

# ПРОЕКТИРОВАНИЕ БИБЛИОТЕКИ ПРОГРАММ ДЛЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

**А. С. Макаревич**

*Белорусский государственный университет, г. Минск;  
anatoly\_mak@yahoo.com;  
науч. рук. – В. И. Малюгин, канд. физ.-мат. наук, доц.*

Приводится общее описание пакета «*Time Series for Data Science*» (сокращенно – *ts4ds*) на языке *Python*, предназначенного для автоматизации процесса анализа временных рядов на основе методов машинного обучения, Код пакета доступен в свободном доступе (Open Source).

**Ключевые слова:** анализ временных рядов; машинное обучение; язык Python.

## ОБЩАЯ ХАРАКТЕРИСТИКА ПРОБЛЕМЫ И ЦЕЛЬ ИССЛЕДОВАНИЯ

Анализ временных рядов занимает особое место в анализе данных на основе методов машинного обучения. Это объясняется большим разнообразием свойств реальных временных рядов и, соответственно, их возможных типов моделей, представимых, как правило, в параметрическом виде. Для адекватного описания моделей необходимо решить ряд обязательных задач [1]:

1. установление свойств временных рядов (стационарный, нестационарный, содержащий детерминированные или стохастические тренды), которые должны учитываться при выборе типа модели и метода оценивания;
2. выбор лучшей спецификации для модели используемого типа, учитывающей дополнительные особенности временных рядов (сезонные и структурные изменения, календарные эффекты и т.д.);
3. выбор критерия оптимальности модели, соответствующего целям ее построения;
4. корректный выбор метода оценивания параметров моделей, обеспечивающего оптимальные статистические свойства оценок параметров и построенной модели в целом (метод максимального правдоподобия, метод наименьших квадратов, метод моментов и их обобщения).
5. При построении моделей должны учитываться также цели ее применения и особенности имеющихся данных, например: число совместно моделируемых временных рядов (одномерные либо многомерные модели), тип зависимых переменных (непрерывные или дискретные) интервал наблюдения и длина временных рядов и др. При построении предиктивных моделей, описывающих зависимости прогнозируемых вре-

менных рядов от внешних факторов требуется также знание свойств соответствующих им временных рядов,

При обработке большого количества временных рядов с целью построения лучших в определенном смысле предиктивных моделей [2] «ручное построение» моделей для каждого временного ряда может занять не приемлемо большое время, что обуславливает актуальность создания пакетов, автоматизирующих поиск оптимальных или приемлемых решений на каждом этапе построения модели. Разработка такого пакета на языке Python и являлось основной целью данного исследования. Приведем общее описание принципов организации разрабатываемого пакета и проиллюстрируем некоторые его возможности.

### ПРИНЦИПЫ РАЗРАБОТКИ И ОРГАНИЗАЦИИ ПАКЕТА

Как отмечалось, процесс моделирования принято разделять на этапы *выбора модели* и *оценки параметров*. На первом этапе (*model selection*) выбирается определенный класс моделей, а на втором (*model estimation, fitting*) производится поиск значений параметров, основываясь на исходных данных. Более сложные модели не обязательно дают более точные прогнозы, поэтому процесс выбора нетривиальный [3]. При построении моделей также приходится применять различные преобразования исходных временных рядов (*transformations*).

При проектировании и написании пакета *ts4ds*, были поставлены следующие задачи:

- использовать методы статистического и машинного обучения для построения эконометрических моделей временных рядов;
- разработать обобщенное представление модели временных рядов, с целью автоматизации процессов её построения и применения;
- реализовать в библиотеке процедуры построения, применения некоторых основных семейств моделей;
- реализовать процедуры автоматизации работы с моделями (например, поиск гипер-параметров);
- разработать утилиты для облегчения создания новых моделей;
- провести тестирование компонент библиотеки (в программном и в статистическом смысле).

**Базовые классы объектов.** Для выполнения поставленных задач потребовался основательный подход к проектированию и разработке библиотеки. Базовые классы формировались автором, основываясь на теории из [4]. Определим базовые классы и соответствующие понятия.

С точки зрения реализации, *модель (model)* – это объект, описывающий поведение временных рядов. В рамках пакета модель представляет-

ся параметризованными уравнениями, описывающими отношения между переменными. Эти переменные можно условно разбить на *эндогенные* (*endogenous* – объясняемые моделью) и *экзогенные* (*exogenous* – не объясняемые, а взятые извне и неизменяемые внутри модели). У моделей есть фиксированный набор *параметров* (*parameters*), значения которых окончательно определяют модель.

*Предиктивные модели* (*predictor*) позволяют строить прогнозы значений временных рядов. Примерами являются авторегрессионные модели, экспоненциальное сглаживание, и рекуррентные нейросети.

*Преобразования* (*transformation*) временных рядов изменяют значения всего временного ряда по определенному принципу, получая при этом новый ряд. Примерами являются взятие разностей, сезонная корректировка, и методы декорреляции. При разработке библиотеки стало понятно что преобразования удобно рассматривать как вид моделей.

*Эстиматор* (от англ. *estimator*) – алгоритм, который ищет оптимальные (по какому-то критерию) модель. Обычно этот поиск происходит в пространстве параметров, т.е. это алгоритмы оценивания параметров. В пакете *ts4ds* эстиматорами также считаются алгоритмы поиска подходящей модели, лучших гиперпараметров, и другие оптимизационные задачи.

Реализация основной части библиотеки следует этой иерархии классов с использованием возможностей языка Python; в подходе моделей от общего к частному [5]. Постоянство высокоуровневого интерфейса позволяет иметь “plug-and-play” стандарт, в котором можно заменять объектами одного рода (например, различные модели) с минимальным количеством изменений.

## ПРИМЕР ПОСТРОЕНИЯ МОДЕЛИ ВРЕМЕННОГО РЯДА

На данный момент реализованы различные способы преобразований временных рядов, а также методы построения некоторых моделей, включая модель сезонной ARIMA (SARIMAX) [1] и модель с маковскими переключениями состояний (MS-ARX) [6, 7]. В библиотеку также включены:

- статистические тесты;
- программные тесты на основании тестовых данных из пакета Stata;
- адаптеры для алгоритмов машинного обучения, предназначенные для анализа перекрестных данных, к задаче прогнозирования;
- утилиты для разработчиков новых моделей.

На графиках, представленных на рис. 1 и 2 приводятся некоторые преобразования, полученные с помощью пакета на основе данных дея-

тельности авиакомпаний США и результаты построения модели SARIMAX на этих же данных.

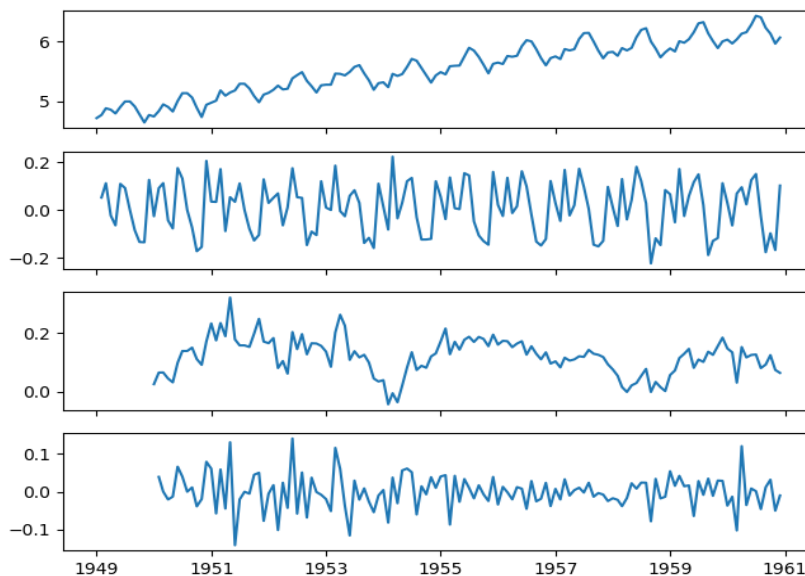


Рис. 1. Сверху вниз: исходный ряд, первая разность, сезонная разность (S=12), композиция сезонной и первой разности

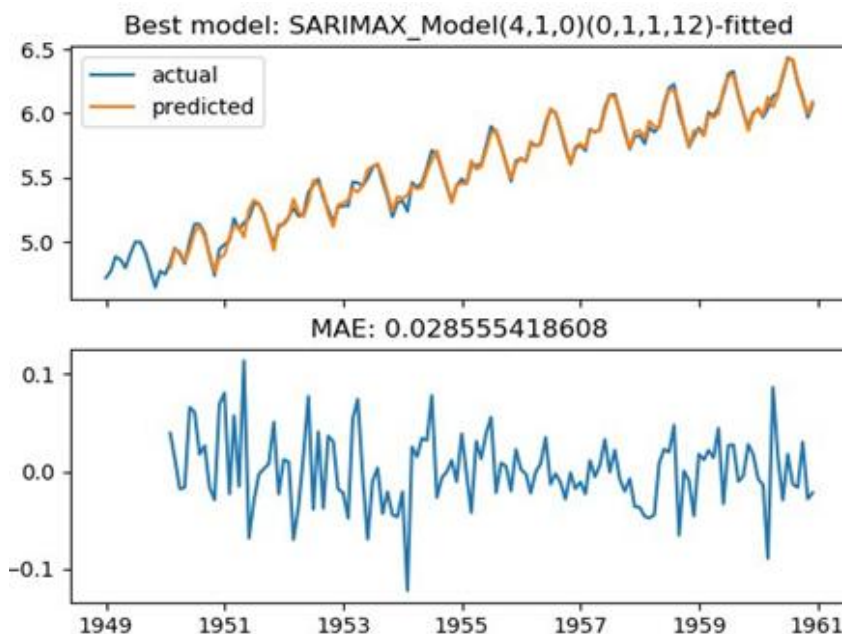


Рис. 2. Результаты автоматического выбора модели SARIMAX

### Библиографические ссылки

1. Харин, Ю. С. Эконометрическое моделирование : учебное пособие / Ю. С. Харин, В. И. Малюгин, А. Ю. Харин. Минск: БГУ, 2004. 313 с.

2. *Makridakis S, Spiliotis E, Assimakopoulos V.* Statistical and Machine Learning forecasting methods: Concerns and ways forward. // PLoS ONE 13(3): e0194889. 2018. <https://doi.org/10.1371/journal.pone.0194889>
3. *Hyndman, R.J., Athanasopoulos, G.* Forecasting: Principles and Practice. 2018. URL: <https://otexts.org/fpp2/>.
4. *Perktold, J.* Python package Statsmodels on statsmodels.org. 2017. URL: <http://www.statsmodels.org/stable/index.html>.
5. *Fulton, C.* State space modeling in Python. 2016. URL: [http://www.chadfulton.com/topics/state\\_space\\_python.html](http://www.chadfulton.com/topics/state_space_python.html).
6. *Malugin, V.* Statistical Estimation and Classification Algorithms for Regime-Switching VAR Model with Exogenous Variables / V. Malugin, A. Novopoltsev // Austrian Journal of Statistics. Vol. 46. 2017. P. 47–56.
7. *Макаревич, А.С.* Сравнительный анализ оценок поворотных точек Экономического цикла на основе алгоритмов Ходрика – Прескотта и Хамильтона / А.С. Макаревич, В.И. Малюгин // Сборник работ 74-ой научной конференции студентов и аспирантов Белорусского государственного университета: В 3 ч. ч.1 БГУ, 2017. С. 54–57.