Национальный исследовательский университет ИТМО
Факультет информационных технологий и программирования
Информационные системы и технологии

Прикладная математика
Отчёт по лабораторной работе №4

Работу выполнили:
Фирсова Д. А., студентка M33081
Деревицкая П. К., студентка M33081
Борздун А. В., студентка M33051

Преподаватель:
Свинцов М. В.

Санкт-Петербург,
2023

Оглавление

Цель работы: исследование регрессии временных рядов

Задачи, решаемые во время выполнения работы:

1. Построение авторегрессионной модели временного ряда AR(3)
2. Построение графиков
3. Определение авторегрессионных параметров
4. Анализ результатов при различных ядрах и других гиперпараметрах модели

Регрессия временных рядов
Краткая теоретическая справка

Временные ряды.

Под **временным рядом** (динамическим рядом, или рядом динамики) подразумевается последовательность наблюдений некоторого признака (случайной величины) x в последовательные моменты времени. Отдельные наблюдения называются уровнями ряда.

Цель анализа временных рядов - определить закономерности в изменениях параметра во времени и сделать прогноз на будущее.

Временной ряд отличается от обычной последовательности тем, что каждое его значение взято на одинаково расположенных точках времени.

Виды временных рядов

Помимо регулярности временные ряды делятся на детерминированные и недетерминированные.

Детерминированный временной ряд – ряд, в котором нет случайных аспектов или показателей: он может быть выражен формулой. Это значит, что мы можем проанализировать, как показатели вели себя в прошлом и прогнозировать их поведение в будущем.

Недетерминированный временной ряд имеет случайный аспект и прогнозирование будущих действий становится сложнее. Природа таких показателей случайна и анализ происходит благодаря средним значениям и дисперсии.

Также ряды делятся по наличию или отсутствию тенденций и сезонных эффектов на стационарные и нестационарные.

В стационарных временных рядах статистические свойства не зависят от времени, поэтому результат легко предсказать. Большинство статистических методов предполагают, что все временные ряды должны быть стационарными. Пример стационарных временных рядов – рождаемость в России. Конечно, она зависит от множества факторов, но ее спад или рост возможно предсказать: у рождаемости нет ярко выраженной сезонности.

В **нестационарных временных рядах** статистические свойства меняются со временем. Они показывают сезонные эффекты, тренды и другие структуры, которые зависят от временного показателя. Пример – международные перелеты авиакомпаний. Количество пассажиров в те или иные направления меняются в зависимости от сезонности.

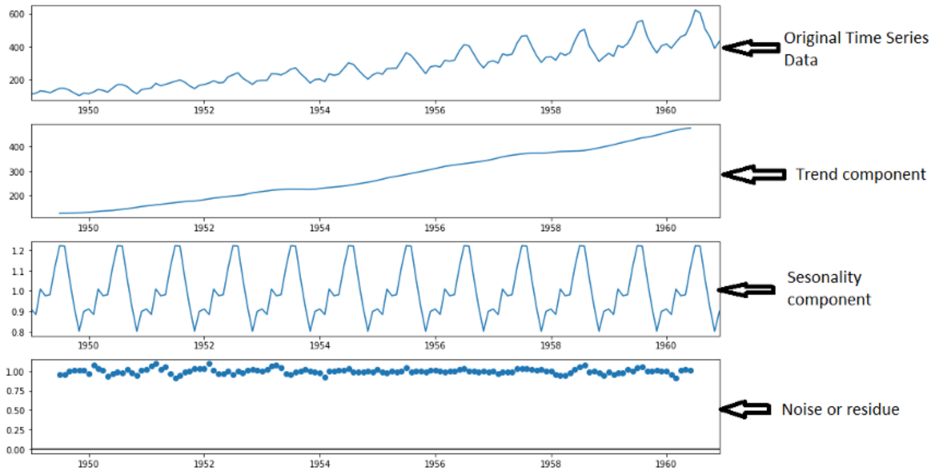
Предполагается, что временные ряды генерируются регулярно, но на практике это не всегда так. Регулярный компонент временного ряда – систематическая составляющая, которая имеет определенную прогнозируемую характеристику. В нерегулярных рядах измерения проходят не через регулярный интервал времени.

Основные понятия в статистическом анализе временных рядов:

- *Тренд* - компонента, описывающая долгосрочное изменение уровня ряда.
- *Сезонность* - компонента, обозначаемая как Q, описывает циклические изменения уровня ряда.
- *Ошибка (random noise)* - непрогнозируемая случайная компонента, описывает нерегулярные изменения в данных, необъяснимые другими компонентами.
- *Автокорреляция* — статистическая взаимосвязь между последовательностями величин одного ряда. Это один из самых важных коэффициентов в анализе временного ряда. Чтобы посчитать автокорреляцию, используется корреляция между временным рядом и её

сдвинутой копией от величины временного сдвига. Сдвиг ряда называется лагом.

- *Автокорреляционная функция*- график автокорреляции при разных лагах



- *Стационарный* ряд - ряд, в котором свойства не зависят от времени. При таких компонентах ряда, как тренд или сезонность, ряд не является стационарным. Это интуитивное понятие. На практике есть два вида стационарности - строгая и слабая. В следующих алгоритмах нам достаточно слабой. Она заключается в постоянной дисперсии (без гетероскедастичности) и в постоянности среднего значения ряда.
- *Гетероскедастичность* - не равномерная дисперсия. То есть, не однородность наблюдений.

Анализ временных рядов.

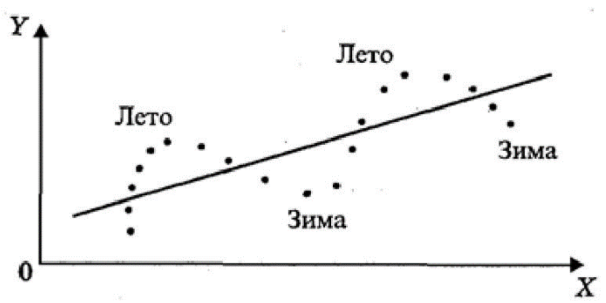
Существует три основных метода анализа временных рядов: стационарность, автокорреляцию и спектральный анализ.

Стационарность - это свойство временного ряда, которое означает, что его средние и стандартные отклонения не меняются со временем. Если временной ряд является стационарным, то его можно легко анализировать и прогнозировать. Нестационарный временной ряд может иметь тренд (постоянный рост или падение), цикличность (повторение циклов) или сезонность (повторение определенных событий в разное время года).

Один из способов проверки стационарности временных рядов - это использование графических методов, таких как график временных рядов, график автокорреляции и график частной автокорреляции. График временного ряда позволяет визуализировать изменения значений ряда с течением времени, график автокорреляции показывает корреляцию между значениями ряда в разные периоды времени, а график частной автокорреляции учитывает корреляцию только между двумя значениями, пропуская все промежуточные значения.

Автокорреляция - это мера корреляции между значениями ряда с разницей во времени. Если временной ряд имеет высокую автокорреляцию, это означает, что значения ряда в разные периоды времени имеют сильную связь между собой. Автокорреляцию можно вычислить с помощью функции корреляции Пирсона, которая вычисляет корреляцию между двумя переменными. Для временных рядов это означает вычисление корреляции между значениями ряда в разные периоды времени. Если автокорреляция является значимой, то это может означать наличие тренда, цикличности или сезонности в ряде.

Пример автокорреляции



Спектральный анализ позволяет исследовать частотную составляющую ряда. Он используется для выявления скрытых цикличных паттернов во временных рядах. Для этого временной ряд разбивается на сигналы разных частот, а затем анализируется спектр этих частот. Спектральный анализ часто используется для анализа финансовых рынков, в которых цены могут изменяться в зависимости от времени и частоты.

Моделирование временных рядов.

Моделирование временных рядов: авторегрессионные модели, скользящее среднее, ARIMA.

Авторегрессионные модели (AR-модели) используют прошлые значения ряда для прогнозирования его будущих значений. Эта модель предполагает, что текущее значение ряда зависит от его предыдущих значений. Наиболее популярной AR-моделью является модель первого порядка (AR(1)), которая предполагает, что текущее значение ряда зависит только от его предыдущего значения.

Примером использования AR-модели может служить анализ финансовых рынков.

Прогнозированная величина в этом случае является суммой прошлых значений, умноженных на числовой множитель.

$$X_t = C + \phi_1 X_{t-1} + \phi_2 X_{t-2} \ldots + \epsilon_t$$

Следующий метод моделирования временных рядов - **скользящее среднее (MA-модели)**. Эта модель использует прошлые значения ошибок - разницу между фактическими значениями ряда и его прогнозируемыми значениями, для прогнозирования будущих значений. В общем случае это расчет для анализа точек данных путем создания ряда средних значений различных подмножеств полного набора данных.

$$SMA_k = \frac{p_{n-k+1} + p_{n-k+2} \cdots + p_n}{k}$$
$$= \frac{1}{k} \sum_{i=n-k+1}^{n} p_i$$

Примером использования MA-модели может служить прогнозирование количества пользователей веб-сайта. Например, если предыдущие прогнозы ошиблись на определенный процент, можно использовать эту информацию в модели прогнозирования для улучшения точности прогноза количества пользователей. Или, например, скользящее среднее часто применяется для анализа временных рядов акций.

ARIMA - это модель, комбинирующая авторегрессионные и скользящие средние модели. ARIMA позволяет моделировать данные, не являющиеся стационарными, как это не требуется для AR- и MA-моделей. ARIMA включает три параметра: параметр авторегрессии (p), параметр скользящего среднего (q) и параметр интегрирования (d).

Примером использования ARIMA может служить прогнозирование месячной выручки продукта на основе ежемесячных данных за прошлый год. Если прошлые данные имеют тренд, сезонность или циклы, можно использовать ARIMA для учета этих факторов в прогнозировании выручки. Существует также модель ARMA. Которая представляет собой сочетание моделей выше. Разница в этих моделях заключается в том, что ARIMA включает в себя этап интегрирования для преобразования временного ряда в стационарный вид перед применением авторегрессии и скользящего среднего. В то время как ARMA используется для моделирования стационарных временных рядов без необходимости интегрирования.

Теорема Волда — утверждение математической статистики, согласно которому каждый временной ряд можно представить в виде скользящего среднего бесконечного порядка.

$$MA(\infty)$$

Так как модель ARMA, согласно теореме о разложении Волда, теоретически достаточна для описания регулярного **стационарного** временного ряда, мы заинтересованы в стационарности ряда.

Почему стационарность так важна? Предполагается, что, если временной ряд ведет себя определенным образом, очень высока вероятность того, что он повторит те же паттерны в будущем.

Преимущества и недостатки

К преимуществам методов анализа и прогнозирования временных рядов можно отнести:
- Позволяют оценить и выявить тенденции и цикличность временного ряда, что может привести к более точному прогнозированию будущих значений ряда;
- Позволяют анализировать корреляцию между данными и выявлять зависимости между ними;
- Позволяют определить влияние различных факторов на изменение ряда.

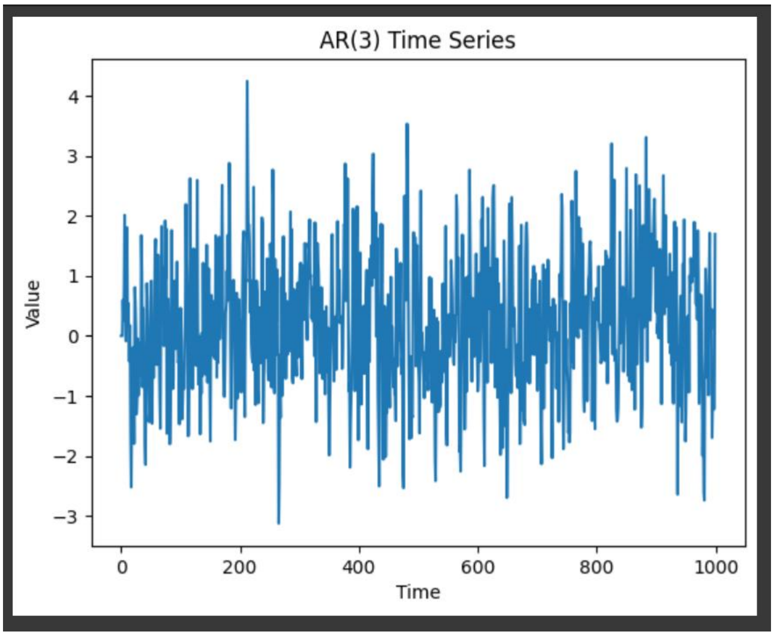
Тем не менее, методы анализа и прогнозирования временных рядов имеют и свои недостатки:
- Недостаточное количество данных может существенно ограничить точность прогнозов;
- Несмотря на составление относительно точных прогнозов на основе исторических данных, случайные факторы и внеплановые обстоятельства могут привести к неверному прогнозу;
- Методы анализа временных рядов могут быть недостаточными, если данные обладают сезонностью и этот фактор не был учтен в анализе.

Программная реализация

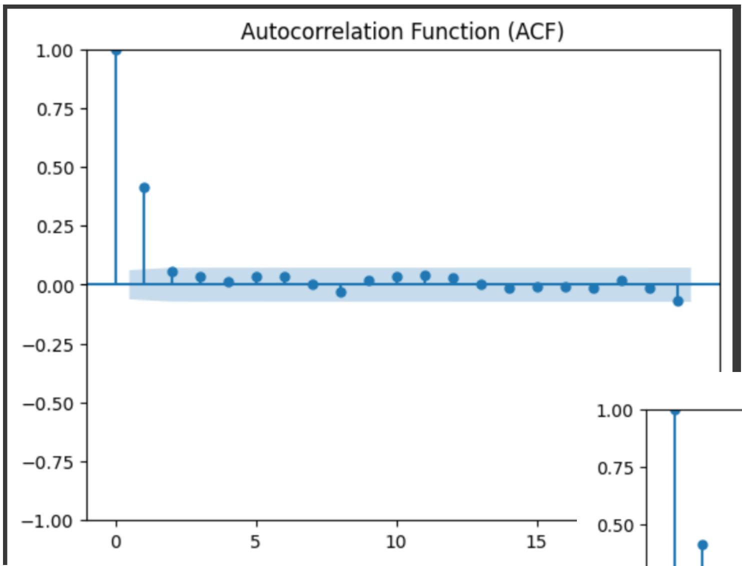
Код предоставлен по ссылке:

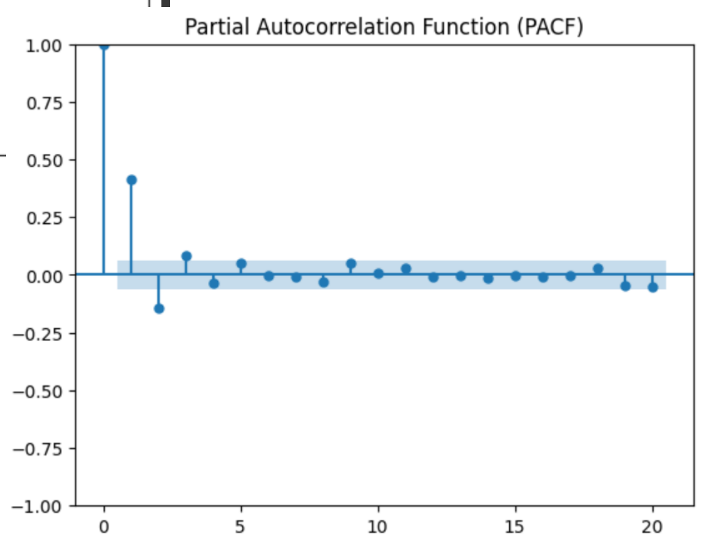
https://colab.research.google.com/drive/1Y99akMiNJt_npdwSrjEqteEhdmvh3qPX?usp=sharing

1. Построили авторегрессионную модель временного ряда AR(3):



2. Строим графики ACF и PACF:





3. Проверяем, что временной ряд с коэффициентами авторегрессии {ai} будет стационарным:

Проверка стационарности с использованием теста Дики-Фуллера

```python
from statsmodels.tsa.stattools import adfuller

# Тест Дики-Фуллера
result = adfuller(x)
print('ADF Statistic:', result[0])
print('p-value:', result[1])
print('Critical Values:', result[4])

# Проверка на стационарность (если p-value меньше уровня значимости, то ряд стационарен)
if result[1] <= 0.05:
    print('Ряд стационарен.')
else:
    print('Ряд нестационарен.')
```

```
ADF Statistic: -15.295552062566694
p-value: 4.383420687353554e-28
Critical Values: {'1%': -3.4369259442540416, '5%': -2.8644432969122833, '10%': -2.5683158550174094}
Ряд стационарен.
```

Проверим с помощью корней характеристического уравнения.

Временной ряд, заданный моделью авторегрессии AR(p), является стационарным тогда и только тогда, когда все корни характеристического уравнения по модулю больше единицы.

Построим график:

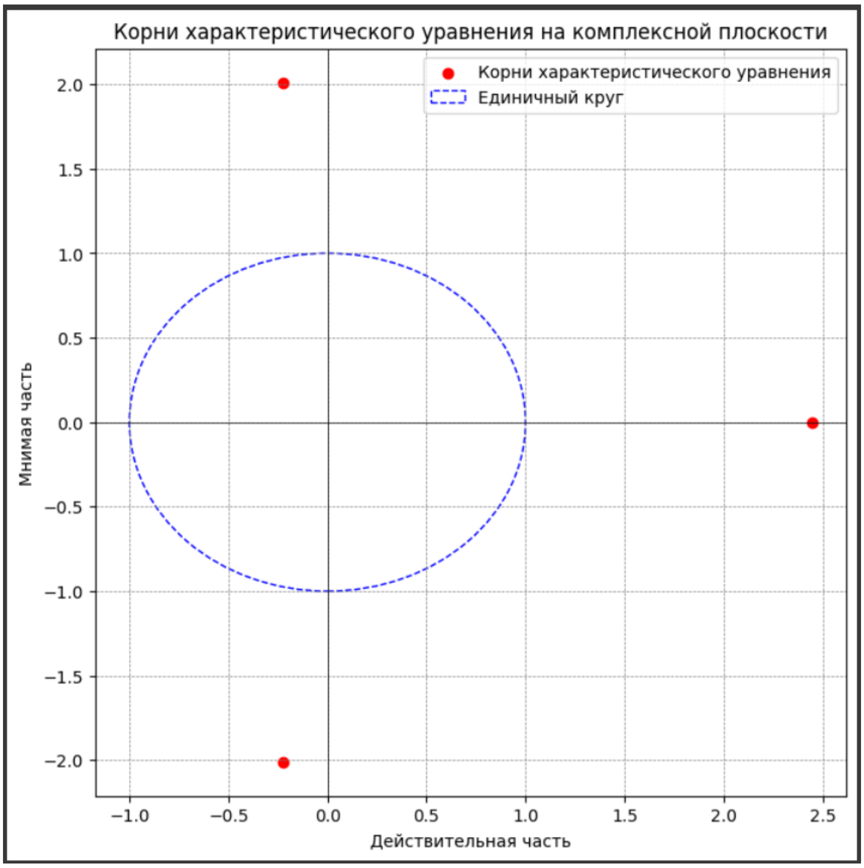


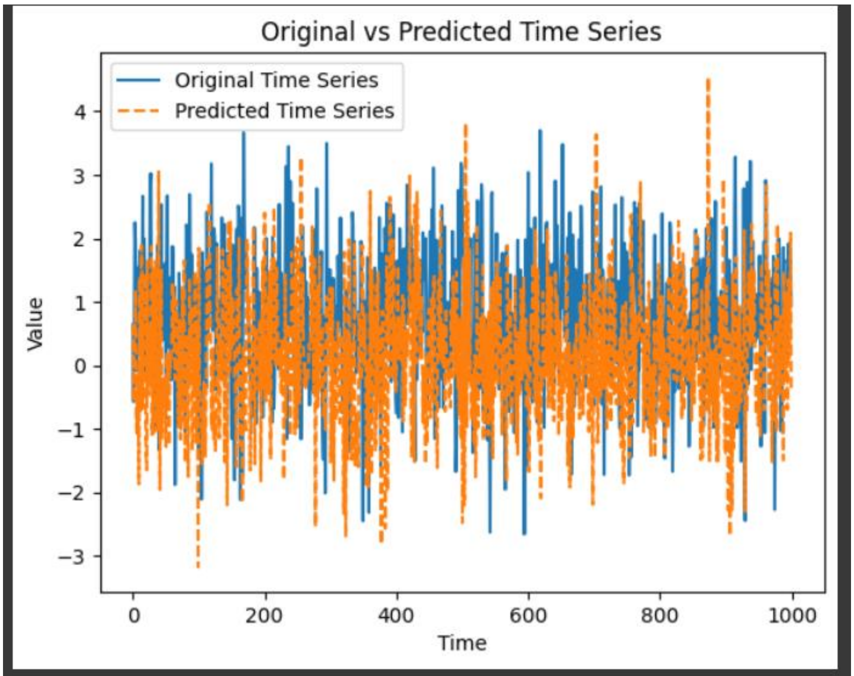
График представляет собой корни характеристического уравнения на комплексной плоскости. Красные точки представляют эти корни. Единичный круг (синяя пунктирная окружность) используется для визуализации единичного круга на комплексной плоскости.

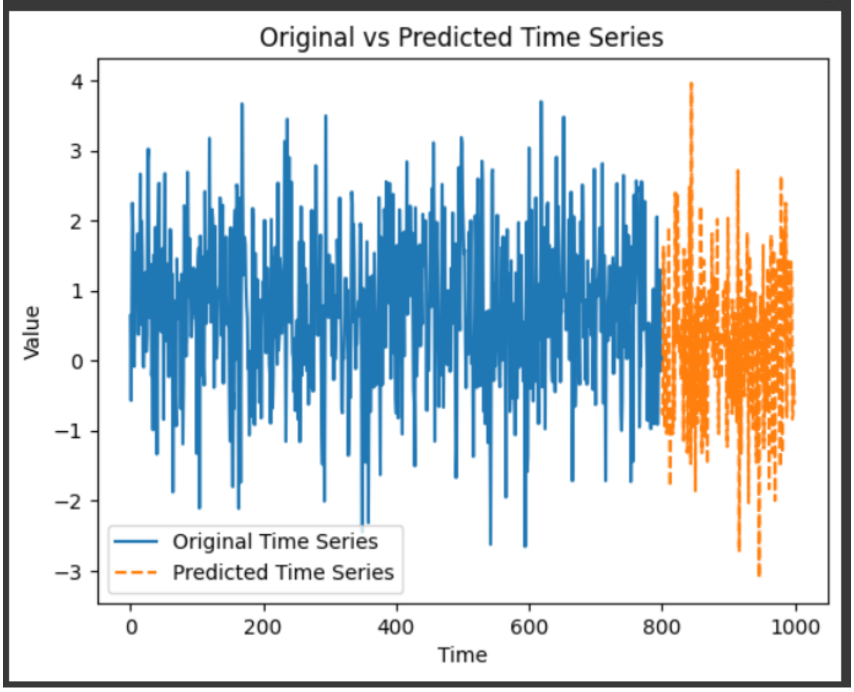
Видим, что наш ряд стационарен.

4. Преобразуем ряд к последовательности векторов задержек. Обучим модель:

```
Mean Squared Error: 1.2346900879279072
Авторегрессионные коэффициенты: [ 0.13472876 -0.15861774  0.21323142]
```

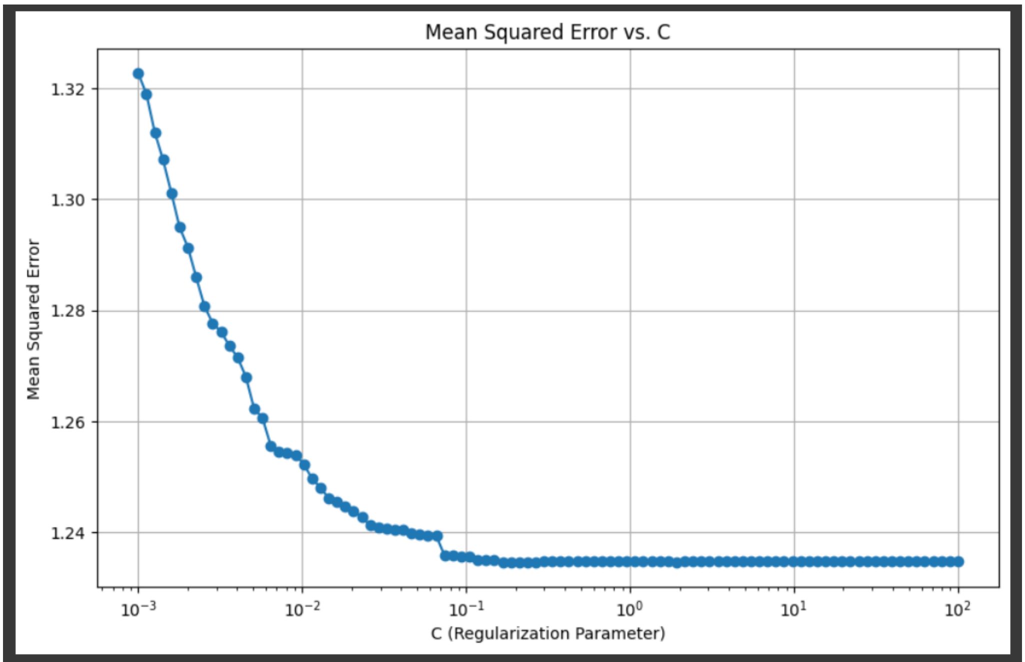
Построим временной ряд:

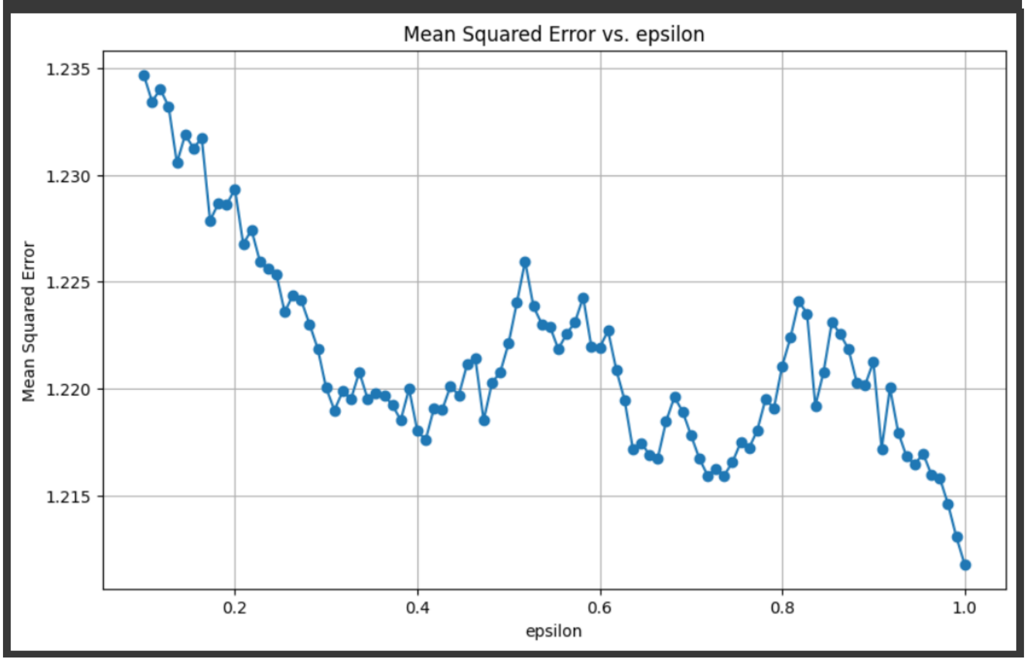




На графике видно, что модель хорошо повторяет особенности оригинального временного ряда.

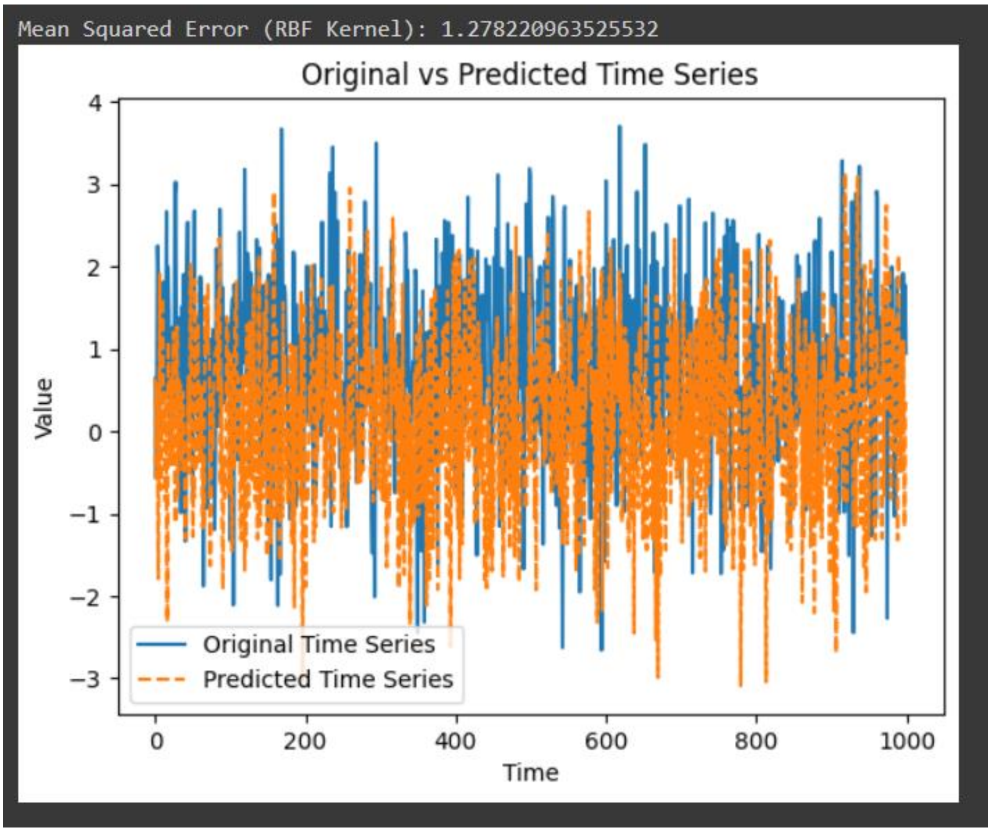
5. Проанализируем результаты при различных ядрах и других гиперпараметрах модели. Ядро linear:



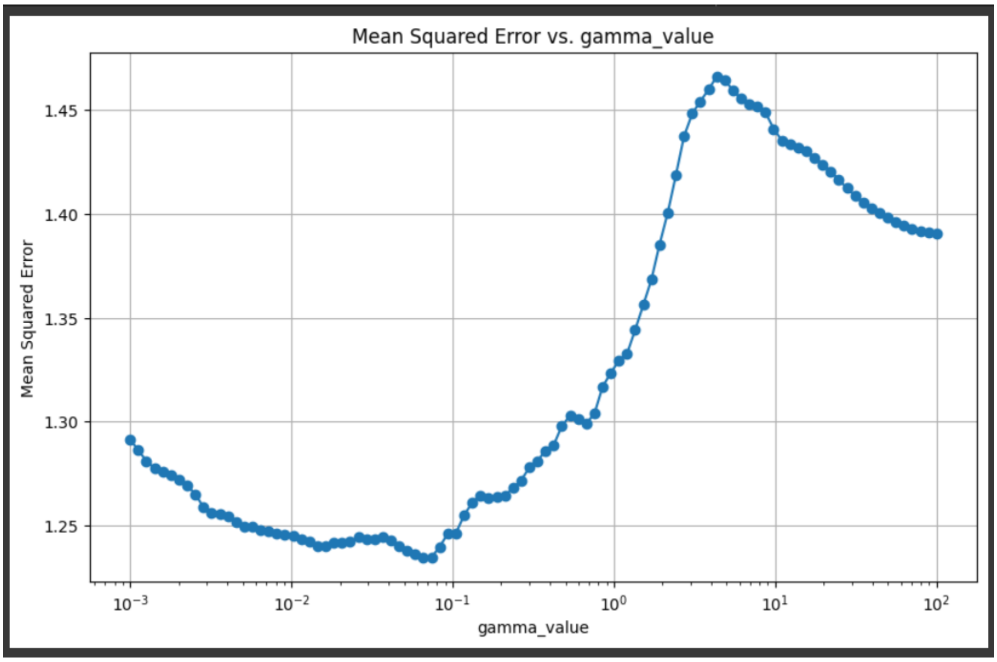


Видно, что с увеличением значений гиперпараметров, модель показывает лучшие результаты, ошибка уменьшается.

Работаем с ядром rbf.
Строим ряд:

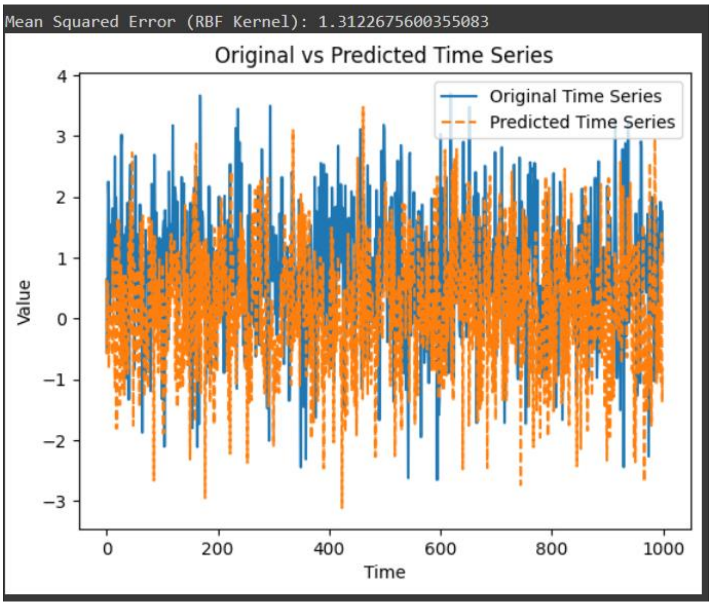


Анализируем:

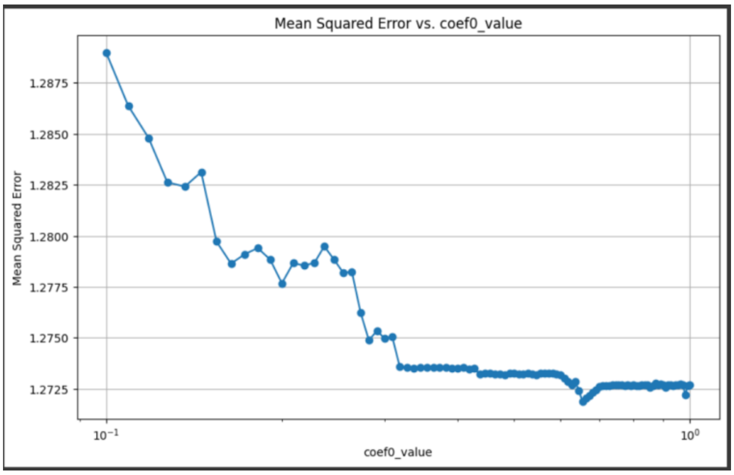


Видно, что лучшие результаты модель получает при $gamma \approx 10^{-1}$, дальше ошибка резко возрастает.

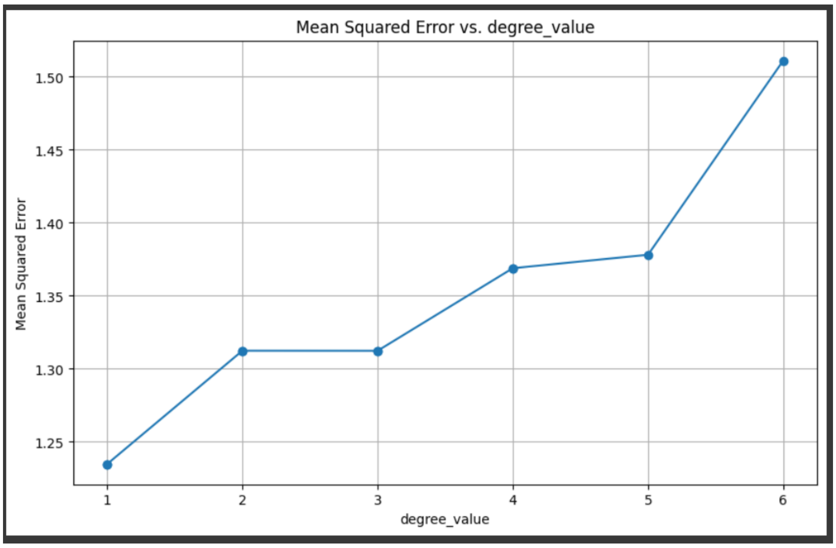
Работаем с ядром poly.
Строим:



Анализируем:

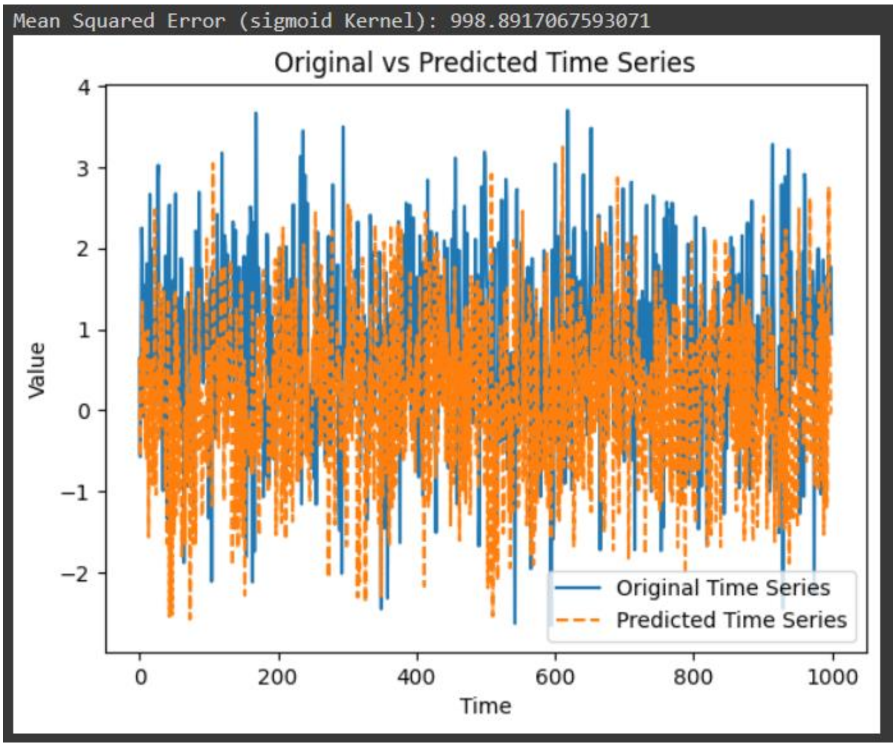


С увеличением значения coef_0 значение ошибки падает, мы улучшаем показатели модели.

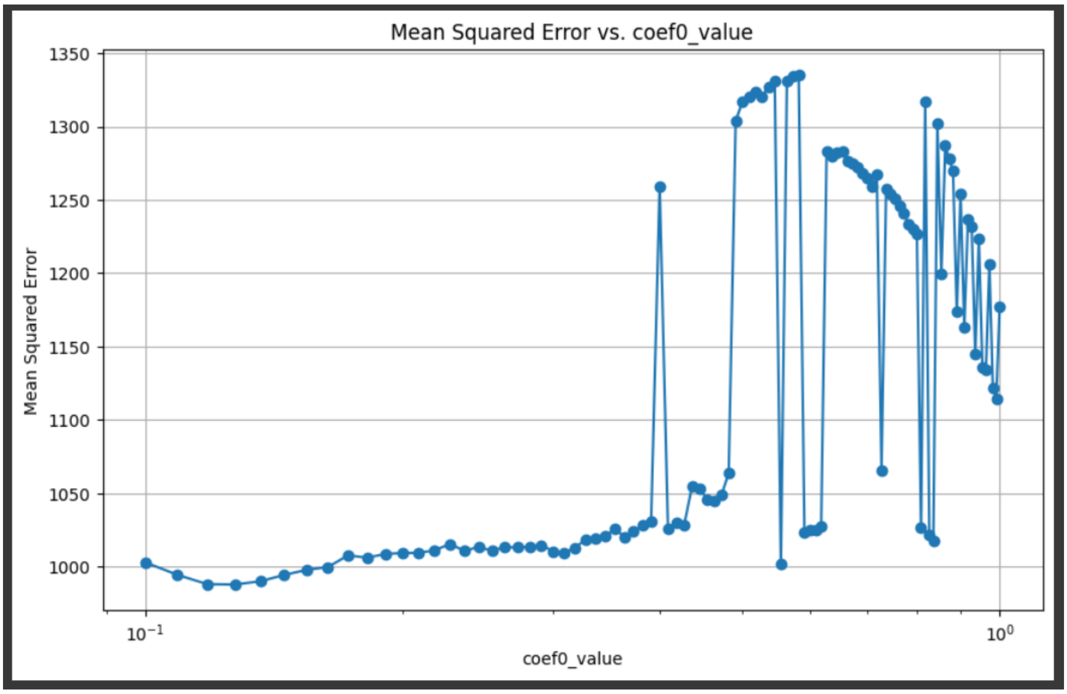


Лучшие показатели модель получает на меньших значениях degree.

Работаем с ядром sigmoid.
Строим график:



Анализируем:



Лучший результат модель показала на низких значениях coef0.

Выводы

В соответствии с поставленной задачей была построена авторегрессионная модель временного ряда AR(3). Перед построением модели была выполнена проверка стационарности временного ряда. Для этого использовались различные статистические тесты, такие как тест Дика-Фуллера и проверка корней характеристического уравнения. С использованием полученных коэффициентов и случайно сгенерированного шума были сгенерированы значения временного ряда. Построенный ряд демонстрирует характерные свойства авторегрессии и стационарности.

Модель машины опорных векторов была успешно обучена на основе преобразованных векторов задержек временного ряда. Найденные параметры использовались для построения продолжения временного ряда.

Сравнение графиков и анализ результатов позволяют сделать вывод о том, насколько хорошо модель SVM аппроксимирует исходный временной ряд AR(3). Качество модели напрямую зависит от выбранных параметров и ядра. Лучше всего себя показала модель, авторегрессионные коэффициенты которой мы принимаем за [0.13472876 -0.15861774 0.21323142], а параметр C за 10^{-2}.