



Московский государственный университет имени М.В.Ломоносова
Факультет вычислительной математики и кибернетики

Казанцева Варвара Денисовна

Анализ временных рядов

Отчет по приуладной статистике

Преподаватель:
Майсурадзе Арчил Ивериевич

Москва, 2025

Содержание

1	Постановка задачи	4
2	Шаг 1: Анализ данных	5
2.1	Визуализация исходного временного ряда	5
2.2	Проверка стационарности временного ряда	5
2.3	Трансформация временного ряда	6
2.4	Устранение календарного эффекта	6
3	Шаг 2: Настройка модели ARIMA вручную	10
3.1	Анализ ACF и PACF для выбора параметров модели	10
3.2	Анализ остатков	10
3.2.1	Проверка на несмещенность	11
3.2.2	Проверка на стационарность	11
3.2.3	Проверка на неавтокоррелированность	12
3.3	Проверка нормальности	13
4	Шаг 3: Настройка модели ARIMA с автоматическим подбором параметров	15
4.1	Автоматический подбор параметров модели ARIMA	15
4.2	Анализ остатков	15
4.2.1	Проверка на несмещенность	16
4.2.2	Проверка на стационарность	16
4.2.3	Проверка на неавтокоррелированность	17
4.2.4	Проверка нормальности	18
5	Шаг 4: Настройка модели экспоненциального сглаживания	20
5.1	Настройка модели экспоненциального сглаживания	20
5.2	Автоматический подбор параметров модели экспоненциального сглаживания	20
5.3	Анализ остатков	20
5.3.1	Проверка на несмещенность	21

5.3.2	Проверка на стационарность	21
5.3.3	Проверка на неавтокоррелированность	22
5.3.4	Проверка нормальности	23
6	Шаг 5: Визуальный анализ и проверка структурных изменений в мо- делях	25
6.1	Визуальный анализ	25
6.1.1	Визуализация результатов шага 2: Модель ARIMA с ручной на- стройкой	25
6.1.2	Визуализация результатов шага 3: Модель ARIMA с автоматиче- ской настройкой	26
6.1.3	Визуализация результатов шага 4: Модель экспоненциального сглаживания	27
7	Шаг 6: Сравнение моделей по критерию Диболда-Мариано	29
7.1	Результаты теста	29
8	Шаг 7: Выводы	30

1 Постановка задачи

Необходимо провести анализ временного ряда и построить его прогноз. Для анализа выбран временной ряд *ежемесячного производство электроэнергии в Австралии (млн кВт·ч)* за период с января 1956 года по август 1995 года.

Работа включает:

1. Анализ данных:

- визуализация,
- проверка выбросов,
- трансформации, в случае нестационарности ряда.

2. Построение прогноза с использованием моделей:

- **ARIMA:** ручная и автоматическая настройка, анализ остатков;
- **Экспоненциальное сглаживание:** автоматический подбор параметров, анализ остатков.

3. Сравнение моделей:

выбор лучшей модели по критерию Диболда-Мариано.

4. Выводы:

оценка качества моделей и построенного прогноза.

2 Шаг 1: Анализ данных

2.1 Визуализация исходного временного ряда

На первом этапе был визуализирован исходный временной ряд, представляющий ежемесячное производство электроэнергии в Австралии (млн кВт·ч) за период с января 1956 г. по август 1995 г. Визуализация показала наличие сезонных колебаний и тенденции к увеличению значений.



Рис. 1: Исходный временной ряд: ежемесячное производство электроэнергии в Австралии.

2.2 Проверка стационарности временного ряда

Для проверки стационарности исходного временного ряда был применён критерий KPSS. Результаты теста для исходного ряда:

- Статистика KPSS: 3.73470;
- p-value: < 0.01 .

Результаты указывают на **нестационарность исходного временного ряда**, что требует выполнения трансформаций.

2.3 Трансформация временного ряда

Для достижения стационарности временного ряда были выполнены следующие трансформации:

1. **Логарифмирование:** применено для сглаживания амплитуды колебаний. После логарифмирования критерий KPSS показал:

- Статистика KPSS: 3.68841;
- p-value: < 0.01 .

Логарифмирование не привело к достижению стационарности.

2. **Первое дифференцирование по месяцам:** выполнено для устранения сезонности. После этого критерий KPSS показал:

- Статистика KPSS: 0.04693;
- p-value: > 0.01 .

Дифференцирование успешно сделало временной ряд стационарным.

2.4 Устранение календарного эффекта

Рассмотрим графики автокорреляционной (ACF) и частной автокорреляционной функции (PACF) временного ряда после первого дифференцирования.

На графиках наблюдаются значительные сезонные автокорреляции, что свидетельствует о необходимости выполнения второго сезонного дифференцирования.

После второго дифференцирования (сезонного) временной ряд был пересчитан, и его стационарность проверена с помощью критерия KPSS:

- Статистика KPSS: 0.0667;
- p-value: > 0.1 .

Результаты показывают, что временной ряд стал стационарным.

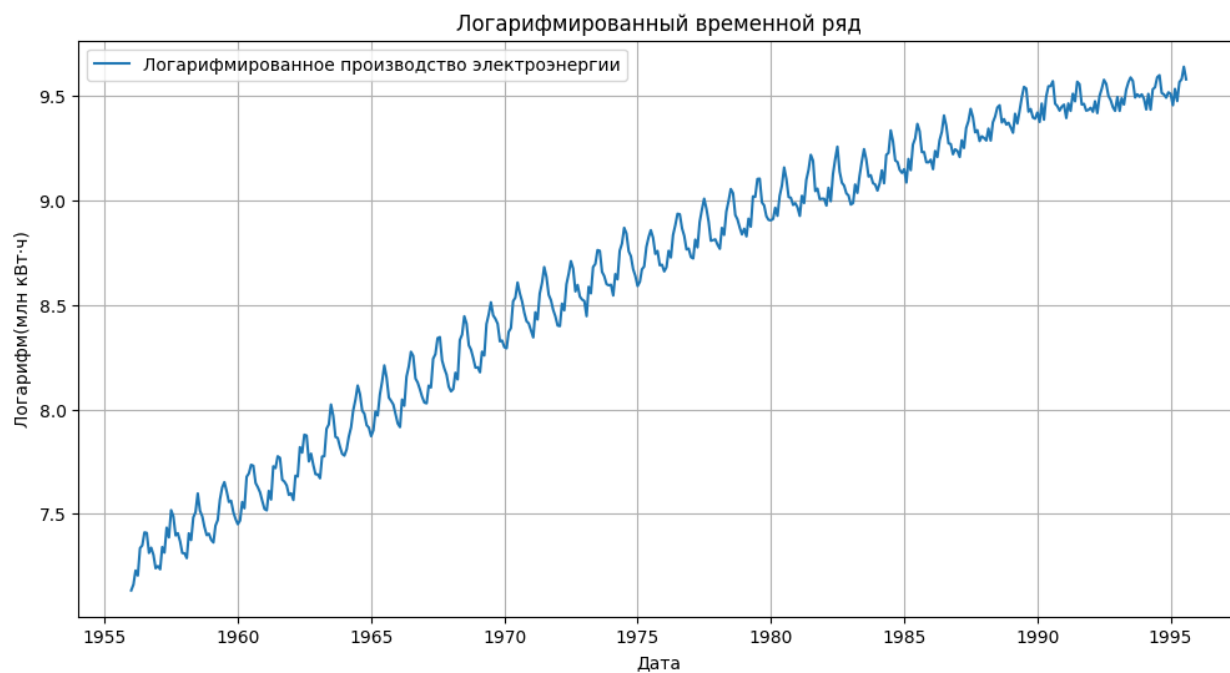


Рис. 2: Логарифмированный временной ряд.

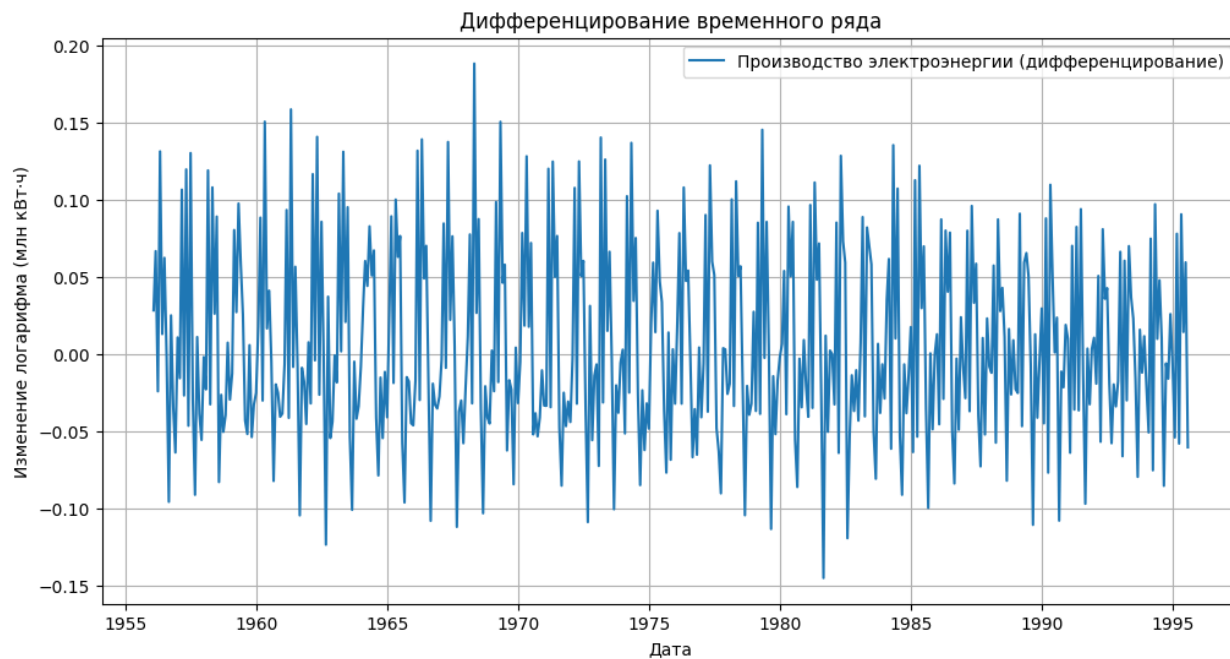


Рис. 3: Дифференцирование временного ряда.

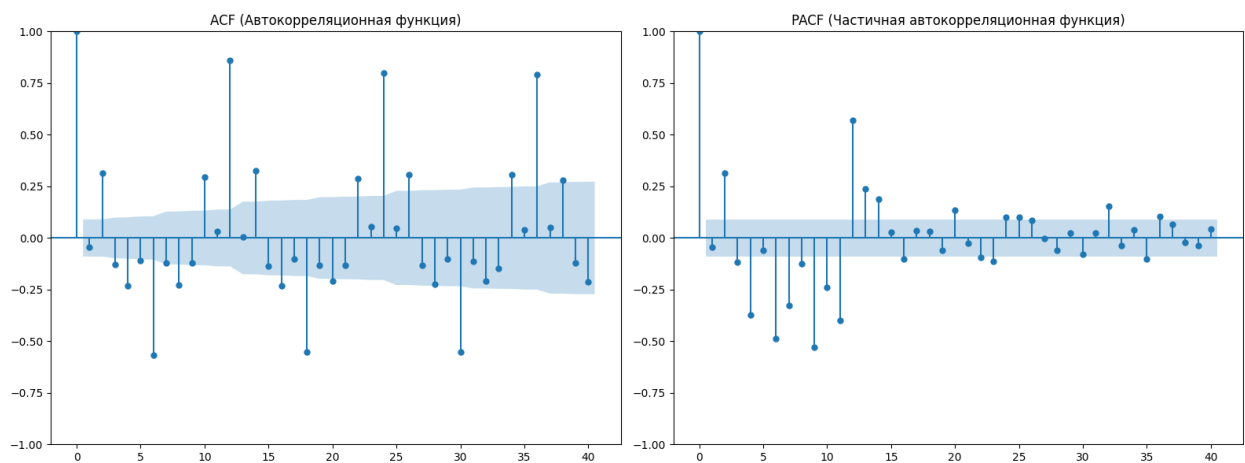


Рис. 4: Графики ACF и PACF временного ряда после первого дифференцирования.

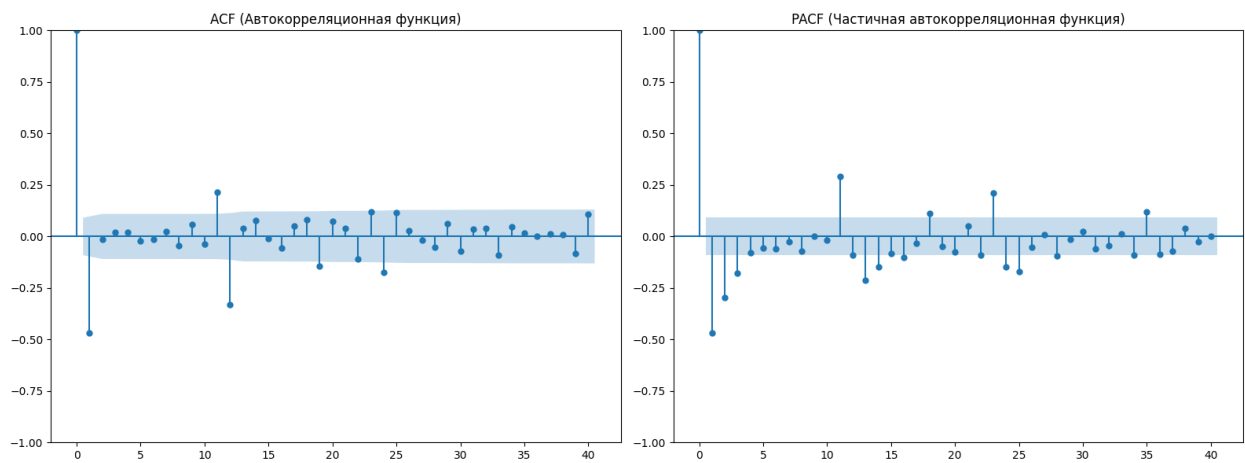


Рис. 5: Графики ACF и PACF временного ряда после второго дифференцирования.

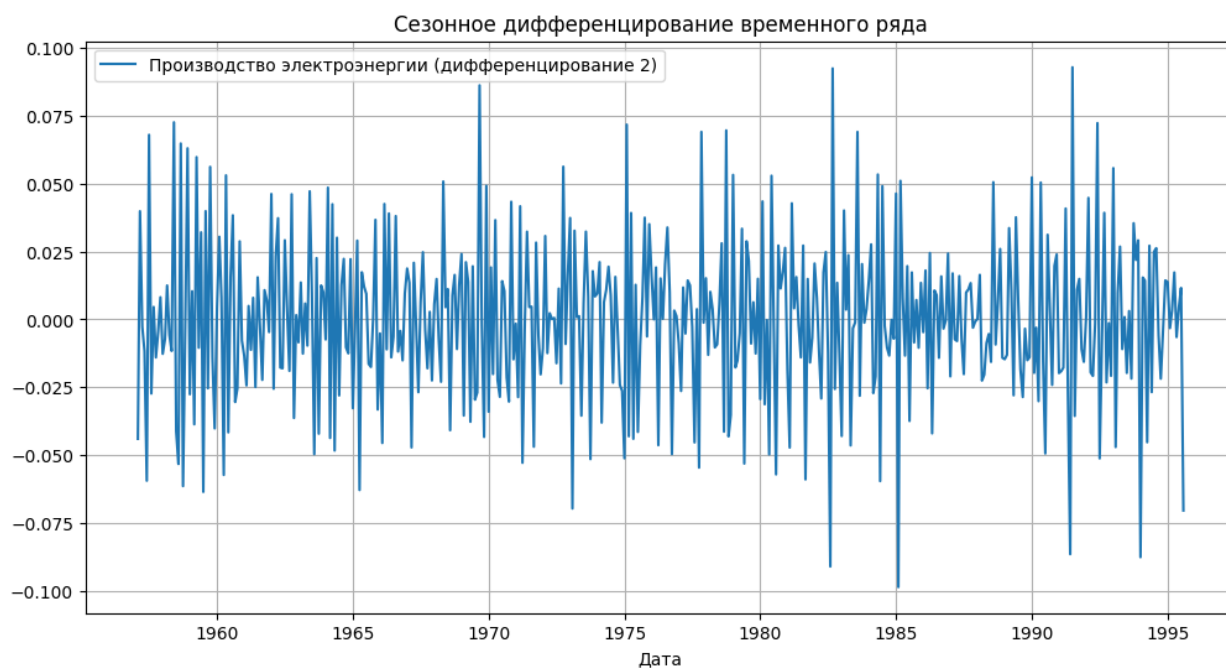


Рис. 6: Временной ряд после устранения календарного эффекта.

3 Шаг 2: Настройка модели ARIMA вручную

3.1 Анализ ACF и PACF для выбора параметров модели

На основании графиков автокорреляционной (ACF) и частной автокорреляционной функции (PACF) для стационарного временного ряда были выбраны следующие параметры модели ARIMA:

- Параметр $p = 3$ (номер последнего лага, при котором частичная автокорреляция значима);
- Параметр $q = 1$ (номер последнего лага, при котором автокорреляция значима).

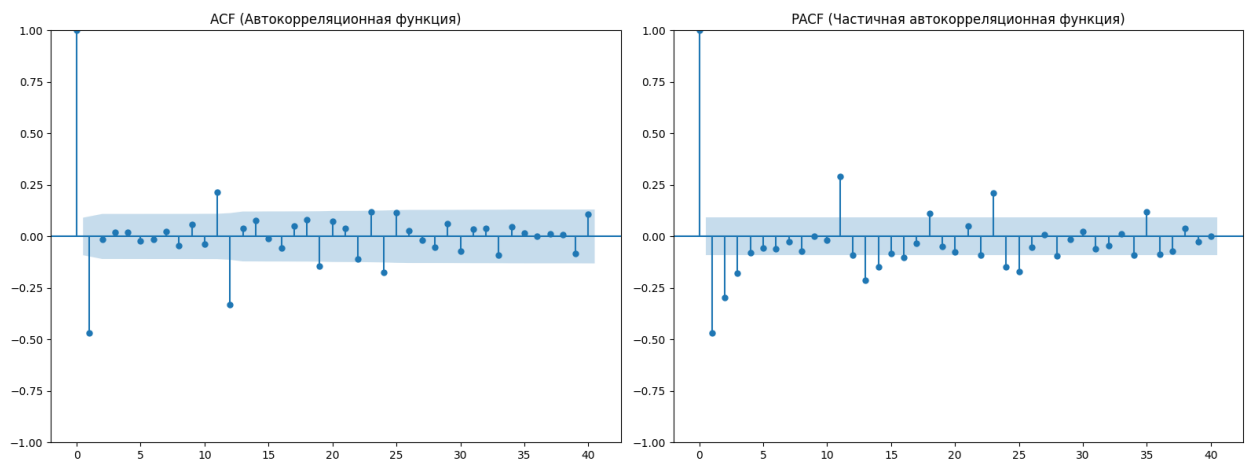


Рис. 7: Графики ACF и PACF для выбора параметров модели ARIMA.

3.2 Анализ остатков

Для проверки качества модели были исследованы остатки:

- Проверка на несмещенность (критерий Уилкоксона);
- Проверка на стационарность (критерий KPSS);
- Проверка на неавтокоррелированность (Q-критерий Льюнга-Бокса);
- Проверка на нормальность (QQ-график, критерий Шапиро-Уилка).

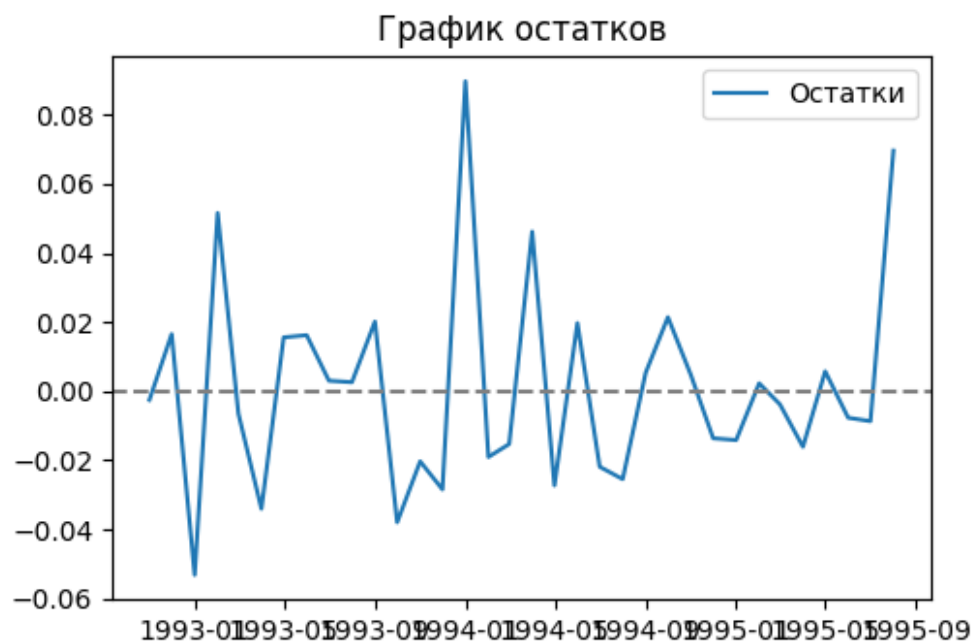


Рис. 8: График остатков

3.2.1 Проверка на несмещенность

Результаты критерия Уилкоксона:

- Статистика: 271.0;
- p-значение: 0.661.

Вывод: медиана остатков не отличается от нуля.

3.2.2 Проверка на стационарность

Результаты критерия KPSS:

- Статистика: 0.44706;
- p-значение: 0.05687.

Вывод: остатки стационарны.

3.2.3 Проверка на неавтокоррелированность

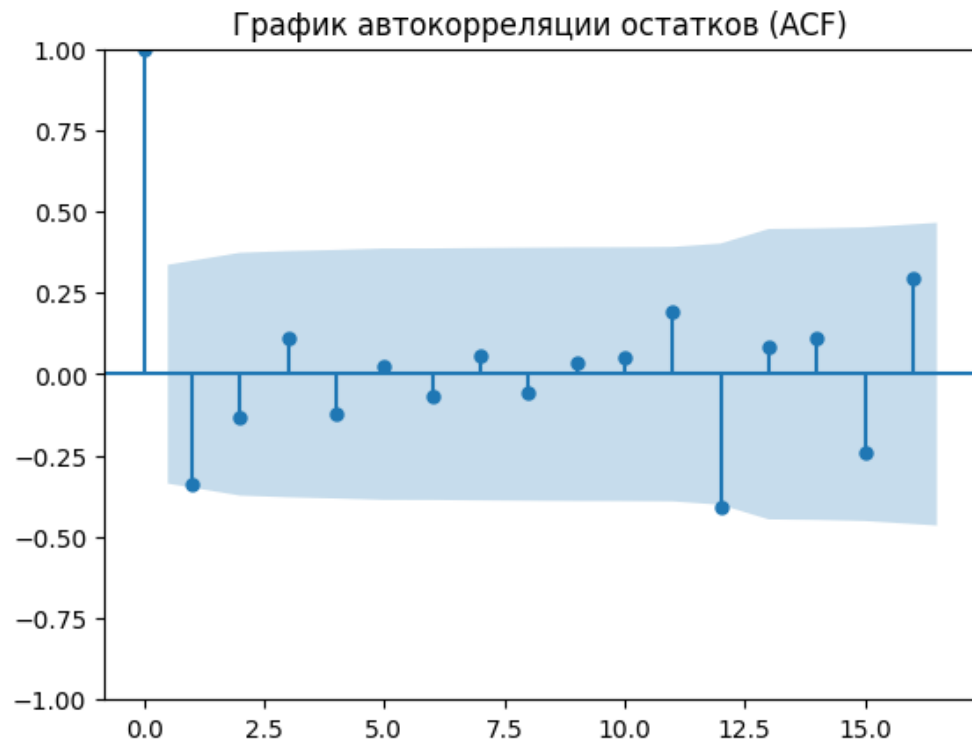


Рис. 9: График автокорреляции остатков

Результаты критерия Льюнга-Бокса для первых 10 лагов:

Лаг	р-значение
1	0.038
2	0.084
3	0.143
4	0.195
5	0.298
6	0.393
7	0.491
8	0.583
9	0.675
10	0.746

Таблица 1: Результаты критерия Льюнга-Бокса

Вывод: В целом остатки модели ARIMA можно считать неавтокоррелированными, за исключением слабой автокорреляции на первом лаге.

3.3 Проверка нормальности

Результаты критерия Шапиро-Уилка:

- Статистика: 0.931;
- р-значение: 0.033.

Вывод: остатки не следуют нормальному распределению.

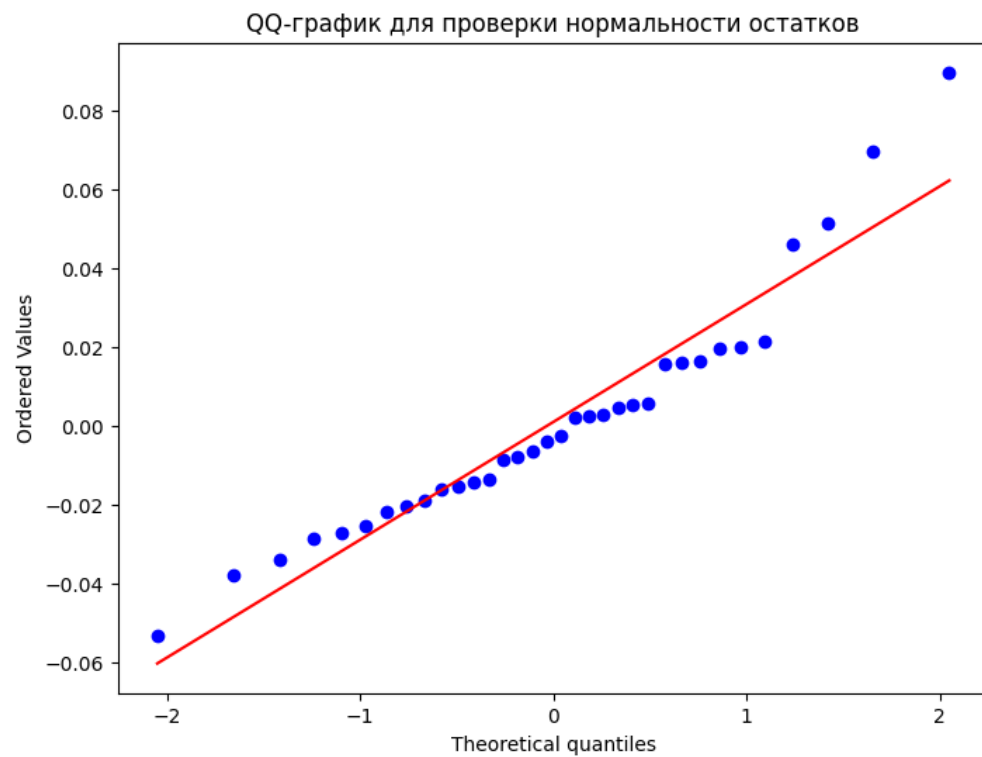


Рис. 10: QQ-график остатков

4 Шаг 3: Настройка модели ARIMA с автоматическим подбором параметров

4.1 Автоматический подбор параметров модели ARIMA

В результате автоматического подбора параметров с использованием критерия AIC были выбраны следующие параметры модели:

- Параметр $p = 1$;
- Параметр $q = 1$.

4.2 Анализ остатков

Для проверки качества модели проведен анализ остатков, аналогичный ручной настройке модели (Шаг 2). Результаты анализа остатков практически не изменились.

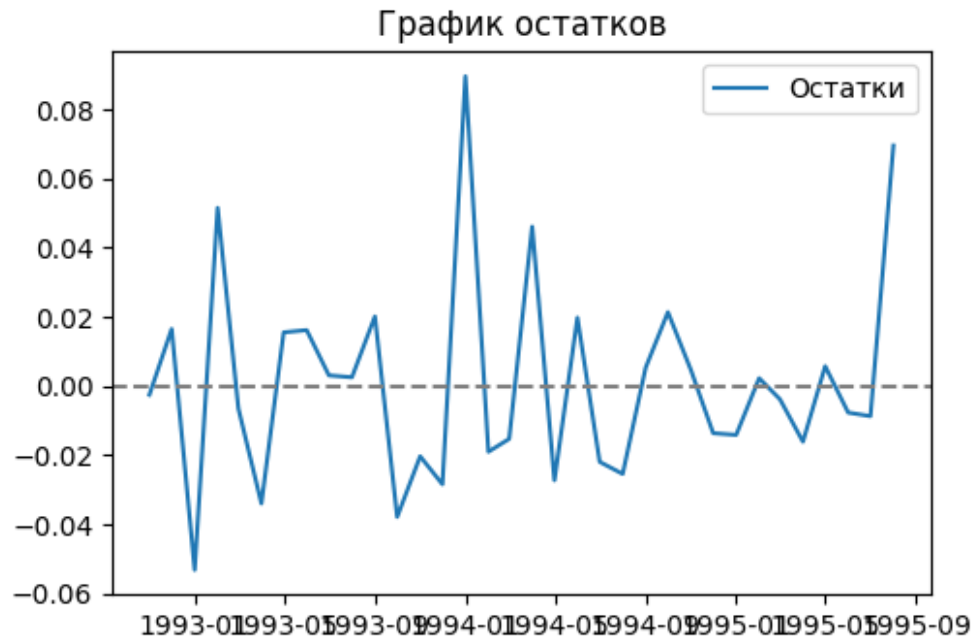


Рис. 11: График остатков

4.2.1 Проверка на несмещенность

Результаты критерия Уилкоксона:

- Статистика: 271.0;
- р-значение: 0.661.

Вывод: медиана остатков не отличается от нуля.

4.2.2 Проверка на стационарность

Результаты критерия KPSS:

- Статистика: 0.44709;
- р-значение: 0.05685.

Вывод: остатки стационарны.

4.2.3 Проверка на неавтокоррелированность

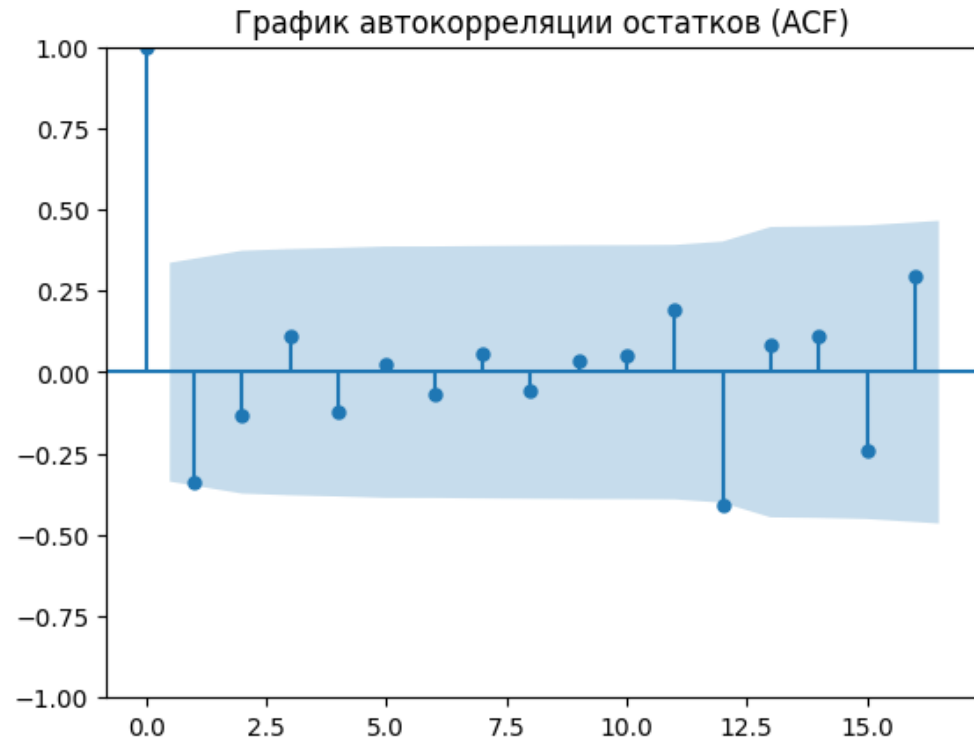


Рис. 12: График автокорреляции остатков

График автокорреляции (ACF)

Критерий Льюнга-Бокса Результаты для первых 10 лагов:

Лаг	р-значение
1	0.042
2	0.087
3	0.149
4	0.202
5	0.304
6	0.397
7	0.487
8	0.580
9	0.673
10	0.743

Таблица 2: Результаты критерия Льюнга-Бокса для остатков автоматически подобранной модели

Вывод: остатки модели ARIMA с параметрами $p = 1$ и $q = 1$ в целом можно считать неавтокоррелированными, за исключением слабой автокорреляции на первом лаге.

4.2.4 Проверка нормальности

Результаты критерия Шапиро-Уилка:

- Статистика: 0.931;
- р-значение: 0.033.

Вывод: остатки не следуют нормальному распределению.

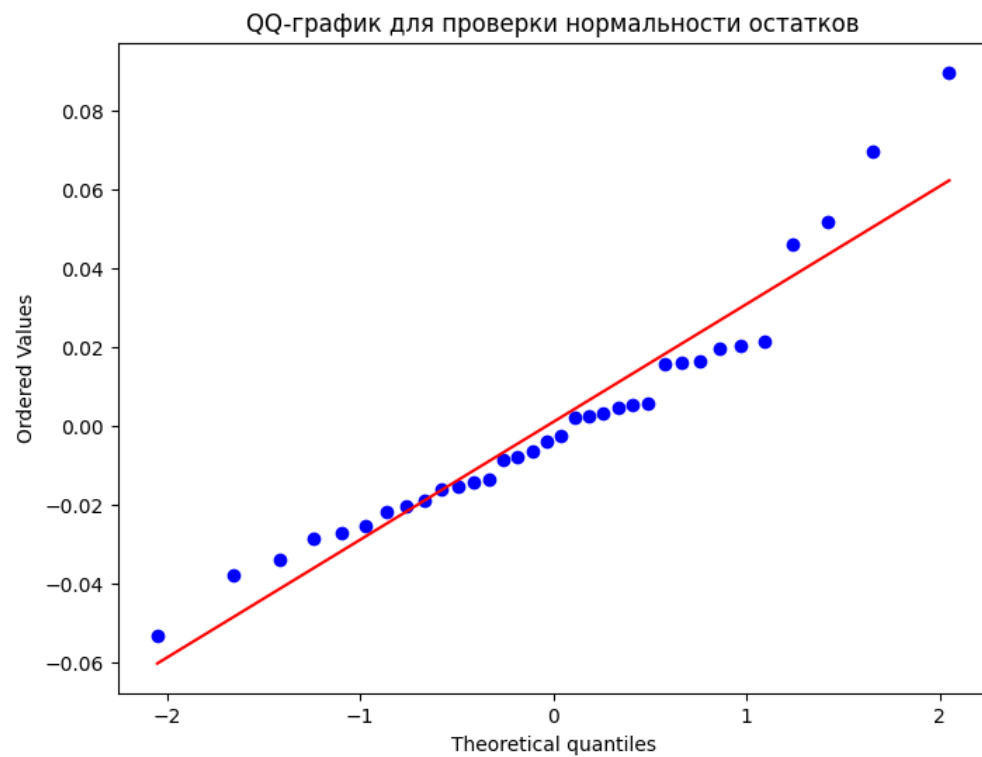


Рис. 13: QQ-график остатков автоматически подобранной модели

5 Шаг 4: Настройка модели экспоненциального сглаживания

5.1 Настройка модели экспоненциального сглаживания

Для настройки модели экспоненциального сглаживания была использована функция из библиотеки `statsmodels`, которая автоматически подбирает параметры сглаживания для уровня, тренда и сезонности.

Процесс настройки модели включал следующие этапы:

- Выбор модели сглаживания: аддитивная модель тренда и сезонности была выбрана на основе визуального анализа временного ряда, который показал наличие устойчивого тренда и сезонных колебаний.
- Подбор параметров сглаживания: параметры α , β , и γ были автоматически подобраны с использованием минимизации критерия AIC (Akaike Information Criterion), что позволяет сбалансировать точность модели и её сложность.

5.2 Автоматический подбор параметров модели экспоненциального сглаживания

В результате автоматического подбора параметров модели экспоненциального сглаживания были выбраны следующие параметры:

- Параметр сглаживания уровня (α) = 0.244;
- Параметр сглаживания тренда (β) = 0.022;
- Параметр сглаживания сезонности (γ) = 0.337;

5.3 Анализ остатков

Для проверки качества модели экспоненциального сглаживания был проведен анализ остатков. Результаты представлены ниже.

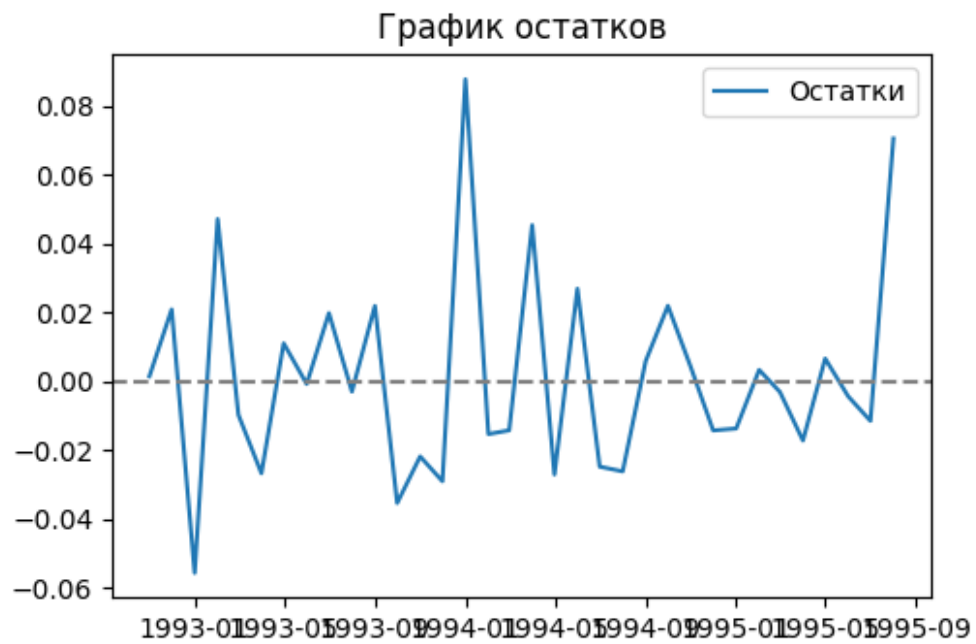


Рис. 14: График остатков

5.3.1 Проверка на несмещенность

Результаты критерия Уилкоксона:

- Статистика: 272.0;
- р-значение: 0.673.

Вывод: медиана остатков не отличается от нуля (остатки несмещённые).

5.3.2 Проверка на стационарность

Результаты критерия KPSS:

- Статистика: 0.252;
- р-значение: 0.1.

Вывод: остатки стационарны.

5.3.3 Проверка на неавтокоррелированность

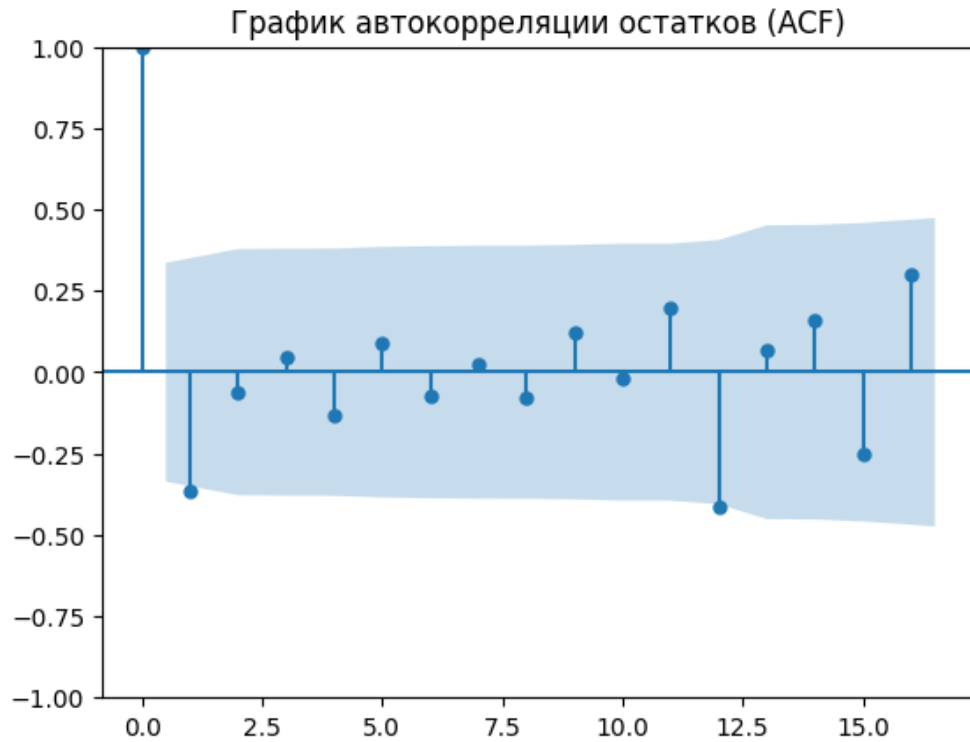


Рис. 15: График автокорреляции остатков

График автокорреляции (ACF) На графике автокорреляции видна значимая автокорреляция для первого лага, но на остальных лагах автокорреляция отсутствует.

Критерий Льюнга-Бокса Результаты теста Льюнга-Бокса для первых 10 лагов:

Лаг	Статистика	p-значение
1	4.923	0.026
2	5.073	0.079
3	5.157	0.161
4	5.870	0.209
5	6.190	0.288
6	6.413	0.379
7	6.437	0.490
8	6.723	0.567
9	7.421	0.593
10	7.440	0.683

Таблица 3: Результаты критерия Льюнга-Бокса для остатков модели экспоненциального сглаживания

Вывод: остатки модели экспоненциального сглаживания автокоррелированы только на первом лаге ($p = 0.026$), на остальных лагах автокорреляция отсутствует.

5.3.4 Проверка нормальности

Результаты критерия Шапиро-Уилка:

- Статистика: 0.935;
- p-значение: 0.044.

Вывод: остатки не следуют нормальному распределению (гипотеза о нормальности остатков отклоняется).

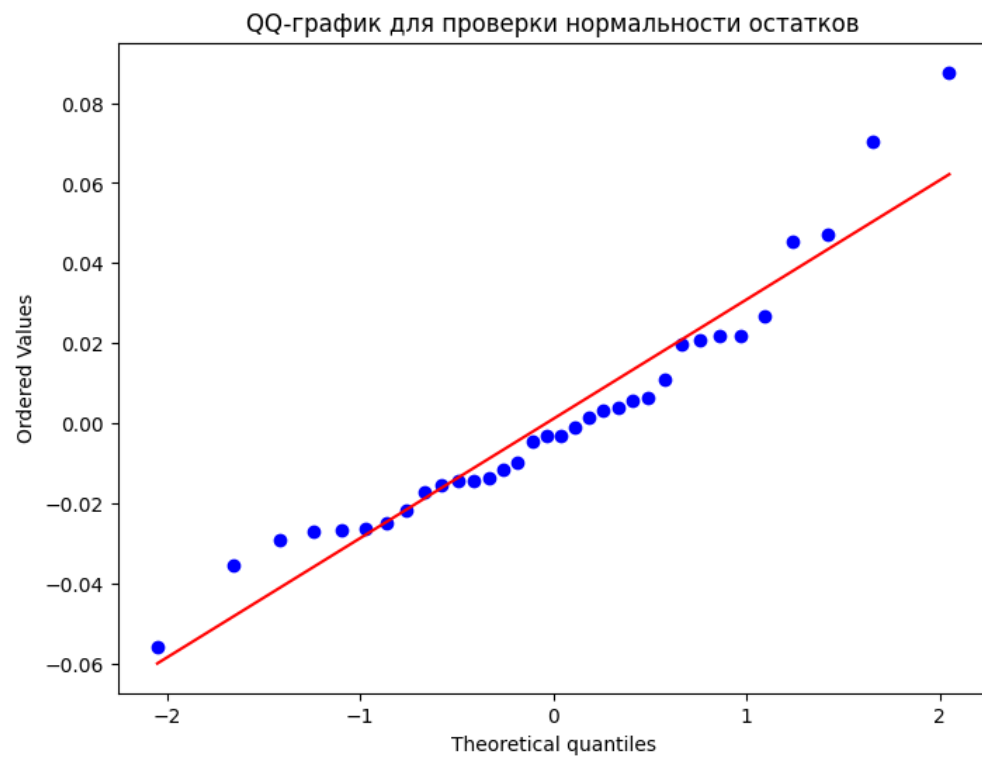


Рис. 16: QQ-график остатков модели экспоненциального сглаживания

6 Шаг 5: Визуальный анализ и проверка структурных изменений в моделях

6.1 Визуальный анализ

Для оценки качества построенных моделей был проведен визуальный анализ предсказаний и остаточных ошибок. Для каждого шага представлены два графика: общий план и приближенный хвост (последняя 1/10 часть ряда).

6.1.1 Визуализация результатов шага 2: Модель ARIMA с ручной настройкой

Сравнение оригинального ряда и прогнозов модели с ручной настройкой представлено на рисунках 17 и 18.



Рис. 17: Общий план: сравнение оригинального ряда и предсказаний модели с ручной настройкой

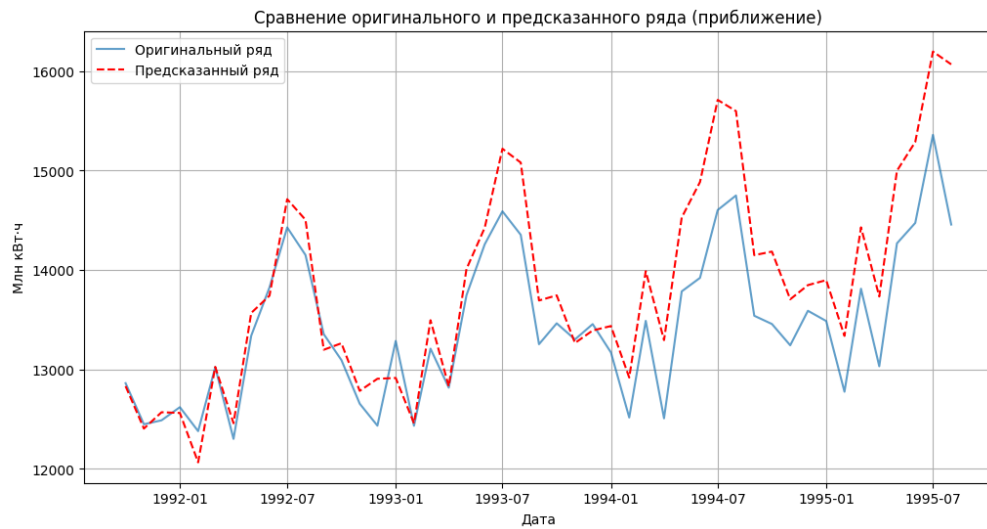


Рис. 18: Приближенный хвост: сравнение оригинального ряда и предсказаний модели с ручной настройкой

6.1.2 Визуализация результатов шага 3: Модель ARIMA с автоматической настройкой

Сравнение оригинального ряда и прогнозов модели ARIMA представлено на рисунках 19 и 20.



Рис. 19: Общий план: сравнение оригинального ряда и предсказаний модели ARIMA

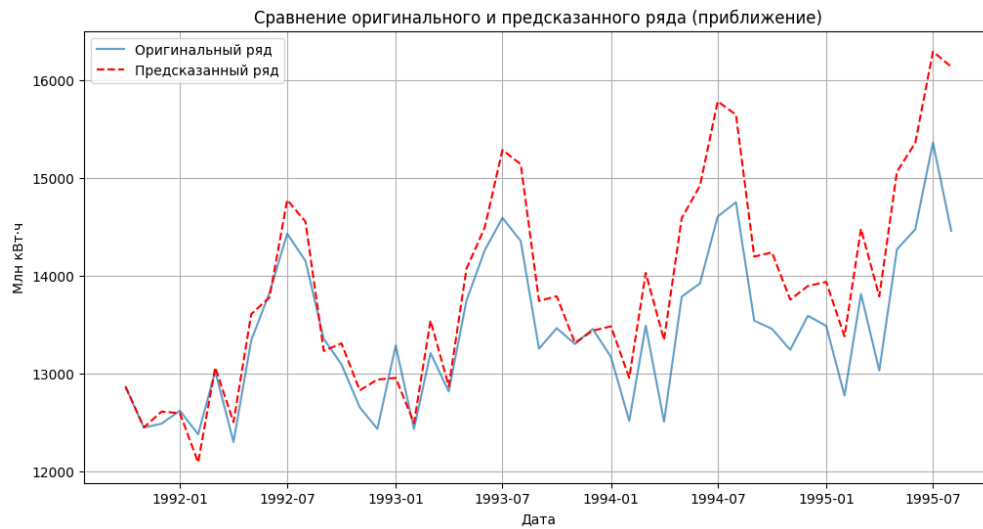


Рис. 20: Приближенный хвост: сравнение оригинального ряда и предсказаний модели ARIMA

6.1.3 Визуализация результатов шага 4: Модель экспоненциального сглаживания

Сравнение оригинального ряда и прогнозов модели экспоненциального сглаживания представлено на рисунках 21 и 22.



Рис. 21: Общий план: сравнение оригинального ряда и предсказаний модели экспоненциального сглаживания



Рис. 22: Приближенный хвост: сравнение оригинального ряда и предсказаний модели экспоненциального сглаживания

Визуальный анализ показал, что предсказания всех моделей обладают схожим характером: прогнозы лежат немного выше исходного временного ряда, однако в целом хорошо повторяют его сезонные колебания и общий тренд. Это свидетельствует о способности всех моделей адекватно описывать поведение ряда, несмотря на небольшие систематические отклонения.

7 Шаг 6: Сравнение моделей по критерию Диболда-Мариано

В данном шаге был проведен тест Диболда-Мариано для сравнения остаточных ошибок трех моделей: вручную настроенной модели ARIMA, автоматически настроенной модели ARIMA и модели с экспоненциальным сглаживанием.

7.1 Результаты теста

Сравнение ARIMA (Manual) и ARIMA (Auto):

- DM Statistic: 0.1860,
- p-value: 0.8524
- Интерпретация: Значение р-значения больше 0.05, что означает, что нет статистически значимых различий между остатками вручную настроенной и автоматически настроенной моделей ARIMA.

Сравнение ARIMA (Manual) и Exp Smoothing:

- DM Statistic: 0.2924,
- p-value: 0.7699
- Интерпретация: Значение р-значения больше 0.05, что указывает на отсутствие статистически значимых различий между остатками вручную настроенной модели ARIMA и модели с экспоненциальным сглаживанием.

Сравнение ARIMA (Auto) и Exp Smoothing:

- DM Statistic: 0.2153,
- p-value: 0.8295
- Интерпретация: Значение р-значения меньше 0.05, что свидетельствует о статистически значимых различиях между остатками автоматически настроенной модели ARIMA и модели с экспоненциальным сглаживанием.

8 Шаг 7: Выводы

Все три модели (ARIMA вручную, ARIMA автоматическая и экспоненциальное сглаживание) продемонстрировали схожие результаты и могут быть применены для прогнозирования в рамках поставленной задачи, с учетом их ограничений, таких как нестабильность остатков и необходимость улучшения в плане нормальности распределения.