ИННОВАЦИОННЫЕ МЕТОДЫ ПРИНЯТИЯ РЕШЕНИЙ

УДК 330.43

JEL C16; G02

DOI: 10.17213/2312-6469-2021-4-233-241

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ ПРИ АНАЛИЗЕ И ПРОГНОЗИРОВАНИИ ИНДЕКСА DOW JONES НА ФОНДОВОМ РЫНКЕ

© Я.В. Федорова*, А.Ф. Рогачев** 2021

**Ростовский государственный экономический университет (РИНХ),
г. Ростов-на-Дону, Россия*

***Волгоградский государственный аграрный университет,
г. Волгоград, Россия*

Рассмотрена задача анализа и прогнозирования фондового индекса на рынке активов. Проведен анализ фондового индекса, отслеживающего развитие промышленной составляющей фондового рынка. Проверена стационарность данных. Реализована предсказательная модель для прогнозирования цены закрытия индекса Доу-Джонса на фондовом рынке, используя модель линейной регрессии.

Ключевые слова: фондовый индекс, фондовый рынок, линейная регрессия, индекс Доу-Джонс, гипотезы статистического теста.

APPLICATION OF STATISTICAL METHODS IN THE ANALYSIS AND FORECASTING OF THE DOW JONES INDEX ON THE STOCK MARKET

© Y.V. Fedorova*, A.F. Rogachev** 2021

**Rostov State University of Economics(RINH), Rostov-on-Don, Russia*

***Volgograd State Agrarian University, Volgograd, Russia*

The problem of analysis and forecasting of the stock index in the asset market is considered. The analysis of the stock index tracking the development of the industrial component of the stock market is carried out. The stationarity of the data was checked. A predictive mod-

el is implemented to predict the closing price of the Dow Jones index on the stock market using a linear regression model

Keywords: stock index, stock market, linear regression, Dow Jones index, statistical test hypotheses
Keywords: stock index, stock market, linear regression, Dow Jones index, statistical test hypotheses.

1. Введение

Индекс Доу-Джонса — это один из самых известных и старых индексов среди американских рыночных индексов. Его задача отслеживать развитие промышленной составляющей американских фондовых рынков [8]. Индекс охватывает 30 крупнейших компаний США. Среди них такие как: Microsoft Corp., Apple Inc., Coca-Cola Co.

Целью настоящей работы являлась реализация предсказательной модели для прогнозирования цены закрытия индекса Доу-Джонса на фондовом рынке, используя модель линейной регрессии [7].

2. Materials and methods

Данные взяты из открытых платформ Quandl и AlphaVantage, обеспечивающих доступ к более чем 9 миллионм бесплатных наборов данных [4-6].

На рис. 1 представлены данные, которые содержит набор данных по индексу Доу-Джонса:

	Open	High	Low	Close	Adj Close	Volume
Date						
2001-01-01	10790.919922	11028.000000	10468.040039	10887.360352	10887.360352	5825960000
2001-02-01	10884.820313	11035.139648	10294.009766	10495.280273	10495.280273	4164820000
2001-03-01	10493.250000	10859.500000	9106.540039	9878.780273	9878.780273	6086470000
2001-04-01	9877.160156	10906.410156	9375.719727	10734.969727	10734.969727	5241100000
2001-05-01	10734.049805	11350.049805	10673.219727	10911.940430	10911.940430	4518100000
...	...	...	...	...	...	...
2020-01-01	28638.970703	29373.619141	28169.529297	28256.029297	28256.029297	6170770000
2020-02-01	28319.650391	29568.570313	24681.009766	25409.359375	25409.359375	6991990000
2020-03-01	25590.509766	27102.339844	18213.650391	21917.160156	21917.160156	15521140000
2020-04-01	21227.380859	24764.769531	20735.019531	24345.720703	24345.720703	9795600000
2020-05-01	24120.779297	25758.789063	22789.619141	25383.109375	25383.109375	7940630000

Рис. 1. Данные по индексу Доу-Джонса

В течение каждого торгового дня (понедельник - пятница) цена акции изменяется и регистрируется в режиме реального времени. Пять значений показывают изменение цены за один день и являются ключевыми торговыми показателями:

Date: дата;

Open: цена открытия;

Close: цена закрытия;

High: самая высокая цена;

Low: самая низкая цена;

Volume: общее количество акций, проданных до закрытия рынка.

При создании модели будем опираться на метод, описанный в [2, р. 59-71].

Перечислим признаки, используемые при предсказании цены закрытия индекса Доу-Джонса:

- цена открытия;
- цена открытия за предыдущий день;
- цена закрытия за предыдущий день;
- самая высокая цена за предыдущий день;
- самая низкая цена за предыдущий день;
- объем продаж за предыдущий день;
- средняя цена закрытия за предыдущие пять дней;
- средняя цена закрытия за предыдущий месяц;
- средняя цена закрытия за предыдущий год;
- соотношение между стандартным отклонением объемов за
- итоговая доходность предыдущего дня;
- итоговая доходность предыдущей недели;
- скользящее среднее итоговой доходности за предыдущую неделю;
- скользящее среднее итоговой доходности за предыдущий месяц;
- скользящее среднее итоговой доходности за предыдущий год и т.п.

Для предсказания цен закрытия индекса Доу-Джонса на преобразованных данных воспользуемся стандартной функцией SGDRegressor из библиотеки sklearn. Для подбора параметров воспользуемся функцией GridSearchCV.

Полученный результат представлен на рис. 2.

MSE: 50415.891.

Точность на обучающем наборе: 0.96.

Точность на тестовом наборе: 0.92.

Для сравнения эффективности SGDRegressor используем RandomForestRegressor и SVR. Для подбора параметров функций воспользуемся GridSearchCV.

SGDRegressor {'alpha': 3e-05, 'eta0': 0.03}:

MSE: 29876.811.

Точность на обучающем наборе: 1.00.

Точность на тестовом наборе: 0.89.

RandomForestRegressor {'max_depth': 50, 'min_samples_split': 10}:

MSE: 46358.319.

Правильность на обучающем наборе: 1.00.

Правильность на тестовом наборе: 0.85.

SVR {'C': 10000, 'epsilon': 0.0001}:

MSE: 26895.198

Точность на обучающем наборе: 1.00

Точность на тестовом наборе: 0.91

Рис. 2. Сравнение цен

Рис. 3. Сравнение 3-х методов с истинным значением

Таким образом, точность метода SGDRegressor ниже SVR, но выше точности метода RandomForestRegressor.

Для увеличения точности предсказания нашей модели проверим являются ли наши данные стационарными, так как нестационарные данные могут привести к ухудшению работы регрессии.

Стационарный временной ряд – это временной ряд, статистические свойства которого, такие как среднее значение, дисперсия и автокорреляция, постоянны во времени.

Для проверки на стационарность используем расширенный тест Дики-Фуллера (ADF). ADF — это тип статистического теста, который определяет, присутствует ли единичный корень в данных временных рядов. Рассмотрим две гипотезы:

- нулевая гипотеза, H_0 : если не удалось отклонить, значит есть вероятность, что временной ряд является нестационарным;
- альтернативная гипотеза, H_1 : если H_0 отклонено, значит временной ряд является стационарным.

Чтобы принять или отклонить нулевую гипотезу пользуемся следующим критическим p -значением :

- $p > 0,05$: нельзя отвергнуть H_0 , значит данные имеют единичный корень и являются нестационарными;
- $p \leq 0,05$: отвергаем H_0 , следовательно, данные не имеют единичного корня и являются стационарными.

Вычисляем скользящее среднее и стандартное отклонение в переменные `df_mean` и `df_std`, соответственно, с периодом окна в один год:

```
df_settle = df['Close'].resample('MS').ffill().dropna()
df_rolling = df_settle.rolling(12)
df_mean = df_rolling.mean()
df_std = df_rolling.std()
```

Выведем графики скользящего среднего относительно исходного временного ряда (рис. 4) и скользящего стандартного отклонения отдельно (рис. 5).

Используя модуль `statsmodels`, выполняем тест ADF для нашего набора данных с помощью метода `adfuller()`:

```
from statsmodels.tsa.stattools import adfuller
result = adfuller(df_settle)
print('ADF statistic: ', result[0])
print('p-value:', result[1])
critical_values = result[4]
for key, value in critical_values.items():
    print('Critical value (%s): %.3f' % (key, value))
Получаем следующие значения:
ADF statistic: 0.21558996270503175
p-value: 0.9731225564902721
```

Critical value (1%): -3.459
Critical value (5%): -2.874
Critical value (10%): -2.574

Рис. 4. График скользящего среднего

Рис. 5. График скользящего стандартного отклонения

Статистическое p -значение теста ADF больше 0,05. Таким образом, нельзя отвергнуть H_0 о существовании единичного корня, следовательно, данные являются нестационарными.

На не стационарность временного ряда, влияет тренд или сезонность. Для того чтобы сделать ряд стационарным, нужно тренд и сезонность устранить. Для этого используют методы: детрендинг, дифференцирование и разложение. Полученные стационарные данные пригодны для статистического прогнозирования.

Произведем детрендирование. Детрендирование – процесс удаления линии тренда из нестационарных данных.

Заметим, что среднее значение и стандартное отклонение не имеют долгосрочной тенденции.

После выполнения теста ADF получаем:

ADF statistic: -14.198292224311526

p-value: 0.00000

Critical value (1%): -3.459

Critical value (5%): -2.874

Critical value (10%): -2.573

p -значение для этих данных составляет менее 0,05. Таким образом, можно отклонить H_0 , следовательно, данные являются стационарными.

Спрогнозируем будущее значение на полученных статистических данных, используя метод авторегрессионного интегрированного скользящего среднего (ARIMA). ARIMA - модель прогнозирования для стационарных временных рядов, основанная на линейной регрессии.

Рис. 6. Результат детрендирования

Поиск параметров для нашей модели будем производить, используя «поиск по сетке», также известный как метод оптимизации гиперпараметров. Подгоним модель ARIMA с помощью функции SARIMAX () модуля statsmodels. На каждой итерации функция возвращает объект класса MLEResults, который содержит атрибут aic для возврата значения AIC.

Модель с наименьшим значением AIC дает нам наиболее подходящую модель, которая определяет наши параметры p , d и q .

Модель сезонного компонента ARIMA (0,1,1,12) даст нам самое низкое значение AIC – 3289.336.

Построим график цен начиная с 2008 года (рис. 7).

ARIMA(1,1,1) RMSE Error: 366.8729197

ARIMA(1,1,1) MSE Error: 134595.7392

MSE: 125080.536

Сплошная линия показывает наблюдаемые значения, в то время, как пунктирные линии отображают пятилетние скользящие прогнозы. По мере того, как прогноз на ближайшие пять месяцев уходит в будущее, доверительный интервал расширяется, чтобы отразить потерю уверенности в прогнозе.

Выводы

Таким образом, реализована предсказательная модель для прогнозирования цены закрытия индекса Доу-Джонса на фондовом рынке, используя модель линейной регрессии.

В рамках поставленной цели в работе поставлены следующие задачи:

- произведено тестирование методов линейной регрессии на стандартном примере;
- разработан метод для предсказания индекса Доу-Джонса.

Рис. 7. Результат прогноза цены на индекс Доу-Джонса
Литература

1. Shalev-Schwartz Sh., Ben-David Sh. Machine Learning Ideas, 2019.
2. Muller, Guido. Introduction to Machine Learning with Python, 2017. (p. 59-71)
3. Documentation for the pandas library [Electronic resource] - URL: https://pandas.pydata.org/docs/user_guide/index.html
4. Documentation on Quandl [Electronic resource] - URL: <https://docs.quandl.com>
5. Weiming - Mastering Python for Finance Implement Advanced State-of-the-art Financial Statistical Applications Using Python-Packt Publishing, 2019
6. Documentation on AlphaVantage [Electronic resource] – URL: <https://www.alphavantage.co/documentation/>
7. Mundt, Philipp & Oh, Ilfan. Asymmetric competition, risk, and return distribution// Economics Letters, Elsevier, 2019. Vol. 179 (C). P. 29-32.
8. Gelman, Sergey & Kliger, Doron. The effect of time-induced stress on financial decision making in real markets: The case of traffic congestion// Journal of Economic Behavior & Organization, Elsevier, 2021. Vol. 185 (C). P. 814-841.

Поступила в редакцию

09.09.2021

Федорова Яна Владимировна – кандидат экономических наук, доцент Департамента Информационной безопасности, Финансовый Университет при Правительстве РФ, г. Москва, Россия.

Fedorova Yana V. – Candidate of Economic Sciences, Associate Professor of the Department of Information Security-News, Financial University under the Government of the Russian Federation, Moscow, Russia.

Рогачев Алексей Фруминович – доктор технических наук, профессор кафедры Математического моделирования и информатики, Волгоградский Государственный Аграрный Университет, г. Волгоград, Россия.

Rogachev Alexey F. – Doctor of Technical Sciences, Professor of the Department of Mathematical Modeling and Computer Science, Volgograd State Agrarian University, Volgograd, Russia.

Россия, 344002, г. Ростов-на-Дону, ул. Большая Садовая, 69
69, Bolshaya Sadovaya str., Rostov-on-Don, 344002, Russia

Россия, 400002, г. Волгоград, Университетский пр-т, 26
26, University Avenue, Volgograd, 400002, Russia

e-mail: fyv21@mail.ru

e-mail: rafr@mail.ru