

Плиска А.Г. Методы прогнозирования временного ряда // Академия педагогических идей «Новация». Серия: Студенческий научный вестник. – 2019. – №6 (июнь). – АРТ 495-эл. – 0,2 п.л. - URL: <http://akademnova.ru/page/875550>

РУБРИКА: ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 519.246.8

Плиска Арсений Геннадьевич
студент 2 курса, отдел магистратуры
Научный руководитель: Кобак В.Г, д.т.н., профессор
ФГБОУ ВО «Донской государственный технический университет»
г. Ростов-на-Дону, Российская Федерация
e-mail: arseniy.pliska@gmail.com

МЕТОДЫ ПРОГНОЗИРОВАНИЯ ВРЕМЕННОГО РЯДА

Аннотация: В статье описывается процесс прогнозирования временного ряда, рассматриваются методы используемы на каждом этапе прогнозирования.

Ключевые слова: Временной ряд, прогнозирование, построение прогноза, информационные критерии, преобразование Бокса-Кокса, тест Харки-Бера.

Pliska Arseniy
2nd year student, department of magistracy
Supervisor: Kobak V.G., Doctor of Technical Sciences, Professor
FGBOU VO "Don State Technical University"
Rostov-on-Don, Russian Federation
E-mail: arseniy.pliska@gmail.com

TIME SERIES FORECASTING METHODS

Abstract: This article describes the process of forecasting a time series, discusses the methods used at each stage of forecasting.

Keywords: time series, forecasting, forecasting, information criteria, Box-Cox transformation, Jarque-Bera test

В статье описываются этапы прогнозирования временного ряда моделью ARIMA и методы применяемы на каждом этапе.

Для начала необходимо стабилизировать дисперсию. Проведем тест Харки — Бера для определения нормальности распределения.

Тест Харке—Бера (Jarque-Bera test) — это статистический тест, проверяющий ошибки наблюдений на нормальность посредством сверки их третьего момента (асимметрия) и четвертого момента (эксцесс) с моментами нормального распределения.

Статистика Jarque – Bera (JB) (Харке – Бера) предназначена для проверки нулевой гипотезы о нормальном законе распределения для значений рассматриваемой переменной. Чем ближе распределение ошибок к нормальному, тем меньше статистика Харке—Бера отличается от нуля. При достаточно большом значении статистики p-value будет мало, и тогда будет основание отвергнуть нулевую гипотезу.

В случае ненормальности распределения. Необходимо избавиться от сильной дисперсии, это можно сделать посредством преобразований Бокса-Кокса.

Для исходной последовательности, однопараметрическое преобразование Бокса-Кокса[1] с параметром λ определяется следующим образом:

$$y = \{y_1, \dots, y_n\}, y_i > 0, i = 1, \dots, n$$

$$y_i^\lambda = \begin{cases} \frac{y_i^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0, \\ \log(y_i), & \text{if } \lambda = 0. \end{cases}$$

Параметр λ можно выбирать, максимизируя логарифм правдоподобия. Еще один способ поиска оптимального значения параметра основан на поиске максимальной величины коэффициента корреляции между квантилями функции нормального распределения и отсортированной преобразованной последовательностью. Преобразования Бокса-Кокса избавят нас от дисперсии в ряде.

Далее необходимо проверить наш ряд на наличие тренда. Мы знаем, что ряд с трендом не может быть стационарным. Проверим ряд на стационарность посредством теста Дикки-Фуллера.

Тест Дики — Фуллера[2] (DF-тест, Dickey — Fuller test) — это методика, которая используется в прикладной статистике и эконометрике для анализа временных рядов для проверки на стационарность. Является одним из тестов на единичные корни.

Данный критерий проверяет гипотезу, что ряд стационарен против того, что он стационарен. Если ряд не стационарен необходимо продифференцировать ряд. Дифференцирование ряда уничтожает тренд и уже после первого преобразования ряд может стать стационарным.

Когда мы добились стационарности ряда можно подбирать параметры модели. Для подбора моделей нам поможет автокорреляция.

Автокорреляция [autocorrelation, serial correlation] — корреляционная связь (см. Корреляция) между значениями одного и того же случайного про

цесса $X(t)$ в моменты времени t_1 и t_2 . Функция, характеризующая эту связь, называется автокорреляционной функцией.

Для построения модели $ARIMA(p,d,q)$ нам нужно определить константы p, P, q, Q, d, D :

Параметры d, D это минимальное необходимое количество дифференцирований необходимых для приведения ряда к стационарному виду. Чем больше дифференцирований, тем больше будет дисперсия итогового прогноза.

Q - значение автокорреляции, последний значимый сезонный лаг.

q -последний значимый несезонный лаг, но не более сезонного периода.

Для подбора параметров p, P используется частичная автокорреляционная функция.

Частичная автокорреляция- автокорреляция после снятия авторегрессии на промежуточные значения.

P - номер последнего значимого сезонного лага, при котором частичная автокорреляция значима

p - номер последнего значимого несезонного лага, при котором частичная автокорреляция значима.

После подбора параметров модели модели с разными p, P, q, Q можно сравнить между собой используя информационный критерий. Чем меньше информационный критерий у модели, тем лучше модель подходит для прогнозирования.

Критерий AIC[3], Исходная расчетная формула критерия имеет вид:

$$AIC=2k-2l.$$

Где l значение логарифмической функции правдоподобия построенной модели, k - количество использованных (оцененных) параметров.

Данный подход позволяет сравнивать модели, оцененные по выборкам разного объема. Чем меньше значение критерия, тем лучше модель. Многие другие критерии являются модификациями АИС.

Критерий ВИС, Байесовский информационный Он разработан исходя из байесовского подхода и является наиболее часто используемой модификацией АИС: $ВИС = SC = k \ln n - 2l$.

Как видно из формулы, данный критерий налагает больший штраф на увеличение количества параметров по сравнению с АИС, так как $\ln n$ больше 2 уже при количестве 8 наблюдений.

Критерий САИС, состоятельный критерий Акаике (Consistent AIC — САИС) предложенный в 1987 году Боздоганом: $САИС = (1 + \ln n)k - 2l$.

Критерий АИСс, скорректированный критерий $AIC_c = AIC + 2k(k+1)/(n-k-1)$. Предназначен для коротких выборок. был предложен для использования в задач маленькой размерности, когда $\frac{n}{K} \leq 40$. В то же время, при больших значениях $\frac{n}{K}$ использование двух критериев равно возможно. Особенность критерия АИСс заключается в том, что функция штрафа умножается на поправочный коэффициент.

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}.$$

$$AIC_c = \ln \frac{SSE}{n} + \frac{n+K}{n-K-2}.$$

Для проверки валидности построенной модели необходимо проверить остатки модели. Остатки - это разность между фактом и прогнозом.

После построения модели необходимо проверить соответствие остатков некоторым свойствам:

- Несмещенность, если остатки в среднем не равны нулю, то прогноз не корректен. Несмещенность можно проверить критерием Стьюдента.
- Стационарность, если остатки не стационарны, значит в них есть структура и модель не смогла выбрать всю полезную для прогнозирования информацию. Если остатки не стационарны, то модель не одинаково точна в разные периоды
- Неавтокоррелированность остатков- отсутствие зависимости от предыдущих наблюдений. Можно проверить с помощью Q-тест Льюнг — Бокса.

В случае невыполнения условий для остатков, необходимо улучшать модель или менять модель прогнозирования.

В рамках данной работы был рассмотрен процесс построения модели. Были проанализированы и изучены методы которые используются при подготовке модели, а также методы анализа и преобразования временного ряда. Подробно рассмотрены популярные информационные критерии.

Список использованной литературы:

1. Порунов А.Н. Бокс-Кокс преобразование и иллюзия "нормальности" макроэкономического ряда. 2010
2. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс. — М.: Дело, 2007. — 504 с.
3. Бидюк П. И., Зворыгина Т. Ф. Структурный анализ методик построения регрессионных моделей по временным рядам наблюдений. 2006

Дата поступления в редакцию: 12.06.2019 г.

Опубликовано: 13.06.2019 г.

© Академия педагогических идей «Новация». Серия «Студенческий научный вестник»,
электронный журнал, 2019

© Плиска А.Г., 2019

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации ЭЛ №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru