Федеральное государственной образовательное учреждение

Высшего профессионального образования

Национальный исследовательский технологический университет «МИСиС»

Институт информационных технологий и компьютерных наук (ИТКН)

Кафедра автоматизированных систем управления (АСУ)

Курсовая работа

по курсу «Прикладной статистический анализ»

Тема:

«Разработка модели прогнозирования объема инвестиций»

Выполнил: студент 2 курса,

БИВТ-21-7 Дорохов Иван

Руководитель: Маркарян А. О.

Москва 2023

Оглавление

[Введение 3](#_Toc153489577)

[АНАЛИЗ ХАРАКТЕРИСТИК ОБЪЕКТА ИССЛЕДОВАНИЯ 4](#_Toc153489578)

[Описание объекта исследования 4](#_Toc153489579)

[Анализ объекта исследования с помощью статистических показателей и выявление причинно-следственных связей 5](#_Toc153489580)

[Показатели динамики 6](#_Toc153489581)

[Постановка задачи моделирования 8](#_Toc153489582)

[Формализация и классификация переменных 9](#_Toc153489583)

[Проверка гипотезы о нормальном распределении выходной величины 12](#_Toc153489584)

[Построение регрессионной модели 13](#_Toc153489585)

[Структурная идентификация модели 14](#_Toc153489586)

# Введение

В современной экономической реальности предвидение играет ключевую роль в стратегическом управлении. Однако сосредоточимся на валовых частных внутренних инвестициях (Gross Private Domestic Investment, GPDI) — стратегическом показателе, оказывающем значительное воздействие на экономическое развитие. Наши усилия направлены на более глубокое осмысление и эффективное предсказание этого параметра, что становится неотъемлемым инструментом формирования макроэкономических стратегий, управления инвестиционным климатом и обеспечения устойчивого роста.

В свете напутствия необходимости глубокого понимания и точного предсказания вложений в частном секторе, актуальность нашего исследования неоспорима. Глобализация, финансовые колебания и изменения в регулировании — все это подчеркивает важность разработки точных прогностических моделей GPDI.

Основной задачей нашего исследования является формирование эффективной математической модели, способной предвидеть динамику валовых частных внутренних инвестиций на основе комплекса влияющих факторов. Для достижения этой цели мы сталкиваемся с рядом задач:

1. **Анализ воздействующих факторов**: Осмысление и выделение основных факторов, влияющих на величину валовых частных внутренних инвестиций.
2. **Создание базы данных**: Сбор и систематизация данных, необходимых для построения модели, включая экономические показатели, финансовые параметры и прочие соответствующие переменные.
3. **Разработка математической модели**: Создание математической модели, учитывающей взаимосвязи между факторами влияния и величиной GPDI.
4. **Тестирование и валидация модели**: Проверка разработанной модели на реальных данных, оценка ее точности и стабильности в различных экономических сценариях.

Объект исследования — валовые частные внутренние инвестиции, а предметом — механизмы прогнозирования данных инвестиций на основе созданной модели.

Полученные результаты обещают стать ценным вкладом в разработку более эффективных стратегий управления экономикой и обоснованных решений в области инвестиционной политики.

# АНАЛИЗ ХАРАКТЕРИСТИК ОБЪЕКТА ИССЛЕДОВАНИЯ

## Описание объекта исследования

Объектом исследования в данном случае являются валовые частные внутренние инвестиции, также известные как Gross Private Domestic Investment (GPDI). GPDI представляет собой общую стоимость всех физических капитальных вложений, осуществляемых частными субъектами (фирмами и домохозяйствами) внутри страны за определенный период времени.

Этот показатель охватывает различные формы инвестиций, такие как строительство новых зданий и сооружений, приобретение нового оборудования, а также изменения в запасах товаров. GPDI существенно влияет на экономическое развитие, поскольку он отражает уровень инвестиционной активности частного сектора, что, в свою очередь, способствует росту производства, увеличению рабочих мест и общему экономическому подъему.

При исследовании GPDI рассматриваются факторы, влияющие на решение частных субъектов инвестировать, такие как уровень процентных ставок, налоговая политика, инновационные возможности, общее экономическое климатическое состояние и динамика рынков. Анализ этих аспектов позволяет понять, какие факторы оказывают наибольшее воздействие на величину валовых частных внутренних инвестиций и как эти инвестиции формируются в рамках экономической системы страны.

## Анализ объекта исследования с помощью статистических показателей и выявление причинно-следственных связей

Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание

Рисунок 3 - Представление данных в виде графика

Для визуализации данных был построен их график. На оси «y» отмечены значения суммы объёма инвестиций на определённый момент времени в миллиардах долларов. На оси «x» отложено время, в которое было проведено измерение.

### Показатели динамики

Изображение выглядит как текст, График, диаграмма, линия

Автоматически созданное описание

Темп роста (или процентный темп роста) представляет собой показатель, который измеряет изменение значения в процентном отношении к предыдущему значению. Этот показатель часто используется для измерения темпа изменения в экономике, финансах, статистике и других областях.

Применительно к временным рядам, таким как данные по выручке, объему продаж, населению и т. д., темп роста используется для измерения процентного изменения этих значений относительно предыдущего периода.

В моём случае каждую пятилетку мы наблюдаем положительный рост GPDI,

1. **Рост экономики:** Увеличение частных внутренних инвестиций может быть индикатором роста экономики. Когда компании и домашние хозяйства увеличивают свои инвестиции, это может способствовать росту производства и созданию новых рабочих мест.
2. **Доверие бизнеса:** Положительный рост GPDI может также свидетельствовать о доверии предпринимателей к экономическим условиям. Когда бизнесы увеличивают инвестиции, это может говорить о том, что они ожидают улучшения перспектив и увеличения спроса на их продукцию или услуги.
3. **Инновации и модернизация:** Увеличение инвестиций может быть связано с внедрением новых технологий, инноваций и модернизацией оборудования. Это может улучшить производительность и эффективность производства.
4. **Развитие капитала:** Увеличение инвестиций может способствовать развитию производственного капитала, включая строительство новых предприятий, расширение существующих, приобретение оборудования и т. д.

Изображение выглядит как текст, График, линия, снимок экрана

Автоматически созданное описание

Темп прироста — это изменение величины в абсолютном выражении за определённый период времени. В контексте экономики, финансов, демографии и других областей этот термин может использоваться для измерения темпа изменения определённых параметров или показателей. Величина темпа прироста может быть положительной или отрицательной в зависимости от направления изменения.

В моём случае наблюдается так же положительный темп прироста, это может указывать на несколько положительных аспектов в экономике:

1. **Долгосрочные перспективы:** Положительный прирост GPDI может свидетельствовать о том, что частные фирмы и домашние хозяйства верят в стабильные и благоприятные долгосрочные перспективы. Увеличение инвестиций может быть результатом доверия к будущему развитию экономики.
2. **Рост производства и занятости:** Частные внутренние инвестиции могут способствовать росту производства и созданию новых рабочих мест. Когда компании инвестируют в расширение или модернизацию своих предприятий, это может стимулировать экономическую активность.
3. **Инновации и технологический прогресс:** Увеличение инвестиций может свидетельствовать о стремлении компаний к инновациям и внедрению новых технологий. Это может повысить производительность и конкурентоспособность.
4. **Долгосрочные активы:** Частные инвестиции часто направлены на создание или приобретение долгосрочных активов, таких как оборудование, здания, и другие капитальные активы, что способствует росту экономического потенциала.

## Постановка задачи моделирования

Мультипликативная модель прогнозирования отличается от аддитивной тем, что компоненты временного ряда считаются зависимыми и их влияние на общий тренд умножается, а не складывается. Вот как можно описать мультипликативную модель, сравнивая ее с аддитивной:

1. Мультипликативная декомпозиция: В мультипликативной модели временной ряд также разделяется на компоненты, такие как тренд, сезонность и остатки. Однако в отличие от аддитивной модели, в мультипликативной эти компоненты умножаются друг на друга, что отражает зависимость между ними.
2. Изменчивость влияния компонентов: Мультипликативная модель предполагает, что влияние компонентов может изменяться в зависимости от уровня временного ряда. Этот подход более подходит, когда изменения в компонентах воспринимаются как пропорциональные уровню ряда.
3. Более сложная интерпретация: Мультипликативные модели могут быть более сложными в интерпретации, так как влияние компонентов проявляется через их взаимодействие и умножение. Это может сделать понимание вклада каждой компоненты менее прямолинейным.
4. Сезонность и тренд умножаются на уровень: В мультипликативной модели предполагается, что эффекты сезонности и тренда умножаются на уровень ряда. Это имеет смысл, когда влияние этих компонентов пропорционально уровню временного ряда.
5. Более сложные временные ряды: Мультипликативные модели могут быть более подходящими для временных рядов, где влияние компонентов масштабируется в зависимости от уровня. Это может быть особенно важно для описания временных рядов с экспоненциальным ростом или убыванием.

Выбор между аддитивной и мультипликативной моделями зависит от характеристик конкретного временного ряда и целей прогнозирования. В некоторых случаях мультипликативные модели могут лучше описывать зависимости между компонентами временного ряда.

МОДЕЛИРОВАНИЕ СТАТИСТИЧЕСКИХ ЗАВИСИМОСТЕЙ

## Формализация и классификация переменных

Динамический ряд может быть представлен в виде оператора

Y = f (T, S, Е)

Y – уровни ряда, Т – тренд, S – сезонные и циклические колебания, Е –

случайные колебания

Виды моделей в зависимости от связи компонент:

- аддитивная модель, в которой значения сезонной компоненты

предполагаются постоянными для различных циклов (амплитуда колебаний

приблизительно постоянна):

Y = T + S + Е

- мультипликативная модель, которая предполагает зависимость

уровней ряда от значений сезонной компоненты (амплитуда колебаний

возрастает или уменьшается):

Y = T \* S \* Е

Давайте подробнее рассмотрим каждый из параметров аддитивной модели прогнозирования:

1. ***Yt*​:** Это само значение временного ряда в момент времени *t*. Это является суммой всех компонентов - тренда, сезонности и остатков.
2. ***Tt*​:** Трендовая компонента представляет общий тренд в данных. Это может быть увеличение или уменьшение во времени, которое не связано с сезонными колебаниями. Оценка тренда помогает выявить общий характер изменений в данных.
3. ***St*​:** Сезонная компонента учитывает периодические колебания, которые повторяются с постоянной периодичностью. Например, если у вас есть месячные данные, то сезонность может отражать изменения, связанные с временем года.
4. ***Et*​:** Остаточная компонента представляет собой случайные или непредсказуемые изменения в данных, которые не объясняются трендом и сезонностью. Это может включать в себя влияние факторов, которые не учтены в модели, или случайные флуктуации.

Процесс мультипликативного прогнозирования временных рядов включает в себя аналогичные шаги, но с некоторыми отличиями в моделировании и интерпретации компонентов. Давайте рассмотрим эти шаги подробнее:

**Оценка параметров:**

Исторические данные: Получение временных данных, которые будут использоваться для прогнозирования.

**Анализ тренда и сезонности:**

Выявление мультипликативных компонентов временного ряда, таких как тренд и сезонность, с использованием визуального анализа или статистических методов.

**Оценка параметров:**

Применение методов оценки параметров, учитывающих мультипликативные взаимодействия между компонентами. Например, использование метода наименьших квадратов для определения параметров тренда и сезонности.

**Прогнозирование компонент:**

Моделирование тренда: Разработка модели для мультипликативного тренда, учитывающего взаимодействие с другими компонентами.

Прогнозирование сезонности: Использование полученных параметров для прогнозирования будущих значений мультипликативной сезонности.

Умножение компонент: Прогнозы компонентов умножаются друг на друга для получения общего прогноза временного ряда.

**Умножение компонент:**

Умножение прогнозов: Прогнозы мультипликативных компонентов (тренда, сезонности) умножаются, а не складываются, как в аддитивной модели.

Учет остатков: Учитываются мультипликативные остатки, которые описывают случайные колебания в данных.

**Получение общего прогноза:**

Общий прогноз: Общий прогноз временного ряда получается путем умножения прогнозов тренда, сезонности и остатков.

**Оценка точности прогноза:**

Оценка качества: Использование метрик для оценки точности прогноза, таких как средняя абсолютная ошибка или среднеквадратичная ошибка.

Таким образом, в мультипликативной модели компоненты временного ряда рассматриваются как взаимозависимые и умножаются друг на друга в процессе прогнозирования.

## Проверка гипотезы о нормальном распределении выходной величины

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

В рамках нашего исследования мы провели анализ распределения данных переменной X. Визуальные инструменты, такие как гистограмма, указывают на отклонение данных от нормальности.

Мы заметили, что распределение данных имеет ярко выраженный хвост и наличие нескольких выбросов. Возможно, это является причиной отклонения от нормальности.

Важно отметить, что нормальность данных является предположением, и в реальных данных отклонения могут быть приемлемыми в зависимости от целей исследования.

## Построение регрессионной модели

Построение регрессионной модели - это процесс разработки математического отображения, которое описывает связь между зависимой переменной (или целевой переменной) и набором независимых переменных (или признаков). Эта модель используется для предсказания значения зависимой переменной на основе значений независимых переменных.

Мультипликативная модель

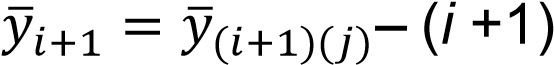
*1. Сглаживание исходного ряда методом скользящей средней Обозначения:*

*h –* количество периодов

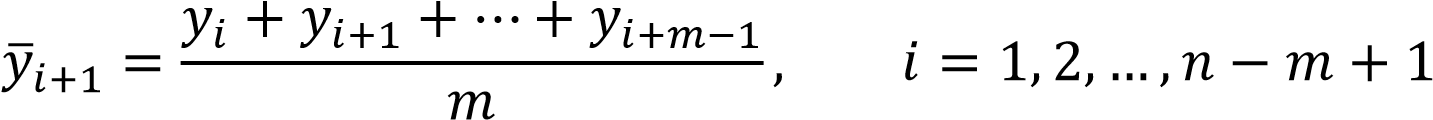
*n* – количество сезонов в одном периоде *m –* число уровней ряда, по которым вычисляется скользящая средняя

𝑦𝑖 = 𝑦𝑖(𝑗) – значение *i*-го уровня ряда *j*-го периода (*i* = 1, …, *n, j* = 1, …, *h*)

𝑌 = (𝑦1(1), 𝑦2(1), … , 𝑦𝑛(1), 𝑦1(2), 𝑦2(2), … , 𝑦𝑛(2), … , 𝑦1(ℎ), 𝑦2(ℎ), … , 𝑦𝑛(ℎ))𝑇

-е скользящее среднее *j*-го периода (*i* = 1, …, *n – m+1*)

*l* – индекс первого вычисленного скользящего среднего *k* – индекс последнего вычисленного скользящего среднего Скользящее среднее вычисляется по формуле:



*При нечетном количестве скользящих средних проводят центрирование, вычисляя среднее значение двух последовательных скользящих средних*

При выравнивании теряется несколько членов исходного ряда с обоих концов. Их можно восстановить путем экстраполяции сглаженного ряда:

Изображение выглядит как черный, темнота

Автоматически созданное описание

*2. Определение сезонной компоненты*

1. Находятся оценки сезонной компоненты как частное от деления между фактических уровней ряда и скользящих средних:

Изображение выглядит как черный, темнота

Автоматически созданное описание

1. Определяются средние значения сезонных компонент за весь период:

Изображение выглядит как черный, темнота

Автоматически созданное описание

*В мультипликативной модели сумма значений сезонных компонент должна быть равна количеству сезонов (или средняя арифметическая равна 1).*

1. Проверяется условие:

Изображение выглядит как черный, темнота

Автоматически созданное описание

Если не выполняется, вводится корректирующий коэффициент:

Изображение выглядит как черный, темнота

Автоматически созданное описание

Рассчитываются скорректированные значения сезонной компоненты:

Изображение выглядит как черный, темнота

Автоматически созданное описание

Таким образом, сезонная компонента *S:*

𝑆 = (𝑠1, 𝑠2, … , 𝑠𝑛, 𝑠1, 𝑠2, … , 𝑠𝑛, … , 𝑠1, 𝑠2, … , 𝑠𝑛)𝑇

|𝑆| = 𝑛 ∙ ℎ

### Структурная идентификация модели

Анализируя первоначальный график данных, мы можем сделать вывод о том, что мультипликативная модель представляется более подходящей для данной ситуации. Визуально график не соответствует форме линейной функции; напротив, он скорее напоминает экспоненциальный тренд.

Мультипликативная модель эффективна в случаях, когда изменения в данных пропорциональны их текущему уровню. Если процентное изменение данных остается постоянным, то предпочтение следует отдать мультипликативной модели.

### Параметрическая идентификация модели

Сглаживание методом скользящей средней используется для уменьшения шума и выявления общих трендов в временных рядах. Этот метод является техникой анализа временных данных, при которой каждое значение заменяется средним значением по некоторому окну предшествующих значений. Основные цели применения скользящего среднего включают:

1. **Удаление шума и флуктуаций:** Сглаживание позволяет устранить краткосрочные колебания в данных, что может облегчить выявление более долгосрочных трендов.
2. **Выявление общих паттернов:** Метод скользящей средней помогает выделить общие паттерны и тенденции, делая их более заметными на графиках и обеспечивая более ясное представление о долгосрочной динамике данных.
3. **Сглаживание краткосрочных колебаний:** Применение скользящего среднего позволяет сгладить краткосрочные изменения, что полезно при анализе данных с выраженной сезонностью или случайными флуктуациями.
4. **Создание прогнозов:** В некоторых случаях скользящее среднее может использоваться для создания прогнозов на основе усредненных значений.
5. **Уменьшение влияния выбросов:** Сглаживание методом скользящей средней может помочь уменьшить влияние выбросов или аномалий, что делает данные менее чувствительными к случайным колебаниям.

**Анализ временных рядов с использованием метода скользящих средних**

В ходе исследования временных рядов, связанных с квартальными данными, я провел разделение данных на группы, с каждой группой включающей 16 сезонов. Для анализа и сглаживания данных в каждой группе был применен метод скользящих средних.

Исходный временной ряд был структурирован по кварталам, что позволило выделить 19 групп, в каждой из которых содержится по 16 сезонов. Это разделение позволило более детально рассмотреть изменения внутри каждой группы и провести анализ трендов и сезонных колебаний.

Для сглаживания данных в каждой группе был применен метод скользящих средних. Этот метод позволяет уменьшить влияние краткосрочных флуктуаций, выделяя общие тенденции в данных. Применение скользящих средних помогло улучшить восприятие временных рядов и выявить общие закономерности.

При выравнивании теряется несколько членов исходного ряда с обоих концов. Их можно восстановить путем экстраполяции сглаженного ряда:

Изображение выглядит как черный, темнота

Автоматически созданное описание

Для этого мы написали функцию calculate\_smooth и экстраполировали с помощью функциии extrapolate

Результатом действий должна быть колонка в наборе данных, заполненная скользящими средними.

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок - Заполненная колонка со скользящими средними

На графиках скользящей средней и изначальных значений при большом количестве данных сложно найти расхождения. Для более детального рассмотрения можно просмотреть фрагмент этого графика за 4 года

Изображение выглядит как текст, снимок экрана, График, диаграмма

Автоматически созданное описание

Рисунок - Заполненная колонка со скользящими средними

**Определение сезонной компоненты.**

Следующим шагом является выделение сезонной компоненты. Мы ищем периодические колебания, которые могут быть связаны с сезонными факторами, такими как месяцы или времена года. Это может быть достигнуто путем вычитания от сглаженных данных из исходного ряда.

Находятся оценки сезонной компоненты как разность между

фактическими уровнями ряда и скользящими средними:

Определяются средние значения сезонных компонент за весь период:

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

Рисунок - Сезонная компонента в процентах

Эта усредненная сезонная компонента будет использоваться для детализации степенной функции.

Для предсказания продаж конкретного месяца будет использоваться функция:

Изображение выглядит как снимок экрана, График, линия, диаграмма

Автоматически созданное описание

Рисунок - Степенная часть композитной модели

Построенная модель GPDI позволяет выполнять несколько важных задач в контексте анализа и прогнозирования инвестиций частного сектора (Gross Private Domestic Investment):

1. **Прогнозирование будущих значений:** Модель позволяет предсказывать будущие значения инвестиций частного сектора на основе предыдущих данных. Это может быть полезно для планирования и принятия управленческих решений.
2. **Идентификация трендов и циклов:** Анализ модели может помочь выделить общие тренды и циклические колебания в инвестициях. Это важно для понимания долгосрочных тенденций и прогнозирования циклических изменений в экономике.
3. **Оценка влияния факторов:** Модель может быть настроена на включение различных факторов, таких как ставки процента, инфляция, изменения валютных курсов и другие экономические показатели. Это позволяет оценить, как эти факторы могут влиять на инвестиции.
4. **Управление рисками:** Анализ модели может помочь идентифицировать потенциальные риски и неопределенности, связанные с инвестициями частного сектора. Это полезно для разработки стратегий управления рисками.
5. **Определение эффективности политики:** Если модель была построена для анализа влияния экономической политики на инвестиции, она может использоваться для оценки эффективности различных мероприятий и мер государственной поддержки.
6. **Экономическое прогнозирование:** Путем интеграции модели в более широкий экономический контекст она может использоваться для формирования более общих экономических прогнозов.

# Исследование модели

## Анализ статистической значимости уравнения регрессии

Критерий F используется для проверки статистической значимости регрессионной модели в целом. Если значение F-критерия значительно, это указывает на то, что объясненная изменчивость моделью является статистически значимой.

Формула для расчета F-критерия в контексте анализа регрессии:

SSR - объясненная сумма квадратов (сумма квадратов регрессии)

SSE - остаточная сумма квадратов (сумма квадратов ошибок)

n – количество наблюдений

m – количество факторов (предикторов) модели

Кроме того, значение P-значения ассоциировано с F-критерием и представляет собой вероятность получения такого или более экстремального значения F при условии, что нулевая гипотеза верна. Низкое P-значение (обычно меньше выбранного уровня значимости, например, 0.05) говорит о том, что можно отклонить нулевую гипотезу и считать модель статистически значимой.

Полученное F-критерия равно 32.72, что сравнивается с критическими значениями для уровней значимости 10% и 5%. Высокое значение F-критерия и низкое P-значение 0.0002868 указывают на статистическую значимость модели. Это означает, что модель в целом имеет существенное воздействие на зависимую переменную, и регрессионные коэффициенты не равны нулю.

## Анализ статистической значимости коэффициентов уравнения регрессии

Связь между F-статистикой Фишера, t-статистикой Стьюдента для коэффициента регрессии, t-статистикой Стьюдента для коэффициента корреляции выражается равенством:

Таким образом, . Знак статистики Стьюдента для коэффициента регрессии и знак собственно коэффициента регрессии совпадают. Табличное значение t-статистики для α=0,05, v=9 составляет 2.26215. Поскольку t факт > t табл., то коэффициент регрессии b статистически значимо отличен от нуля.

Определим доверительный интервал для коэффициента регрессии b с надежностью 95%:

= 0.17825

Границы доверительного интервала, в котором с надежностью 95% лежит истинное значение коэффициента регрессии, составят:

В границы интервала ноль не попадает, следовательно, оцениваемый параметр статистически значим на уровне α=0,05 и сформировался под влиянием систематически действующего фактора х.