



МИНОБРНАУКИ РОССИИ

**федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный университет экономики и управления «НИНХ»
(ФГБОУ ВО «НГУЭУ», НГУЭУ)**

Кафедра статистики

КУРСОВАЯ РАБОТА

Многофакторные модели прогнозирования и их основные типы

Дисциплина: Анализ временных рядов и прогнозирование

Ф.И.О студента: Калюжин Максим Александрович

Направление: 38.03.01 Экономика

Направленность (профиль): Бизнес-аналитика и экономическое
прогнозирование

Номер группы: БА201

Номер зачетной книжки: 220465

Проверил: Глинский Владимир Васильевич, доктор эконом.наук, профессор

Новосибирск 2024



МИНОБРНАУКИ РОССИИ

**федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный университет экономики и управления «НИНХ»
(ФГБОУ ВО «НГУЭУ», НГУЭУ)**

Кафедра статистики

**ЗАДАНИЕ
на курсовую работу**

Тема: Многофакторные модели прогнозирования и их основные типы

ФИО студента: Калюжин Максим Александрович

Группа: БА201

Перечень подлежащих разработке вопросов и календарный график

№ п/п	Наименование вопросов, подлежащих разработке (этапы работы)	Срок выполнения
1	Выбор темы курсовой работы	15.10.2024
2	Согласование содержания курсовой работы	12.11.2024
3	Поиск источников литературы	30.11.2024
4	Написание введения	12.12.2024
5	Написание первой главы	12.12.2024
6	Написание второй главы	14.12.2024
7	Написание практической части	18.12.2024
8	Проверка работы на антиплагиат	21.12.2024
9	Сдача работы на проверку руководителю	23.12.2024

Дата выдачи задания 15 октября 2024 г.

Срок сдачи работы 23 декабря 2024 г.

Руководитель

Глинский В.В

Задание получил студент

Калюжин М.А.



МИНОБРНАУКИ РОССИИ

**федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный университет экономики и управления «НИНХ»
(ФГБОУ ВО «НГУЭУ», НГУЭУ)**

Кафедра статистики

**ЗАЯВЛЕНИЕ
о самостоятельном характере выполненной работы**

Я, Калюжин Максим Александрович, студент(ка) группы БА201, направления подготовки 38.03.01 Экономика направленности (профиля) Бизнес-аналитика и экономическое прогнозирование, заявляю, что в моей курсовой работе, выполненной на тему «Многофакторные модели прогнозирование и их основные типы» не содержится элементов плагиата.

Все заимствования из печатных и электронных источников, а также из защищенных ранее письменных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

21 декабря 2024 г.

М.А.Калюжин

Результаты проверки в системе «Антиплагиат»

Оригинальность авторского текста с учетом цитирования в результате автоматизированной проверки составила 69,78 %.

Руководитель курсовой работы Глинский Владимир Васильевич, доктор эконом.наук, профессор

21 декабря 2024 г.



МИНОБРНАУКИ РОССИИ

**федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный университет экономики и управления «НИНХ»
(ФГБОУ ВО «НГУЭУ», НГУЭУ)**

Кафедра статистики

РЕЦЕНЗИЯ
на курсовую работу

Тема: Многофакторные модели прогнозирования и их основные типы

ФИО студента: Калюжин Максим Александрович

Группа: БА201

№ п/п	Критерии оценки	Оценочные баллы	Баллы по результатам работы
Итого		100	

Шкала итоговой оценки:

Неудовлетворительно

Удовлетворительно

Хорошо

Отлично

0-40

41-60

61-80

81-100

Оценочное заключение: Выберите элемент.

Руководитель

Место для ввода даты.

И.О. Фамилия

Содержание

Введение	6
1. Основы многофакторного прогнозирования	8
1.1 Понятие и необходимость многофакторных моделей	8
1.2. Классификация методов прогнозирования.....	9
1.3. Преимущества и недостатки многофакторного подхода.....	12
2. Типы многофакторного прогнозирования.....	13
2.1. Регрессионные модели.....	13
2.2. Модели временных рядов.....	15
2.3. Машинное обучение в прогнозировании.....	17
3. Методы оценки параметров модели	21
4. Перспективные расчёты ключевой ставки РФ с 2013-2023 гг.	23
4.1. Построение модели ARIMA	23
4.2. Оценка качества и точности модели	30
Вывод.....	33
Список использованной литературы.....	34
ПРИЛОЖЕНИЕ	37

Введение

В настоящее время значительная часть экономических и финансовых решений основана на прогнозировании. Прогнозирование позволяет минимизировать риски, а также планировать стратегии реагирования на них и принимать обоснованные бизнес-решения. Одним из ключевых инструментов для прогнозирования являются многофакторные модели - модели, которые, в отличие от однофакторных моделей, учитывают влияние различных факторов на исходную величину. Данный тип моделей предоставляет более точные и надежные результаты по сравнению с однофакторными моделями, поскольку включают в себя широкий спектр данных, что позволяет учесть взаимосвязи между различными переменными.

В рамках данной курсовой работы будут рассмотрены основные типы многофакторных моделей прогнозирования, их типы и особенности и методология построения моделей. Особое внимание будет уделено линейным и нелинейным моделям и методам машинного обучения, которые активно используются в современных экономических и финансовых областях исследования для анализа сложных зависимостей и работы с большими объемами данных.

Актуальность темы многофакторных моделей прогнозирования связана с стремительным развитием таких технологий, как искусственный интеллект, и расширение доступности данных открыли новые возможности для повышения качества прогнозов. В сложной и нестабильной экономической и финансовой среде традиционные методы и инструменты прогнозирования часто оказываются недостаточно эффективными, что подчеркивает необходимость применения более сложных методов анализа данных.

Цель данной работы заключается в исследовании многофакторных моделей прогнозирования, их основных типов, а также в выявлении особенностей их применения в различных областях. Для достижения поставленной задачи необходимо решить следующие задачи:

- Проанализировать основные теоретические подходы к построению многофакторных моделей прогнозирования;
- Классифицировать и описать основные типы моделей, включая линейные и нелинейные методы;
- Изучить роль методов машинного обучения и нейронных сетей в прогнозировании на основе многофакторных моделей;
- Рассмотреть области практического применения многофакторных моделей на основе прогнозирования показателя ключевой ставки в РФ.

В рамках данной курсовой работы будут изучены как классические методы прогнозирования, так и современные подходы, и технологии, что позволит комплексно оценить перспективы и потенциал применения многофакторных моделей в различных областях.

1. Основы многофакторного прогнозирования

1.1 Понятие и необходимость многофакторных моделей

Многофакторное прогнозирование - это математическое описание объекта или явления. Многофакторное прогнозирование – это более сложный и трудоёмкий инструмент, так как в отличие от статистической экстраполяции, многофакторное моделирование подразумевает использование информации для построения моделей не только об объекте или явлении, но и о факторах внешней среды, влияющих на динамику и структуру изучаемого объекта или явления – множества независимых переменных.

Необходимость многофакторных моделей заключается в их способности учитывать влияние множества факторов одновременно, что позволяет более точно прогнозировать результаты и анализировать сложные взаимосвязи между переменными. Эти модели помогают глубже понять структуру изучаемого объекта или явления, что помогает выявить скрытые закономерности и минимизировать ошибки прогноза, возникающие при использовании простых моделей, углубляя понимание структуры изучаемых объектов или явлений.

Стоит отметить, что многофакторное прогнозирование не способно учитывать непредвиденные события, так называемые “черные лебеди”. Эти непредсказуемые события выходят за рамки привычных статистических моделей, так как они не поддаются прогнозированию на основе известных факторов и закономерностей.

1.2. Классификация методов прогнозирования

В зависимости от целей, характера данных и условий применения прогнозируемых моделей, методы прогнозирования можно разделить на 2 основные группы: качественные и количественные методы. Каждая группа включает в себя отдельные подходы, которые имеют различия в принципах и применяемом инструменте.

1. Качественные методы прогнозирования

Качественные методы прогнозирования основаны на экспертных оценках и интуиции, они зачастую используются в ситуациях, когда отсутствует достаточно количество данных для применения математических моделей. Данные методы ориентированы на субъективные суждения и выводы, основанные на опыте и знаниях экспертов, проводящих оценку.

- Метод экспертных оценок. Данный метод заключается в сборе и обобщении мнения группы экспертов для получения прогноза;
- Метод сценариев. Этот метод строится следующим образом: строятся несколько возможных сценариев развития событий, которые впоследствии анализируются на предмет вероятности реализации;
- Метод “Дельфи”. Данный метод предполагает проведение нескольких раундов (сессий) опросов экспертов с целью достижения консенсуса относительно будущих изменений.

Чаще всего качественные методы применяются в том случае, когда количественные данные отсутствуют или их недостаточно для применения более точных методов прогнозирования.

2. Количественные методы прогнозирования.

Количественные методы можно классифицировать следующим образом:

2.1 Экстраполяционные методы

Методы экстраполяции основываются на предположении, что закономерности и исторические тренды, наблюдаемые в прошлом, будут сохраняться и в будущем. Прогноз строится путём выявления тенденций.

Метод аналитического выравнивания. Прогноз осуществляется с помощью уравнения тренда, описывающего долгосрочную динамику. В простейшем случае используется линейный тренд, выраженный уравнением:

$$y_t = a + b * t$$

где:

- y_t – прогнозируемое значение в момент времени t ;
- a – константа (среднее значение);
- b – коэффициент наклона – показывает скорость изменения переменной;
- t – временной параметр.

В зависимости от характера данных могут применяться также экспоненциальные или степенные уравнения тренда.

Метод скользящего среднего. Данный метод используется для сглаживания временных рядов и исключения случайных колебаний. Скользящие средние могут быть 3-хзвенные, 5-звенные и т.д.

$$SMA = \frac{1}{n} \sum_{i=1}^n x_i,$$

где:

- SMA (Simple Moving Average) – значение скользящего среднего;
- x_i – значение ряда динамики;
- n – количество наблюдений в рассматриваемом периоде.

2.2 Регрессионные методы

Регрессионные методы применяются для установления количественных зависимостей между одной зависимой переменной и несколькими факторными переменными. Множественная линейная регрессия описывается следующим образом:

$$y = a + b_1x_1, b_2x_2, \dots, b_nx_n + e,$$

где:

y – зависимая переменная;

x_1, x_2, \dots, x_n – независимые переменные;

b_1, b_2, \dots, b_n – коэффициенты регрессии, отражающие вклад каждой независимой переменной;

e – случайная ошибка.

Регрессионные методы будут подробно разобраны в следующих главах, посвящённых типам и методам многофакторного прогнозирования.

1.3. Преимущества и недостатки многофакторного подхода

Несмотря на свою универсальность и эффективность, многофакторное прогнозирование обладает как преимуществами, так и недостатками.

Преимуществом многофакторного прогнозирования можно выделить учёт множества факторов. Например, в экономическом прогнозировании можно одновременно учитывать такие факторы, как экономический рост, инфляция, уровень безработицы и т.д., что даёт наиболее полное представление о возможных сценариях развития будущего. Это особенно важно в сложных системах, в которых воздействие одного фактора невозможно оценить без учёта влияния других факторов.

Одним из главных недостатков многофакторного прогнозирования является сложность реализации и необходимость большого объёма данных для корректного прогнозирования и работы моделей. Для точного прогнозирования требуется качественная информация по каждому из факторов – её сбор может быть трудоёмким и достаточно затратным процессом. Также стоит отметить, что использование большого количества независимых переменных повышает вероятность возникновения мультиколлинеарности – явления, когда отдельные факторы оказываются зависимыми друг от друга, что приводит к искажению результатов и снижению точности прогноза.

Многофакторное прогнозирование требует применения сложных математических и статистических методов. Это требует высокого уровня технической и теоретической подготовки и использования специального ПО.

2. Типы многофакторного прогнозирования

2.1. Регрессионные модели

Регрессионные модели – один из наиболее распространённых типов многофакторного прогнозирования, применяемых для анализа и предсказания зависимости между зависимой переменной и несколькими независимыми переменными. Основная задача – определить, как изменение одной или нескольких независимых переменных (факторов) влияет на зависимую переменную, также построить математическую модель, которая бы описывала эти отношения.

Регрессионные модели разделяются на несколько основных групп: простая и множественная регрессия, линейная, степенная, экспоненциальная, полиномиальная и т.д. Рассмотрим подробнее простую и множественную регрессию:

1. Простая линейная регрессия – используется для прогнозирования значений зависимой переменной y на основе одной независимой переменной x . Модель просто линейной регрессии имеет следующий вид:

$$y = \beta_0 + \beta_1 x + e,$$

Где:

- y – зависимая переменная;
- x – независимая переменная;
- β_0 – свободный член;
- β_1 – коэффициент регрессии, показывающий изменение y при изменении x на одну единицу;
- ε – ошибка прогноза (всё, что не объясняется моделью).

Простая линейная регрессия используется крайне редко, поскольку зачастую явления, требующие предсказания, зависят от 2 и более факторов, как, например, значение ключевой ставки в РФ.

2. Множественная линейная регрессия – обобщение простой линейной регрессии, при котором зависимая переменная y прогнозируется на основе нескольких независимых переменных x_1, x_2, \dots, x_n . Данная модель имеет следующий вид:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + e,$$

Здесь x_1, x_2, \dots, x_n – независимые переменные, каждая из которых влияет на зависимую переменную y , а $\beta_1, \beta_2, \dots, \beta_n$ – коэффициенты регрессии, характеризующие силу и направление соответствующего фактора.

2.2. Модели временных рядов

Модели временных рядов (рядов динамики) – это статистические модели, которые используются для анализа последовательных наблюдений данных, собранных через равные промежутки времени. Их задача заключается в выявлении структуры данных и создания прогноза на основе прошлого поведения ряда. Существуют различные модели временных рядов, каждая из которых используется в зависимости от свойств данных и задач анализа.

1. Автокорреляция и частичная автокорреляция

Для анализа рядов динамики важно понимать зависимость между текущими значениями и их прошлыми значениями. Это выражается через функцию автокорреляции (ACF) и частичной автокорреляции (PACF).

Функция автокорреляции имеет следующий вид:

$$r_k = \frac{E[(X_t - \mu)(X(t - k) - \mu)]}{\sigma^2}$$

Где r_k – автокорреляция порядка k , E – математическое ожидание, X_t и $X(t-k)$ – значения ряда динамики, а σ^2 – дисперсия.

Частичная автокорреляция отображает, как текущее значение ряда динамики связано с его прошлыми значениями, исключая влияние промежуточных лагов. В математическом виде частичная автокорреляция на k -м лаге может быть записана через условные автокорреляции:

$$\alpha_k = \text{Corr}(X_t, X(t - k) | X(t - 1), X(t - 2), \dots, X(t - k + 1))$$

Автокорреляция и частичная автокорреляция часто применяются для прогнозирования рядов динамики. Эти инструменты активно используются при построении статистических моделей ARIMA и SARIMA.

Статистические модели прогнозирования рядов динамики SARIMA и ARIMA будут подробно рассмотрены в главе 3.3 “Машинное обучение в прогнозировании”.

После построения модели важно провести её диагностику для оценки качества.

Основные методы диагностики:

- Анализ остатков – остатки должны быть белым шумом, т.е. не должны проявлять автокорреляции;
- Q-тест Бокса-Пирса – проверка гипотезы о некоррелированности остатков;
- AIC/BIC критерии – данные критерии используются для сравнения различных моделей и выбора по критерию сложности и качества;
- Тест Дики-Фуллера – проверка стационарности ряда динамики;
- Кросс-валидация по временным рядам – метод разделения данных на тренировочную и тестовую выборки с учётом временной структуры, что позволяет оценить способность модели делать прогноз.

2.3. Машинное обучение в прогнозировании

“Искусственный интеллект – совокупность методов и инструментов решения различных сложных прикладных задач, использующих принципы и подходы, аналогичные размышлениям человека над их решением” [2, 2 с.].

“Машинное обучение – методика анализа данных, которая позволяет компьютерам самостоятельно обучаться посредством решения массива сходных задач” [3, 1 с.]. Для прогнозирования используются следующие методы машинного обучения:

1. Рекуррентные нейронные сети (RNN) – один из наиболее распространенных методов машинного обучения (глубокого обучения), суть которого заключается в последовательном использовании информации, то есть той информации, которая имеет строгую упорядоченность и взаимосвязи на основе данных (в случае прогнозирования – временных рядов).

2. Градиентный бустинг – метод, используемый для решения задач классификации и регрессии путём построения моделей предсказания в виде набора слабых моделей предсказания (обычные деревья решений). Такие модели очень слабы, но благодаря последовательному построению решений ошибки минимизируются, а результаты становятся лучше, говоря более математическим языком, минимизируется функция потерь.

3. Гауссово распределение – гауссовы процессы успешно применяются для решения задач машинного обучения, таких как классификация, регрессия и задачи уменьшения размерности. Зачастую их используют для автоматической сложности модели и оценки неопределённости прогноза.

4. Методы опорных векторов (SVM) – суть метода заключается в отображении векторов пространства признаков в пространства более высокой размерности [1, 84-85 с.]. Он связан с линейной разделимостью данных: линейная разделимость множества ниже, чем в пространстве большей

размерности. Ожидаемое качество распознавания выше, когда для распознавания используется больше признаков.

5. Случайный лес – “метод основан на построение большого числа(ансамбля) деревьев решений (это число является параметром метода), каждое из которых строится по выборке, получаемой из исходной выборки с помощью выборки с возвращением” [5, 4 с.].

В рамках практической части курсовой работы будут использованы статистические модели прогнозирования временных рядов SARIMA и ARIMA.

ARIMA - статистическая модель, которая объединяет три компонента: авторегрессию (AR), интегрированность (I) и скользящее среднее (MA).

AR – компонент авторегрессии предполагает, что текущее значение временного ряда зависит от нескольких его предыдущих значений. Модель AR(p) задаётся следующим образом:

$$X_t = \varphi_1 X(t - 1) + \varphi_2 X(t - 2) + \dots + \varphi_p X(t - p) + e_t,$$

Где:

X_t – значение ряда динамики в момент времени t ;

φ_i – коэффициент авторегрессии;

e_t – ошибка модели.

Интегрированность (I) – интегрированный компонент отвечает за стационарность данных. Стационарный ряд имеет постоянные средние и дисперсию. Для преобразование нестационарного ряда динамики в стационарный применяет дифференцирование (интеграция). Если ряд необходимо дифференцировать d -раз, то модель называется интегрированной на порядок d .

Скользящее среднее (МА) – компонент скользящего среднего моделирует зависимость текущего значения от прошлых ошибок. Модель МА(q) выражается как:

$$X_t = e_t + \theta_1 e(t-1) + \theta_2 e(t-2) + \dots + \theta_q e(t-q),$$

Где:

θ_i – коэффициенты скользящего среднего;

e_t – случайная ошибка.

Полная модель ARIMA (p, d, q) объединяет эти три компонента в следующее уравнение:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^d X_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) e_t,$$

Где B – оператор сдвигов(лагов), например, $BX_t = X_{t-1}$

SARIMA

SARIMA моделирует как сезонные, так и несезонные аспекты ряда динамики. Она имеет такие же параметры, как и ARIMA, плюс сезонные параметры.

Полная модель SARIMA (p, d, q) (P, D, Q, s) включает:

- (p, d, q) – это параметры ARIMA для несезонных данных.
- (P, D, Q, s) – сезонные параметры, где:
 - P – порядок авторегрессии для сезонности;
 - D – количество дифференцирований для устранения сезонности;
 - Q – порядок скользящего среднего для сезонности;
 - s – длина сезонного цикла (например, 12 - для месячных данных с годовой сезонностью).

Формально модель SARIMA записывается как:

$$\Phi_P(B^s)\varphi_p(B)(1 - B)^d(1 - B^s)^D X_t = \Theta_Q(B^s)\theta_q(B) e_t$$

Где:

- B – оператор сдвига, который уменьшает время на один шаг;
- $\Phi_p(B^s)$ – полином авторегрессии для сезонных данных;
- $\phi_p(B)$ – полином авторегрессии для несезонных данных
- $\Theta_Q(B^s)$ – полином скользящего среднего для сезонных данных;
- $\Theta_q(B)$ – полином скользящего среднего для несезонных данных;
- e_t – ошибка модели.

Стоит отметить, что SARIMA объединяет как сезонные, так и трендовые составляющие ряда динамики, что делает данную модель достаточно мощным инструментом для прогнозирования рядов динамики с циклическими колебаниями.

3. Методы оценки параметров модели

Для оценки коэффициентов $\beta_1, \beta_2, \dots, \beta_n$ зачастую используют метод наименьших квадратов (МНК), задача которого заключается в минимизации суммы квадратов отклонений наблюдаемых значений от прогнозируемых моделью:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min$$

Где:

- y_i – фактическое значение зависимой переменной для наблюдения i ;
- \hat{y}_i – прогнозируемое моделью значение для наблюдения i .

Решение задачи МНК приводит к системе нормальных уравнений:

$$X^T * X\beta = X^T * Y,$$

где X – матрица исходных переменных, Y – вектор значений зависимой переменной, β – вектор оценок коэффициентов регрессии.

Для проверки статистической значимости модели используют критерий Фишера, который имеет вид:

$$F = \frac{R^2}{1 - R^2} * \frac{n - m - 1}{m},$$

Где R^2 – коэффициент детерминации, n – число наблюдений, m – число факторов в уравнении регрессии.

Значение F-критерия Фишера сравнивается с критическим значением из таблицы распределения Фишера для заданного уровня значимости. Если расчётное значение превышает критическое, то уравнение считается статистически значимой.

Для оценки качества модели (регрессионного уравнения) также используется коэффициент детерминации R^2 , который показывает какая доля общей вариации зависимой переменной объясняется моделью:

$$R^2 = 1 - \frac{\sum ei^2}{\sum (yi - \hat{yi})^2}$$

Где $\sum ei^2$ – сумма квадратов остатков регрессии, $\sum (yi - \hat{yi})^2$ – сумма квадратов отклонений точек данных от среднего значения.

Недостатком регрессионных моделей является предположение о линейности зависимости между переменными. В действительности зависимости могут быть нелинейными, что требует использование более сложных моделей, таких как полиномиальная регрессия или методы машинного обучения.

4. Перспективные расчёты ключевой ставки РФ с 2013-2023 гг.

4.1. Построение модели ARIMA

Практическая часть курсовой работы будет реализована на языке программирования Python, поскольку данный язык программирования обладает широким спектром аналитических инструментов, а также библиотек для построения и обучения моделей машинного обучения, которые нельзя реализовать, используя другие инструменты, например, Microsoft Excel.

Перед построением модели необходимо собрать данные о ключевой ставке в РФ, все данные были взяты из официальных источников, таких как ЦБ РФ, Минфин, Росстат и т.д. Таблица с данными представлена в приложении А.

Процесс построения и оценки модели был разделён на 4 этапа:

1. Сбор и предобработка данных;
2. Первичный анализ и визуализация данных;
3. Построение модели прогнозирования;
4. Оценка качества модели;

1. Сбор и предобработка данных:

Для многофакторного прогнозирования ключевой ставки в РФ за 2013 – 2023 гг. было отобрано 6 признаков: инфляция, % г/г, денежная масса(M2), млрд руб., уровень безработицы по РФ, %, курс USD/RUB, объем выданных кредитов, млн руб., торговый баланс, млрд \$. После сбора данных была проведена предобработка данных: все данные преобразовались в нужный формат, а даты проиндексированы.

2. Первичный анализ и визуализация данных:

На следующем этапе был выполнен тест Дики-Фуллера для определения стационарности ряда динамики. Тест показал следующие результаты:

- ADF статистика: -2.98;
- p-value: 0.04;
- Критические значения:
 - Для уровня значимости 1%: -3.49;
 - Для уровня значимости 5%: -2.89;
 - Для уровня значимости 10%: -2.58.

Так как p-value меньше 0,05, а ADF- статистика меньше критических значений на 5 и 10%, то отвергается нулевая гипотеза, следовательно, данные стационарны.

Для многофакторного прогноза ключевой ставки в РФ была выбрана статистическая модель ARIMA, для построения этой модели необходимо определить параметры p , q , d . Для определения параметров p , q построим график автокорреляции (ACF) и частичной автокорреляции (PACF):

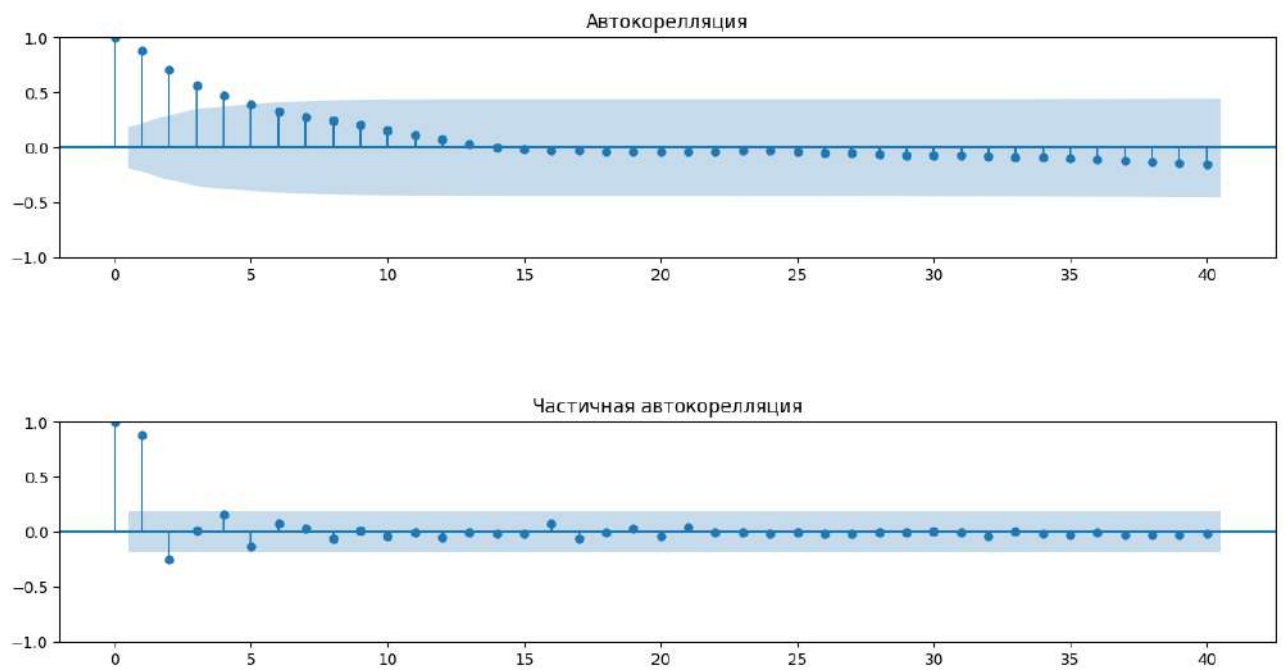


Рисунок 1 - Визуализация автокорреляции и частичной автокорреляции

На графиках было отображено 40 лагов, поскольку 40 лагов – это достаточное количество для выявления ключевых закономерностей, но при этом график достаточно информативный и не перегруженный.

Для определения мультиколлинеарности построим корреляционную матрицу:

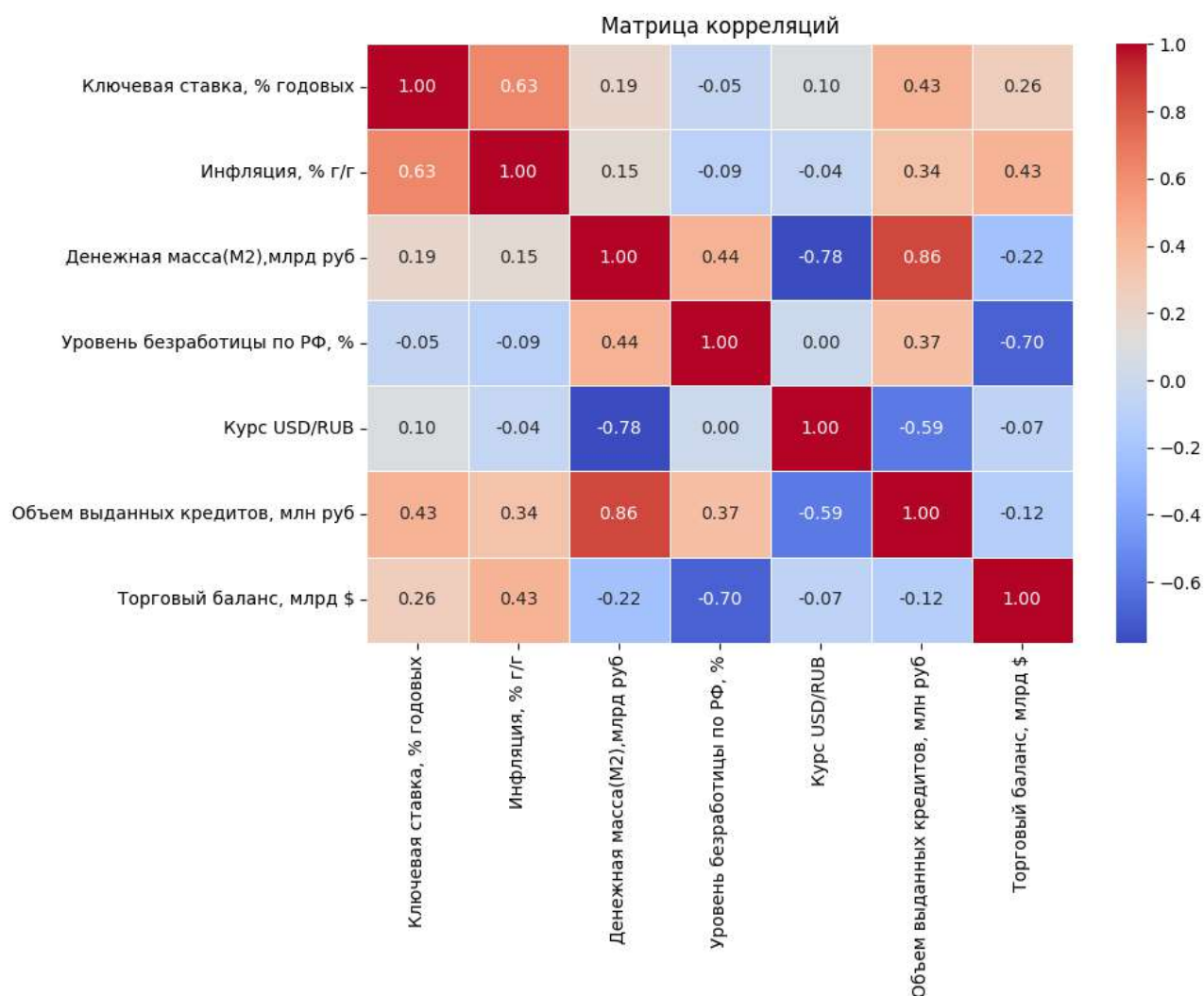


Рисунок 2 - Матрица корреляций

Исходя из визуального анализа можно выявить следующие пары с высокой корреляцией:

- Курс USD/RUB – Денежная масса(M2), млрд руб.;
- Объём выданных кредитов, млн руб. – Денежная масса(M2), млрд руб.;
- Торговый баланс, млрд \$ - Уровень безработицы по РФ, %.

В итоговую модель войдут следующие экзогенные переменные:

1. Инфляция, % г/г;
2. Курс USD/RUB;
3. Объем выданных кредитов, млн руб.;
4. Торговый баланс, млрд \$.

3. Построение модели прогнозирования:

Следующим шагом необходимо стандартизировать экзогенные переменные, это необходимо сделать для масштабирования признаков.

Для многофакторного прогнозирования с помощью модели ARIMA необходимо определить оптимальные параметры p , q , d . Определим оптимальные параметры с помощью `auto_arima` из статистической библиотеки в Python. `Auto_arima` ищет оптимальную конфигурацию параметров, используя следующие методы и подходы:

1. Систематический перебор различных комбинаций параметров;
2. Информационный критерий Акаике (AIC)/Информационный критерий Байеса(BIC);
3. Дифференцирование;
4. Тестирование стационарности;
5. Интеграция сезонности;
6. Оценка ошибок остатков.

По результатам анализа графика автокорреляции и частичной автокорреляции были определены первичные параметры: $p = 1$, $q = 0$.

Далее для определения оптимальных параметров модели после перебора, воспользуемся информационным критерием Акаике, который выражается в следующем виде:

$$AIC = 2k - \ln(L)$$

Где k – число параметров в статистической модели, в нашем случае 3, а L – максимизированное значение функции правдоподобия модели.

AIC при параметрах: $p = 1$, $q = 0$ и $d = 3$ составил 405,690, что является минимальным значением.

Также воспользуемся критерием BIC (Байесовский информационный критерий), который выражается в следующем виде:

$$BIC = -2 \ln(L) + k * \ln(n),$$

Где:

- L – значение функции правдоподобия модели;
- k – количество параметров в модели;
- n – количество наблюдений.

После определения системой оптимальных параметров получаем следующие результаты:

$$p = 1, q = 0, d = 3$$

Далее в модели ARIMA определяем параметры: $p = 1$, $q = 0$, $d = 3$, передаём стандартизированные экзогенные переменные и строим прогноз.

На графике представлены фактические данные о ключевой ставке в России за период с 2013 по 2023 годы, фактические значения за 2024 год и прогнозируемые значения:



Рисунок 3- Визуализация прогноза, фактических значений

Исходя из прогноза на 2025-2026 годы, можно заключить, что ключевая ставка скорректируется до 16,5% к 2025 году, а к 2026 году достигнет уровня 13-15% годовых.

Однако стоит отметить, что модель не является идеальной и нуждается в доработке. В процессе построения модели не были учтены геополитические факторы, что и исказило точность прогноза на валидационных данных, учёт этих факторов в будущем позволит повысить качество модели и точность прогноза , а также адаптировать модель к изменяющимся внешним условиям.

4.2. Оценка качества и точности модели

Для оценки качества и точности модели для начала проанализируем график остатков:



Рисунок 4 - График остатков

Выбросы в остатках можно охарактеризовать событиями, которые происходили в периодах с аномальными отклонениями:

1. 2015-2017 гг. — финансовый кризис в РФ, характеризующийся резкой девальвацией рубля, начавшийся во второй половине 2014 года.
2. 2021-2022 гг. — российско-украинский кризис 2021-2022 гг.

Остатки не имеют систематических отклонений, нет автокорреляции, ошибки случайны.

Далее рассмотрим распределение остатков:

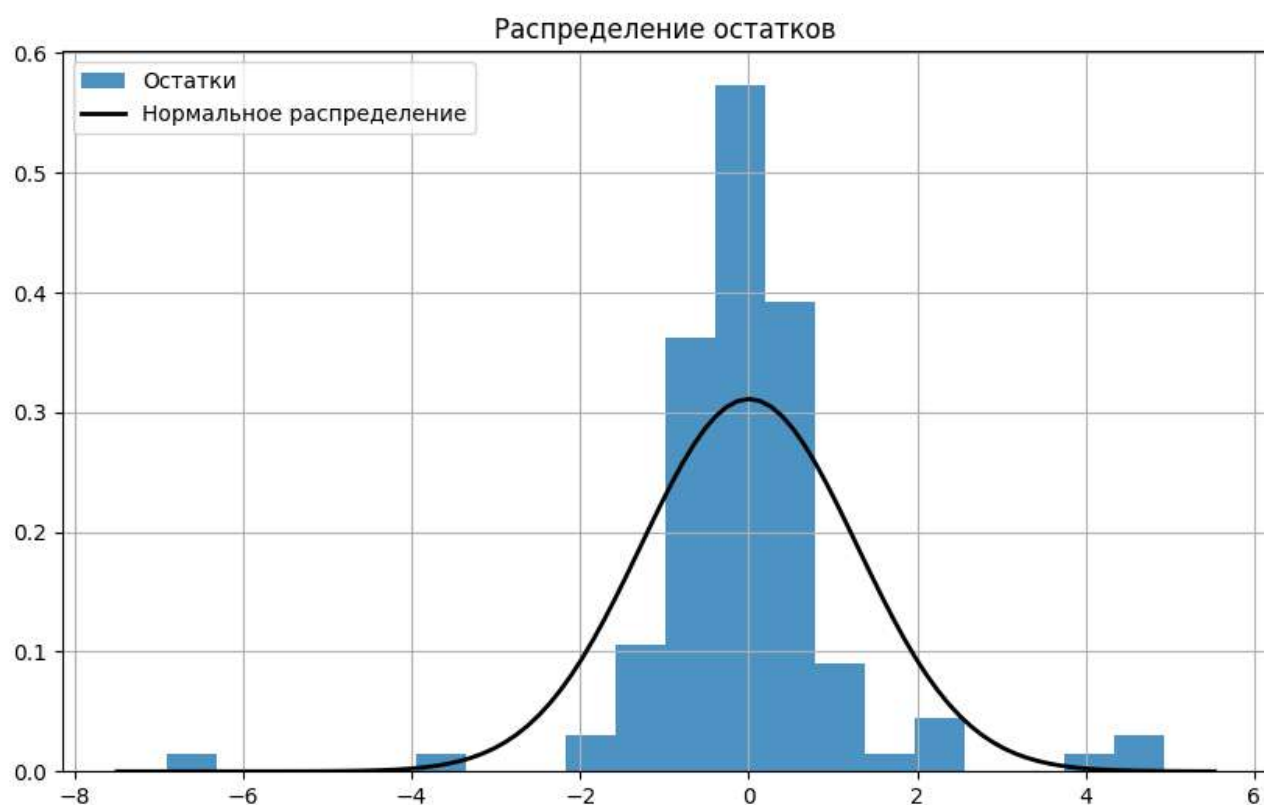


Рисунок 5 - График распределения остатков

Для подтверждения или опровержения гипотезы о нормальном распределении остатков на уровне значимости 0,05, воспользуемся тестом Колмогорова-Смирнова, получаем следующие результаты:

- $p\text{-value} = 0,1690$

Так как значение $p\text{-value}$ больше критического значения можно сделать вывод, что не отвергается гипотеза о нормальном распределении остатков на уровне значимости 0,05.

Для оценки модели были собраны фактические данные за 2024 год – валидационные данные – с которыми сравниваются прогнозные данные для определения качества прогноза, получаем следующие результаты:

```
Средняя квадратическая ошибка: 3.48
Средняя абсолютная ошибка: 2.75
Средняя ошибка аппроксимации: 19.11 %
Детерминация равна 57.51 %

F-статистика : 2.0842
F-критическое значение: 1.5906
```

Рисунок 6 - Оценка модели в Python

- Средняя квадратическая ошибка равна 3.48;
- Средняя абсолютная ошибка равна 2.75;
- Средняя ошибка аппроксимации равна 19.11%;
- Коэффициент детерминации(R^2) равен 0.5751;
- F-критерий фактический равен 2.0842
- F-критерий критический равен 1.5906.

В связи с тем, что средняя ошибка аппроксимации больше 15%, можно сделать вывод, что модель плохо описывает изучаемое явление.

Так как F-критерий (фактический) больше F-критерий (критический), можно сделать вывод, что модель объясняет значимую долю вариации зависимой переменной.

Вывод

В рамках данной курсовой работы были рассмотрены понятие и основные аспекты многофакторного прогнозирования, его роль и значимость в прогнозировании экономических и финансовых процессов и явлений.

В ходе выполнения практической части курсовой работы была реализована модель многофакторного прогнозирования ARIMA. Процесс моделирования включал в себя несколько ключевых этапов: сбор и предобработка данных, первичный анализ рядов динамики, тест на стационарность, выбор и оценку модели с учётом экзогенных факторов.

Использование разного рода макро- и микроэкономических показателей, таких как инфляция, курс доллара к рублю, объём выданных кредитов и денежная масса (M2), позволило смоделировать динамику ключевой ставки более точно, что подтвердило значимость использования многофакторного подхода в прогнозировании экономических явлений.

Наличие выбросов в остатках можно обосновать внешними и внутренними экономическими кризисами. Модель неплохо отразила общие тенденции развития ключевой ставки в РФ.

Для улучшения точности модели необходимо более глубоко проанализировать влияние макро- и микроэкономических факторов на динамику ключевой ставки, а также учесть геополитическую обстановку и её влияние на динамику ключевой ставки в РФ.

Список использованной литературы

1. Себастьян Рашка “Python и машинное обучение” / пер. с англ. А.В.Логунова – М.: ДМК Пресс, 2017 – 418 с.: ил.
2. Дворянкин Олег Александрович ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ - БУДУЩАЯ НОВЕЙШАЯ ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ИНТЕРНЕТА // URL: <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-buduschaya-noveyshaya-informatsionnaya-tehnologiya-interneta>
3. Полетаева Наталья Григорьевна Классификация систем машинного обучения // Вестник Балтийского федерального университета им. И. Канта. Серия: Физико-математические и технические науки. 2020. №1. URL: <https://cyberleninka.ru/article/n/klassifikatsiya-sistem-mashinnogo-obucheniya>
4. Морозова, В. И. Прогнозирование методом машинного обучения / В. И. Морозова, Д. И. Логунова. — Текст: непосредственный // Молодой ученый. — 2022. — № 21 (416). — С. 202-204.
5. Чистяков Сергей Павлович Случайные леса: обзор // Труды КарНЦ РАН. 2013. №1. URL: <https://cyberleninka.ru/article/n/sluchaynye-lesa-obzor>
6. Машины опорных векторов, [Электронный ресурс], URL: <https://wiki.loginom.ru/articles/support-vector-machines.html/>
7. Антохонова И.В. “Методы прогнозирования социально-экономических процессов” Учебное пособие. – Улан-Удэ: Изд-во ВСГТУ, 2004. - 212 с.
8. Щелкалин В. Н. Гибридные математические модели и методы прогнозирования временных рядов с учётом внешних факторов // ВЕЖИТ. 2014. №4 (72). URL: <https://cyberleninka.ru/article/n/gibridnye-matematicheskie-modeli-i-metody-prognozirovaniya-vremennyh-ryadov-s-uchyotom-vneshnih-faktorov>
9. Алжеев Андрей Вадимович, Кочкаров Расул Ахматович Сравнительный анализ прогнозных моделей ARIMA и LSTM на примере акций российских компаний // Финансы: теория и практика. 2020. №1. URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-prognoznyh-modeley-arima-i-lstm-na-primere-aktsiy-rossiyskih-kompaniy>

10. Трегуб А. В., Трегуб И. В. Методика построения модели ARIMA для прогнозирования динамики временных рядов // Вестник МГУЛ – Лесной вестник. 2011. №5. URL: <https://cyberleninka.ru/article/n/metodika-postroeniya-modeli-arima-dlya-prognozirovaniya-dinamiki-vremennyh-ryadov>
11. Введение в анализ временных рядов: учебное пособие для вузов /Н. В. Артамонов, Е. А. Ивин, А. Н. Курбацкий, Д. Фантаццини; Московский государственный университет имени М. В. Ломоносова, Московская школа экономики, Кафедра эконометрики и математических методов экономики. – Вологда: ВолНЦ РАН, 2021. – 134 с.: ил., табл.
12. Сизиков Д.О. Метод анализа временных рядов и его математическая модель в программном обеспечении / Д.О. Сизиков // Международный научно-исследовательский журнал. — 2024. — №3 (141). — URL: <https://research-journal.org/archive/3-141-2024-march/10.23670/IRJ.2024.141.5>
13. Фелькер Мария Николаевна, Чеснов Владислав Вадимович ИССЛЕДОВАНИЕ ВЛИЯНИЯ ИЗМЕНЕНИЯ ПАРАМЕТРОВ МОДЕЛИ ARIMA НА КАЧЕСТВО ПРОГНОЗА ДЛЯ КОРОТКИХ НАБОРОВ ДАННЫХ // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. 2021. №3. URL: <https://cyberleninka.ru/article/n/issledovanie-vliyaniya-izmeneniya-parametrov-modeli-arima-na-kachestvo-prognoza-dlya-korotkih-naborov-dannyh>
14. Макаров Дмитрий В., Кантор Евгений А., Красулина Наталья А., Греб Андрей В., Бережнова Зульфия З. Прогнозирование значений цветности питьевых и исходных вод с помощью ARIMA-модели и нейронной сети // Юг России: экология, развитие. 2019. №1. URL: <https://cyberleninka.ru/article/n/prognozirovanie-znacheniy-tsvetnosti-pitievyyh-i-ishodnyh-vod-s-pomoschyu-arima-modeli-i-neyronnoy-seti>
15. Глинский В.В., Ионин В. Г. Статистический анализ. Учебное пособие – 3-е изд., перераб. и доп. – Москва, 2002, 238 с.

16. Эконометрика: учебник для вузов / Санкт-Петербургский государственный экономический университет; редактор И. И. Елисеева. — Москва: Юрайт, 2023. — 449 с. — ISBN 978-5-534-00313-0.
17. Анализ временных рядов и прогнозирование: учебник /В. Н. Афанасьев; Ай Пи Ар Медиа – Саратов, Оренбургский гос. ун-т. –Оренбург: 2020. – 286 с.
18. Попова, И. Н. Анализ временных рядов: учебник для вузов / И. Н. Попова; ответственный редактор В. В. Ковалев. — Москва: Издательство Юрайт, 2024. — 74 с. — (Высшее образование). — ISBN 978-5-534-18394-8. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/534918>
19. Саженкова, Т.В. Методы анализа временных рядов [Текст]: учебно-методическое пособие / Т.В. Саженкова, И.В. Пономарёв, С.П. Пронь. – Барнаул: Издво Алт. ун-та. – 2020. – 60 с.
20. О В. Натроби́на, А Н. Рожкова АНАЛИЗ ВРЕМЕННЫХ РЯДОВ В ЭКОНОМИКЕ: МЕТОДЫ И ПРИЛОЖЕНИЯ // Экономика и бизнес: теория и практика. 2024. №9-1 (115). URL: <https://cyberleninka.ru/article/n/analiz-vremennyh-ryadov-v-ekonomike-metody-i-prilozheniya>

ПРИЛОЖЕНИЕ

Приложение А

Дата	Ключевая ставка, % годовых	Инфляция, % г/г	Денежная масса(M2), млрд руб.	Уровень безработицы по РФ, %	Курс USD/ RUB	Объем выданных кредитов, млн руб.	Торговый баланс, млрд \$
01.12.2022	7,50	11,94	28.352	0,04	69,90	776353	25,1
01.11.2022	7,50	11,98	28.276	0,04	60,99	736822	18,8
01.10.2022	7,50	12,63	28.873	0,04	61,48	898756	18,9
01.09.2022	7,50	13,68	31.155	0,04	58,45	523223	23,1
01.08.2022	8,00	14,30	29.861	0,04	60,23	637611	25,4
01.07.2022	8,00	15,10	30.169	0,04	61,62	725301	24,3
01.06.2022	9,50	15,90	29.519	0,04	51,45	784935	32,7
01.05.2022	11,00	17,10	29.879	0,04	61,50	705541	30,8
01.04.2022	17,00	17,83	29.890	0,04	70,96	706996	31,6
01.03.2022	20,00	16,69	30.073	0,04	83,20	767997	37,5
01.02.2022	20,00	9,15	30.178	0,04	94,60	707854	24,4
01.01.2022	8,50	8,73	30.338	0,04	77,38	712450	22,9
01.12.2021	8,50	8,39	30.297	0,05	74,65	728359	27,3
01.11.2021	7,50	8,40	29.831	0,05	74,08	648899	21,5
01.10.2021	7,50	8,13	30.141	0,05	70,95	807187	20,3
01.09.2021	6,75	7,40	31.615	0,05	72,75	325991	19,9
01.08.2021	6,50	6,68	31.033	0,05	73,23	353548	17,2
01.07.2021	6,50	6,50	31.225	0,05	73,14	407736	17,3
01.06.2021	5,50	6,50	31.028	0,05	73,15	448760	17,9
01.05.2021	5,00	6,00	31.696	0,05	73,43	429966	10,3
01.04.2021	5,00	5,50	31.820	0,05	75,21	508624	11,4
01.03.2021	4,50	5,80	31.936	0,05	75,70	530626	9,68
01.02.2021	4,25	5,70	32.072	0,05	74,62	525404	9,35
01.01.2021	4,25	5,20	32.361	0,05	75,74	518987	10,9
01.12.2020	4,25	4,90	32.052	0,06	74,41	669219	7,38
01.11.2020	4,25	4,40	32.161	0,06	76,40	399139	6,7
01.10.2020	4,25	4,00	32.745	0,06	79,53	528925	10,5
01.09.2020	4,25	3,70	35.179	0,06	77,63	555217	3,31
01.08.2020	4,25	3,60	33.966	0,06	74,07	589176	4,9
01.07.2020	4,25	3,40	34.309	0,06	74,41	557174	5,34
01.06.2020	4,50	3,20	34.689	0,06	71,17	589930	3,76
01.05.2020	5,50	3,00	35.105	0,06	70,14	575684	7,42
01.04.2020	5,50	3,10	35.642	0,06	74,38	653630	9,48
01.03.2020	6,00	2,50	35.856	0,06	78,44	619426	10,2
01.02.2020	6,00	2,30	36.031	0,06	66,88	643242	13,5
01.01.2020	6,25	2,40	36.169	0,06	63,92	627326	14,40
01.12.2019	6,25	3,00	36.148	0,05	61,99	761738	15,3
01.11.2019	6,50	3,50	36.051	0,05	64,32	487764	12,5
01.10.2019	6,50	3,80	36.433	0,05	64,13	553367	13,2

01.09.2019	7,00	4,00	38.418	0,05	64,86	713108	14,6
01.08.2019	7,25	4,30	38.016	0,05	66,77	660244	12,4
01.07.2019	7,25	4,60	38.462	0,05	63,63	745693	11,1
01.06.2019	7,50	4,70	38.555	0,05	63,23	754820	12,4
01.05.2019	7,75	5,10	38.663	0,05	65,43	771758	12,6
01.04.2019	7,75	5,20	39.222	0,05	64,64	809646	14,7
01.03.2019	7,75	5,30	39.623	0,05	65,63	782134	16,1
01.02.2019	7,75	5,20	39.275	0,05	65,92	875194	16,6
01.01.2019	7,75	5,00	39.419	0,05	65,41	913388	14,5
01.12.2018	7,75	4,30	39.571	0,05	69,83	1065393	18,8
01.11.2018	7,50	3,80	39.667	0,05	66,95	731708	19
01.10.2018	7,50	3,50	40.114	0,05	65,88	790444	19,8
01.09.2018	7,50	3,40	42.442	0,05	65,55	951031	18,9
01.08.2018	7,25	3,10	41.597	0,05	67,55	995975	15,9
01.07.2018	7,25	2,50	42.045	0,05	62,51	1012175	13,1
01.06.2018	7,25	2,30	42.377	0,05	62,73	997607	15,3
01.05.2018	7,25	2,40	43.122	0,05	62,40	1056818	15,2
01.04.2018	7,25	2,40	43.257	0,05	62,94	1109692	15
01.03.2018	7,25	2,40	44.126	0,05	57,14	1056196	15
01.02.2018	7,50	2,20	43.910	0,05	56,33	1177042	12,2
01.01.2018	7,75	2,20	44.369	0,05	56,21	1177836	16,9
01.12.2017	7,75	2,50	44.254	0,05	57,61	1310135	13
01.11.2017	8,25	2,50	44.218	0,05	58,44	897835	11,4
01.10.2017	8,25	2,70	44.891	0,05	58,33	979799	10
01.09.2017	8,50	3,00	47.109	0,05	57,49	1113757	10,2
01.08.2017	9,00	3,30	45.721	0,05	58,01	1203017	6,61
01.07.2017	9,00	3,90	46.212	0,05	59,77	1075957	3,79
01.06.2017	9,00	4,40	46.141	0,05	58,94	1115059	8,7
01.05.2017	9,25	4,10	46.435	0,05	56,56	1191238	8,59
01.04.2017	9,75	4,10	46.735	0,05	56,93	1225916	7,79
01.03.2017	9,75	4,30	47.349	0,05	56,27	1253724	12,3
01.02.2017	10,00	4,60	47.351	0,05	58,31	1267992	10,3
01.01.2017	10,00	5,00	47.584	0,05	60,16	1235804	11,8
01.12.2016	10,00	5,40	48.266	0,06	61,27	1446046	11,7
01.11.2016	10,00	5,80	48.082	0,06	64,14	1052210	8,92
01.10.2016	10,00	6,10	49.195	0,06	63,38	1169929	6,57
01.09.2016	10,00	6,40	51.660	0,06	62,83	1394068	7,41
01.08.2016	10,50	6,90	50.622	0,06	65,33	744822	4,66
01.07.2016	10,50	7,20	51.314	0,06	66,04	847285	6,3
01.06.2016	10,50	7,50	52.327	0,06	63,97	1150350	8,06
01.05.2016	11,00	7,30	52.951	0,06	66,70	1347441	7,61
01.04.2016	11,00	7,30	53.068	0,06	64,65	1433661	6,64
01.03.2016	11,00	7,30	54.392	0,06	67,03	1525016	7,83
01.02.2016	11,00	8,10	54.687	0,06	75,17	1565832	7,27
01.01.2016	11,00	9,80	55.294	0,06	75,46	1554991	7,3
01.12.2015	11,00	12,90	56.023	0,06	73,60	1700440	11,2
01.11.2015	11,00	15,00	55.871	0,06	66,49	1162340	8,95
01.10.2015	11,00	15,60	56.122	0,06	63,95	1393258	10

01.09.2015	11,00	15,70	58.652	0,06	65,45	1741578	9,51
01.08.2015	11,00	15,80	57.598	0,06	64,28	1827917	8,8
01.07.2015	11,50	15,60	58.178	0,06	61,74	1666159	10,6
01.06.2015	11,50	15,30	58.261	0,06	55,28	1846309	13,9
01.05.2015	12,50	15,80	59.206	0,06	52,32	1715811	15,4
01.04.2015	14,00	16,40	59.194	0,06	51,65	1806081	14,5
01.03.2015	14,00	16,90	59.583	0,06	58,20	1809028	15,8
01.02.2015	15,00	16,70	59.379	0,06	61,62	1795547	14
01.01.2015	17,00	14,96	59.816	0,06	68,86	1845852	15,7
01.12.2014	17,00	11,35	60.606	0,05	55,91	2060165	13,6
01.11.2014	9,50	9,06	60.813	0,05	50,41	1455223	13,6
01.10.2014	8,00	8,29	62.312	0,05	43,03	1692068	14,7
01.09.2014	8,00	8,03	66.252	0,05	39,57	1298097	12
01.08.2014	8,00	7,55	65.310	0,05	37,07	859296	16,1
01.07.2014	8,00	7,45	66.659	0,05	35,68	985994	17
01.06.2014	7,50	7,81	68.203	0,05	33,99	1276829	13,9
01.05.2014	7,50	7,59	68.475	0,05	34,92	1472199	17,8
01.04.2014	7,50	7,33	68.993	0,05	35,63	1704845	19,8
01.03.2014	7,00	6,92	69.623	0,05	35,05	1734817	19,6
01.02.2014	5,50	6,21	70.825	0,05	36,05	1623712	12,3
01.01.2014	5,50	6,07	73.333	0,05	35,18	1758385	18,6
01.12.2013	5,50	6,47	75.096	0,06	32,90	2095882	16,9
01.11.2013	5,50	6,50	75.636	0,06	33,15	1400793	16,9
01.10.2013	5,50	6,27	76.873	0,06	32,08	1606116	12,7
01.09.2013	5,50	6,14	82.388	0,06	32,39	2060513	16