90	40,8
Июль	45,5	π	–1,00	0,00	–45,50	0,07	44,7
Август	45,8	7π/6	–0,87	–0,50	–39,71	–22,83	47,9
Сентябрь	43,8	4π/3	–0,50	–0,86	–21,98	–37,89	49,4
Октябрь	51,8	3π/2	0,00	–1,00	–0,12	–51,80	48,9
Ноябрь	48	5π/3	0,50	–0,87	23,89	–41,63	46,5
Декабрь	52	11π/6	0,86	–0,50	44,96	–26,13	42,9
Сумма	501,80		–0,01	0,00	–17,48	–42,45	501,8

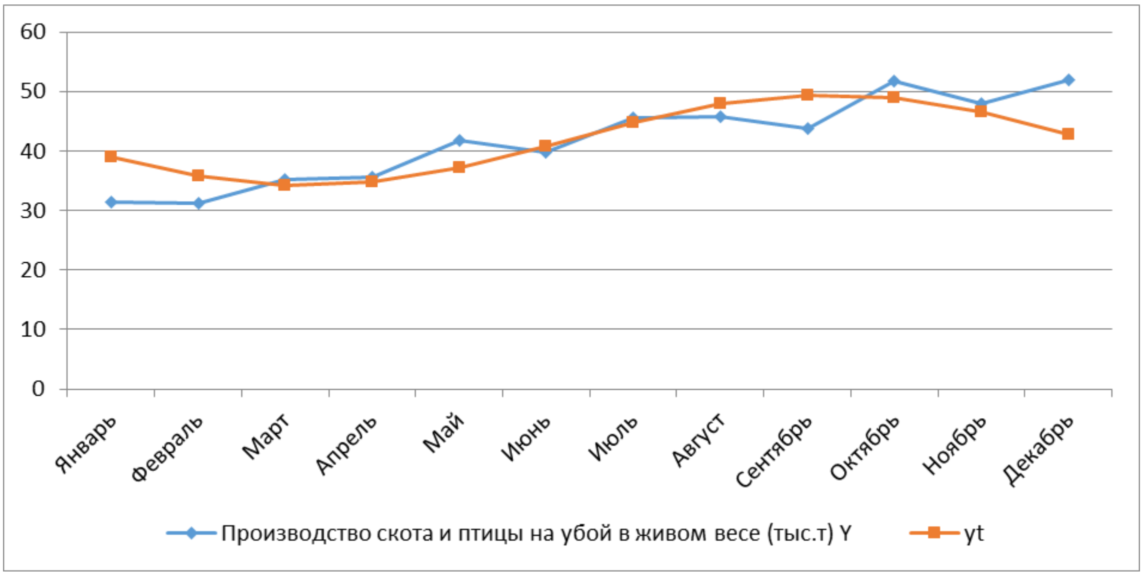


Рис. 23. Фактические и теоретические уровни ряда динамики Фурье производства скота и птицы на убой в живом весе (тыс. т)

На рисунке 23 представлены фактические и теоретические уровни ряда динамики Фурье.

Прогнозирование уровней временного ряда на основе кривых роста. Прогнозирование с использованием кривой роста (модели тенденции) включает следующие этапы:

1. Выбор кривой роста: на основе качественного анализа динамики временного ряда выбирается одна или несколько кривых, которые подходят для описания характера изменения уровней ряда.

2. Оценка параметров выбранной кривой: определяются параметры выбранной кривой роста.

3. Проверка точности и адекватности: оценивается точность подобранной кривой роста и ее соответствие ожидаемому поведению временного ряда.

4. Расчет прогноза: производится расчет точечных и интервальных прогнозов.

Чтобы по имеющемуся временному ряду $y_1, y_2, ..., y_n$ осуществить прогноз на L шагов вперед, необходимо в построенную модель тенденции (кривую роста) $y = f_t$ подставить значение аргумента, соответствующее интервалу прогноза:

$$y_{n(+L)} = f\,(t_{(n+L)}).$$

«Значение $y_{n(+L)}$, которое было получено, называют точечным прогнозом. Далее необходимо определить доверительный интервал прогноза, т.е. пределы, внутри которых находится истинное значение уровня явления с определенной вероятностью (уровнем доверия). Этот процесс называется расчетом интервального прогноза. Интервальный прогноз устанавливает границы возможных изменений прогнозируемого показателя».[42]

Расхождение между реальными данными и точечным прогнозом, который был сделан методом экстраполяции тенденции по кривым роста, может объясняться несколькими важными причинами. Эти расхождения могут существенно влиять на качество прогнозирования, поэтому их анализ является необходимым этапом при работе с временными рядами.

Во-первых, одной из наиболее частых причин подобных несоответствий является *субъективная ошибка при выборе типа кривой*. При проведении экстраполяции аналитики часто сталкиваются с необходимостью выбора определенного типа математической или статистической кривой для моделирования тенденции. Однако, если выбор окажется неправильным или не будет соответствовать реальному поведению данных, это может привести к неточным прогнозам. Чаще всего подобные ошибки происходят из-за недостаточного понимания свойств данных или игнорирования предварительных анализов.

Во-вторых, существенным источником ошибок является *неточность оценки параметров кривых*. Даже если была выбрана правильная кривая, неточные значения ее параметров могут исказить результаты экстраполяции. Неправильные параметры могут возникнуть из-за недостаточного объема данных для анализа или же из-за применения некорректных методов оценки, что опять же приводит к ненадежным прогнозам. Это подчеркивает важность качественного подхода к сбору данных и методологии их обработки.

В-третьих, *ошибки, вызванные отклонениями отдельных наблюдений от тренда*, который отражает средний уровень временного ряда для каждого момента времени, также могут сыграть значительную роль. Такие отклонения могут происходить в силу различных внешних факторов, которые не были учтены при моделировании. Они могут возникнуть из-за сезонных колебаний, экономических кризисов или других неожиданных событий, приводящих к резким изменениям в данных. Эти движения могут скрыть истинную тенденцию и затруднить анализ.

Важно отметить, что ошибки, обусловленные вторым и третьим источниками, могут быть учтены через *доверительный интервал прогноза*. Доверительные интервалы позволяют аналитикам учитывать неопределенность отдельных прогнозов и предоставляют более широкий контекст для анализа. Вместо аппелирования лишь к точечным прогнозам, использование довери-

[42] Яковлев, В. П. Эконометрика: учебник для бакалавров / В. П. Яковлев // Электроннобиблиотечная система IPR BOOKS : [сайт]. – М. : Дашков и К, 2019. – 384 с. – URL : https://www.iprbookshop.ru/85674.html

тельных интервалов дает представление о диапазоне возможных значений, которые могут возникнуть в будущем. Это особенно актуально в условиях растущей неопределенности и динамично меняющихся рынков, где подготовка к возможным отклонениям может оказаться решающей.

Таким образом, понимание причин расхождений между прогнозируемыми и реальными данными является критически важным для совершенствования методов прогнозирования и повышения их точности. Это способствует более качественному принятию решений и адекватной реакции на изменения в окружающей среде, что делает процесс экстраполяции более надежным и целенаправленным.

Доверительный интервал для линейной тенденции по аналогии с парной регрессией вычисляется по формуле:

$$\hat{y}_{n(+L)} \pm t_{1-\alpha,\, n-2} S_{\hat{y}} \sqrt{1 + \frac{1}{n} + \frac{(t_{n+1} - \bar{t})^2}{\sum_{t=1}^{n} (t - \bar{t})^2}},$$

где n – длина временного ряда; L – период упреждения; $y_n(+_L)$ – точечный прогноз на момент $n + L$; t_{1-an-2} – значение t-статистики Стьюдента при уровне значимости a и числе степеней свободы n – 2; S_y – средняя квадратическая ошибка оценки прогнозируемого показателя.

$$S_{\hat{y}} = \sqrt{\frac{1}{n-m} \sum_{t=1}^{n} (\hat{y}_t - y_t)^2},$$

m – число параметров модели кривой роста (для линейной модели $m = 2$).

Для линейной модели доверительный интервал также можно определить следующим образом:

$$y_{n(+L)} \pm t_{1-\alpha,n-2} S_{\hat{y}} \sqrt{1 + \frac{1}{n} + \frac{3(n + 2L - 1)^2}{n(n-1)}}.$$

Доверительный интервал для кривой роста в виде полинома второго или третьего порядка вычисляется по формуле:

$$y_{n(+L)} \pm t_{1-\alpha,n-2} S_{\hat{y}} \sqrt{1 + \frac{1}{n} + \frac{t_{n+L}^2}{\sum_{t=1}^{n} t^2} + \frac{\sum_{t=1}^{n} t^4 - 2t_{n+L}^2 \sum_{t=1}^{n} t^2 - n t_{n+L}^4}{n \sum_{t=1}^{n} t^4 - \left(\sum_{t=1}^{n} t^2 \right)^2}},$$

где m – число параметров модели кривой роста. Для полинома второго порядка $m = 3$, для полинома третьего порядка $m = 4$.

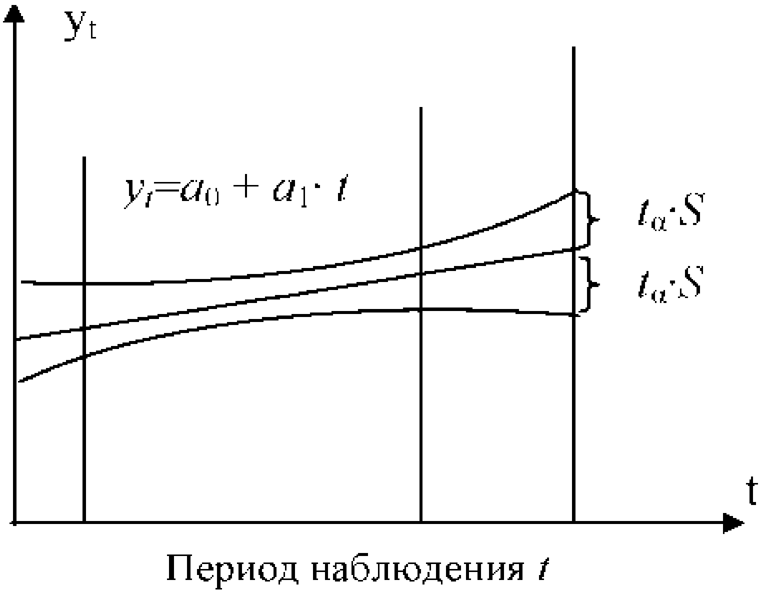


Рис. 24. Доверительные интервалы прогноза для линейного тренда

Ширина доверительного интервала зависит от уровня значимости, периода упреждения, среднего квадратического отклонения временного ряда от тренда и степени полинома (рис. 24).

При увеличении степени полинома доверительный интервал расширяется при одинаковом значении S_y, поскольку дисперсия уравнения тренда рассчитывается как взвешенная сумма дисперсий соответствующих параметров этого уравнения.

Доверительные интервалы для прогнозов, основанных на экспоненциальном уравнении, определяются схожим способом. Однако отличие заключается в том, что как при расчете параметров кривой, так и при вычислении среднеквадратичной ошибки используются не исходные значения уровней временного ряда, а их логарифмы.

Адаптивные модели прогнозирования. При анализе временных рядов зачастую важнее оказывается текущая тенденция (т.е. та, которая определяется на основе последних нескольких наблюдений), чем долгосрочная тенденция, сформировавшаяся за длительный период. Поэтому наибольшую значимость приобретает информация о последнем периоде. В связи с этим в последние годы особую популярность приобрели так называемые адаптивные методы прогнозирования.

Адаптивными методами прогнозирования называются те, которые позволяют создавать самокорректирующиеся (или самообучающиеся) экономико-математические модели, способные быстро реагировать на изменения условий путем учета результатов предыдущих прогнозов и различий в информационной ценности уровней ряда.

Основные особенности адаптивных методов прогнозирования заключаются в следующем:

– возможность учитывать информационную ценность различных уровней временного ряда (например, с помощью системы весовых коэффициентов, присваиваемых этим уровням);

– применение рекуррентных процедур для уточнения параметров модели по мере поступления новой информации, что позволяет адаптировать модель под новые условия развития явления.

Скорость реакции модели на изменения в процессе характеризуется параметром адаптации. Этот параметр следует выбирать таким образом, чтобы он обеспечивал адекватное отражение текущей тенденции, одновременно фильтруя случайные отклонения. Значение параметра адаптации может быть определено на основе эмпирических данных, выведено аналитически либо найдено методом проб.

Критерием оптимальности при выборе параметра адаптации чаще всего служит минимизация среднего квадрата ошибок прогноза. Благодаря своим характеристикам, адаптивные методы особенно эффективны при краткосрочном прогнозировании (прогнозирование на один или несколько шагов вперед). Обычно такие методы основываются на процедуре экспоненциального сглаживания.

Экспоненциальное сглаживание временного ряда y_t осуществляется с помощью рекуррентной формулы.

$$S_t = \alpha y_t + \beta S_{t-1},$$

где S_t – значение экспоненциальной средней в момент t; y_t – значение временного ряда в момент t; α – параметр сглаживания, α = const, $0 < \alpha < 1$; $\beta = 1 - \alpha$. Совокупность значений S_t образует сглаженный временной ряд.

Приведенное соотношение позволяет выразить экспоненциальную среднюю S_t через предшествующие значения уровней временного ряда y_t. При $n \to \infty$

$$S_t = \alpha \sum_{i=0}^{n} \beta^i y_{t-i}.$$

Таким образом, величина S_t оказывается взвешенной суммой всех членов ряда. Причем веса отдельных уровней ряда $\alpha\beta^i$ убывают по мере их удаления в прошлое соответственно экспоненциальной функции (в зависимости от «возраста» наблюдений).

Например, при $\alpha = 0{,}4$ вес текущего наблюдения y_t будет равен $\alpha = 0{,}4$, вес предыдущего уровня y_{t-1} будет соответствовать $\alpha\beta^i = 0{,}4 \ 0{,}6 = 0{,}24$; для уровня y_{t-2} вес составит $\alpha\beta^2 = 0{,}144$; для y_{t-3} – $\alpha\beta^3 = 0{,}0864$ и т.д.

Доказано, что математические ожидания исходного ряда и экспоненциальной средней совпадают. В то же время дисперсия экспоненциальной средней $D(S_t)$ меньше дисперсии временного ряда σ^2. Чем меньше α, тем это отличие больше.

Таким образом, с одной стороны, желательно увеличивать вес более свежих наблюдений, что может быть достигнуто повышением α, с другой стороны, для сглаживания случайных отклонений величину a нужно уменьшить. Выбор параметра сглаживания, а с учетом этих двух противоречивых требований составляет задачу оптимизации модели.

В качестве начального значения S_0 используется среднее арифметическое значение из всех имеющихся уровней временного ряда или из какой-то их части. Из выражения следует, что вес, приписываемый этому значению, уменьшается по экспоненциальной зависимости по мере удаления от первого уровня. Поэтому для длинных временных рядов влияние неудачного выбора S_0 погашается.

При использовании *экспоненциальной средней для краткосрочного прогнозирования* предполагается, что модель ряда имеет вид

$$y_t = \alpha_{1,t} + e_t,$$

где $\alpha_{1,t}$ – варьирующий во времени средний уровень ряда; e_t – случайные неавтокоррелированные отклонения с нулевым математическим ожиданием и дисперсией σ^2.

Прогнозная модель определяется соотношением

$$\hat{y}_\tau(t) = \hat{\alpha}_{1,t},$$

где $\hat{y}_\tau(t)$ – прогноз, сделанный в момент t на τ единиц времени (шагов) вперед; $\hat{\alpha}_{1,t}$ –оценка $\hat{\alpha}_{1,t}$. Величина параметра модели $\hat{\alpha}_{1,t}$ принимается равной экспоненциальной средней S в момент t:

$$\hat{\alpha}_{1,t} = S_t;$$

$$\hat{\alpha}_{1,0} = S_0.$$

Прогнозирование предполагает следующую последовательность действий:

– на основании исходного временного ряда y_i, y_2, y_n вычисление сглаженных уровней ряда S_i, S_2, ..., S_n;

– вычисление $a_{ln} = S_n$;

осуществление прогноза на τ шагов вперед $\hat{y}_\tau(n) = \hat{\alpha}_{1,n}$.

Перегруппировав члены выражение можно записать по-другому:

$$S_t = S_{t-1}\alpha(y_t - S_{t-1}).$$

Если величину $(y_t - S_{t-1})$ рассматривать как погрешность прогноза, то новый прогноз S_t получается как результат корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит адаптация модели.

Экспоненциальное сглаживание представляет собой пример одной из самых простых самообучающихся моделей. Расчеты проводятся итерационно, при этом вся предшествующая информация содержится в единственном значении S_{t-1}.

Адаптивные полиномиальные модели. Если для прогнозирования временного ряда с ярко выраженной линейной тенденцией применять метод, основанный на модели экспоненциального сглаживания, то такая модель, скорее всего, будет выдавать смещенные прогнозы, т.е. содержать систематическую ошибку. Для подобных временных рядов лучше использовать модели линейного роста, где процедуре экспоненциального сглаживания подвергаются оценки коэффициентов адаптивной модели.

$$\hat{y}_\tau(t) = \hat{\alpha}_{1,t} + \hat{\alpha}_{2,t}\tau,$$

где $\hat{\alpha}_{1,t}$ и $\hat{\alpha}_{2,t}$ – текущие оценки коэффициентов; τ – время упреждения прогноза.

Наиболее часто применяются три модели данного типа, отличающиеся рекуррентными выражениями для пересчета текущих оценок коэффициентов (параметры адаптации или параметры экспоненциального сглаживания $0 < \alpha_1$, α_2, α_3, $\beta < 1$):

– двухпараметрическая модель Ч. Хольта

$$\hat{\alpha}_{1,t} = \alpha_1 y_t + (1-\alpha)(\hat{\alpha}_{2,t-1} + \hat{\alpha}_{2,t-1});$$

– однопараметрическая модель Р. Брауна

$$\hat{\alpha}_{1,t} = \alpha_{1,t=1} + (\hat{\alpha}_{2,t=1} + (1-\beta^2)e_t;$$

$$\hat{\alpha}_{2,t} = \alpha_{2,t=1} + (1-\beta^2)e_t;$$

– трехпараметрическая модель Дж. Бокса и Г. Дженкинса

$$\hat{\alpha}_{1,t} = \alpha_1 y_t + (1-\alpha_1)(\hat{\alpha}_{1,t-1} + \hat{\alpha}_{2,t-1})\alpha_3(e_t - e_{t-1});$$

$$\hat{\alpha}_{2,t} = \alpha_2(\hat{\alpha}_{1,t-1} + \hat{\alpha}_{1,t-1}) + (1-\alpha_2)\hat{\alpha}_{2,t=1}.$$

Начальные значения коэффициентов a_{lt} и a_{2t} принимаются равными коэффициентам уравнения регрессии, построенного по начальным уровням ряда. В эконометрическом программном обеспечении чаще встречается модель Чарльза Холта, позволяющая выбрать оптимальные параметры по критерию минимизации среднеквадратической ошибки посредством перебора возможных значений на сетке. Рекуррентные формулы для оценки коэффициентов в данной модели могут быть представлены в форме, наглядно демонстрирующей зависимость корректировки от величины ошибки,

$$\hat{\alpha}_{1,t} = \hat{\alpha}_{1,t-1} + \hat{\alpha}_{2,t-1} + \alpha_1 e_t;$$

$$\hat{\alpha}_{2,t} = \hat{\alpha}_{2,t-1} + \alpha_1\alpha_2 e_t,$$

где $e_t = y_t - \hat{y}_1(t-1)$ – ошибка прогноза.

Из последних выражений видно, что модель Р. Брауна можно считать частным случаем модели Ч. Хольта. При этом единственный параметр β играет роль коэффициента дисконтирования наблюдений.

Исследование взаимосвязи двух временных рядов. Модели, построенные на основе данных, собранных о объекте на протяжении ряда временных моментов (периодов), называются временными рядами. Анализ взаимосвязей между переменными, представленными временными рядами, имеет свои специфические особенности.[43]

Наличие тенденций и периодических компонентов в временных рядах может вызвать ложную корреляцию или ложную регрессию при использовании традиционных методов корреляционного и регрессионного анализа. Это означает, что высокий коэффициент корреляции между переменными *x* и *y*, которые на самом деле не влияют друг на друга, может возникнуть из-за их зависимости от времени. Аналогично, коэффициент детерминации может показывать высокую точность регрессии между этими переменными. Чтобы избежать таких искажений, перед исследованием взаимосвязи между переменными *x* и *y* важно сначала удалить влияние тенденций и периодических компонентов

[43] Мхитарян, В. С. Эконометрика : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин // Цифровой образовательный ресурс IPR SMART : [сайт]. – М. : Евразийский открытый институт, 2012. – 224 с. – URL : https://www.iprbookshop.ru/11125.html

из уровней временных рядов. Для удаления периодических компонентов можно использовать метод скользящего среднего.[44]

Чтобы устранить тенденцию, применяются различные подходы, такие как метод последовательных разностей, метод отклонений от тренда, а также интеграция временного фактора в модель регрессии для временных рядов.

Метод отклонений от тренда. Рассмотрим два временных ряда x_t и y_t, каждый из которых содержит трендовую компоненту T и случайную компоненту S. Предположим, что проведено аналитическое выравнивание этих рядов и найдены параметры соответствующих уравнений тенденций $x_t = f_1(t)$ и $y_t = f_2(t)$. Вычитание расчетных значений уровней ряда X_t и y_t из фактических x_t и y_t позволяет устранить влияние тенденции в обоих рядах. Дальнейший анализ взаимосвязи рядов проводят с использованием отклонений от тренда $(x_t - \hat{x}_t)$ и $(y_t - \hat{y}_t),$ т.е. уравнение регрессии строится в виде

$$(y_t - \hat{y}_t) = \alpha + b(x_t - \hat{x}_t).$$

Метод последовательных разностей. Если временной ряд демонстрирует четко выраженную полиномиальную тенденцию (которая выглядит как полином относительно времени t, для ее устранения можно воспользоваться методом последовательных разностей. Этот метод предполагает замену исходных уровней ряда на последовательные разности соответствующего порядка (где порядок разности соответствует порядку полинома).

Последовательными разностями первого порядка называются величины: $\Delta y_t = y_t - y_{t-1}$. Последовательными разностями второго порядка называются величины: $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$.

Замена исходных уровней ряда последовательными разностями первого порядка позволяет устранить линейную тенденцию, задаваемую уравнением $v = \alpha + bt.$

Замена исходных уровней ряда последовательными разностями второго порядка позволяет устранить параболическую тенденцию, задаваемую уравнением в виде полинома второго порядка $y = \alpha + bt + ct^2$, и т.д.

Если тенденция временного ряда характеризуется экспоненциальной зависимостью, то временной ряд из логарифмов исходных уровней будет иметь линейную тенденцию, что позволяет применить метод последовательных разностей к ряду логарифмов.

[44] Яковлев, В. П. Эконометрика : учебник для бакалавров / В. П. Яковлев // Электроннобиблиотечная система IPR BOOKS: [сайт]. – М. : Дашков и К, 2019. – 384 с. – URL : https://www.iprbookshop.ru/85674.html

С использованием первых разностей Δy_t, Δx_t уравнение регрессии находится в виде:

$$\Delta y_t = \alpha + b\Delta x_t \text{ или } y_t - y_{t-1} = \alpha + b(x_t - x_{t-1}).$$

Включение в модель регрессии фактора времени. Добавление временного фактора в качестве независимой переменной позволяет зафиксировать динамику, что помогает исключить ее воздействие на параметры модели. Уравнение парной регрессии в данной ситуации выглядит следующим образом:

Уравнение парной регрессии в этом случае принимает следующий вид

$$y_t = \alpha + b_1 x + b_2 t + \varepsilon_t.$$

Этот метод может быть также применим в ситуациях, когда количество факторов больше одного. Параметры α, b_1 и b_2 модели, в которой учитывается время как фактор, вычисляются с использованием стандартного метода наименьших квадратов (МНК).

Интерпретация параметров регрессионного уравнения следующая:

Параметр b_1 показывает, насколько в среднем изменяется значение зависимой переменной y_t при увеличении фактора x_t на единицу, при условии, что остальные факторы остаются постоянными.

Параметр b_2 указывает, насколько в среднем изменится значение зависимой переменной y_t за период наблюдения под влиянием всех факторов, исключая фактор x_t.

Автокорреляция в остатках может возникать по различным причинам, которые разнообразны по своему характеру. Она может быть вызвана исходными данными, например, наличием ошибок измерений в значениях зависимого переменного. Иногда автокорреляция происходит из-за некорректной спецификации модели. Например, если модель не учитывает ключевой фактор, существенно влияющий на результат, это воздействие может проявляться в остатках, что приводит к автокорреляции. Часто таким важным фактором оказывается время t.

Необходимо различать истинную автокорреляцию остатков от случаев, когда автокорреляция возникает из-за неправильно выбранной функциональной формы модели. В таких ситуациях следует корректировать модель, а не использовать специальные методы для оценки параметров регрессионного уравнения, если в остатках наблюдается автокорреляция. Одним из распространенных методов для обнаружения автокорреляции в остатках является вычисление критерия Дарбина–Уотсона.

$$d = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}, \qquad 0 \le d \le 4.$$

Коэффициент автокорреляции остатков первого порядка:

$$r_1^{\varepsilon} = \frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sum_{t=2}^{n} \varepsilon_t^2}, \qquad -1 \le r_1^{\varepsilon} \le 1.$$

Алгоритм выявления автокорреляции остатков с использованием данного критерия включает следующие шаги. Сначала выдвигается гипотеза об отсутствии автокорреляции. Альтернативные гипотезы обозначают наличие соответственно положительной или отрицательной автокорреляции в остатках. Затем, опираясь на статистические таблицы, определяются критические значения критерия Дарбина–Уотсона в зависимости от заданного числа наблюдений, количества независимых переменных модели и уровня значимости. На основе этих значений числовой интервал разбивается на пять сегментов.

Принятие или отклонение каждой из гипотез с вероятностью $1-\alpha$ осуществляется следующим образом:

– $0 < d < d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $1-\alpha$ принимается H_1;

– $d_L < d < d_U$ – зона неопределенности;

– $d_U < d < 4-d_U$ – нет оснований отклонять H_0, т.е. автокорреляция остатков отсутствует;

– $4-d_U < d < 4-d_L$ – зона неопределенности;

– $4-d_L < d < 4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $1-\alpha$ принимается H_1^*.

Если фактическое значение критерия Дарбина–Уотсона оказывается в зоне неопределенности, то в практической деятельности предполагается наличие автокорреляции остатков, и гипотеза об отсутствии автокорреляции отвергается.

При использовании критерия Дарбина–Уотсона имеются некоторые ограничения:

1. Он не подходит для моделей, в которых независимыми переменными являются лаговые значения зависимой переменной.

2. Метод расчета и применение критерия Дарбина–Уотсона направлен только на выявление автокорреляции остатков первого порядка.

3. Критерий Дарбина–Уотсона предоставляет надежные результаты только при анализе больших выборок.

ТЕСТ

1. Что такое временной ряд?

а) набор случайных данных;

б) единичный показатель;

в) группа объектов;

г) последовательность значений.

2. Как называется модель, которая представляет временной ряд как сумму его компонентов?

а) мультипликативная модель;

б) квадратичная модель;

в) аддитивная модель;

г) линейная модель.

3. Какой компонент временного ряда отражает общее долгосрочное воздействие различных факторов?

а) циклические колебания;

б) тренд;

в) случайный фактор;

г) сезонные изменения.

4. Что характерно для циклических колебаний в экономических временных рядах?

а) случайные изменения;

б) отсутствие закономерностей;

в) влияние времени года;

г) сезонные изменения.

5. Что следует учитывать при проверке исходных данных перед разработкой модели временного ряда?

а) круглосуточный анализ;

б) методы сбора и расчета данных;

в) случайные выбросы;

г) температурные колебания.

6. Какой компонент временного ряда может иметь как положительные, так и отрицательные значения?

а) сезонные колебания;

б) тренд;

в) цикл;

г) случайный фактор.

7. Какое количество наблюдений должно превышать количество параметров модели для устойчивой модели временного ряда?

а) 2 – 4 раза;

б) 1 – 3 раза;

в) 7 – 10 раз;

г) 3 – 5 раз.

8. Какова основная цель эконометрического анализа временного ряда?

а) сбор дополнительных данных;

б) обнаружение и описание компонентов;

в) построение новых моделей;

г) создание графиков.

9. Как называют временные ряды, в которых нет ни тренда, ни цикличности?

а) регулярные ряды;

б) случайные ряды;

в) значимые ряды;

г) статистические ряды.

10. Какой из следующих факторов НЕ является компонентом временного ряда?

а) циклы;

б) сезонные колебания;

в) тренд;

г) факторы риска.

11. Что такое адаптивные методы прогнозирования?

а) методы, не учитывающие предыдущие данные;

б) методы, использующие фиксированные параметры;

в) методы, корректирующиеся на основе предыдущих результатов;

г) методы для долгосрочного прогнозирования.

12. Как определяется параметр адаптации в адаптивных моделях?

а) методом оценки ошибок;

б) случайным образом;

в) на основе предыдущих результатов;

г) заранее зафиксированным значением.

13. Какой из следующих критериев оптимальности используется при выборе параметра адаптации?

а) увеличение веса наблюдений;

б) минимизация среднего квадрата ошибок прогноза;

в) максимизация вариации;

г) снижение числа наблюдений.

14. Какое начальное значение S_0 используется в моделях экспоненциального сглаживания?

а) максимум временного ряда;

б) минимум временного ряда;

в) произвольное значение;

г) среднее арифметическое значение уровней ряда.

15. Как адаптивные методы реагируют на изменения условий?

а) служат для долгосрочного анализа;

б) не реагируют вообще;

в) поддерживают постоянные параметры;

г) самокорректируются по мере поступления новой информации.

16. Как определяется прогностическая модель $\hat{y}_\tau(t)$?

а) на основе исторических данных;

б) как оценка текущего уровня ряда;

в) с фиксированным значением параметров;

г) как случайное значение.

17. Какие модели рекомендуется использовать для временных рядов с линейной тенденцией?

а) модели с фиксированными коэффициентами;

б) линейные модели;

в) случайные модели;

г) экспоненциальные модели.

18. Каково основное преимущество экспоненциального сглаживания на краткосрочных интервалах?

а) минимизация случайных ошибок;

б) способность учитывать только последние наблюдения;

в) долгосрочное прогнозирование;

г) устойчивость к изменению параметров.

ЗАКЛЮЧЕНИЕ

Учебное пособие «Эконометрика» охватывает ключевые аспекты данной дисциплины, включая построение и оценку регрессионных моделей, анализ временных рядов и другие важные темы. В книге представлены шесть глав, каждая из которых подробно рассматривает различные методы эконометрического анализа.

В ходе изучения материалов учебного пособия, посвященного основам эконометрики, ознакомились с ключевыми концепциями и методами, используемыми для анализа экономических данных.

В темах, посвященных нелинейной парной корреляции и множественной регрессии, были рассмотрены более сложные модели, позволяющие учитывать множественные факторы, влияющие на изучаемые явления. Анализ регрессионных моделей с переменной структурой, в частности использование фиктивных переменных, представил возможности для работы с изменчивыми данными и специфическими случаями.

В разделе о моделировании временных рядов и прогнозировании были представлены изучили подходы, позволяющие предсказывать поведение экономических переменных на основе прошлых наблюдений. Освоенные темы формируют комплексное представление о том, как эконометрические инструменты могут быть использованы для анализа, интерпретации и предсказания экономических явлений.

Данное пособие станет полезным ресурсом для обучения и вдохновит на дальнейшее углубление знаний в области эконометрики. Принимая во внимание постоянные изменения в экономической среде, углубленное понимание эконометрических методов поможет принимать более обоснованные решения и проводить качественный анализ данных в профессиональной деятельности.

СПИСОК ЛИТЕРАТУРЫ

1. **Агаларов, З. С.** Эконометрика : учебник / З. С. Агаларов, А. И. Орлов // Электронно-библиотечная система IPR BOOKS : [сайт]. – М. : Дашков и К, 2021. – 380 с. – URL: https://www.iprbookshop.ru/107834.html

2. **Буравлев, А. И.** Эконометрика : учебное пособие / А. И. Буравлев // Электронно-библиотечная система IPR BOOKS : [сайт]. – 4-е изд. – М. : Лаборатория знаний, 2021. – 165 с. – URL : https://www.iprbookshop.ru/109431.html

3. **Практикум** по эконометрике (+CD) : учебное пособие / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 344 с.

4. **Эконометрика** : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд. перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.

5. **Еремеева, Н. С.** Эконометрика: лабораторный практикум в Excel : учебное пособие / Н. С. Еремеева, Т. В. Лебедева // Цифровой образовательный ресурс IPR SMART : [сайт]. – Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2016. – 159 с. – URL : https://www.iprbookshop.ru/61426.html

6. **Ивченко, Ю. С.** Эконометрика в MS EXCEL : лабораторный практикум / Ю. С. Ивченко // Электронно-библиотечная система IPR BOOKS : [сайт]. – Саратов : Ай Пи Эр Медиа, 2018. – 94 с. – URL : https://www.iprbookshop.ru/70785.html

7. **Ивченко, Ю. С.** Эконометрика: курс лекций / Ю. С. Ивченко // Электронно-библиотечная система IPR BOOKS : [сайт]. – Саратов : Вузовское образование, 2018. – 121 с. – URL : https://www.iprbookshop.ru/73609.html

8. **Кремер, Н. Ш.** Эконометрика : учебник для вузов / Н. Ш. Кремер, Б. А. Путко ; под ред. проф. Н. Ш. Кремера. – М. : ЮНИТИ-ДАНА, 2006. – 311 с.

9. **Мхитарян, В. С.** Эконометрика : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин // Цифровой образовательный ресурс IPR SMART : [сайт]. – М. : Евразийский открытый институт, 2012. – 224 с. – URL : https://www.iprbookshop.ru/11125.html

10. **Никитин, Б. Е.** Теория игр, эконометрика: модели, алгоритмы, компьютерная реализация : учебное пособие / Б. Е. Никитин, М. Н. Ивлиев // Электронно-библиотечная система IPR BOOKS : [сайт]. – Воронеж : Воронежский государ-

ственный университет инженерных технологий, 2019. – 92 с. – URL : https://www.iprbookshop.ru/95379.html

11. **Новиков, А. И.** Эконометрика : учебное пособие / А. И. Новиков // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – М. : Дашков и К, 2019. – 224 с. – URL: https://www.iprbookshop.ru/85184.html

12. **Орлов, А. И.** Эконометрика : учебное пособие / А. И. Орлов // Электронно-библиотечная система IPR BOOKS : [сайт]. – 3-е изд. – М., Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. – 676 с. – URL : https://www.iprbookshop.ru/89481.html

13. **Реннер, А. Г.** Основы эконометрики : учебное пособие / А. Г. Реннер, О. И. Стебунова, Л. М. Туктамышева // Цифровой образовательный ресурс IPR SMART : [сайт]. – Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2009. – 156 с. – URL: https://www.iprbookshop.ru/30069.html

14. **Рожков, И. М.** Эконометрика : учебное пособие / И. М. Рожков, И. А. Ларионова // Электронно-библиотечная система IPR BOOKS : [сайт]. – М. : Издательский Дом МИСиС, 2018. – 154 с. – URL : https://www.iprbookshop.ru/ 84429.html

15. **Скляров, Ю. С.** Эконометрика. Краткий курс : учебное пособие / Ю. С. Скляров. – 2-е изд., испр. – СПб. : ГУАП, 2007. – 140 с.

16. **Тихомиров, Н. П.** Эконометрика : учебник / Н. П. Тихомиров, Е. Ю. Дорохина. – 2-е изд., стереотип. – М. : Изд-во «Экзамен», 2007. – 512 с.

17. **Яковлев, В. П.** Эконометрика : учебник для бакалавров / В. П. Яковлев // Электронно-библиотечная система IPR BOOKS : [сайт]. – М. : Дашков и К, 2019. – 384 с. – URL : https://www.iprbookshop.ru/85674.html

18. **Яковлева, А. В.** Эконометрика : учебное пособие / А. В. Яковлева // Электронно-библиотечная система IPR BOOKS : [сайт]. – 2-е изд. – Саратов : Научная книга, 2019. – 223 с. – URL : https://www.iprbookshop.ru/81090.html

19. **Frisch R.** Editorial. Econometrica. – 1933. – № 1.

ОГЛАВЛЕНИЕ

Учебное электронное издание

МЕНЬЩИКОВА Вера Ивановна

ЭКОНОМЕТРИКА

Учебное пособие

Редактирование Е. С. Мордасовой
Графический и мультимедийный дизайнер Т. Ю. Зотова
Обложка, упаковка, тиражирование Е. С. Мордасовой

ISBN 978-5-8265-2846-4

9 785826 528464

Подписано к использованию 19.12.2024.
Тираж 50 шт. Заказ № 136

Издательский центр ФГБОУ ВО «ТГТУ»
392000, г. Тамбов, ул. Советская, д. 106, к. 14
Тел./факс (4752) 63-81-08.
E-mail: izdatelstvo@tstu.ru