

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра математического моделирования и анализа данных

Аннотация к дипломной работе

Обнаружение и анализ выбросов во временных рядах

Пацкевич
Анастасия Андреевна

Научный руководитель:
кандидат физ.-мат. наук,
доцент кафедры ММАД
Сталевская Светлана Николаевна

Минск 2022

Реферат

Дипломная работа, 50 страниц, 40 рисунков, 11 источников.

Ключевые слова: ВРЕМЕННЫЕ РЯДЫ, СТАТИСТИЧЕСКИЙ АНАЛИЗ, АДДИТИВНЫЕ ВЫБРОСЫ, ИННОВАЦИОННЫЕ ВЫБРОСЫ, СДВИГИ УРОВНЯ, ARIMA-МОДЕЛЬ.

Объектом исследования в рамках данной работы являются ARIMA-модели.

Предметом исследования являются выбросы во временных рядах и их влияние на прогнозирование и модель в целом.

Целью работы является классификация выбросов, их идентификация, а также прогнозирование с помощью выбросов. Оценка влияния их на различные модели данных.

Методы исследования – методы теории вероятности и математической статистики, эконометрическое моделирование.

Результат – для достижения указанной цели в работе решаются следующие задачи:

- 1) обзор литературы, посвященной статистическому анализу выбросов во временных рядах.
- 2) классификация выбросов в рамках ARIMA модели. Моделирование временных рядов и их с выбросами, чтобы наглядно рассмотреть влияние.
- 3) изучение процедур идентификации и оценки выбросов. Реализация алгоритма и проведение компьютерных экспериментов на реальных данных для демонстрации выбросов и их влияния на модель.

Область применения – статистический анализ временных рядов.

Рэферат

Дыпломная работа, 50 старонак, 40 малюнкаў, 11 крыніц.

Ключавыя словы: ЧАСОВЫЯ ПАСЛЯДОЎНАСЦІ, СТАТЫСТЫЧНЫ АНАЛІЗ, АДЫТЫЎНЫЕ ВЫКІДЫ, ІНАВАЦЫЙНЫЯ ВЫКІДЫ, СДВІГІ ЎЗРОЎНЮ, ARIMA-МАДЭЛІ.

Аб'ектам даследавання ў межах гэтай работы з'яўляюцца arima-мадэлі.

Прадметам даследавання з'яўляюцца выкіды ў часовых паслядоўнасцях і іх уплыў на прагназаванне і мадэль у цэлым.

Мэтай работы з'яўляецца класіфікацыя выкідаў, іх ідэнтыфікацыя, а таксама прагназаванне пры дапамозе выкідаў. Ацэнка ўплыву іх на розныя мадэлі.

Метады даследавання – метады тэорыі верагоднасці і матэматычнай статыстыкі, эканаметрычнае мадэляванне.

Вынік – дзеля дасягнення пазначанай мэты ў рабоце вырашаюцца наступныя задачы:

1) агляд літаратуры, прысвечанай статыстычнаму аналізу выкідаў у часовых паслядоўнасцях.

2) класіфікацыя выкідаў у рамках ARIMA мадэлі. Мадэляванне часовых паслядоўнасцяў і іх з выкідамі, каб наглядна разгледзець уплыў.

3) вывучэнне працэдур ідэнтыфікацыі і адзнакі выкідаў. Рэалізацыя алгарытму і правядзенне кампутарных эксперыментаў на рэальных дадзеных для дэманстрацыі выкідаў і іх уплыву на мадэль.

Вобласць прымянення – статыстычны аналіз часовых паслядоўнасцяў.

Abstract

Graduate work, 50 pages, 41 drawings, 11 sources.

Key words: TIME SERIES, STATISTICAL ANALYSIS, ADDITIVE OUTLIERS, INNOVATIVE OUTLIERS, LEVEL SHIFTS, ARIMA MODEL.

Object of research within this work are ARIMA models.

Subject of research are outliers in time series and their impact on forecasting and the model as a whole.

Purpose of the work is the classification of outliers, their identification, as well as forecasting using outliers. Evaluation of their impact on various data models.

Research methods – methods of probability theory and mathematical statistics, econometric modeling.

Result – to achieve the goal of this work several tasks are solved:

- 1) a review of the literature on the statistical analysis of outliers in time series.
- 2) classification of outliers within the ARIMA model. Modeling time series and them with outliers to visually consider the impact.
- 3) research of procedures for identifying and evaluating outliers. Implementing the algorithm and running computer experiments on real data to demonstrate outliers and their impact on the model.

Scope – statistical time series analysis.

Введение

В данных временных рядов иногда есть значения, сильно отличающиеся от других данных, и они не отражают характеристики набора данных. Значения данных называются выбросами. При анализе данных временных рядов часто обнаруживаются выбросы. Данные о выбросах существенно влияют на данные прогнозирования, если мы используем в прогнозировании такие модели, как ARIMA, ARMA и другие. Прогнозирование — это деятельность по предсказанию того, что произойдет в будущем. Для прогнозирования временного ряда мы можем использовать метод прогнозирования, например, ARIMA Box-Jenkins метод. Прогнозирование будет далеко от того, что мы предскажем, если данные содержат выбросы. Исследование по обнаружению выбросов очень важно, потому что наличие выбросов может привести к неправильному прогнозу. Если выброс данных не идентифицируется правильно, то это повлияет на прогноз, а не на фактические данные. Аддитивные выбросы являются наиболее распространенным типом выбросов, обнаруживаемых во временных рядах. Если наблюдения во временном ряду могут быть предсказаны с уверенностью и не требуют дальнейшего исследования, это называется детерминированным временным рядом, и, если наблюдения могут показать только структуру из вероятностного состояния, временной ряд называется стохастическим. В моделировании анализа временных рядов предполагается, что данные являются стационарными.

К изучению выбросов подошли с двух точек зрения. Первый подход — это диагностический подход, при котором методы диагностики применяются к остаткам оцениваемой модели для выявления возможных выбросов, которые затем проверяются. Когда выделяются выбросы, предлагается модель, которая их учитывает. Далее эффекты выбросов и параметры модели оцениваются совместно. Таким образом, мы получаем эффект выбросов и оценку параметров, которые становятся основной целью анализа. Второй подход: метод оценки изменен таким образом, чтобы оценки не были искажены по наличию выбросов. Когда у нас есть надежная оценка параметра, выбросы могут быть легко идентифицированы и протестированы. Обе методологии дополняют друг друга, и идеи одного подхода могут быть использованы для улучшения другого.

В работе рассматриваются модели с одним выбросом или с несколькими изолированными выбросами, и все рассмотренные методы работают только для таких моделей, но существуют модели с множеством (целыми пятнами)

выбросов, которые рассмотрены не будут. Работа организована следующим образом. В главе 1 представлены модель ARIMA и различные типы выбросов. В главе 2 обсуждаются некоторые процедуры идентификации и оценки выбросов. Также рассмотрена задача прогнозирования с помощью выбросов. Проведена работа по классификации и идентификации выбросов на реальных данных. Проведены тесты для нахождения выбросов.