



**Московский Государственный Университет имени
М. В. Ломоносова**

Научно-исследовательская работа

Кафедра: Финансов и кредита.

Тема: «Построение автоматизированных торговых алгоритмов и оценка их
эффективности»

Руководитель

_____ В. И. Байбаков

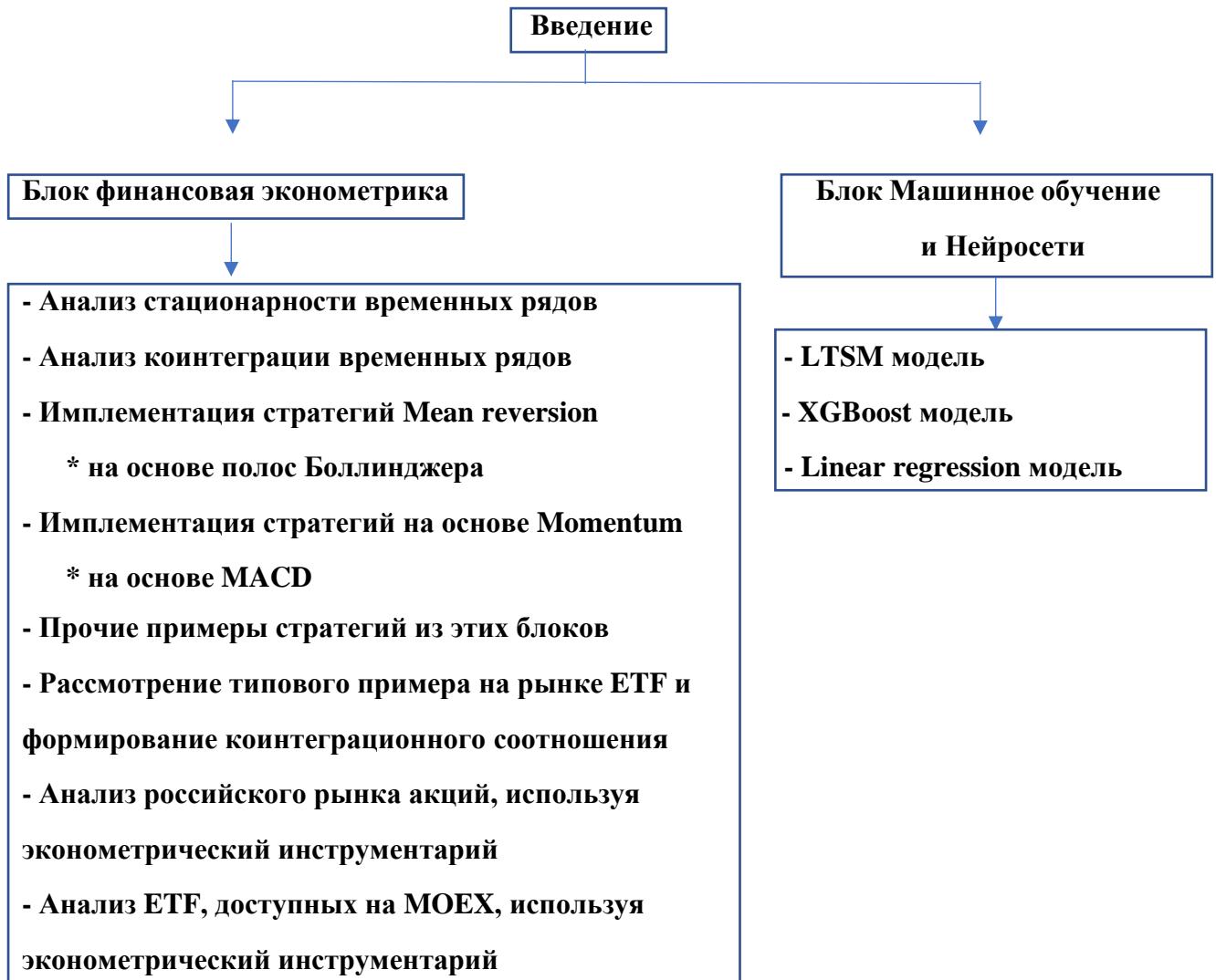
Выполнил студент группы э – 301

_____ С. Д. Петраков

Оглавление

План работы	3
Введение	4
Цель.....	4
Подмножества стратегий, в рамках которых я буду работать.....	4
Основа.....	5
Исследовательские вопросы.....	5
Задачи.	5
Глава 1. Эконометрический анализ.	6
Теоретическая база. Финансовая эконометрика.....	6
Глава 2. Имплементация стратегий блока mean reversion и momentum на Python.	12
1) Стратегия моментум, основанная на использовании технического показателя – средней взвешенной цены по объёму (VWAP Momentum trade).....	12
2) Стратегия возврата к среднему значению (Mean reversion) используя полосы Боллинджера (Bollinger Bands).	13
3) Стратегия моментум MACD (Moving average convergence indicator).....	21
Итоги соревнования MACD и Bollinger Bands на CBOM.	24
Глава 3. Машинное обучение и нейросети.	25
Стратегия на основе алгоритма XGBoost.....	25
XGBoost Regressor.....	25
Настройка параметров модели XGBoost.....	29
XGBoost Classifier.....	39
Линейная регрессия (linear regression).....	48
Нейросеть. LSTM model.....	50
Настройка параметров модели.....	50
Итоги.....	57
Приложения.	59
Библиографический список:.....	63

План работы.



Введение.

В современном мире торговли на мировых финансовых рынках всё больше и больше операции выполняют торговые роботы. Уже в 2010 доля, занимаемая алгоритмической торговлей, превышала 60% от всей торговли акциями, для фьючерсов эти показатели были выше 40%. 30% опционов торговалось алгоритмически. Что очень важно тенденция монотонно растёт, причём по всему миру¹.

Сейчас трейдер не может справиться с конкуренцией, которую оказывают торговые роботы, если работает старыми методами. Также с увеличением доли алгоритмической торговли на мировых финансовых рынках повышается качество торговли. Это выражается в ряде аспектов: уменьшаются спреды, сокращается неблагоприятный отбор, а также увеличивается ликвидность (Hendershott T., Jones C. M., Menkeld A. J.)², (Hendershott T., Riordan R.)³. Поэтому внедрение технических компонентов торговли в совокупности приводят не только к большей скорости операций, их точности, но и к большей эффективности функционирования рынка в целом. Необходимо осваивать достижения техники, если стоит цель успешной торговли. Более того, в современных условиях появляются программы позволяющие использовать современные инструменты для трейдинга, например патент (Kemp I. I. G. A., Ebersole J., Kline R. J.)⁴

В рамках данной работы мною будут рассмотрены разные стратегии и подходы, которые применяются в рамках алгоритмической торговли. Каждую из них я буду имплементировать в Python или R.

Цель.

Моей основной целью в рамках данной работы является выделение торговых алгоритмов из ряда подмножеств и оптимизация их параметров, а также сравнение их результатов между собой для рынков акций и ETF.

Подмножества стратегий, в рамках которых я буду работать.

1. Стратегии средней реверсии (Mean reversion).
2. Стратегии, построенных на гипотезе о сохранении тренда (Momentum).
3. Стратегии на базе машинного обучения и нейросетей.

¹ <https://www.quora.com/How-much-trading-in-the-stock-market-is-algorithmic-trading-and-how-much-is-non-algorithmic>

² Hendershott T., Jones C. M., Menkeld A. J. Does algorithmic trading improve liquidity? //The Journal of finance. – 2011. – Т. 66. – №. 1. – С. 1-33.

³ Hendershott T., Riordan R. Algorithmic trading and the market for liquidity //Journal of Financial and Quantitative Analysis. – 2013. – Т. 48. – №. 4. – С. 1001-1024.

⁴ Kemp I. I. G. A., Ebersole J., Kline R. J. System and method for automated trading : пат. 9805417 США. – 2017.

Основа.

Моей теоретической основой будут являться методы финансовой математики, финансовой эконометрики, машинное обучение и методы глубокого обучения, а также их комбинации.

Исследовательские вопросы.

Существует ли значимое воздействие свойств временного ряда на результативность стратегий из разного блока, иными словами, существуют ли отдельные свойства временного ряда, которые могут говорить в пользу применения стратегий возврата к среднему значению, или стратегий Momentum? Можно ли говорить об устойчивости результатов стратегий? Верно ли утверждение, что стратегии из блока машинного обучения и нейросетей систематически обыгрывают Mean reversion и Momentum? Насколько велико преимущество рассмотренных торговых алгоритмов по сравнению с базовой стратегией «купи и держи»?

Задачи.

Для того, чтобы справиться с поставленной целью я поставлю ряд задач, которые необходимо осуществить:

- 1) Проведение эконометрического анализа временных рядов с целью выдвижения гипотезы о наиболее оптимальном подмножестве торговых стратегий.
- 2) Осуществить имплементацию стратегий на временных рядах акций и ETF.
- 3) Сравнить результаты стратегий между собой по метрике доходности, в ряде случаев будет сравниваться коэффициент Шарпа. Расчёт дополнительных характеристик результативности стратегий. Оптимизация различных параметров стратегий.
- 4) Сравнение результатов традиционных стратегий (Mean reversion, Momentum) с результатами более современных стратегий из подмножества машинного обучения и нейросетей.
- 5) Сравнение результатов моделей с дефолтной стратегией купи и держи (buy and hold)
- 6) Формирование выводов.

Глава 1. Эконометрический анализ.

Для решения первой задачи, необходимо определить, какие свойства временного ряда могут быть нам полезны для реализации конкретного класса стратегий.

Теоретическая база. Финансовая эконометрика.

- Понятия: временной ряд, стационарность⁵, интеграция, коинтеграция⁶.

Тесты:

- Расширенный тест Дики – Фуллера⁷ (Augmented Dickey – Fuller test).

Применяется для проверки стационарности временного ряда.

- Cadf тест (Engle – Granger test)⁸

Тест для анализа коинтеграции временных рядов и поиска коэффициента хеджирования (hedge ratio).

- Тест Йохансена (Johansen test)⁹

Продвинутый тест для анализа коинтеграции временных рядов.

- Показатель Хёрста (Hurst exponent)¹⁰.

Величина, которая имеет значения от 0 до 1. Смысл данного показателя заключается в том, что он уменьшается, если растёт в случае, когда лаг между парами одинаковых значений временного ряда растёт. Согласно теории, если показатель Хёрста равен 0.5, то временной ряд описывается броуновским движением, если $1 > H > 0.5$, то движение временного ряда называется персистентным (persistent), или другими словами, сохраняющим тенденцию. Если же верно $0.5 > H > 0$, тогда процесс характеризуется как антиперсистентный (anti-persistent), что можно объяснить как стремление любой тенденции смениться на противоположную.

В рамках данного анализа мы можем:

- 1) Проверить, является ли временной ряд стационарным. Если временной ряд не является интегрированным порядка 0, то можно перейти к разностям и изучать их

⁵ Gagniuc, P. A. (2017). *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons.

⁶ Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1), 121-130.

⁷ Fuller, W. A. (2009). *Introduction to statistical time series* (Vol. 428). John Wiley & Sons.

⁸ Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

⁹ Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3), 231-254.

¹⁰ Carbone, Anna, Giuliano Castelli, and H. Eugene Stanley. "Time-dependent Hurst exponent in financial time series." *Physica A: Statistical Mechanics and its Applications* 344.1-2 (2004): 267-271.

поведение. Как правило, так и делают. Часто получается, что первые разности являются стационарным временным рядом, а они могут быть описаны моделями ARIMA, ARCH, GARCH. Однако в рамках своей работы я буду использовать аналог коинтеграции для получения стационарного временного ряда. Разница лишь в том, что коинтеграция подразумевает наличие причинно – следственных связей, в то время как меня будет интересовать лишь взаимосвязь. Благодаря коинтеграции временных рядов я буду получать стационарный временной ряд.

- 2) Получив стационарный временной ряд существует эконометрическое основание для стратегий из класса возвращения к среднему значению, поскольку постоянство среднего значения, как одно из необходимых условий слабой стационарности даёт понимание того, к чему ряд будет возвращаться. В таком случае появляется гипотеза, что для таких коинтегрированных временных рядов подмножество mean reverting стратегий должны показывать лучшие результаты (например, доходность), чем стратегии из блока momentum или даже современные подходы машинного обучения.
- 3) Основным инструментарием, которым я буду пользоваться для анализа коинтеграции и поиска стационарных временных рядов - это расширенный тест Дики-Фуллера, тест Энгла-Грэнджера и тест Йохансена.
- 4) Очень большим подспорьем является показатель Хёрста в определении того, у какого блока стратегий скорее будет преимущество. Если $H > 0.5$, то скорее всего лучше себя покажут стратегии momentum, если же $H < 0.5$, то mean reverting стратегии. Помимо того, поскольку это не просто булевская переменная, то можно определить степень склонности к определённому классу и дополнительно проконтролировать на это.
- 5) В случае стационарности временного ряда есть возможность получить показатель периода среднего возврата к среднему значению, рассчитанное по формуле $half\ life = -\frac{\ln(2)}{\theta}$, где θ – коэффициент при лагированном значении зависимой переменной оценённой по МНК. Это значение позволяет ответить на такой нетривиальный вопрос как «Через сколько в среднем показатель вернётся к своему среднему значению»

Для решения задачи анализа временного ряда для первоначальной предрасположенности ряда к той или иной стратегии используется следующий код на Python.

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.formula.api as sm
import statsmodels.tsa.stattools as ts
import statsmodels.tsa.vector_ar.vecm as vm

df=pd.read_csv('data\inputData_EWA_EWC_IGE.csv')
df['Date']=pd.to_datetime(df['Date'], format='%Y%m%d').dt.date # remove HH:MM:SS
df.set_index('Date', inplace=True)

df.plot()
df.plot.scatter(x='EWA', y='EWC')
plt.xlabel('EWA share price')
plt.ylabel('EWC share price')
plt.xlabel('April 4,2006, to April 9,2012')
plt.ylabel('Share price $')
plt.show()
results=sm.ols(formula="EWC ~ EWA", data=df[['EWA', 'EWC']]).fit()
print(results.params)
hedgeRatio=results.params[1]
print('hedgeRatio=%f' % hedgeRatio)

```

Рис. 1. Импорт необходимых библиотек. Загрузка данных (ETF: EWA, EWC, IGE)¹¹ и вычисление коэффициента хеджирования по МНК (коинтеграционное соотношение).

```

# cadf test
coint_t, pvalue, crit_value=ts.coint(df['EWA'], df['EWC'])
print('t-statistic=%f' % coint_t)
print('pvalue=%f' % pvalue)
print(crit_value)

# Johansen test
result=vm.coint_johansen(df[['EWA', 'EWC']].values, det_order=0, k_ar_diff=1)
print('Johansen test')
print(result.lr1)
print(result.cvt)
print(result.lr2)
print(result.cvm)

# Add IGE for Johansen test
result=vm.coint_johansen(df.values, det_order=0, k_ar_diff=1)
print('Add IGE for Johansen test')
print(result.lr1)
print(result.cvt)
print(result.lr2)
print(result.cvm)

print('eigenvalues')
print(result.eig) # eigenvalues
print('eigenvectors')
print(result.evec) # eigenvectors

```

¹¹ Конкретный код сделан для примера данных ETF, в общем случае можно анализировать динамику любого временного ряда.

Рис. 2. Имплементация CADF и теста Йохансена.

Поскольку интересующей меня областью применения алгоритмов является ранок акций и ETF, то и эконометрический анализ соответственно производится именно этих временных рядов.

На практике выявилось интересная особенность: стационарных рядов довольно много получилось именно из ETF, а не из акций¹².

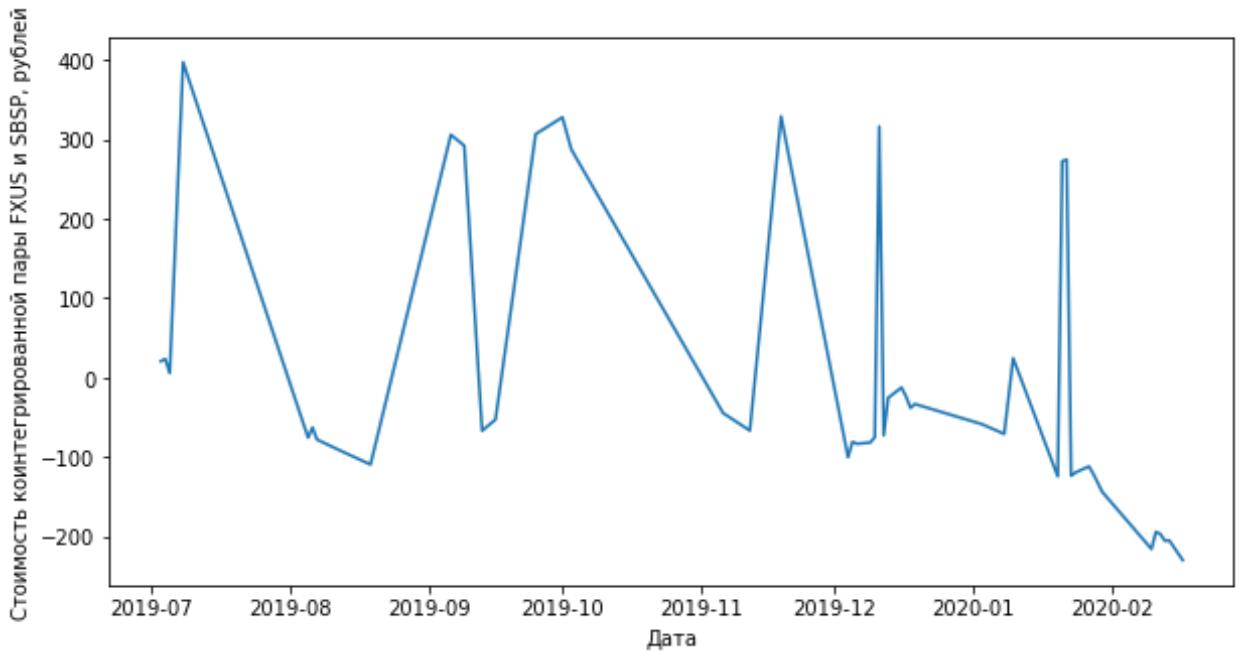


Рис. 3. Поведение пары ETF: FXUS и SBSP.

Для данной пары $\text{halflife} = 1.14$, $H = 0.098$, $P\text{-value}$ (для экспоненты Хёрста) = 0.41

$P\text{-value}$ для теста Энгла-Гренджера близко к нулю при любом разумном уровне значимости, значит ряды кointегрированы.

¹² Речь идёт о стационарности без перехода к разностям.

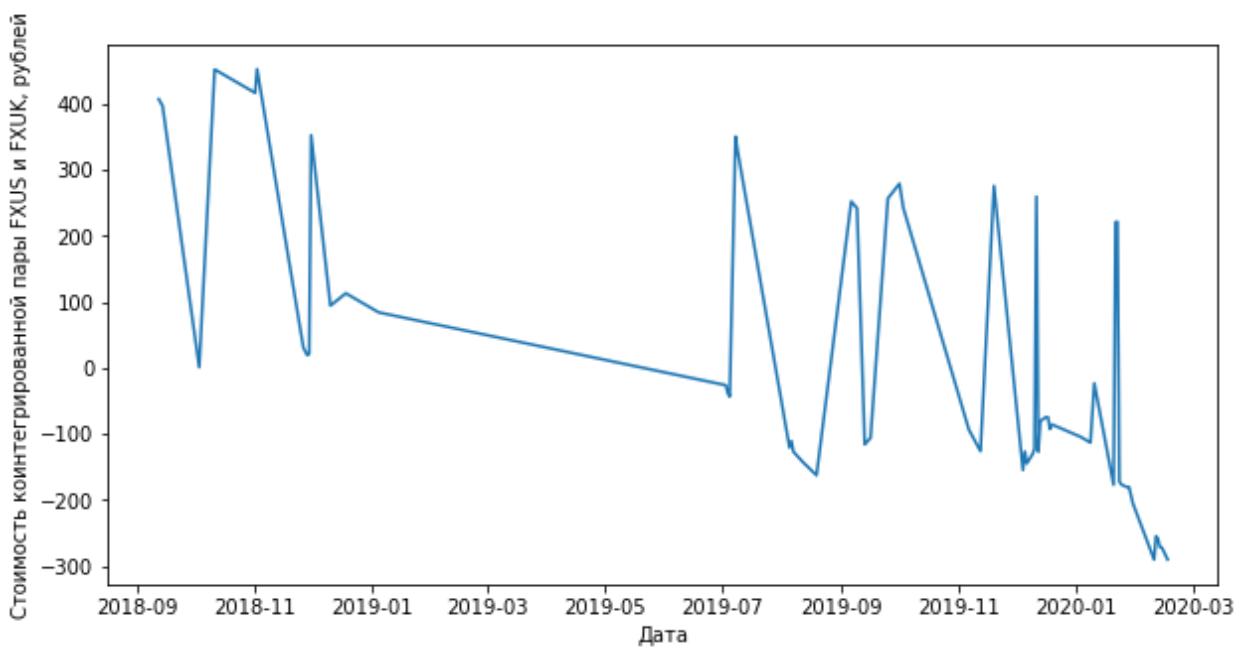


Рис. 4. Поведение пары ETF: FXUK и FXUS.

Для данной пары $\text{halflife} = 1.40$, $H = 0.054$, $P\text{-value}$ (для экспоненты Хёрста) = 0.45

$P\text{-value}$ для теста Энгла-Гренджера близко к нулю при любом разумном уровне значимости, значит ряды коинтегрированы.

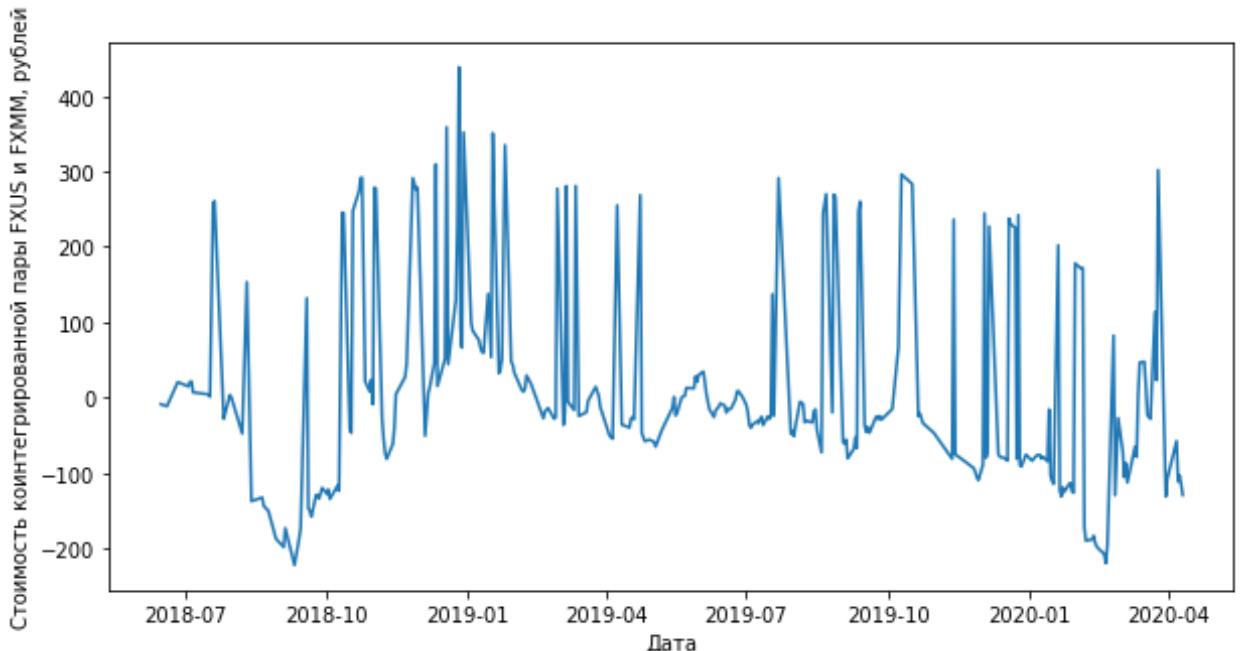


Рис. 5. Поведение пары ETF: FXUS и FXMM.

Для данной пары $\text{halflife} = 1.22$, $H = 0.058$, $P\text{-value}$ (для экспоненты Хёрста) = 0.0

$P\text{-value}$ для теста Энгла-Гренджера близко к нулю при любом разумном уровне значимости, значит ряды коинтегрированы.

Как можно заметить, визуально гипотеза об использовании стратегии возврата к среднему значению подтверждается на графике, также в добавок к этому это подтверждается тестом Энгла-Гренжера и экспонентой Хёрста. Можно предположить, что для стратегий Momentum экспонента Хёрста будет играть аналогичную роль, это моя гипотеза.

В качестве примера анализа временных рядов акций мною были взяты данные с московской фондовой биржи.

	Hurst exponent	P-value
AFLT	0.593384	5.965600e-277
ALRS	0.453789	2.104739e-250
CHMF	0.437660	1.286855e-210
GAZP	0.412806	0.000000e+00
GMKN	0.526297	5.636702e-275
LKOH	0.486462	1.347549e-305
MGNT	0.515925	3.826590e-253
MTSS	0.385923	3.219767e-157
СВОМ	0.384622	4.370995e-128

Рис. 6. Значение экспоненты Хёрста и её p-значения для разных компаний, представленных тикерами.

Заметно, что значение в основном не сильно отклоняется от показателя 0.5, что говорит о схожести данных с белым шумом, тем не менее можно утверждать, что гипотезу о лучшей работе стратегий momentum для ценных бумаг с экспонентой Хёрста больше 0.5, и соответственно стратегий mean reverting для ценных бумаг с экспонентой Хёрста меньше 0.5 можно проверить на практике.

Более подробный анализ выявил сложности в обнаружении коинтегрированных временных рядов между представленными компаниями. Для поиска я производил попарное сравнение отдельных временных рядов.

В качестве результата проведённого анализа я могу сказать, что на рынке ETF на московской фондовой бирже оказалось проще найти коинтегрированные временные ряды, чем на рынке акций. Это является гипотезой в пользу того, что возможно на блоке коинтегрированных временных рядов ETF стратегия Mean Reversion будет показывать результат лучше, чем другие стратегии.

Для проверки этого утверждения, а также анализа разных стратегий и их сравнения перейду к следующей главе.

Глава 2. Имплементация стратегий блока `mean reversion` и `momentum` на Python.

Были имплементированы следующие стратегии:

1) Стратегия моментум, основанная на использовании технического показателя – средней взвешенной цены по объёму (VWAP Momentum trade).

Смысл данной стратегии заключается в том, что VWAP – это volume weighted average price, то есть цена, взвешенная объёмом. Данный технический показатель зачастую используется для определения тренда. Основная идея - если цена выше VWAP, то тренд – положительный, а если ниже, то отрицательный.

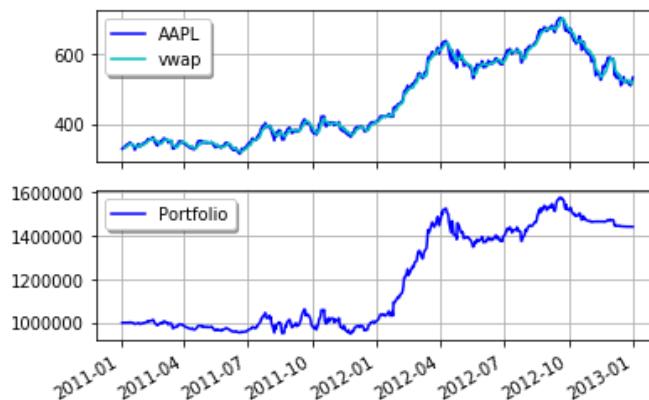


Рис. 7. Результаты имплементации стратегии из блока `momentum – VWAP` на базовом примере – акциях Apple. На верхнем графике – динамика цен акции и показателя VWAP, на нижнем изменение стоимость портфеля, который изначально насчитывал \$1 млн.

Коэффициент Шарпа для данной стратегии составил 0.89 (не очень высокий показатель)

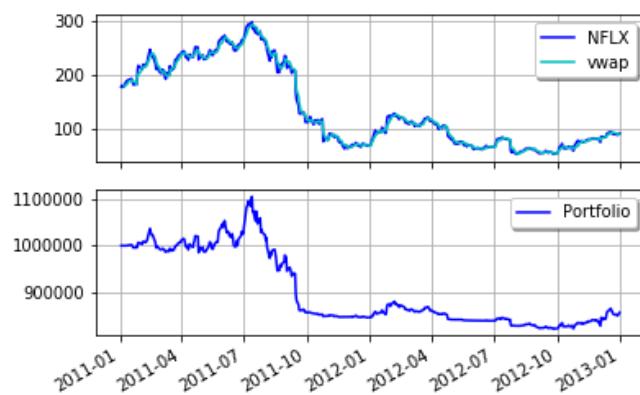


Рис. 8. Результаты имплементации стратегии из блока `momentum – VWAP` на новом примере – акциях Netflix. На верхнем графике – динамика цен акции и показателя VWAP, на нижнем изменение стоимость портфеля, который изначально насчитывал \$1 млн. Коэффициент Шарпа для данной стратегии составил -1.22 (отрицательный показатель, свидетельство убыточности стратегии).

2) Стратегия возврата к среднему значению (Mean reversion) используя полосы Боллинджера (Bollinger Bands).

Торговля по принципу средней реверсии предполагает, что цены вернутся к среднему значению. Таким образом, всякий раз, когда цены слишком далеки от среднего значения, мы будем занимать позицию (длинную или короткую) и закрывать позицию, когда цена возвращается к среднему значению. Полосами Боллинджера называются следующие три объекта: центральная полоса - это скользящее среднее за несколько последних торговых дней. Верхняя полоса - это сумма произведения стандартного отклонения за несколько последних торговых дней с коэффициентом, который определяется заранее, и средней полосы. Нижняя полоса, соответственно получается за счёт вычитания из средней полосы стандартного отклонения за несколько последних торговых дней, домноженное на коэффициент. Как правило коэффициенты для верхней и нижней полосы совпадают.

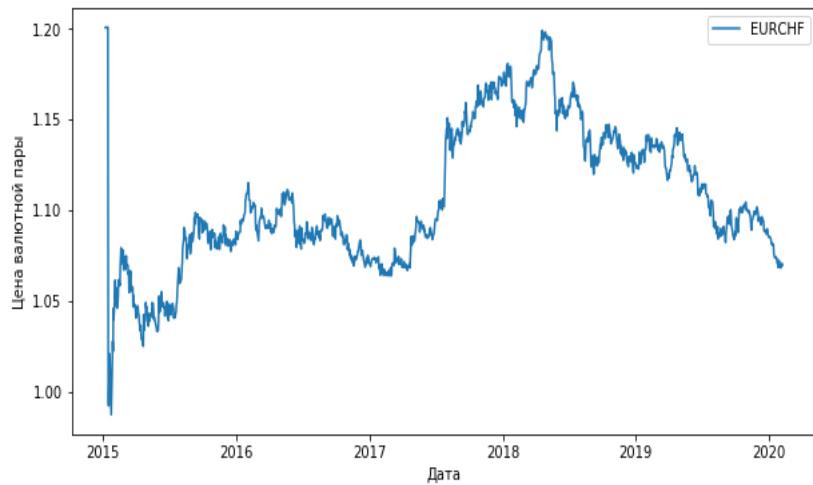


Рис. 9. Динамика валютной пары EURCHF (Евро - Швейцарский Франк).

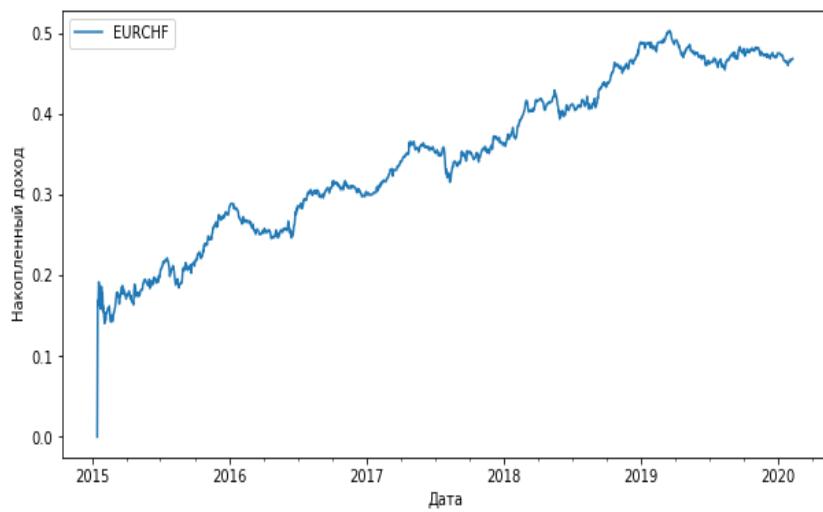


Рис. 10. Накопленный доход стратегии Bollinger Bands на примере валютной пары EURCHF.

Резюмируя подготовку к ответу на 2 задачу, скажу:

Существует много подходов к алгоритмической торговле. Одним из них является финансовая эконометрика.

Рассматривая временные ряды, можно с помощью специальных тестов обнаружить свойства ряда или комбинаций рядов, на основе которых можно составить торговую стратегию. Например, группа моделей, которые основаны на свойстве возврата к среднему значению (Mean reversion).

Рассмотрим стратегию Bollinger Bands в применении к коинтегрированной паре ETF: FXUK и SBSP. Тогда получим:

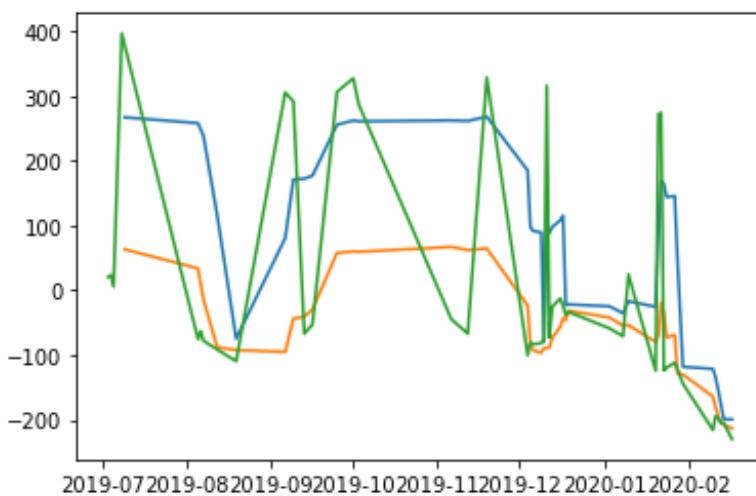


Рис. 11. Две полосы Боллинджера и коинтегрированный временной ряд FXUK и SBSP.

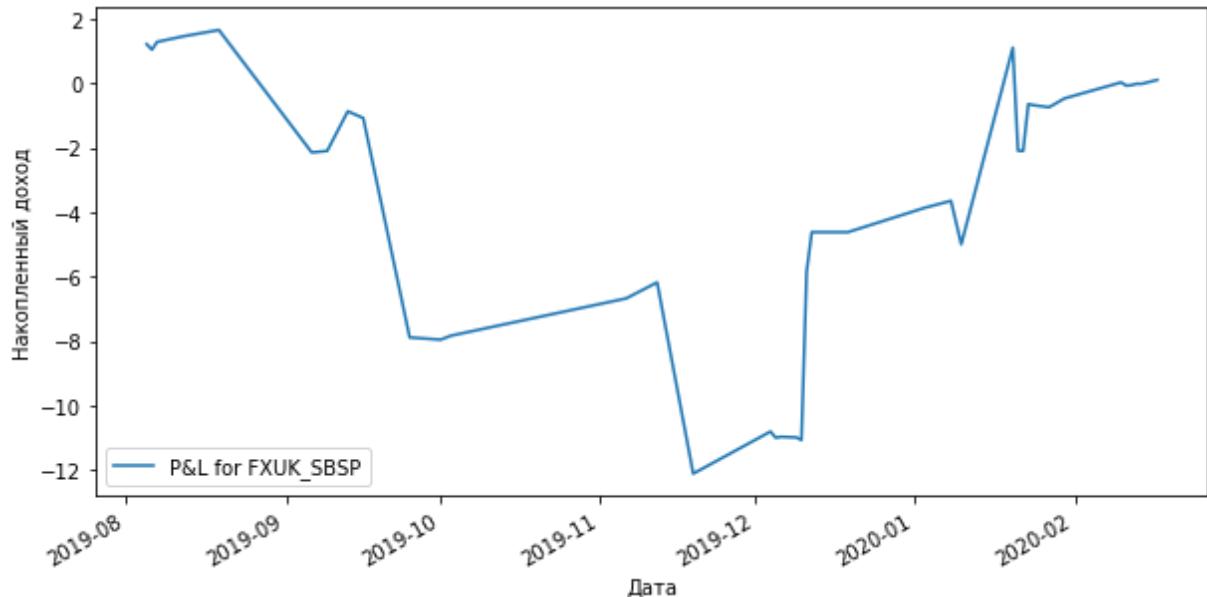


Рис. 12. P&L для стратегии Bollinger Bands для коинтегрированной пары FXUK и SBSP. Хотя накопленный P&L в итоге и составил 9.24%, говорить об удачности данной стратегии не приходится так как в течение практически всего периода она была убыточна.

Проводя подсчёт P&L стратегий, в данном случае это Bollinger Bands, ранее это была стратегия моментум – VWAP, я готовлю решение для своей третьей задачи, то есть для сравнения результатов.

Также, чтобы решить часть оптимизации параметров стратегии, входящих в третью задачу проведу следующий блок.

Для проверки модели будем менять значения параметров модели:

Для данной стратегии мы можем изменить значение окна (lookback), в рамках которого берётся стандартное отклонение и среднее.

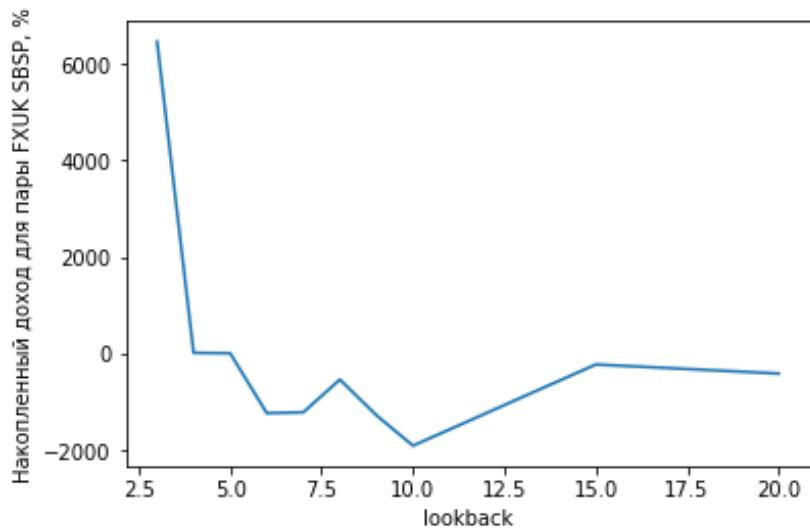


Рис. 13. Изменения значений параметра окна показали следующие результаты.
Параметр, отвечающий за ширину полосы Боллинджера полагаем равным 0.5.

В целом, можно сказать, что стратегия убыточна, случай, когда $\text{lookback} = 3$ стратегия ловит шок, который в самом начале очень сильно увеличивает доходность. Произведём оптимизацию параметра, соответствующего широте полосы Боллинджера¹³.

¹³ Это множитель для стандартного отклонения, произведение множителя и самого стандартного отклонения и формирует полосу (верхнюю или нижнюю соответственно).

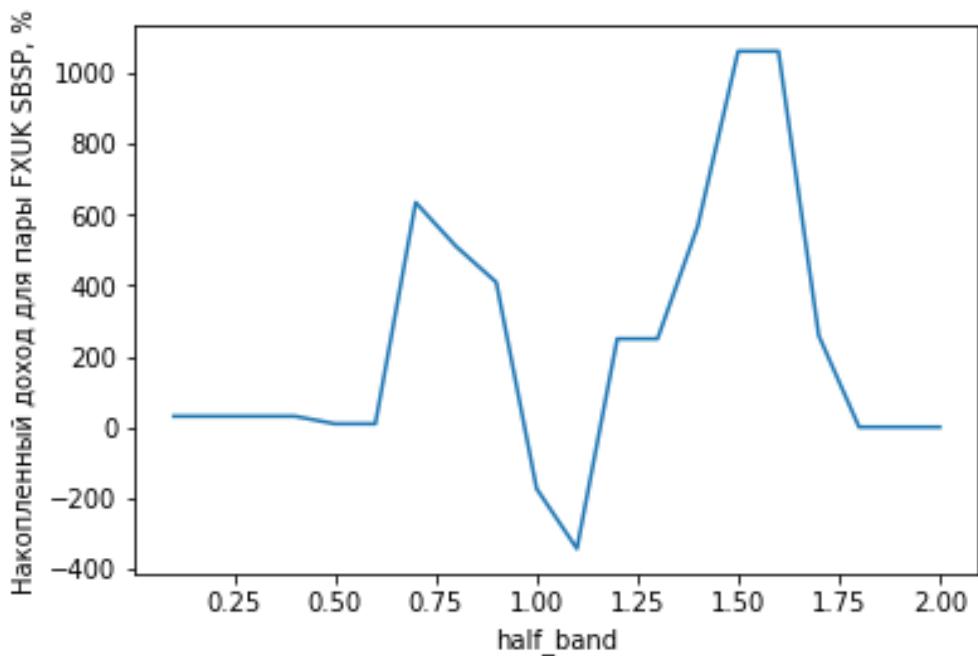


Рис. 14. Меняя параметр, который отвечает за ширину полосы Боллинджера получаем следующие значения P&L, полагаем, что *lookback* = 5.

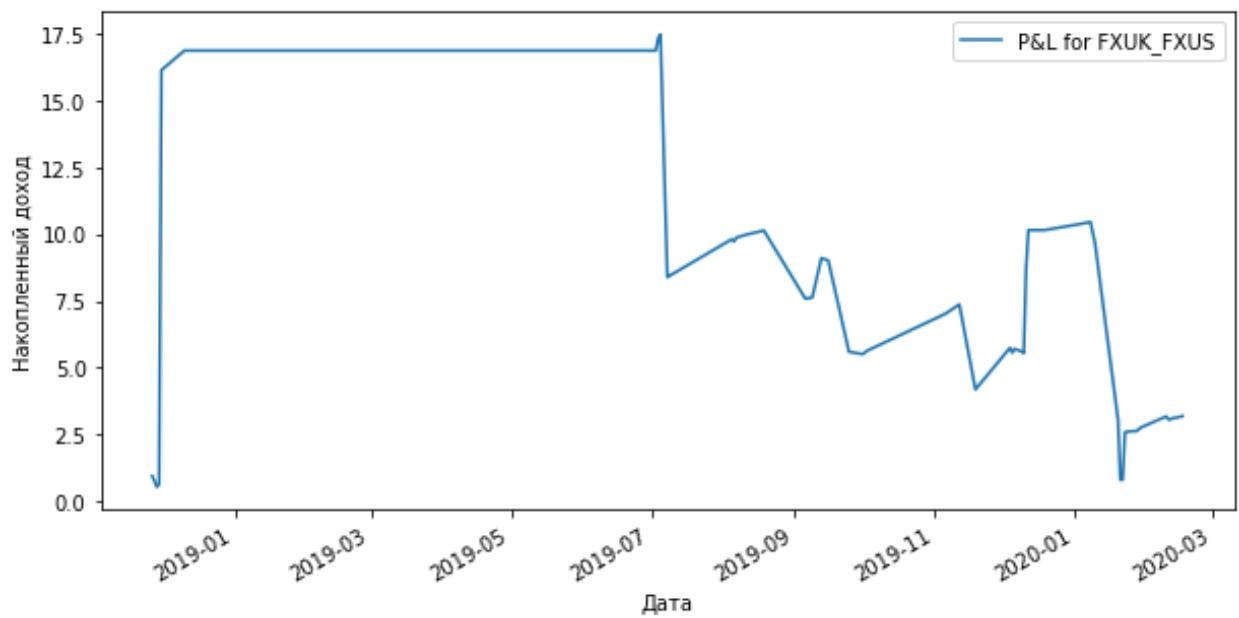


Рис. 15. Динамика P&L для пары ETF FXUK и FXUS. Итоговый результат составляет 317,89%.

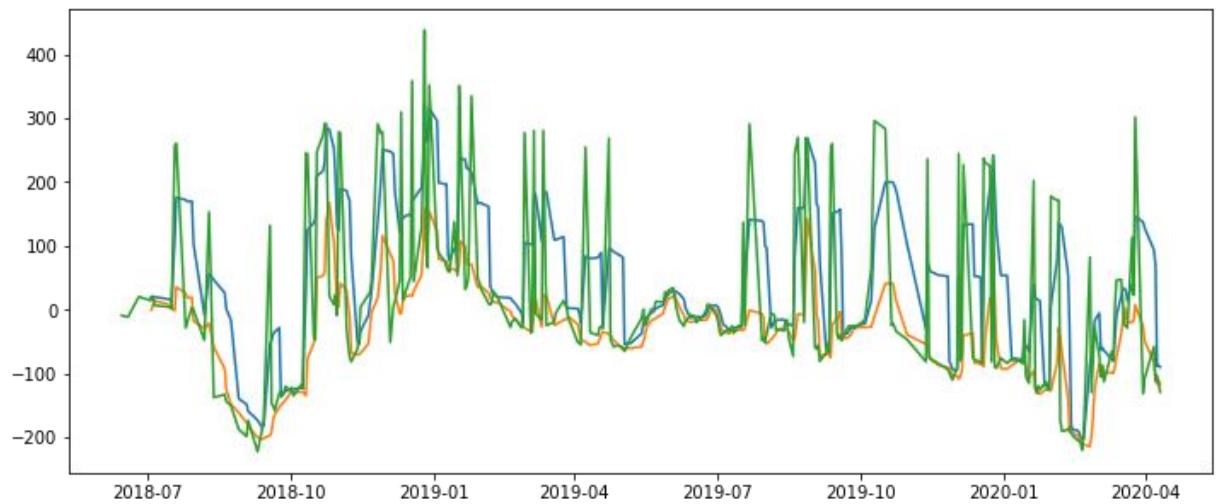


Рис. 16. Полосы Боллинджера для пары ETF FXMM и FXUS.

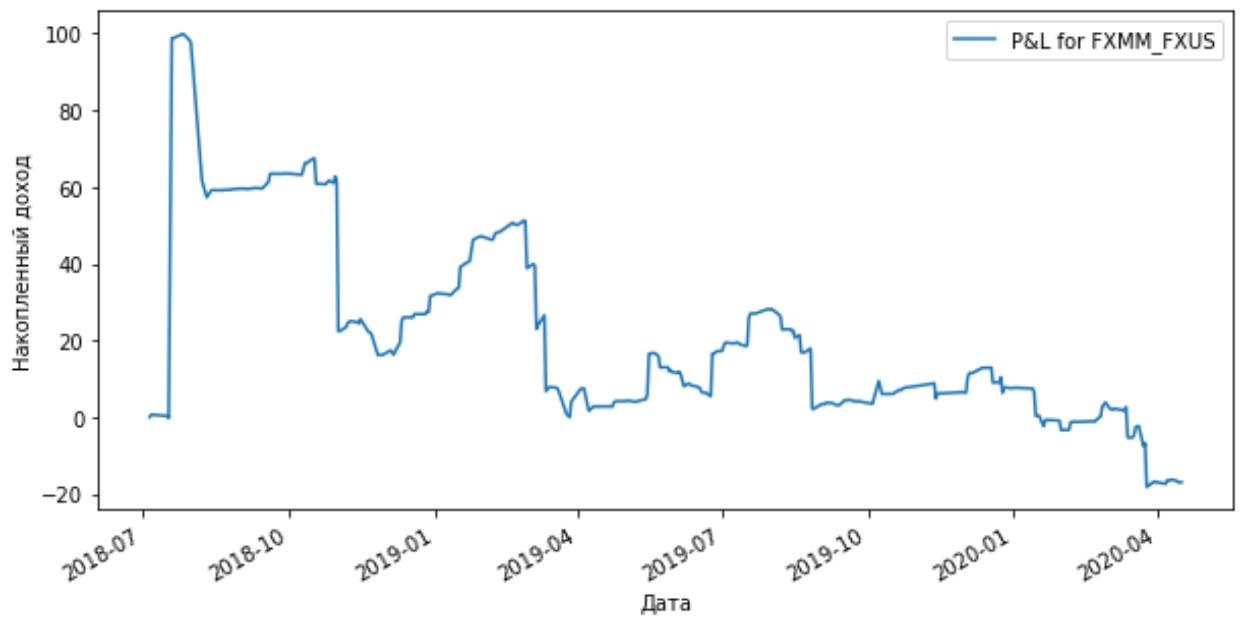


Рис. 17. Динамика накопленного P&L для пары ETF FXMM и FXUS. Итоговый результат составил -1700,2%.

Ключевыми терминами в рамках подхода возврата к среднему значению являются: стационарность ряда, коинтеграция, тесты, которые проводят анализ стационарности (Расширенный тест Дики-Фуллера (Augmented Dickey-Fuller test) или просто ADF test, тест Дики-Фуллера на коинтеграцию CADF test, тест Йохансена (Johansen test)), выполнение расчёта дополнительных значений, которые помогают более глубоко проанализировать свойства ряда, такие как экспонента Хёрста (Hurst exponent). В частности, значение экспоненты Хёрста помогает дать ответ на вопрос, к какому свойству ряд относится: группа рядов со свойством возвращения к среднему, белый шум или группа рядов со свойством трендовости.

Согласно анализу пар ETF, которые построены таким образом, чтобы их линейная комбинация давала стационарный ряд, то есть добивались коинтеграции¹⁴, можно сказать что итоговый P&L был весьма нестабилен, и показывал при этом весьма разные результаты, то есть как очень неплохие, так и очень плохие. Можно сделать вывод о неустойчивости результатов модели Bollinger Bands для групп ETF, доступных на МОЕХ.

Однако, стоит проверить, как работает данная стратегия на рынке акций. Выстроим стратегию на базе акций Московского кредитного банка.

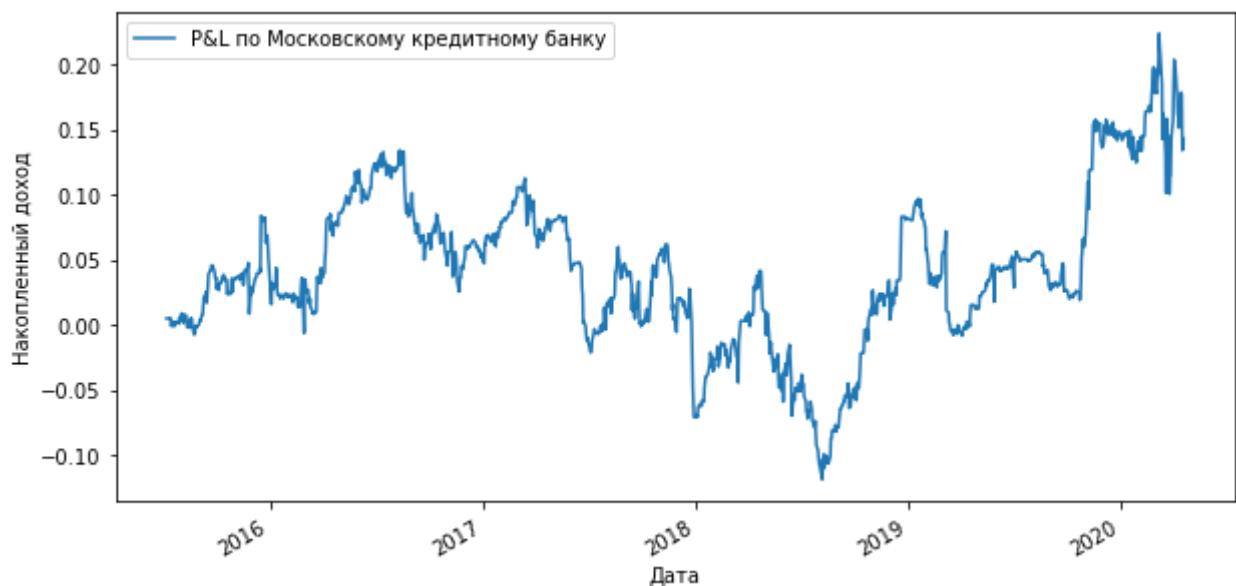


Рис. 18. Накопленный P&L стратегии Bollinger Bands на базе акций Московского кредитного банка (СБОМ). Среднегодовая доходность составляет 2.1%. Итоговый накопленный доход равен 10.5%.

Результат стратегии Mean Reversion Bollinger Bands¹⁵:

Cumulative returns	10.5%
Annual volatility	13.4%
Sharpe ratio	0.22
Calmar ratio	0.09
Stability	0.01
Max drawdown	-23.4%
Omega ratio	1.05
Sortino ratio	0.31
Skew	-0.47
Kurtosis	8.72
Tail ratio	1.05
Daily value at risk	-1.7%
Worst drawdown periods	
Net drawdown in %	
Peak date	
Valley date	
Recovery date	
Duration	
0	
23.44	
2016-08-11	
2018-08-06	
2020-02-12	
915	
1	
12.15	
2020-03-06	
2020-03-24	
NaN	
2	
8.82	
2015-12-18	
2016-02-29	
2016-05-05	
100	
3	
3.87	
2015-11-25	
2015-11-27	
2015-12-17	
17	
4	
2.50	
2016-06-02	
2016-06-07	
2016-06-27	
18	
Stress Events	
mean	
min	
max	
Fall2015	
0.10%	
-0.82%	
0.96%	
New Normal	
0.01%	
-6.09%	
5.71%	

¹⁴ Имеется в виду коинтеграция в слабом смысле, то есть взаимозависимость, без наличия утверждений о причинно следственных связях.

¹⁵ Данные статистики были подсчитаны с помощью пакета pyfolio в Python.

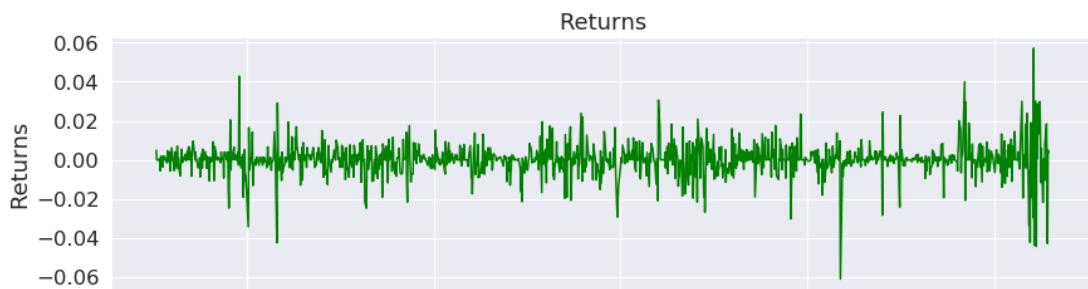
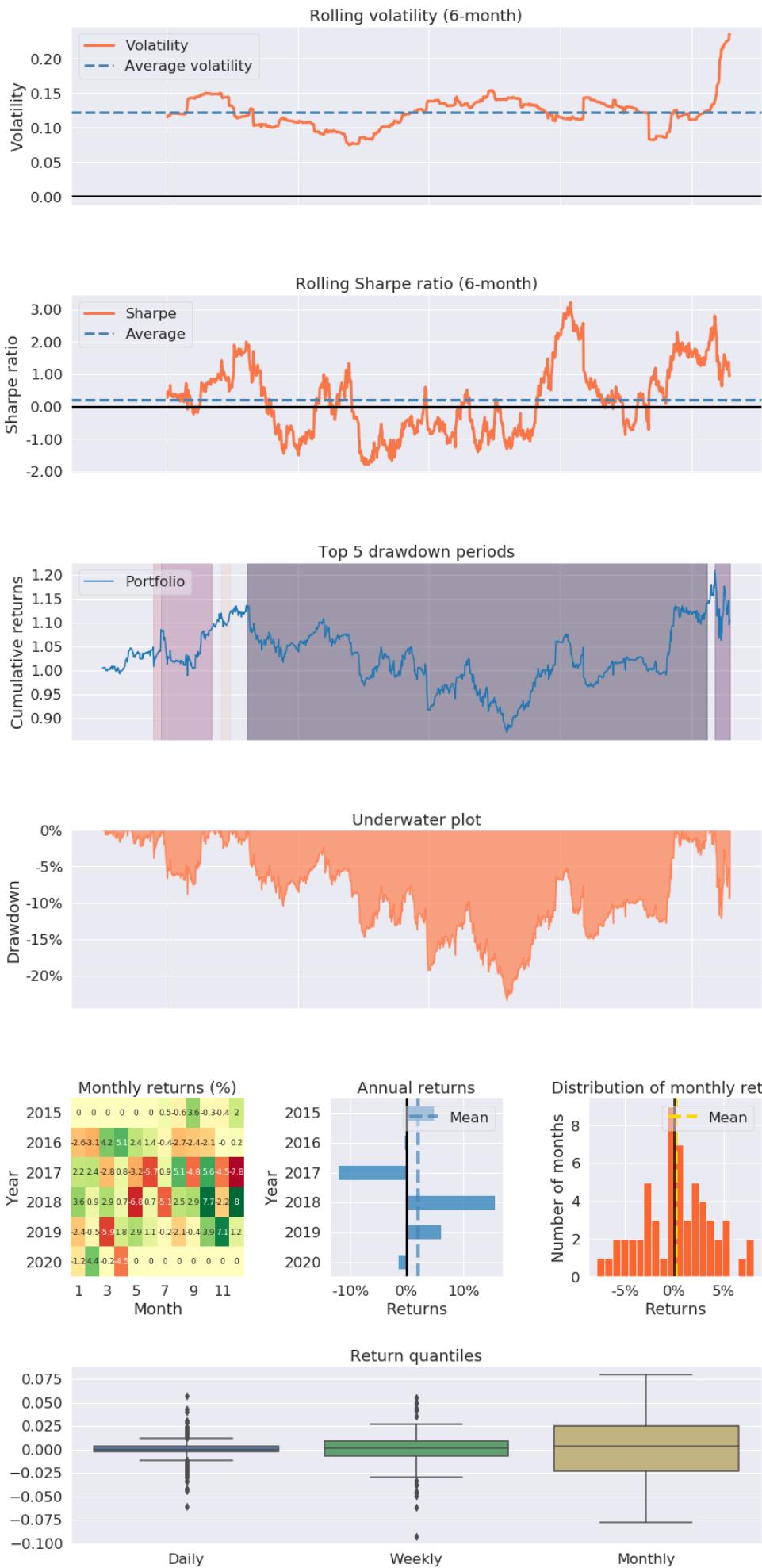


Рис. 19. Графическая иллюстрация статистик по стратегии ВВ на базе СВОМ.



3) Стратегия моментум MACD (Moving average convergence indicator).

Теперь стоит рассмотреть стратегию из блока моментум на базе той же акции московского кредитного банка (СВОМ), в частности с целью реализации 3 и 4 задачи. Стратегия MACD выстроена на базе одного из самых известных трендовых индикаторов (trend following momentum indicator). Чтобы результаты стратегии были сопоставимы я взял тот же временной промежуток.

Отмечу, что экспонента Хёрста (0.36) говорит больше в пользу проведения стратегий из блока Moving Average. Тем не менее стратегия показала неплохие результаты.

Результат стратегии Momentum MACD:

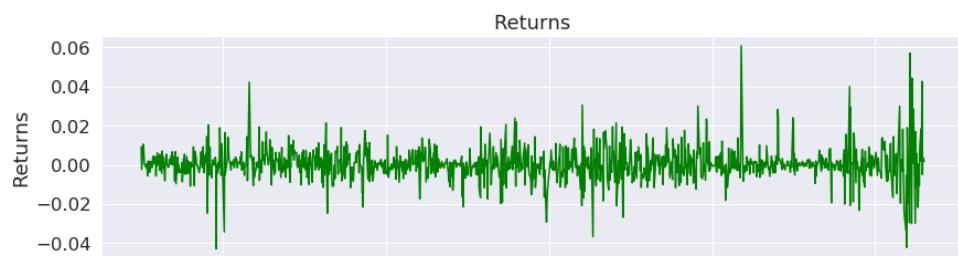
Annual return	17.1%
Cumulative returns	113.2%
Annual volatility	13.9%
Sharpe ratio	1.21
Calmar ratio	0.85
Stability	0.79
Max drawdown	-20.0%
Omega ratio	1.26
Sortino ratio	1.89
Skew	NaN
Kurtosis	NaN
Tail ratio	1.20
Daily value at risk	-1.7%

Рис. 20. Основные метрики качества для алгоритмической стратегии MACD.

Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
0	20.03	2017-05-17	2018-04-09	2018-12-20	417
1	14.56	2020-02-25	2020-03-18	NaT	NaN
2	5.96	2019-11-15	2020-01-15	2020-01-23	50
3	5.70	2016-08-19	2016-09-08	2016-10-10	37
4	5.43	2019-01-24	2019-02-12	2019-03-07	31

Stress Events	mean	min	max
Fall2015	0.04%	-0.96%	0.70%
New Normal	0.07%	-4.29%	6.09%

Рис. 21. Показатели просадки стратегии и их анализ.



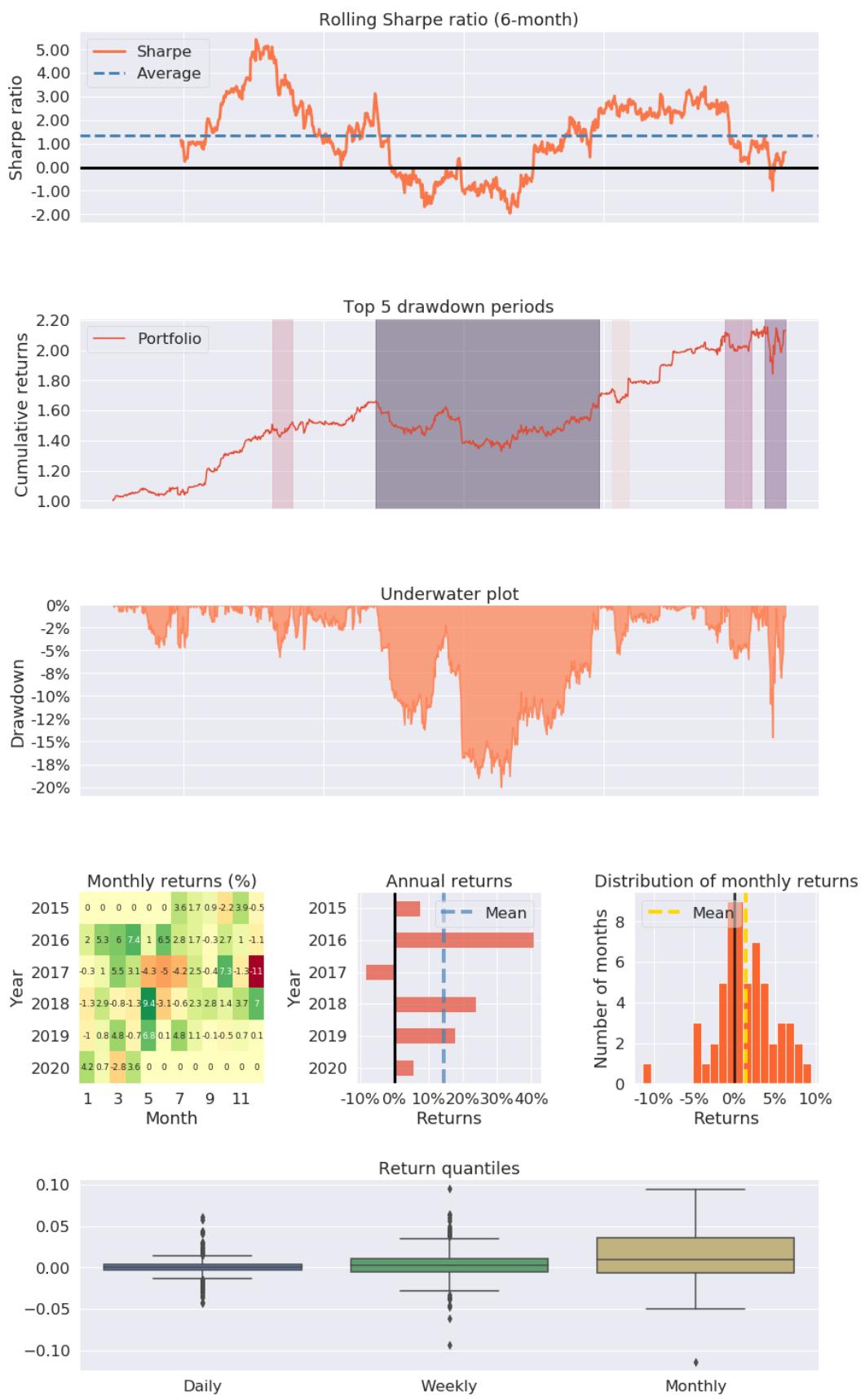


Рис. 22. Графическая иллюстрация статистик по стратегии MACD на базе CBOM.

Итоги соревнования MACD и Bollinger Bands на СВОМ.

В результате мы получили, что стратегия из блока momentum обыграла стратегию Mean-Reversion на данном примере и по доходности, и по коэффициенту Шарпа, а также по величине максимальной абсолютной просадки и максимальной продолжительности. Это ставит под сомнение гипотезу о том, что показатель Хёрста и характер временного ряда способны вносить решающую роль в результаты стратегии. Также я счёл интересным сравнить результативность стратегии с дефолтной опцией – стратегией «Купи и держи» (Buy and hold), на графике ниже я продемонстрировал, что данную стратегию не так уж просто превзойти, что подтверждается общепризнанным мнением.

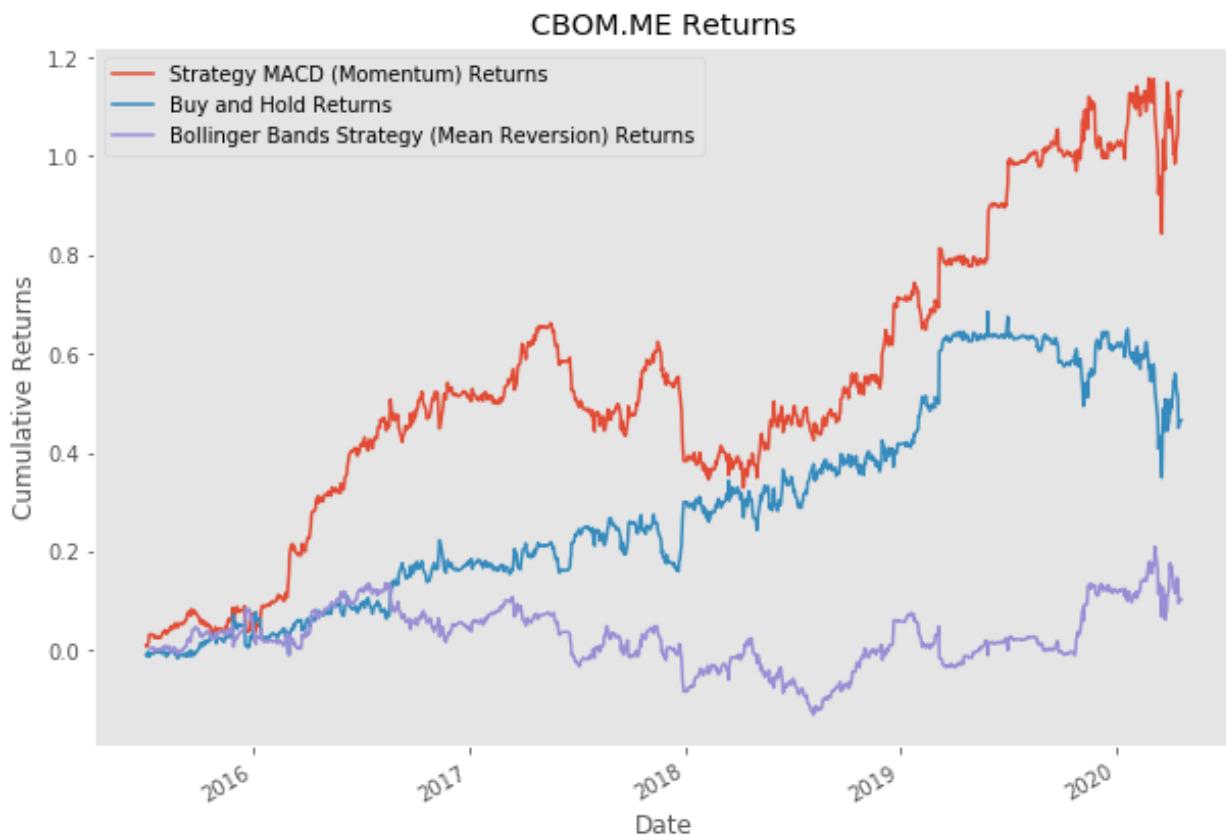


Рис. 22. Графическая иллюстрация накопленной доходности по стратегии MACD (momentum), Bollinger Bands (mean reversion) и ‘Buy and hold’ на базе СВОМ. Стоит заметить, что издержки мною не учитывались при проведении стратегий, поэтому надо представлять, что реальное преимущество стратегии ‘Купи и держи’ несколько выше в силу практически полного отсутствия транзакционных издержек¹⁶.

¹⁶ Разумным действием является проверка устойчивости результатов моделей на других ценных бумагах. В приложения я вынес дополнительные примеры, которые были мною проанализированы (ALRS, GMKN, YNDX)

Глава 3. Машинное обучение и нейросети.

Стратегия на основе алгоритма XGBoost.

XGBoost Regressor.

Чтобы справиться с 4 задачей нам нужно продемонстрировать работу алгоритмов машинного обучения.

Сейчас мною будет реализован алгоритм XGBoost на примере акций Московского кредитного банка (тикер на МОЕХ - CBOE), я импортировал данные с yahoo finance. Взглянем на скоректированную цену закрытия (adjusted close price).

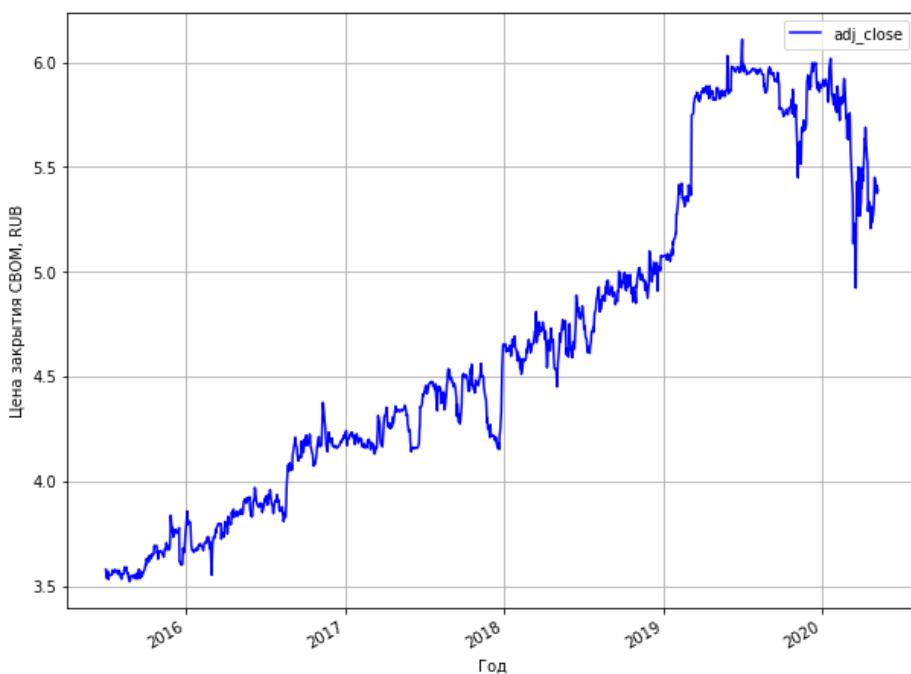


Рис. 23. Демонстрация графика скоректированных цен закрытия акций СВОМ.

Для тренировки модели необходимо наличие различных признаков. Создам несколько признаков. Этот процесс в англоязычной литературе принимает название Feature Engineering. Например, одним из дополнительных признаков я создал разницу между ценой открытия и ценой закрытия за каждый день, а также разницу между минимальной и максимальной ценой за день. Также мною были сделаны различные пролагированные признаки из уже созданных мною признаков, таким образом была собран набор признаков, на которых можно тренировать модель.

Затем был произведён процесс разделения выборки для подвыборки тренировки модели и для подвыборки тестирования (Бектестирование осуществлялось мною впоследствии именно на тестовой части, что методологически верно).

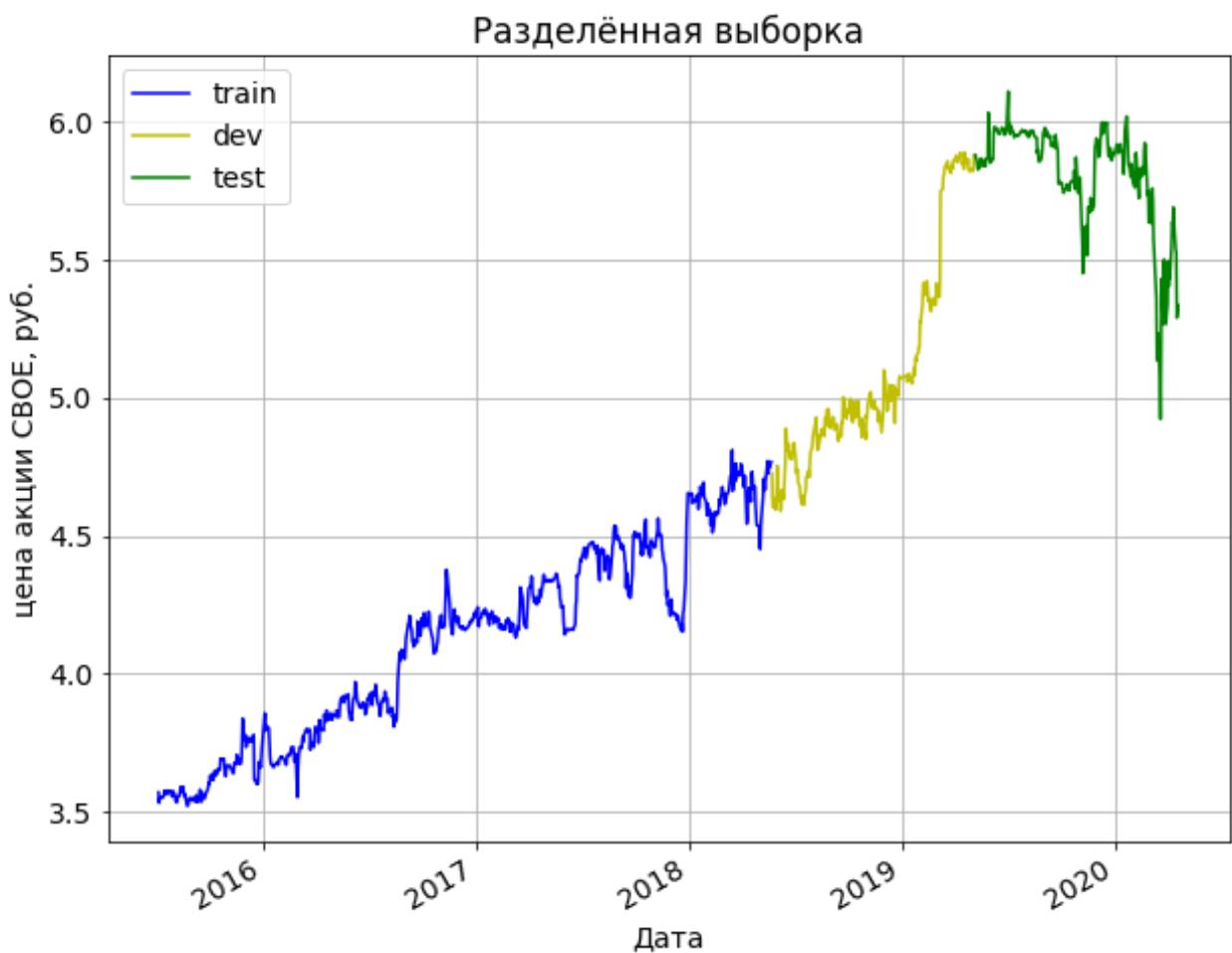


Рис. 24. Разделение выборки¹⁷

Соответственно выборка, на которой будет обучаться XGBoost модель выделена голубым цветом.

Тренировка модели начинается с задания базовых параметров и инициализации алгоритма.

¹⁷ В нашем случае выборка – это временной ряд

Train the model using XGBoost

```
In [25]: # Create the model
model = XGBRegressor(seed=model_seed,
                      n_estimators=n_estimators,
                      max_depth=max_depth,
                      learning_rate=learning_rate,
                      min_child_weight=min_child_weight,
                      subsample=subsample,
                      colsample_bytree=colsample_bytree,
                      colsample_bylevel=colsample_bylevel,
                      gamma=gamma)

# Train the regressor
model.fit(X_train_scaled, y_train_scaled)

Out[25]: XGBRegressor(base_score=0.5, booster=None, colsample_bylevel=1,
                      colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
                      importance_type='gain', interaction_constraints=None,
                      learning_rate=0.1, max_delta_step=0, max_depth=3,
                      min_child_weight=1, missing=nan, monotone_constraints=None,
                      n_estimators=100, n_jobs=0, num_parallel_tree=1,
                      objective='reg:squarederror', random_state=100, reg_alpha=0,
                      reg_lambda=1, scale_pos_weight=1, seed=100, subsample=1,
                      tree_method=None, validate_parameters=False, verbosity=None)
```

Рис. 25. Инициализация модели XGBoost.

После обучения модели на тренировочной выборке (train set), получаем, что метрики качества показывают следующие значения:

RMSE on train set = 0.021
MAPE on train set = 0.368%

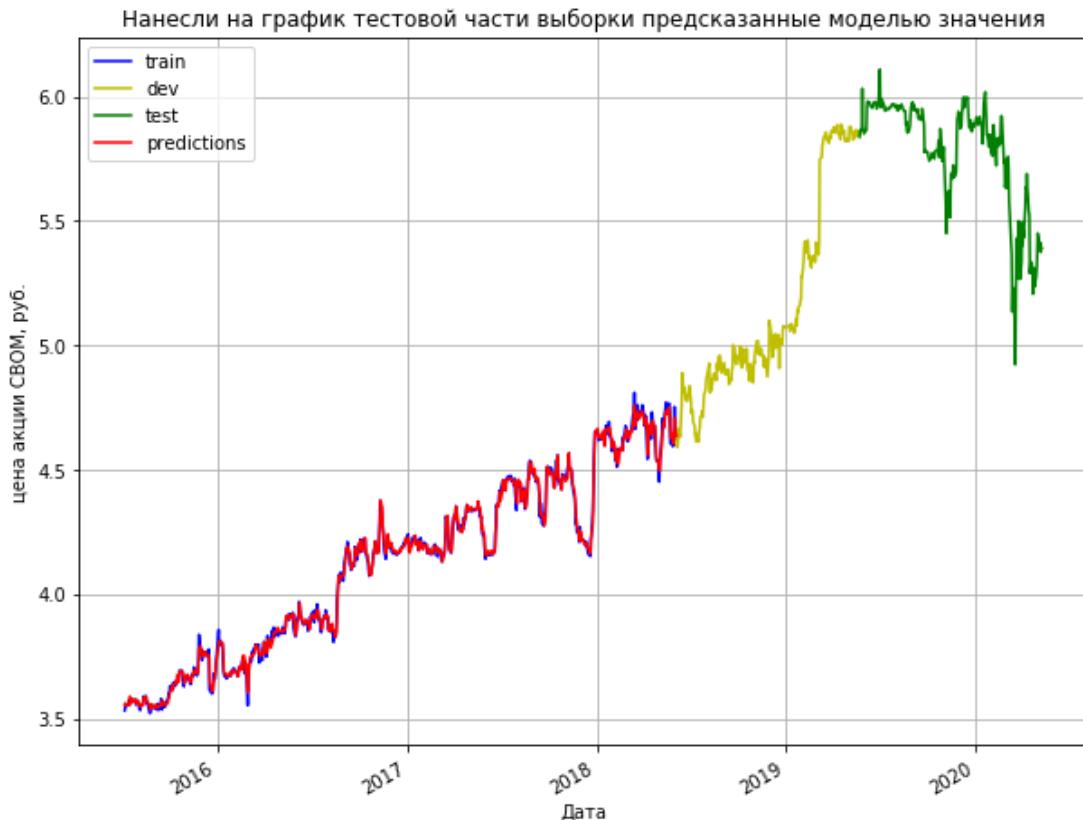


Рис. 26. Иллюстрация предсказания на части train.

Предсказания отражены на графике красным цветом. Визуально кажется, что поведение предсказанных очень неплохо описывается нашей моделью.

Теперь протестируем работу алгоритма на “dev” подвыборке.

RMSE on dev set = 0.043

MAPE on dev set = 0.572%

Метрики качества закономерно ухудшились

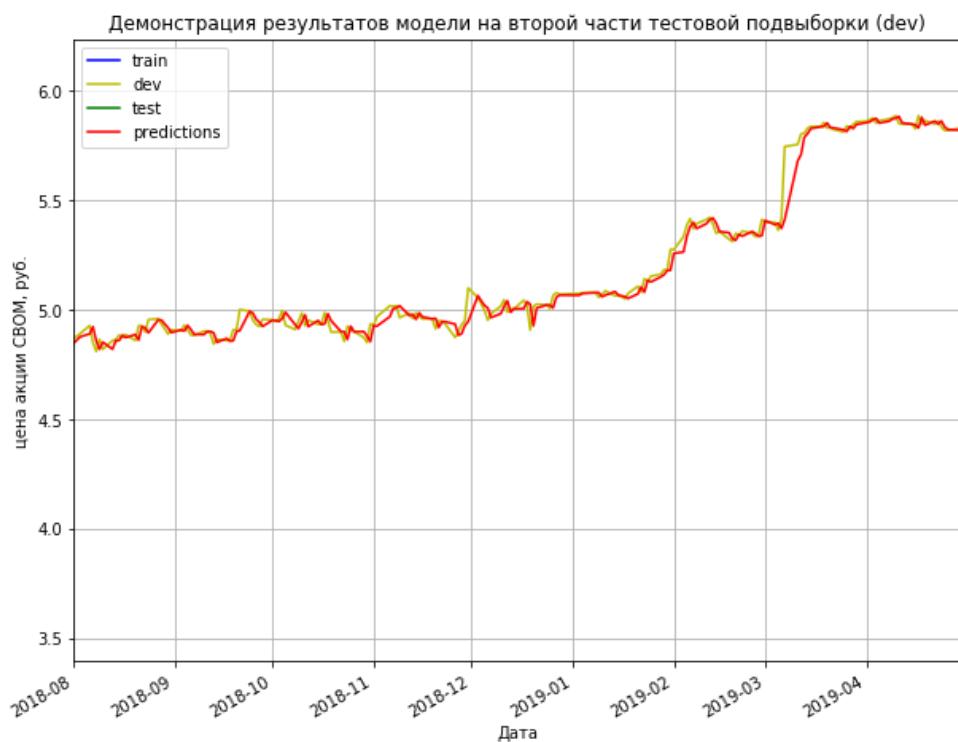


Рис. 27. Иллюстрация предсказания на второй части тренировочной выборки - dev.

Анализ значимости признаков выдал следующие результаты:

(Здесь приведены 10 наиболее значимых признаков)

```
[('range_hl_lag_2', 0.001044068),  
 ('range_oc_lag_1', 0.0010487086),  
 ('range_oc_lag_2', 0.0010558855),  
 ('range_hl_lag_1', 0.001057535),  
 ('range_hl_lag_3', 0.0011338568),  
 ('volume_lag_1', 0.0012817393),  
 ('volume_lag_3', 0.0017048698),  
 ('adj_close_lag_2', 0.032863766),  
 ('adj_close_lag_3', 0.35987738),  
 ('adj_close_lag_1', 0.5975027)]
```

Это значит, что самым значимым оказалось значение цены закрытия на прошлый день, что вполне закономерно. Разумно полагать, что наиболее близкие к текущему моменту цены закрытия более важны по сравнению с теми, что были раньше. Также стоит отметить, что важность других признаков существенно меньше.

Далее проведём настройку разных параметров и выясним, при каких из них модель показывает более точные результаты:

Настройка параметров модели XGBoost.

- 1) Настройка параметра N.

Для начала продемонстрируем перформанс модели для разного параметра N, то есть для параметра окна, за которого вычисляется среднее и среднеквадратическое отклонение.

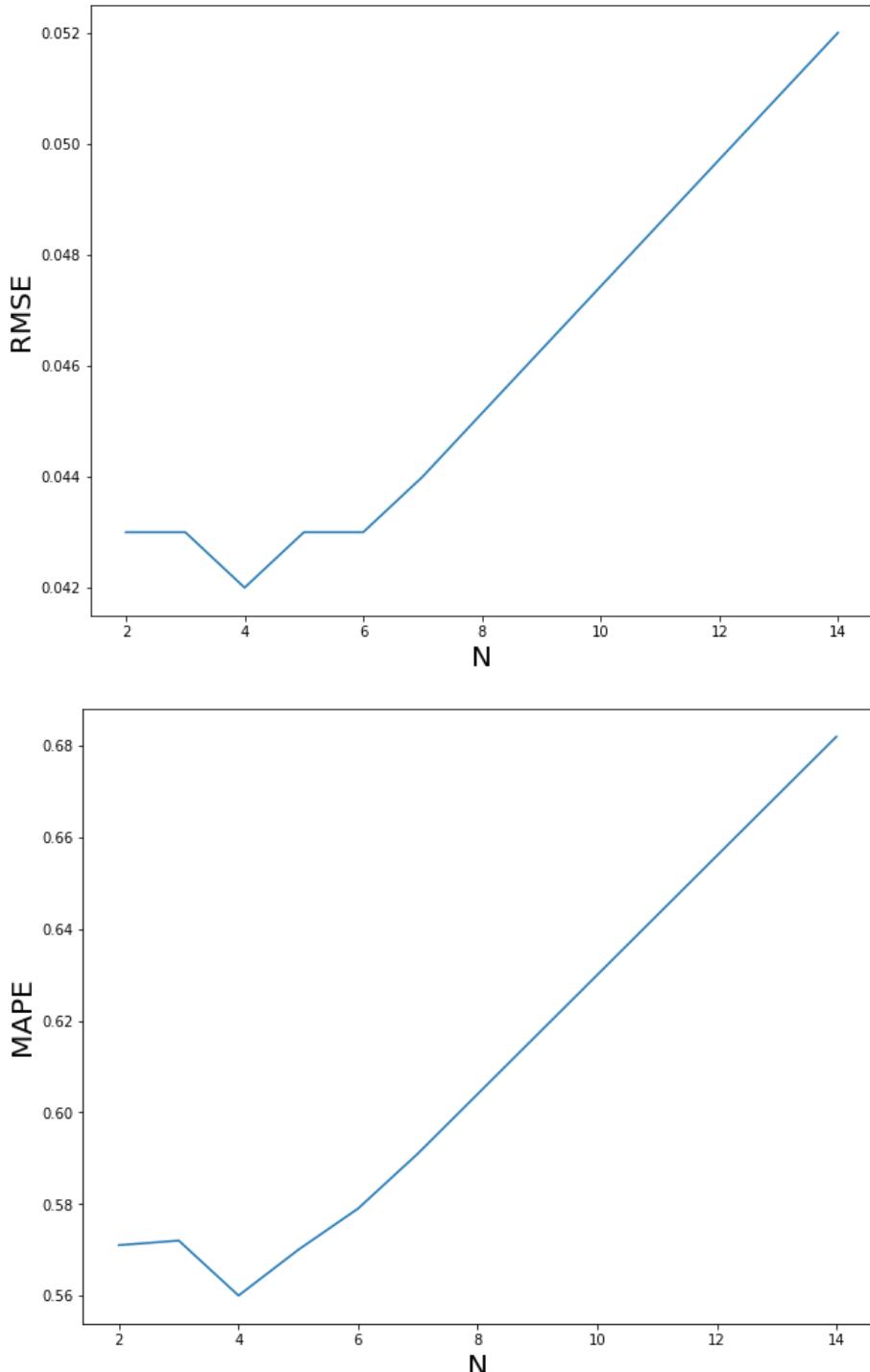


Рис. 28. Получаем, что наиболее точная модель при $N = 4$, то есть окно в предсказательной модели стоит установить на уровне 4 дней.

- 2) Теперь прозведём настройку некоторых внутренних параметров нашей XGBoost модели:

Настроим n_estimators (по умолчанию 100) и max_depth (по умолчанию 3).

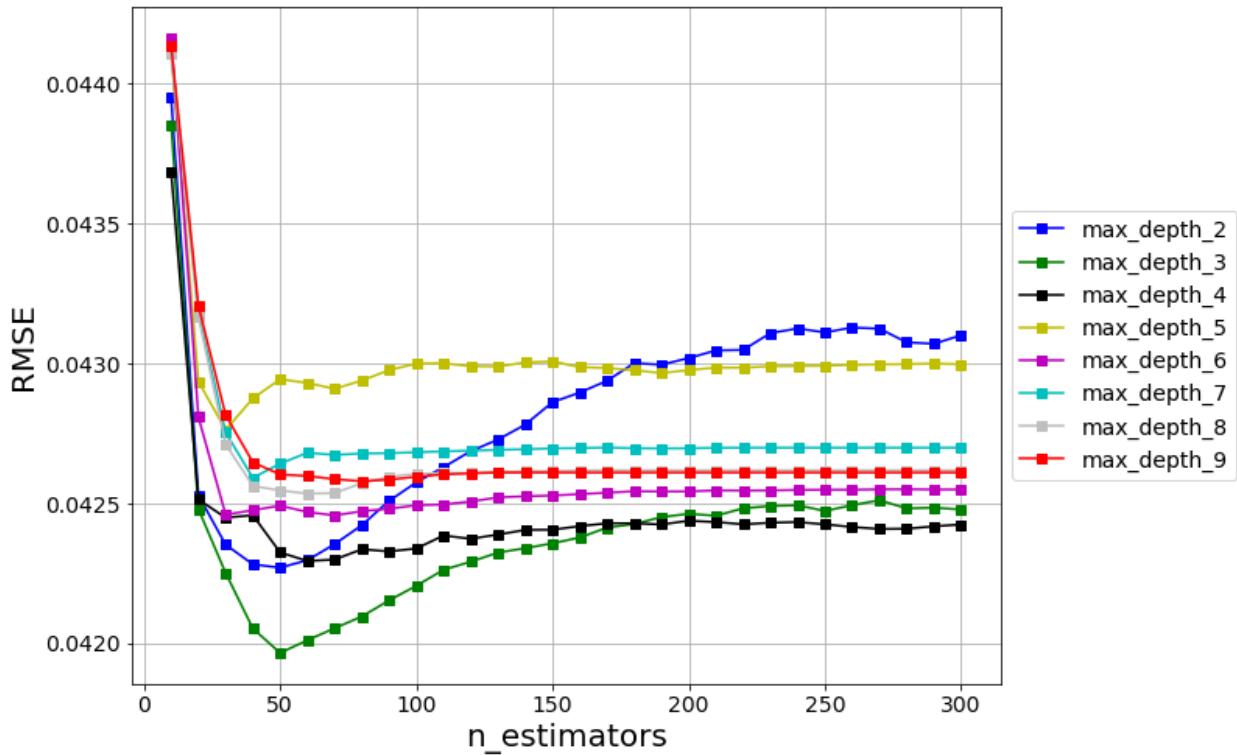


Рис. 29. согласно RMSE оптимальное сочетание параметров ($n_estimators$, max_depth) = (50, 3); $\min RMSE = 0.042$; согласно MAPE оптимальное сочетание параметров ($n_estimators$, max_depth) = (50, 3); $\min MAPE = 0.557\%$.

Настроим `learning_rate` (по умолчанию 0.1) и `min_child_weight` (по умолчанию 1).

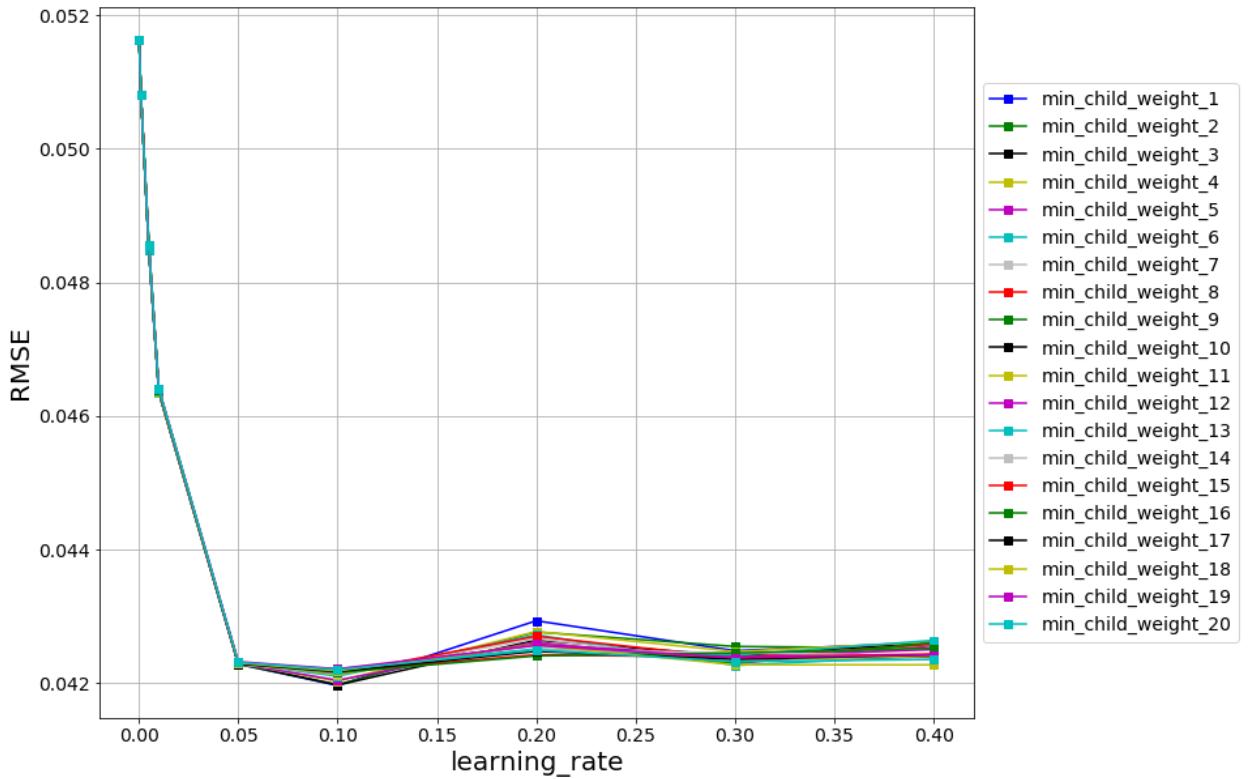


Рис. 30. согласно RMSE оптимальная пара параметров (`learning_rate`, `min_child_weight`) = (0.1, 1); $\min \text{RMSE} = 0.042$; согласно MAPE оптимальная пара параметров (`learning_rate`, `min_child_weight`) = (0.1, 3); $\min \text{MAPE} = 0.557\%$.

Настроим параметры `subsample` (по умолчанию 1) и `gamma` (по умолчанию 0).

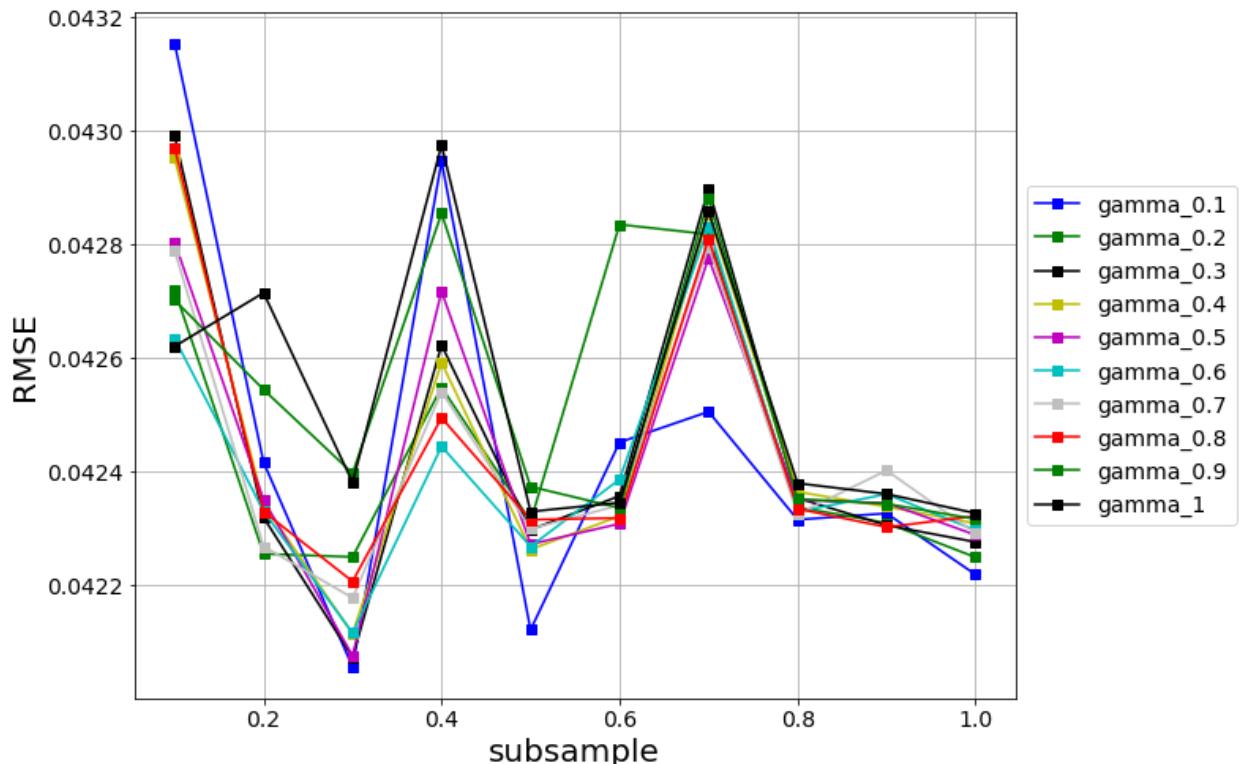


Рис. 31. согласно RMSE оптимальная пара параметров (*subsample*, *gamma*) = (0.3, 0.1);
 $\min \text{RMSE} = 0.042$; согласно MAPE оптимальная пара параметров (*learning_rate*,
min_child_weight) = (0.3, 0.3); $\min \text{MAPE} = 0.559\%$.

Настроим параметры `colsample_bytree` (по умолчанию 1) и `colsample_bylevel` (по умолчанию 1).

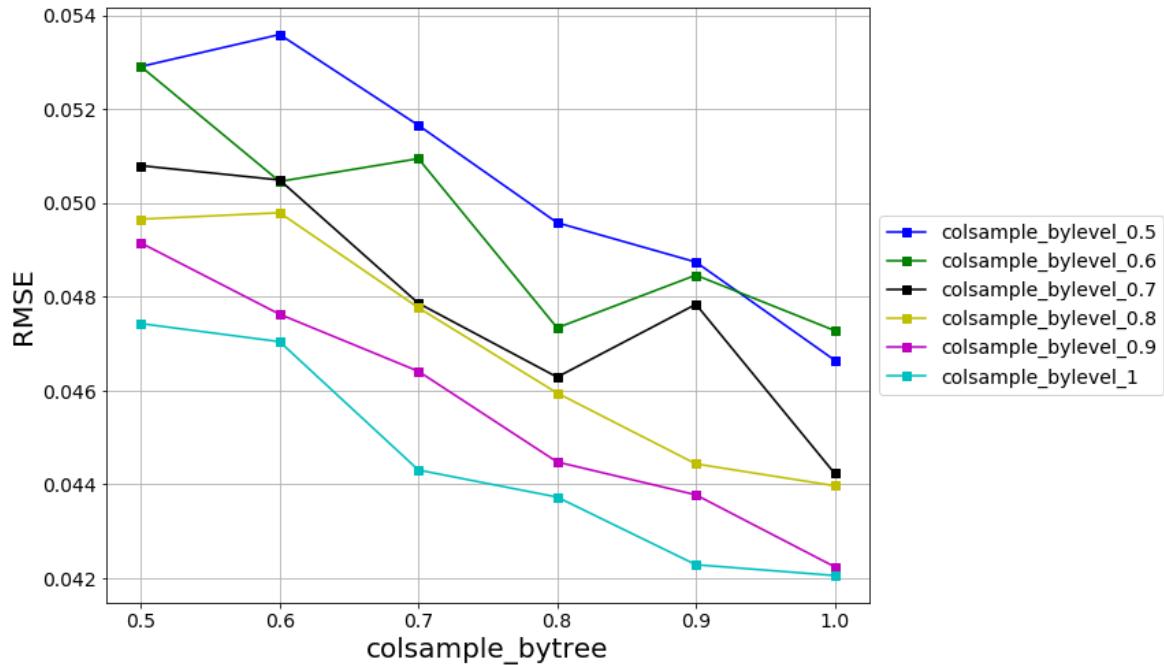


Рис. 32. согласно RMSE оптимальная пара параметров (`colsample_bytree`,
`colsample_bylevel`) = (1, 1); $\min \text{RMSE} = 0.042$; согласно MAPE оптимальная пара
параметров (`colsample_bytree`, `colsample_bylevel`) = (1, 1); $\min \text{MAPE} = 0.560\%$.

param	original	after_tuning
n_estimators	100.00000	50.00000
max_depth	3.00000	3.00000
learning_rate	0.10000	0.10000
min_child_weight	1.00000	1.00000
subsample	1.00000	0.30000
colsample_bytree	1.00000	1.00000
colsample_bylevel	1.00000	1.00000
gamma	0.00000	0.10000
rmse	0.04220	0.04206
mape_pct	0.56042	0.55963

Рис. 33. Результатирующая таблица с параметрами и метриками качества до и после
настройки параметров.

Повезло изначально, что значения параметров, которые установлены в модели по умолчанию близки к оптимальным, поэтому мы не наблюдаем значительного изменения метрик качества, однако оно есть.

Проводя проверку модели на тестовой подвыборке мы получаем следующие результаты:

RMSE on test set = 0.065

MAPE on test set = 0.695%

Это значительно превышает показатели на тренировочной выборке (что плохо в данном случае, но ухудшение в данной ситуации – это довольно обычное явление для финансовых данных, поскольку появлялись новые шоки, которые не описываются теми взаимосвязями, которые установились в рамках модели на тренировочной подвыборке)



Рис. 34. Иллюстрация результата модели на тестовой части выборки.

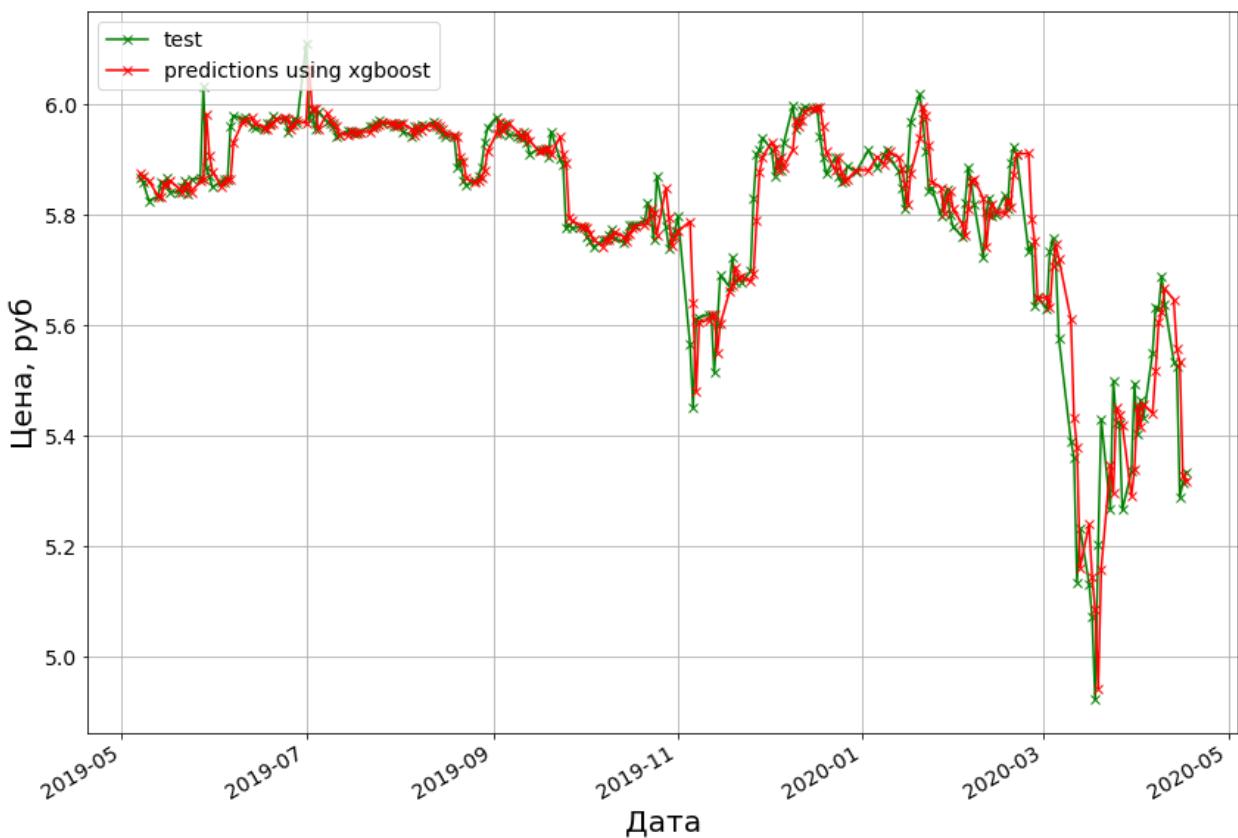


Рис. 34. Иллюстрация тестовых данных и работы модели XGBoost.

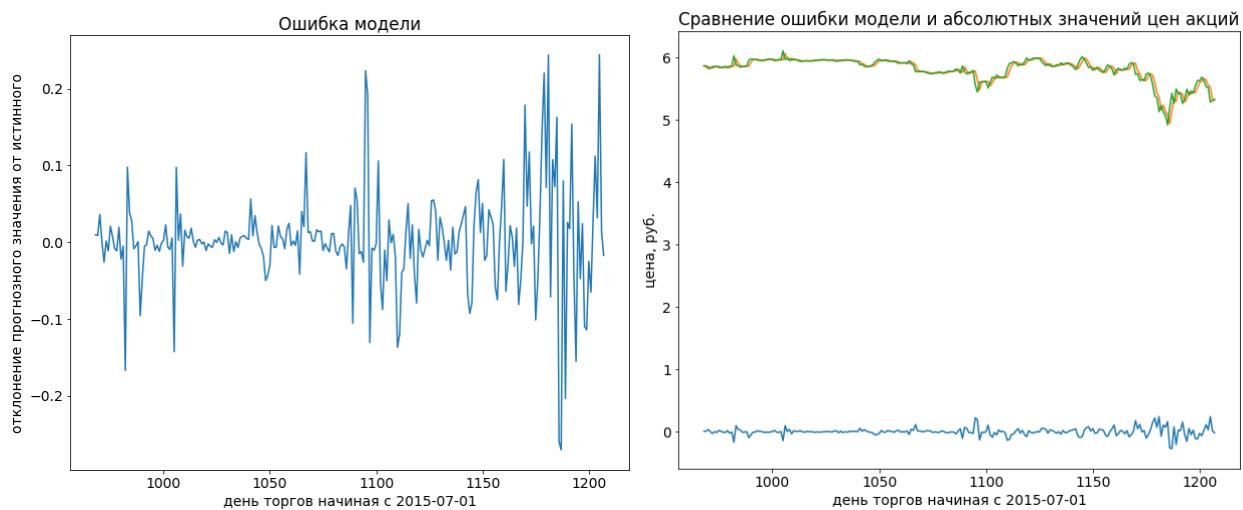
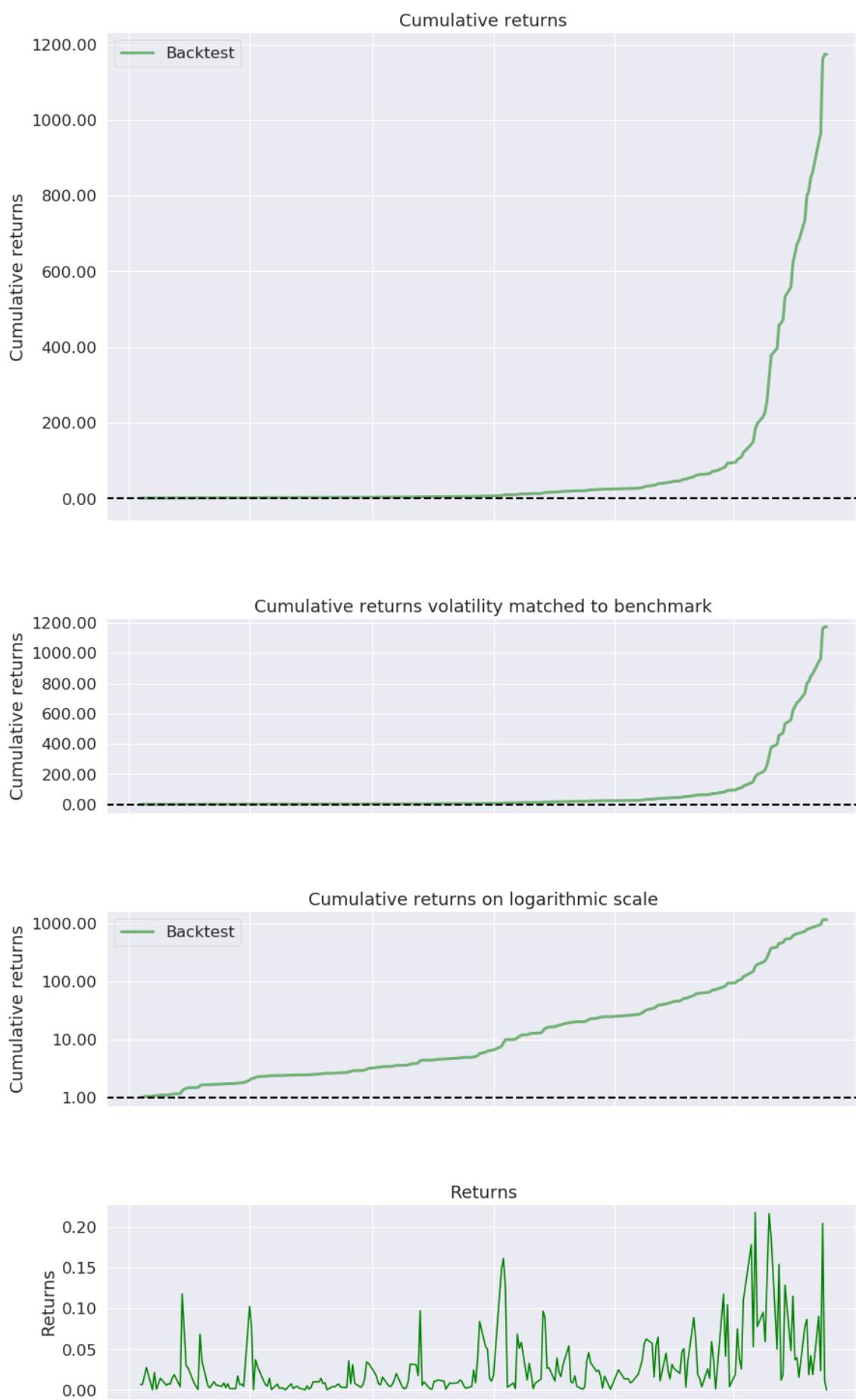


Рис. 35. Иллюстрация ошибки модели (левый график). Дополнительно демонстрация погрешности относительно курса акции (правый график). Можно сказать о незначительности ошибок модели, обращая внимание на этот график.

Annual return	167017.2%
Cumulative returns	117266.4%
Annual volatility	64.6%
Sharpe ratio	11.96
Calmar ratio	NaN
Stability	0.91
Max drawdown	0.0%
Omega ratio	NaN
Sortino ratio	inf
Skew	2.37
Kurtosis	6.02
Tail ratio	112.24
Daily value at risk	-5.1%

Рис. 36. Основные метрики качества для алгоритмической стратегии на базе алгоритма машинного обучения XGBoost.



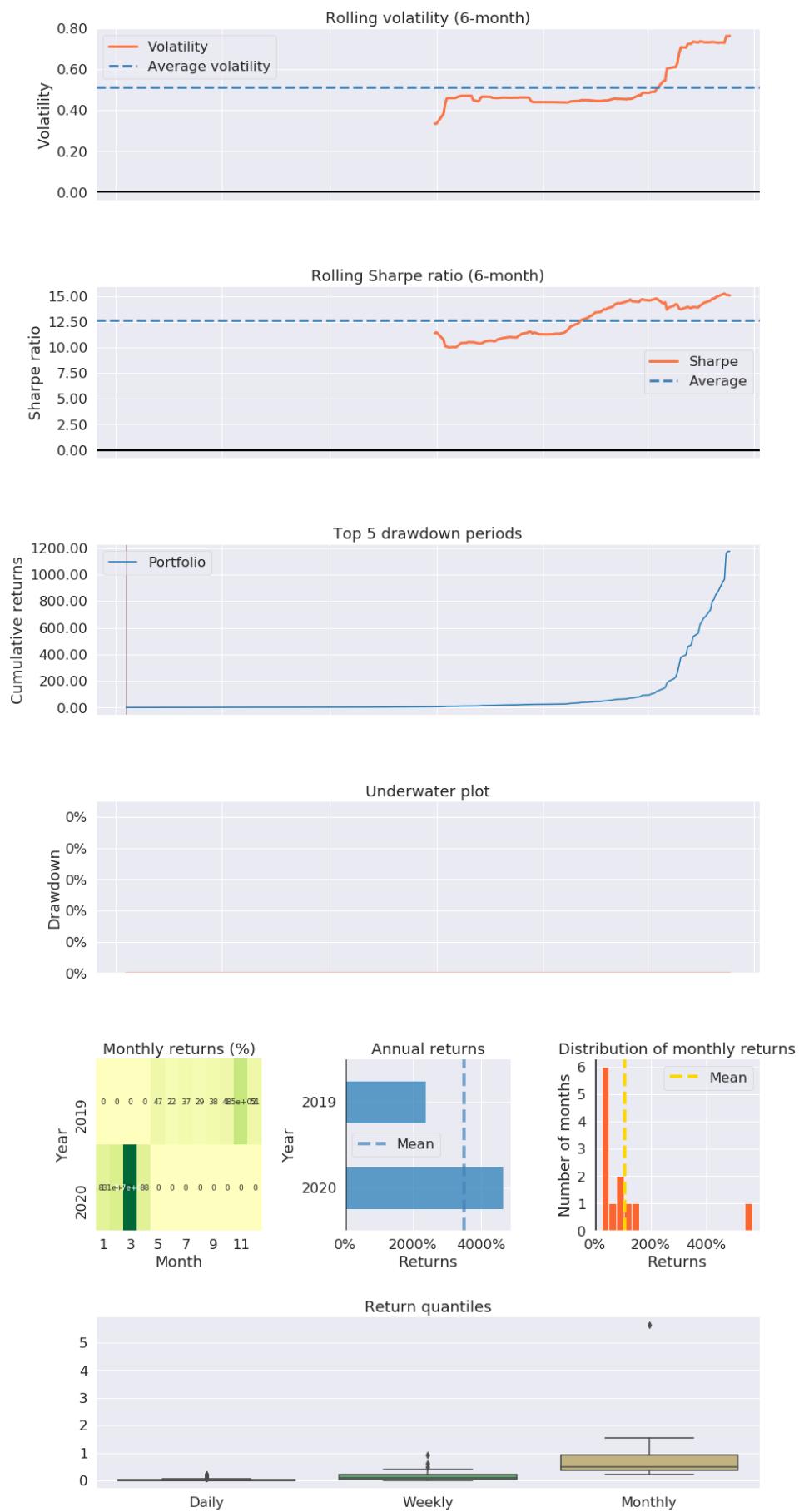


Рис. 37. Графическая иллюстрация статистик по стратегии XGBoost на базе CBOM.

Если убрать предпосылку реинвестирования денежных средств:

Pnl = 735.41%

Annual P&L = 222.38% годовых

Sharpe = 158.95



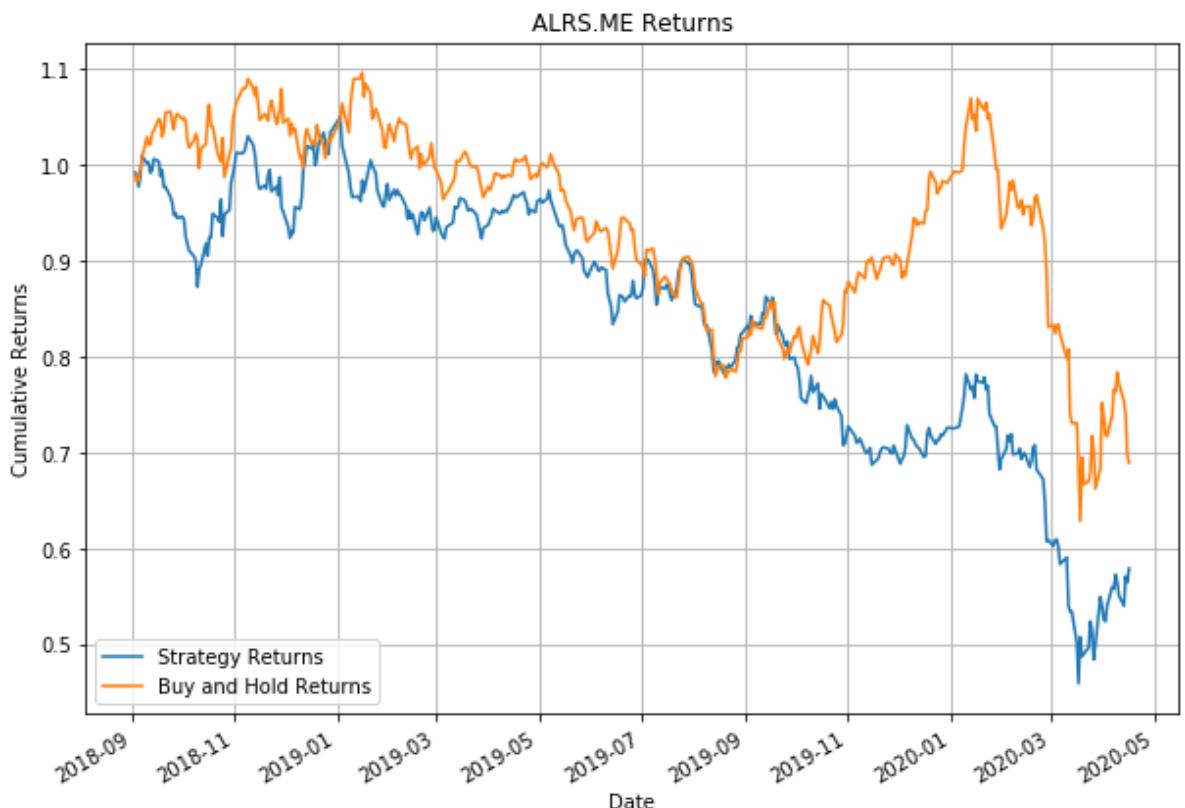
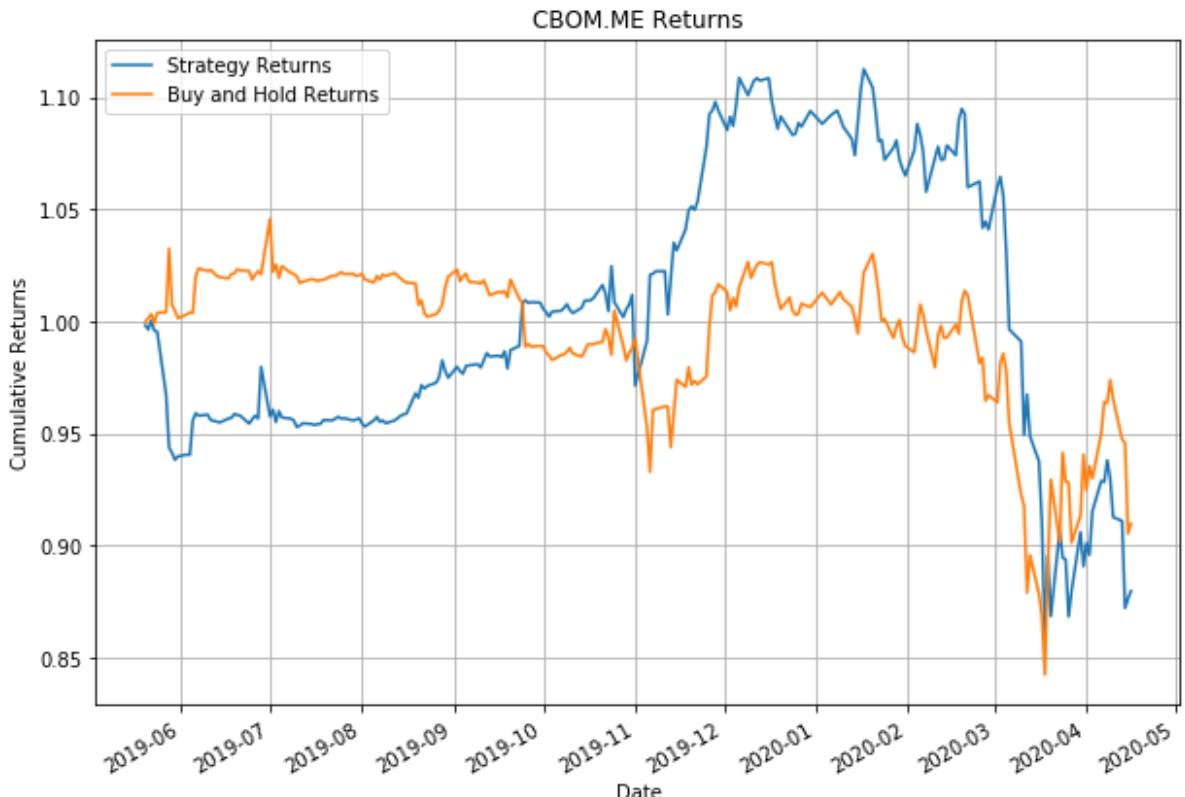
Рис. 38. Доходность стратегии XGBoost Regressor без реинвестиций.

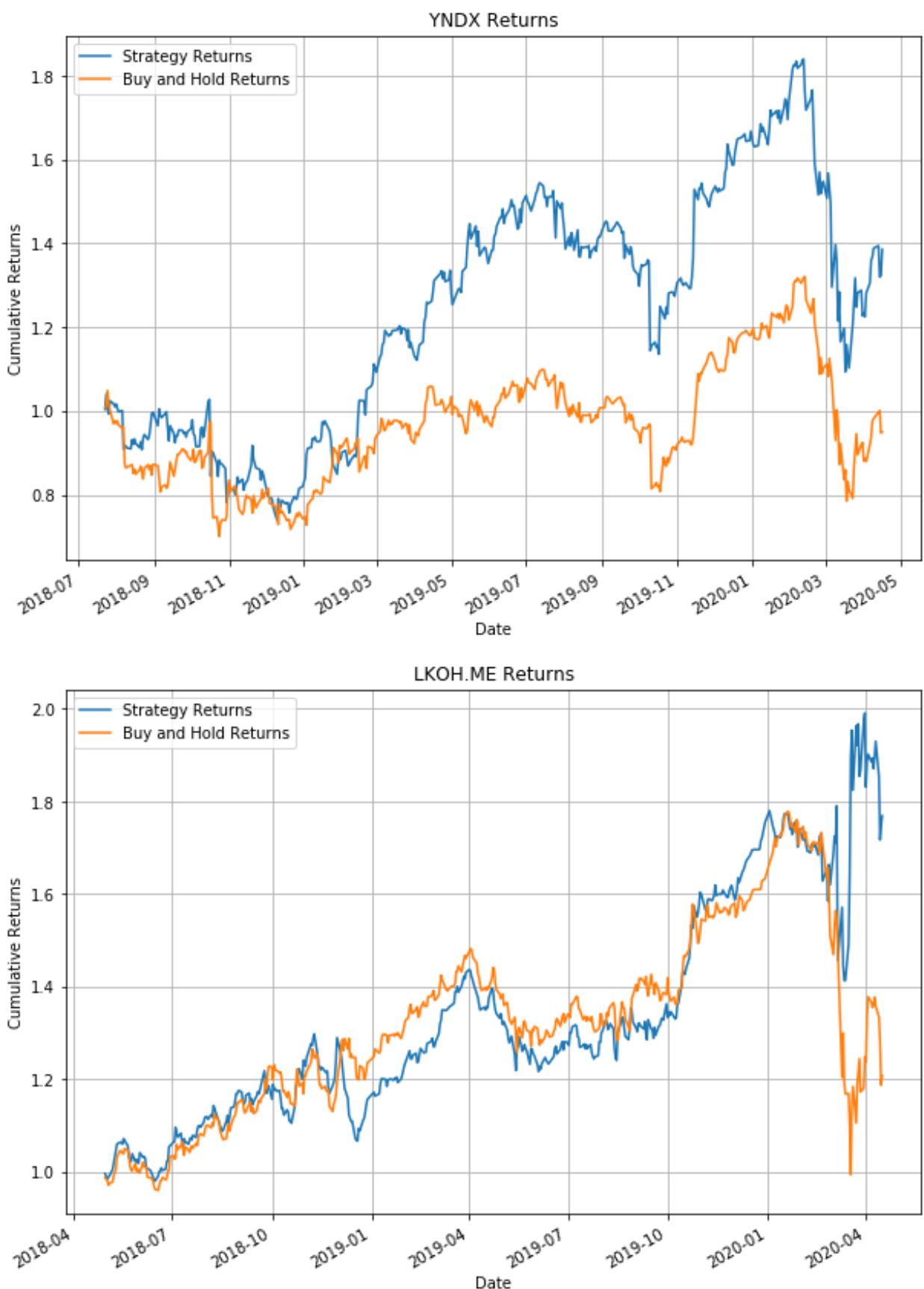
Важные предпосылки:

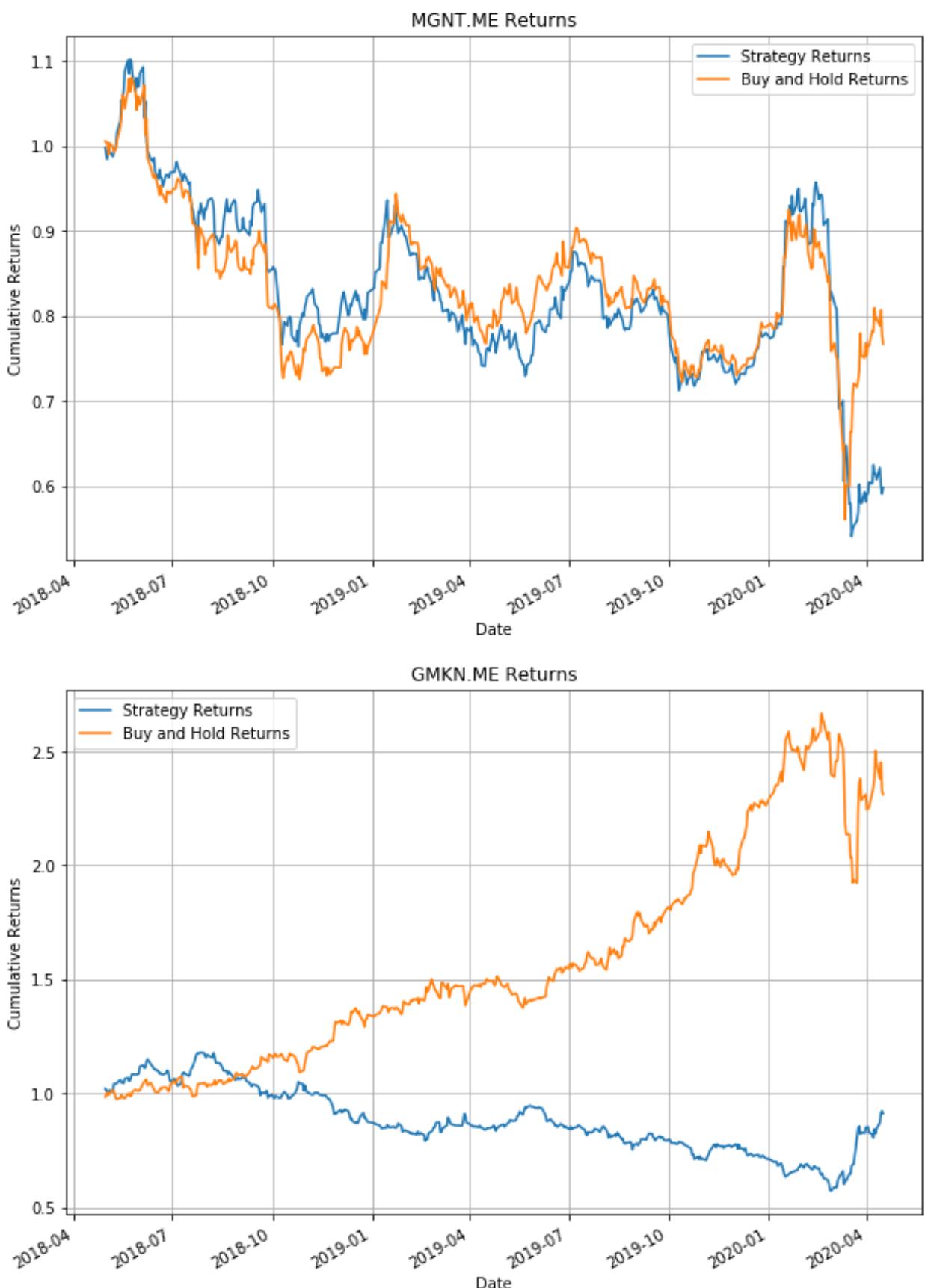
1. Отсутствие транзакционных издержек
2. Все заработанные средства автоматически реинвестируются обратно в стратегию.
3. Подсчёт P&L ведётся по правилу:
 - 3.1.Модель предсказывает цену на следующий торговый день, если она больше текущей, то она ставит позицию на покупку, иначе на короткую продажу.
 - 3.2.На следующий день открытая позиция автоматически закрывается.

XGBoost Classifier.

В этом разделе я рассмотрю алгоритм классификации. Бустинг позволяет решать как задачи регрессии (как в предыдущем разделе), так и задачи классификации. То есть если речь идёт о регрессии, то ставится задача предсказания непосредственного значения цены, в случае классификации достаточно ответить на вопрос направления движения цены актива. Это более общая задача.







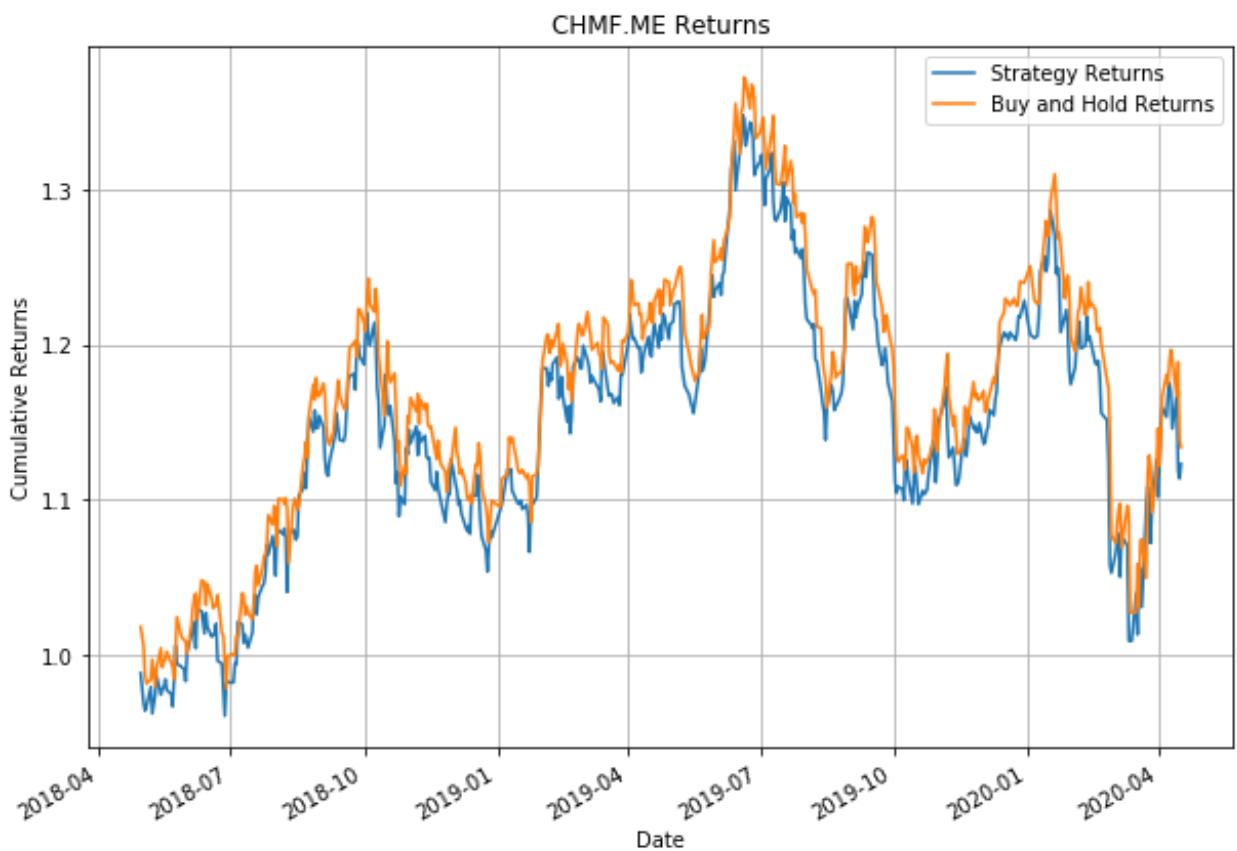


Рис. 39 – 45. Результативность стратегии XGBoost Classifier для разных компаний.

Совокупная результативность стратегии XGBoost Classifier (средневзвешенные результаты).

Backtest		Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
Annual return	-4.0%						
Cumulative returns	-3.6%						
Annual volatility	19.6%						
Sharpe ratio	-0.11						
Calmar ratio	-0.18						
Stability	0.07						
Max drawdown	-22.5%	0	22.49	2020-01-21	2020-03-17	NaT	NaN
Omega ratio	0.98	1	10.20	2019-07-08	2019-10-11	2020-01-10	135
Sortino ratio	-0.14	2	1.06	2019-06-19	2019-06-24	2019-06-26	6
Skew	-1.41	3	0.93	2019-06-26	2019-06-27	2019-06-28	3
Kurtosis	16.55	4	0.87	2019-06-10	2019-06-14	2019-06-18	7
Tail ratio	0.83	Stress Events	mean	min	max		
Daily value at risk	-2.5%	New Normal	-0.01%	-8.45%	6.79%		



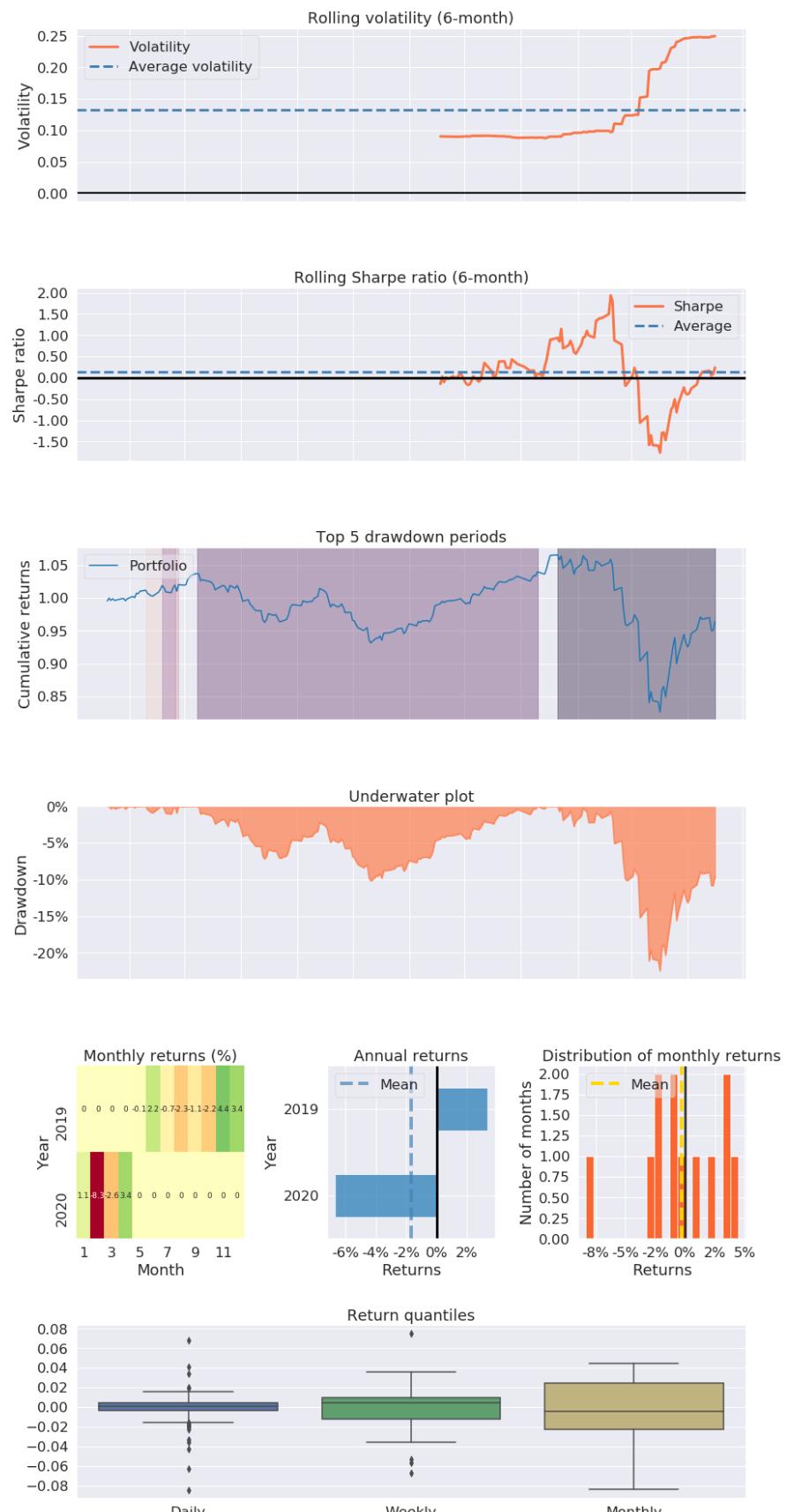


Рис. 46. Сводная статистика по результатам XGBoost Classifier для средневзвешенного среди акций Северстали, Норильского никеля, Магнита, Лукойла, Яндекса, Московского кредитного банка и Алросы. Заметим, что в целом результат – убыток. Однако, как можно было замечено, результативность сильно варьируется от бумаг, входящих в состав.

Если взять в портфель только акции компаний Московского кредитного банка, Яндекса и Алросы, то результаты становятся совершенно иными, они очень сильно улучшаются. Продемонстрирую на статистиках ниже.

Annual return	35.4%
Cumulative returns	30.7%
Annual volatility	21.5%
Sharpe ratio	1.52
Calmar ratio	2.35
Stability	0.55
Max drawdown	-15.1%
Omega ratio	1.35
Sortino ratio	2.35
Skew	0.04
Kurtosis	9.43
Tail ratio	1.37
Daily value at risk	-2.6%

	Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
Max drawdown	0	15.06	2020-01-23	2020-03-12	2020-03-23	43
Omega ratio	1	9.49	2019-08-07	2019-10-11	2019-11-18	74
Sortino ratio	2	4.49	2019-05-22	2019-06-17	2019-07-22	44
Skew	3	3.33	2020-03-30	2020-04-02	2020-04-07	7
Kurtosis	4	2.56	2019-11-18	2019-12-02	2019-12-09	16

	Stress Events	mean	min	max
Daily value at risk	New Normal	0.13%	-7.71%	7.39%

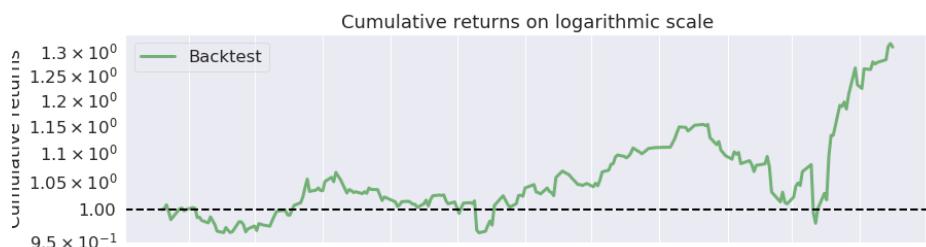




Рис. 47. Результативность стратегии XGBoost Classifier выросла значительно на другой выборке компаний (Московский кредитный банк, Яндекс и Алроса).

Заключение по стратегии XGBoost Classifier: Данная стратегия может показывать довольно привлекательные результаты. Во многих случаях она обыгрывает стратегию Buy and hold. Однако, стоит допускать, что в целом применив данную стратегию к портфелю бумаг, может получиться отрицательная доходность в результате.

Для случая акций Московского кредитного банка продемонстрирую важность каждого из регрессоров по метрике F1 меры. В англоязычной терминологии это представление носит название feature importance.¹⁸

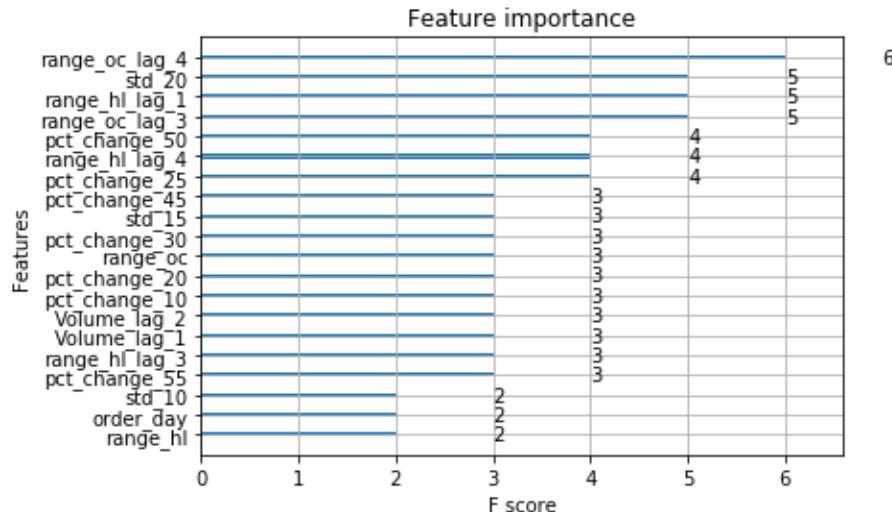


Рис. 48. Важность регрессоров для случая анализа акций Яндекса, Московского кредитного банка и Алросы.

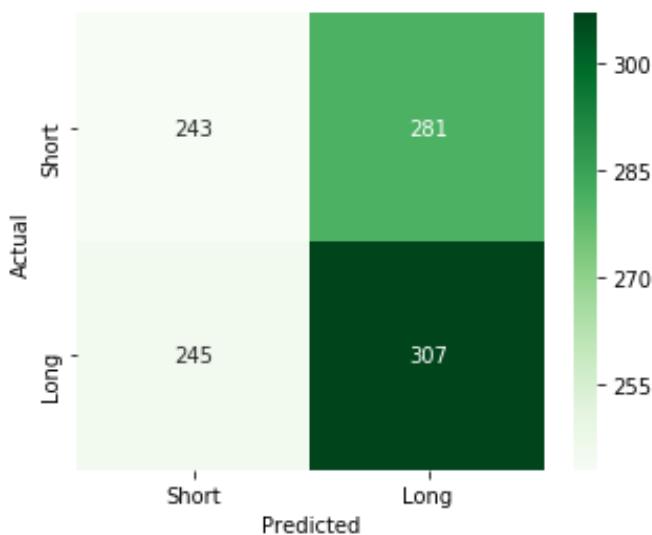


Рис. 49. Матрица точности для алгоритма XGBoost Classifier на базе акций Яндекса, Московского кредитного банка и Алросы.

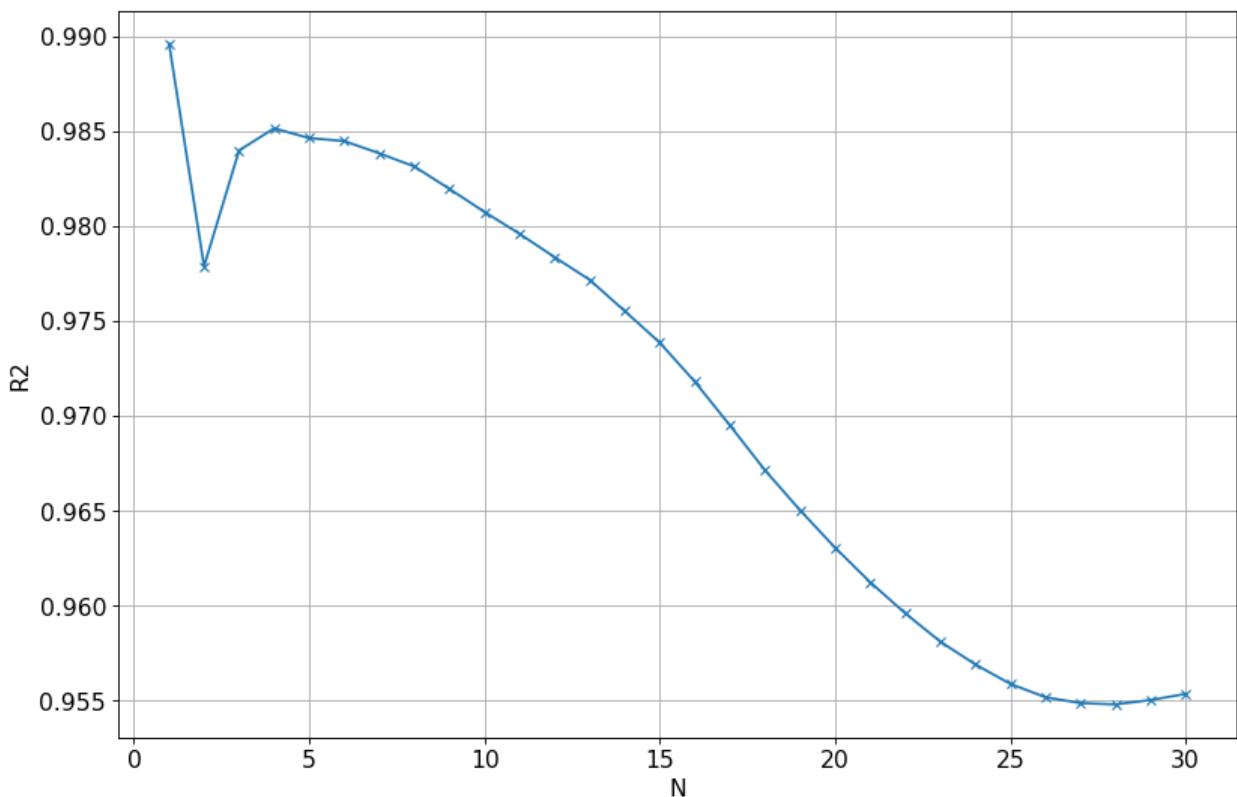
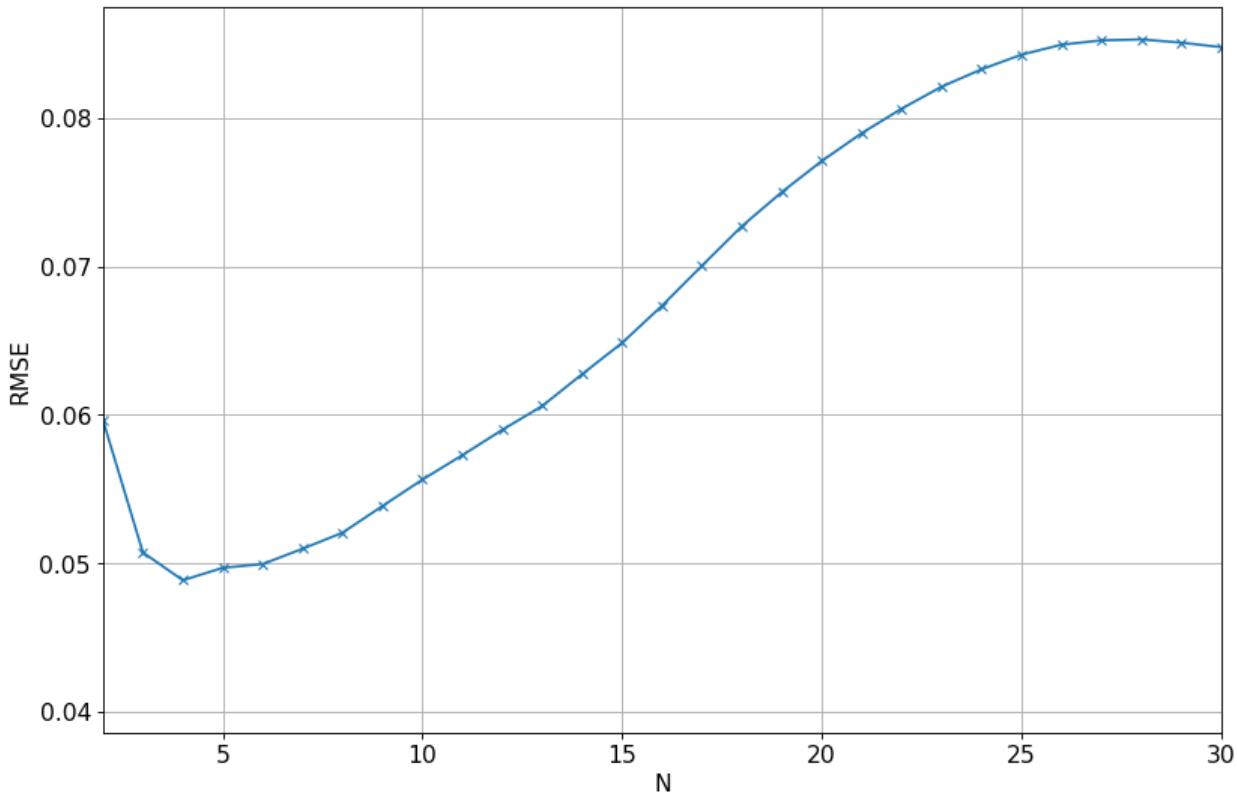
Теперь построим более простую модель Машинного обучения:

¹⁸ Разумеется можно далее продолжать и продолжать оптимизацию параметров модели, например с помощью библиотеки GridsearchCV, на уровне логики это машинный перебор заданных значений для набора параметров, которым задаются вектора оптимизируемых значений. Также можно использовать библиотеку StratifiedKfold, чтобы осуществлять рандомизацию выборки, на которой происходит настройка параметров.

Линейная регрессия (linear regression).

Я коснулся оптимизации лишь величины окна (N), по которому идёт предсказание на следующие торговый день. Как и ранее здесь я работаю с акцией Московского кредитного банка.

Получаем такие результаты



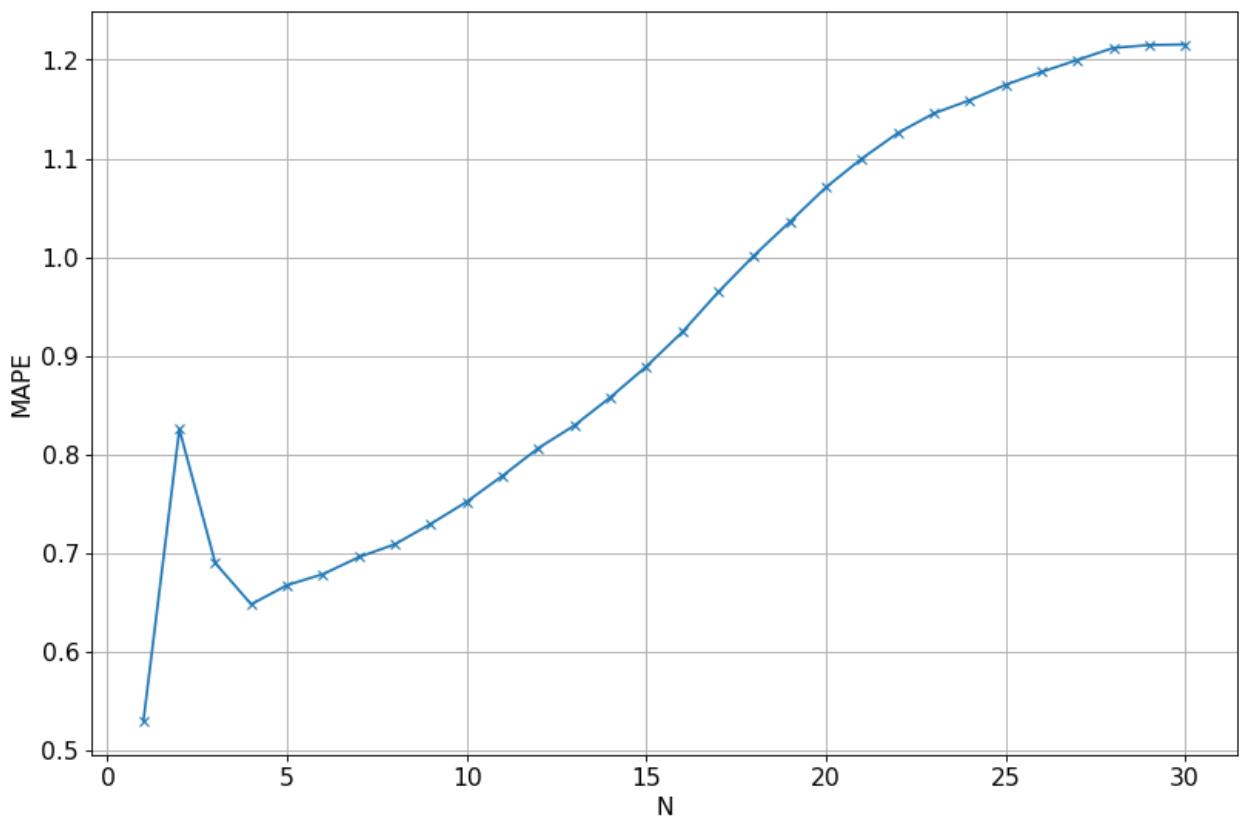


Рис. 50 – 52. Иллюстрация оптимизации N .

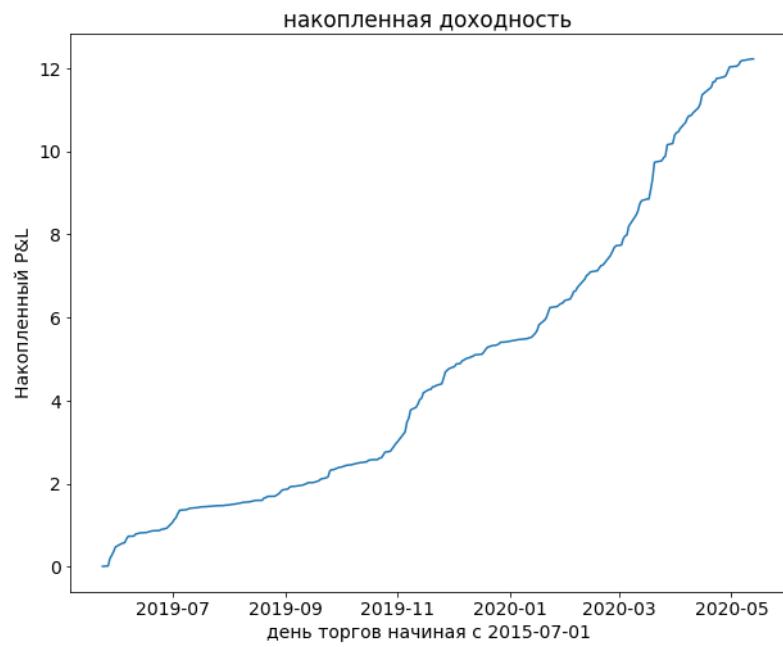


Рис. 53. Результаты стратегии на основе линейной регрессии.

На этом заканчивается блок, связанный с машинным обучением и теперь я перейду к рассмотрению стратегии на основе нейросети.

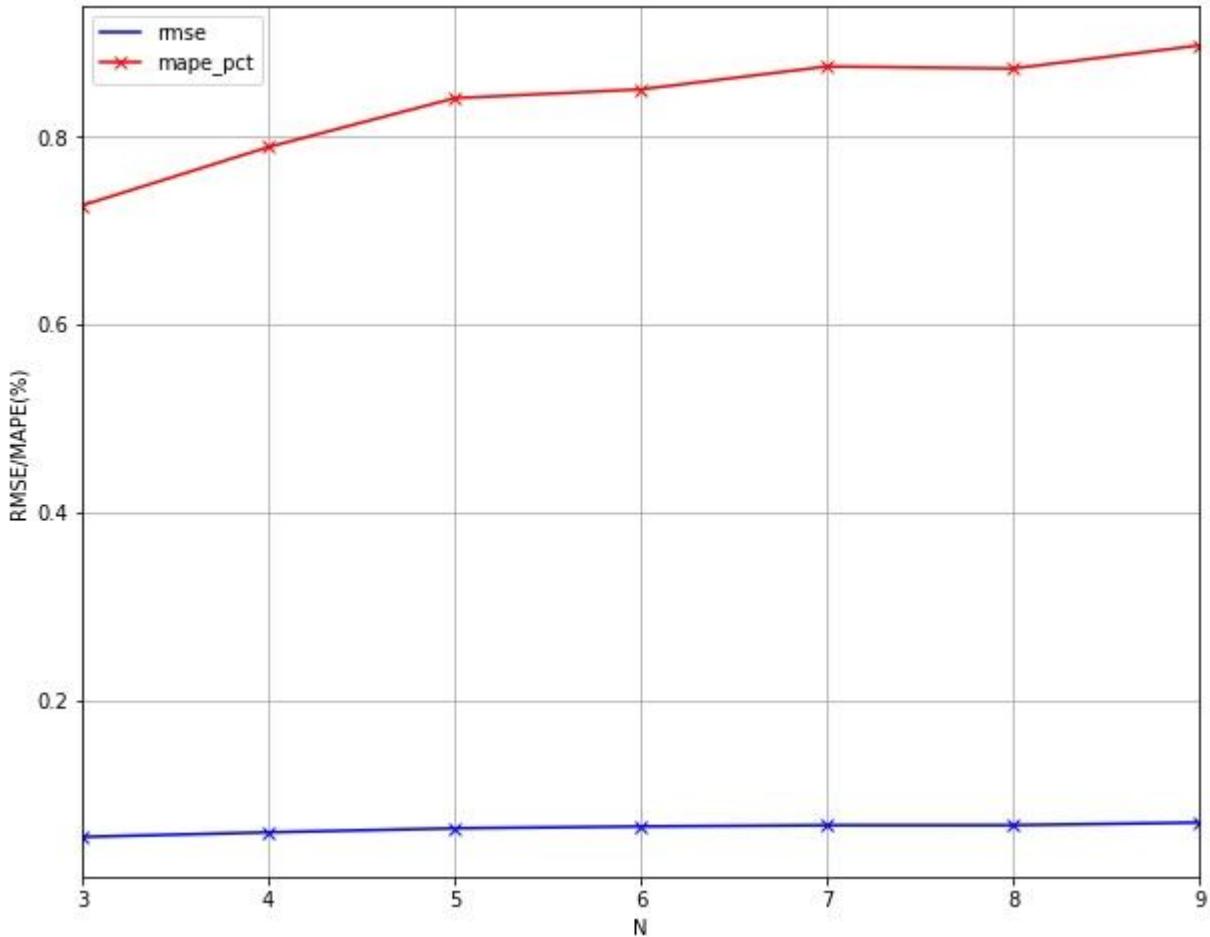
Нейросеть. LSTM model.

Нашей следующей задачей будет работа с нейросетью LSTM (Long Short Term Memory model).

Настройка параметров модели.

В процессе построения стратегии можно были оптимизированы ряд параметров:

- 1) Количество дней, используемые в подсчёте скользящих величин (N):



Ruc. 54. min RMSE = 0.055, min MAPE = 0.727%, optimum N = 3

Для более корректной работы нейросети следует произвести настройку внутренних параметров.

- 2) Настройка эпох и размера выборки.

Настроим другие параметры модели (будем также делать попарно): В рамках нейросети очень большое значение имеет количество эпох, которое обучается сеть, а также размер серии (batch), поэтому настройку имеет смысл начать именно с них:

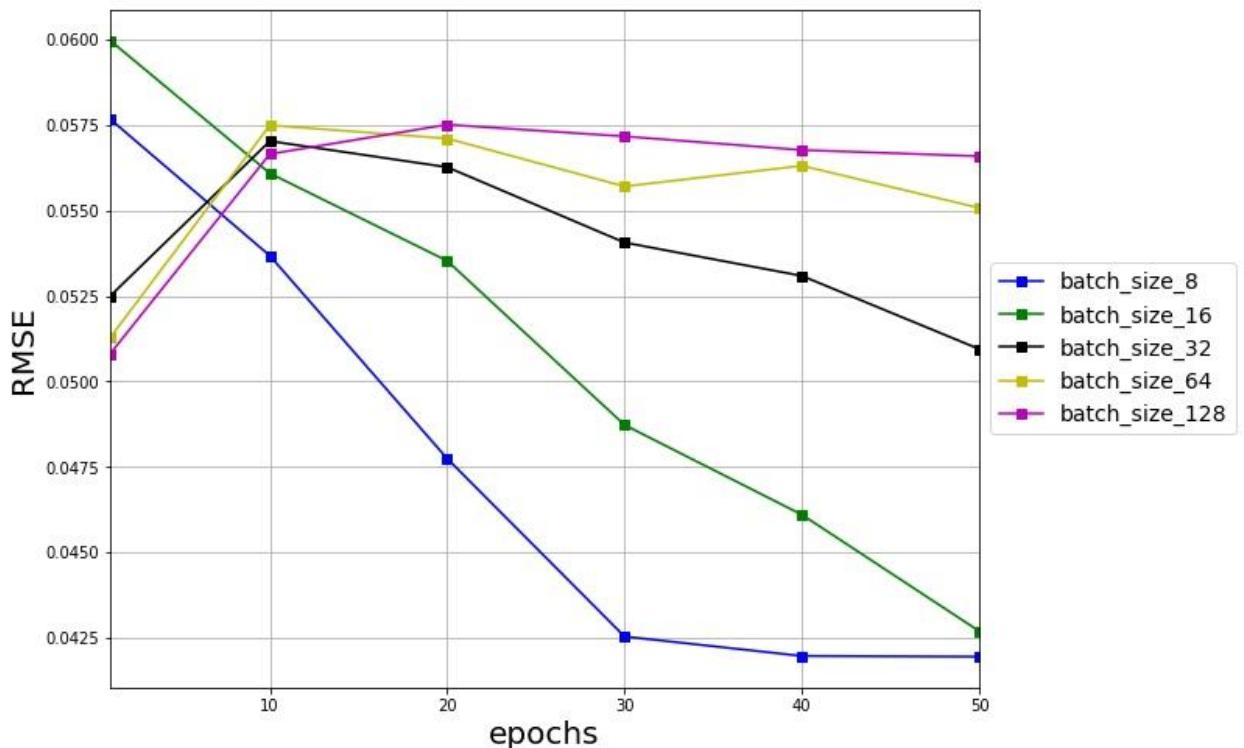


Рис. 55. $\min \text{RMSE} = 0.042$, $\min \text{MAPE} = 0.553\%$, optimum epochs = 50, optimum batch_size = 8

3) Далее сделаем настройку Units и Dropout prob.

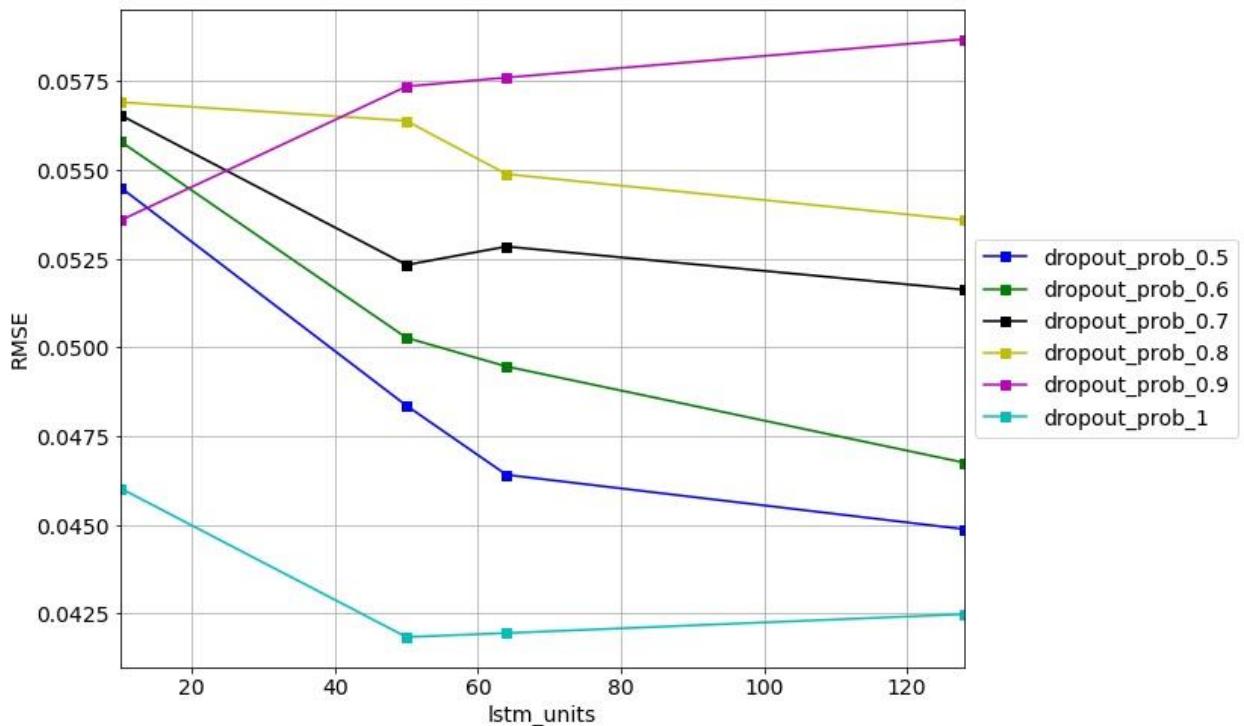


Рис. 56. $\min \text{RMSE} = 0.042$, $\min \text{MAPE} = 0.552\%$, optimum lstm_units = 50, optimum dropout_prob = 1.0

4) Настройка оптимизатора модели.

Существует ряд возможных оптимизаторов, которые можно использовать, в этом разделе я произведу выбор наилучшего конкретно для случая акций Московского кредитного банка.

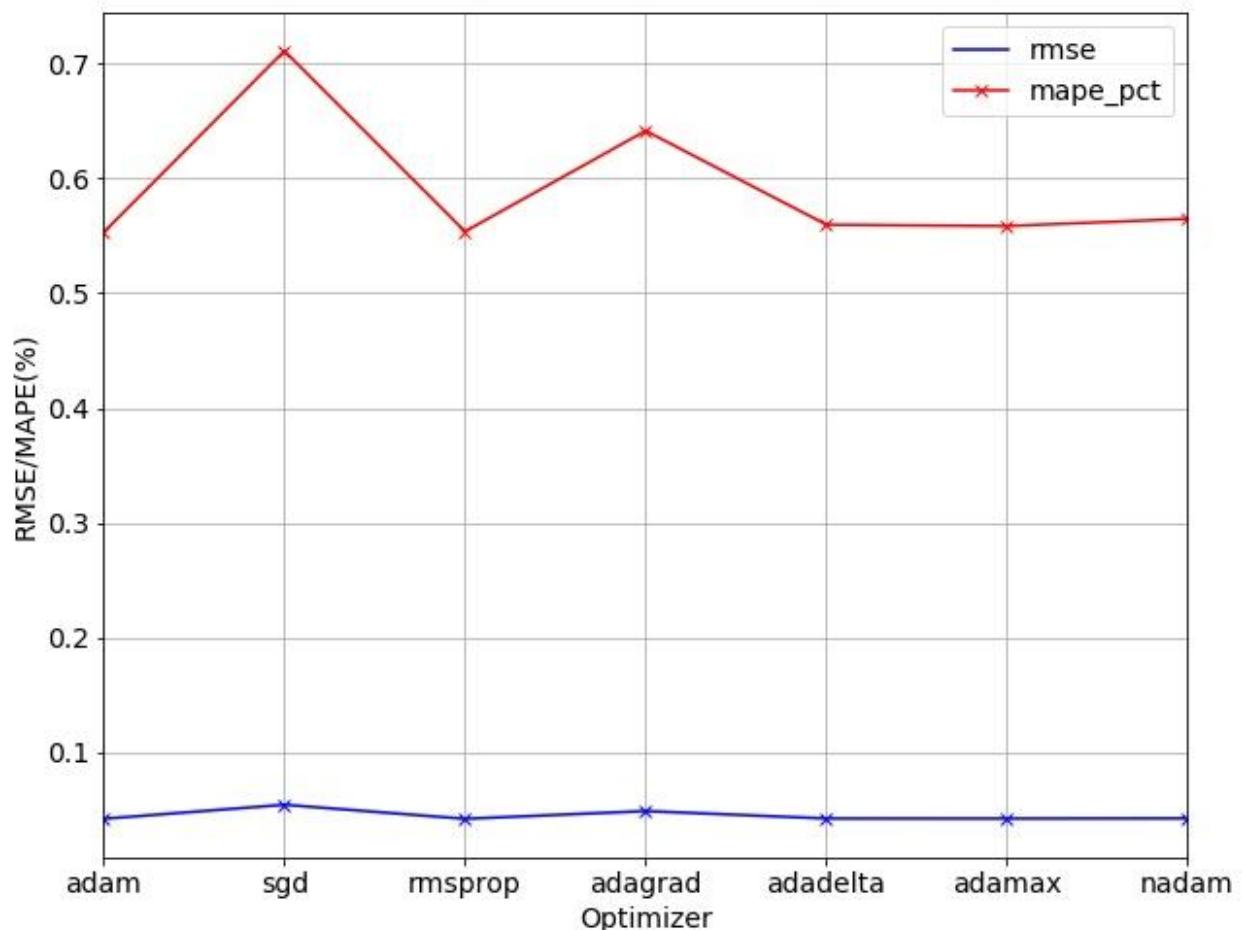


Рис. 57. $\min \text{RMSE} = 0.042$, $\min \text{MAPE} = 0.553\%$, optimum optimizer = rmsprop

Анализ совокупных результатов метрик качества после проведения настройки параметров модели:

параметр	исходный вариант	после настройки
N	3	3
lstm_units	50	50
dropout_prob	1	1
optimizer	adam	rmsprop
epochs	1	50
batch_size	1	8
rmse	0.0574201	0.0419105
mape_pct	0.756746	0.553131

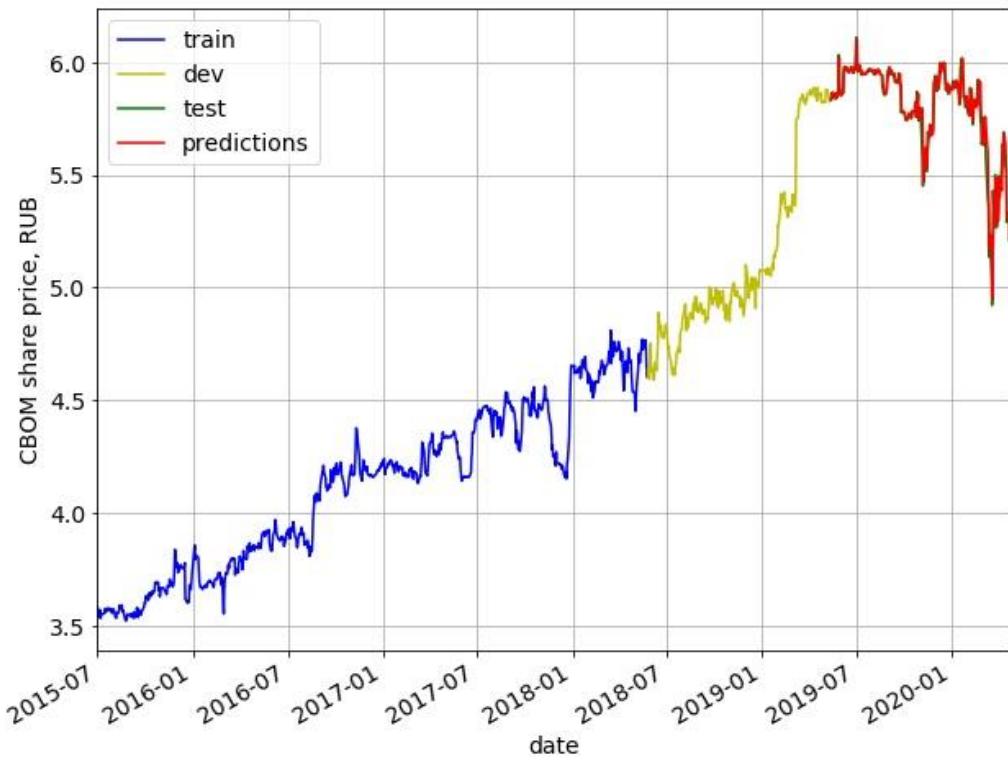
Рис. 58. Параметры и ключевые метрики качества LSTM модели.

Результаты модели на тестовой выборке:

RMSE on test set = 0.065

MAPE on test set = 0.710%

Как и следовало ожидать, метрики качества ухудшаются на новых данных.



Zoom in to test set



Рис. 59. Результаты LSTM модели на тестовой части выборки.

Результаты в финансовой форме:

Annual return	584616.6%
Cumulative returns	414341.7%
Annual volatility	72.9%
Sharpe ratio	12.44
Calmar ratio	NaN
Stability	0.91
Max drawdown	0.0%
Omega ratio	NaN
Sortino ratio	inf
Skew	2.15
Kurtosis	4.68
Tail ratio	93.83
Daily value at risk	-5.6%

Рис. 60. Основные метрики качества для алгоритмической стратегии на базе нейросети LSTM.



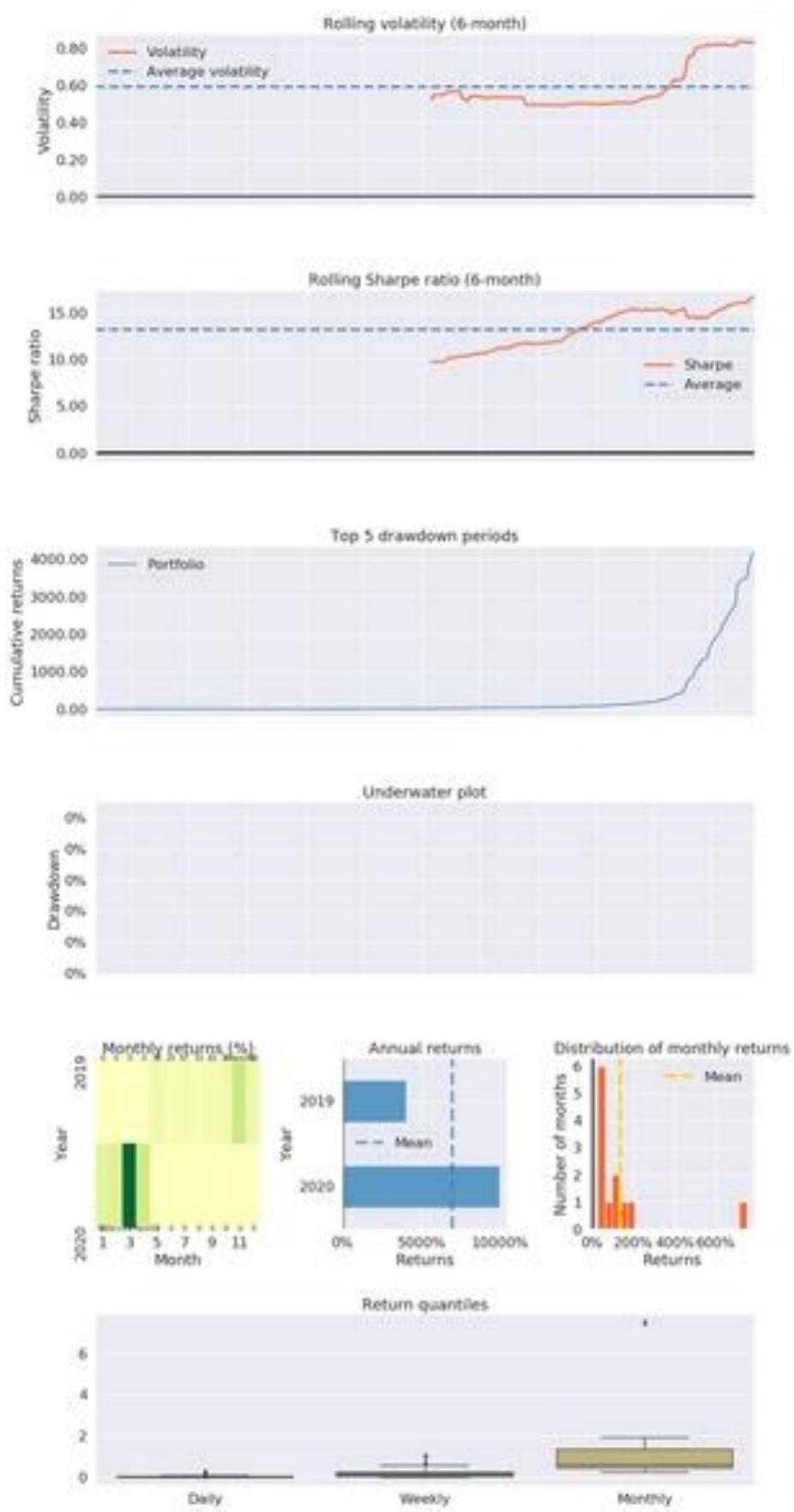


Рис. 61. Графическая иллюстрация статистик по стратегии LSTM на базе CBOM.

Итоги.

В рамках моей работы я рассмотрел ряд алгоритмов и подходов к созданию алгоритмической стратегии. Главной моей задачей было рассмотрение различных алгоритмов, стратегий, а также их реализации в Python. Я показал несколько из возможных способов анализа, которые уже активно применяются. На заре эры искусственного интеллекта я конечно же включил в работу отдельные примеры анализа на базе Машинного обучения и применения нейросетей. Помимо этого, одним из ключевых инструментов анализа была эконометрика и такие понятия, как стационарность временного ряда, коинтеграция, полезные следствия этих явлений, такие как возможность применения стратегий из семейства mean reversion.

Предрасположенность к лучшим результатам для стратегии из блока Mean Reversion по сравнению с блоком Momentum не была подтверждена на практике акций Московского кредитного банка (СВОМ), это значит что для временного ряда, который скорее соотносится по своей природе к Mean Reversion поведению может лучше подходить стратегия из блока Momentum, примером может служить сравнительно лучшие результаты MACD (Momentum) над Bollinger Bands (Mean Reverting), критерием оптимальности является накопленная доходность.

Касательно результативности стратегии на основе XGBoost Classifier можно сказать, что результаты получаются весьма неустойчивые, однако модель неплохо приближена к жизни и в целом справляется с задачей довольно неплохо, поскольку довольно часто, сравнивая её результативность с базовой стратегией Buy and Hold, она побеждает, тогда как на практике не раз подтверждалось, что Buy and Hold обгоняет и MACD (momentum) и Bollinger Bands (mean reversion). Таким образом можно сказать, что при прочих равных стоить отдать предпочтение бустинговому классификатору.

Что касается задач регрессии, как XGBoost Regressor, Linear regression, так и нейросеть LSTM model выдают чрезвычайно хорошие результаты, что довольно странно, и наталкивает меня на поиски ошибки в эти алгоритмах, там присутствует техническая неточность, которая не очень связана с подходом к построению стратегии и методологией, которую я применил, мною будут в скором будущем поправлены технические недочёты, и сейчас стоит оценивать остальные достоинства этих моделей, например методологию настройки параметров и оптимизацию.

Резюмируя проделанную работу могу сказать, что я справился с поставленной во введении целью: я выделил ряд стратегий из различных блоков, провёл оптимизацию их параметров, а также сравнил их результаты между собой для рынка акций и ETF. Я получил ответы на поставленные изначально гипотезы: на практике я не нашёл подтверждения тому, что стратегия Bollinger Bands из блока mean reversion превосходит результаты стратегии из блока momentum MACD для ситуации, когда временной ряд стационарен, коинтегрирован. Данная гипотеза не подтвердилась. Также была выдвинута гипотеза, что показатель Хёрста будет определять преимущество для блока mean reversion или momentum. Эта гипотеза также не была подтверждена на практике. Довольно часто MACD обгоняла ВВ для ситуации, когда показатель Хёрста был в промежутке [0.4, 0.5]. Не менее важно то, что часто обе стратегии MACD (momentum) и ВВ (mean reversion) с настроенными параметрами проигрывали дефолтной стратегии Buy and hold в бэктестах, где было опущено условие транзакционных издержек, то есть реально этим стратегиям очень не часто удаётся обыграть относительно беззатратную по времени силам и дополнительным средствам (комиссионные) Buy and Hold. Тем не менее, проведя анализ в блоке “Машинное обучение и нейросети” я получил результат, что часто XGBoost

Classifier побеждает Buy and Hold, что делает её более выигрышной стратегией среди всех рассмотренных мною.

Опровержение моих гипотез даёт мне возможность сказать, что весьма не факт, что если временной ряд обладает стационарностью, показатель Хёрста меньше 0.5 или есть возможность коинтеграции временных рядов, то блок Mean reversion будет выигрышнее. На практике это оказалось далеко не всегда так. Справедливо также, что нельзя утверждать, что если показатель Хёрста больше, чем 0.5, то будет преимущество у блока моментум. Также стоит сказать о неустойчивости результатов стратегий. Не было найдено устойчивости результатов не для одного алгоритма и стратегии из рассмотренных мною ни для ETF ни для акций.

Буду считать задачи выполненными и продолжу дальнейшую исследовательскую работу уже вне рамок данной работы.

Приложения.



Рис. 47. Проверка устойчивости стратегий на примере акций Алросы. Экспонента Хёрста показала значение 0.42 для цен акций Алросы, моя гипотеза о том, что данный временной ряд больше подходит для стратегии возврата к среднему значению не подтвердилась, и стратегия BB и MACD убыточны и сильно уступают дефолтной стратегии Buy and hold.

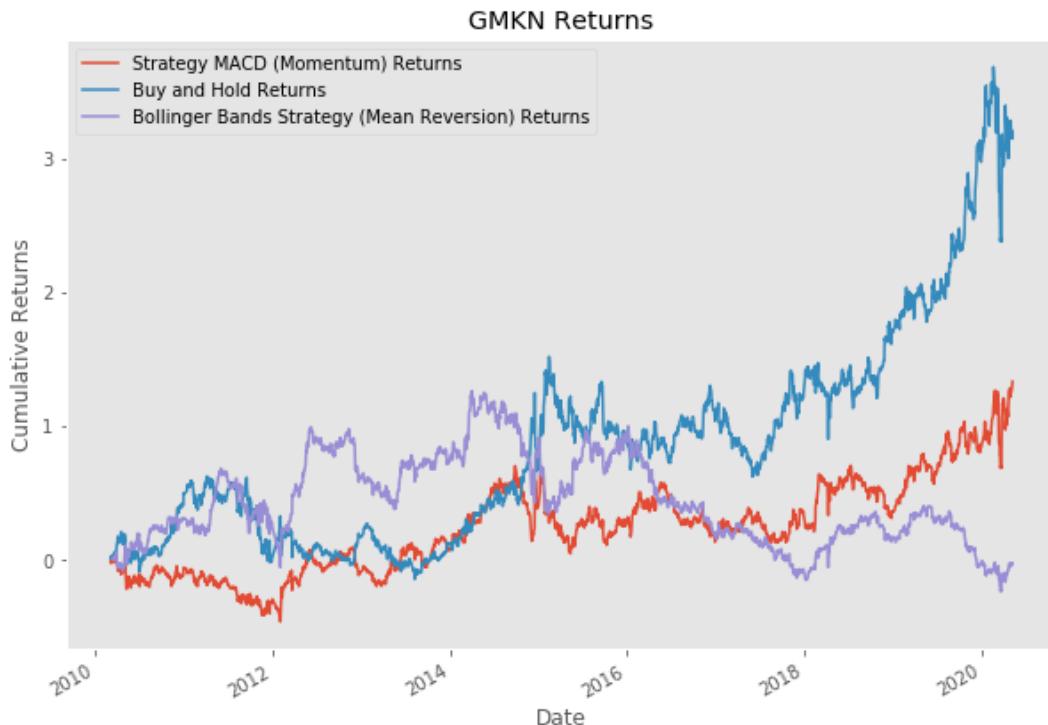


Рис. 48. Проверка устойчивости стратегий на примере акций Норильского Никеля. Экспонента Хёрста показала значение 0.499, то есть не должно быть явных преимуществ у какого-то блока, и стратегия BB и MACD сильно уступают дефолтной стратегии Buy and hold. Однако здесь MACD превзошла BB.



Рис. 49. Проверка устойчивости стратегий на примере акций Норильского Никеля. Экспонента Хёрста показала значение 0.43, гипотеза, что ВВ лучшие MACD не подтвердилась. И стратегия ВВ и MACD сильно уступают дефолтной стратегии Buy and hold и убыточны. ВВ проигрывает MACD.

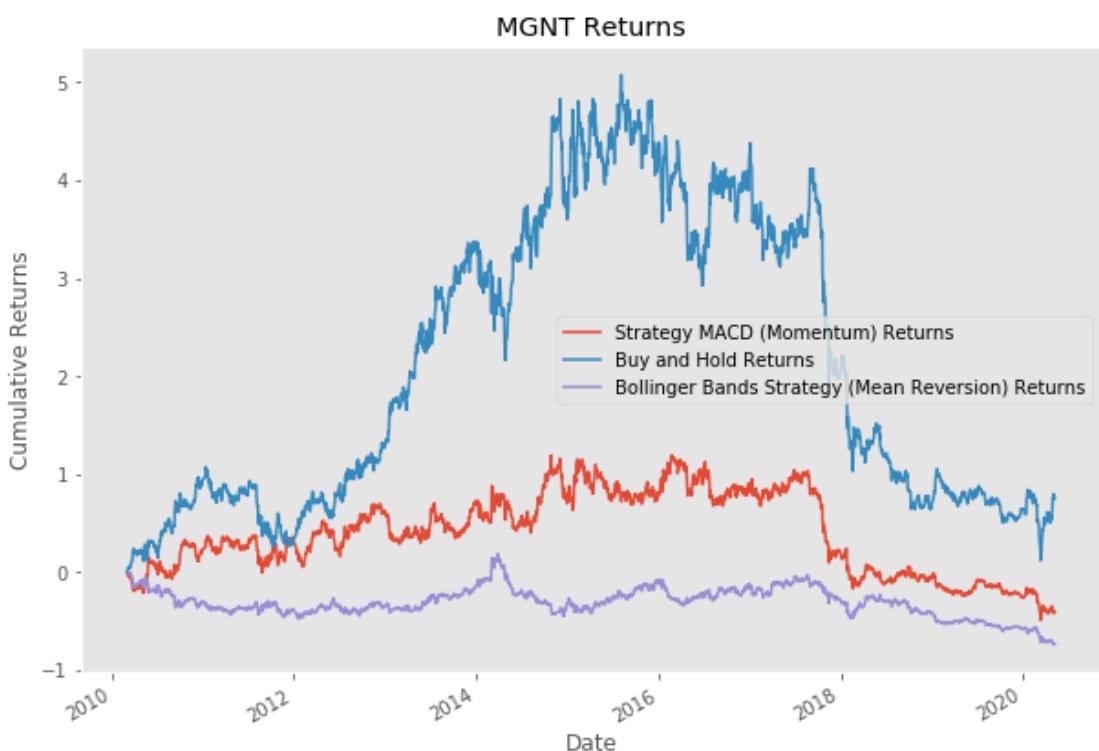


Рис. 50. Проверка устойчивости стратегий на примере акций Норильского Никеля. Экспонента Хёрста показала значение 0.52, гипотеза, что ВВ лучше MACD не подтвердилась. И стратегия ВВ и MACD сильно уступают дефолтной стратегии Buy and hold и убыточны. ВВ проигрывает MACD.

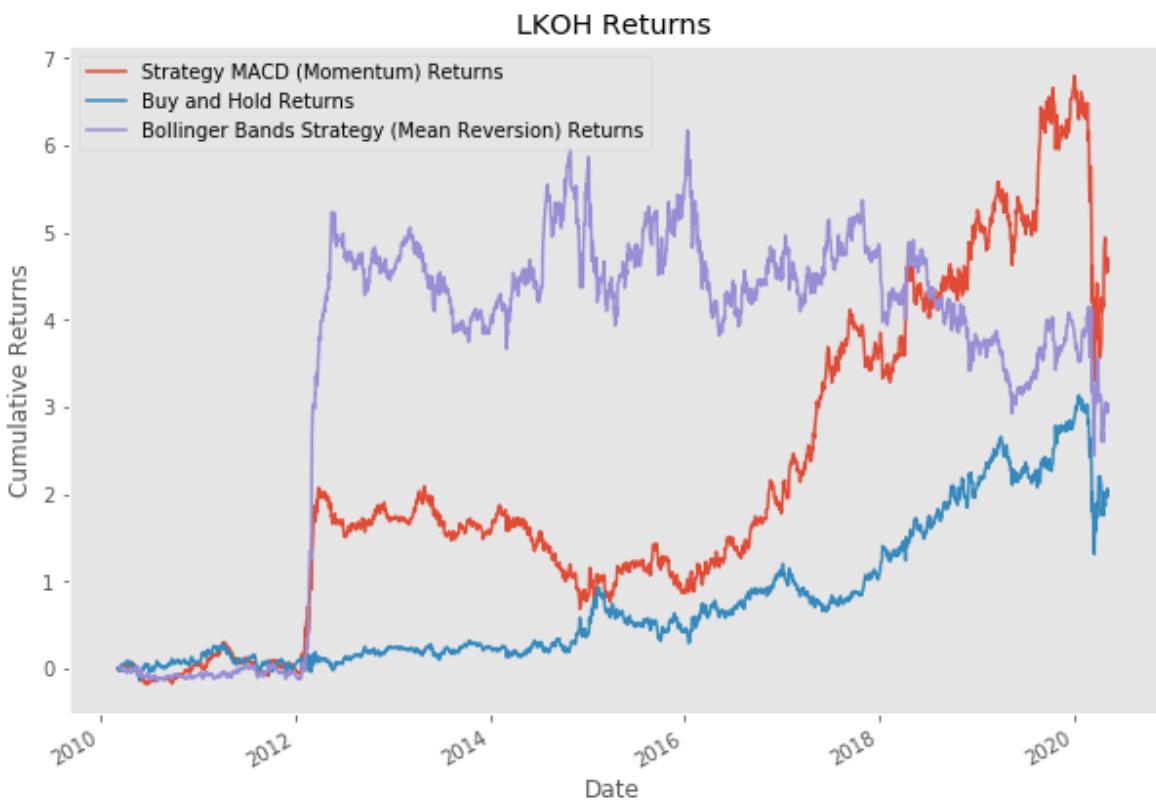


Рис. 51. Проверка устойчивости стратегий на примере акций Норильского Никеля. Экспонента Хёрста показала значение 0.36, В данном случае долгое время лидировала стратегия Bollinger Bands, однако в результате всё равно MACD победила. Заметно, что и стратегия ВВ и стратегия MACD значительно выросли в 2012 году в близкие моменты времени. Этот пример примечателен тем, что хуже всего проявила себя дефолтная стратегия Buy and hold.

По стратегиям XGBoost Regressor, Linear Regression из блока Машинного обучения и по стратегии LSTM model необходимо сказать, что чрезвычайная точность моделей скорее всего связана с технической реализацией стратегий, это не уменьшает исследовательского интереса и осмыслинности данных подходов, в частности процесс оптимизации параметров и бэктестирования несёт очень большую смысловую нагрузку.

Названия файлов ipynb, которые были использованы мною для имплементации результатов:

- XGBoost Classifier.ipynb
- StockPricePrediction_v2_lin_reg.ipynb
- StockPricePrediction_v1c_xgboost_working_version.ipynb
- Momentum strategies and mean reverting strategies-(GMKN).ipynb
- Momentum strategies and mean reverting strategies-(LKOH).ipynb
- Momentum strategies and mean reverting strategies-(MGNT).ipynb
- Momentum strategies and mean reverting strategies-(YNDX).ipynb
- Также можно заменить значения в файле для любого интересующего доступного тикера
- all together (cadf, johansen, half-life).ipynb
- Стратегия возврата к среднему значению (Mean reversion) использую полосы Боллинджера (Bollinger Bands)-CBOM.ipynb
- Стратегия возврата к среднему значению (Mean reversion) использую полосы Боллинджера (Bollinger Bands)-CopyETF_FXMM_FXUS.ipynb
- Momentum strategies and mean reverting strategies-change (alrs).ipynb (VWAP momentum trade здесь)
- StockPricePrediction_v4a_lstm_version_with_finance_results.ipynb

Библиографический список:

1. Hendershott, Terrence, and Ryan Riordan. "Algorithmic trading and the market for liquidity." *Journal of Financial and Quantitative Analysis* 48.4 (2013): 1001-1024.
<https://doi.org/10.1017/S0022109013000471>
2. Hendershott, Terrence, Charles M. Jones, and Albert J. Menkeld. "Does algorithmic trading improve liquidity?." *The Journal of finance* 66.1 (2011): 1-33.
<https://doi.org/10.1111/j.1540-6261.2010.01624.x>
3. Chaboud, Alain P., et al. "Rise of the machines: Algorithmic trading in the foreign exchange market." *The Journal of Finance* 69.5 (2014): 2045-2084.
<https://doi.org/10.1111/jofi.12186>
4. Johansen, Søren. "Statistical analysis of cointegration vectors." *Journal of economic dynamics and control* 12.2-3 (1988): 231-254.
[https://doi.org/10.1016/0165-1889\(88\)90041-3](https://doi.org/10.1016/0165-1889(88)90041-3)
5. Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen. "Time series momentum." *Journal of financial economics* 104.2 (2012): 228-250.
<https://doi.org/10.1016/j.jfineco.2011.11.003>
6. Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan. "The short of it: Investor sentiment and anomalies." *Journal of Financial Economics* 104.2 (2012): 288-302.
<https://doi.org/10.1016/j.jfineco.2011.12.001>
7. Ledoit, Oliver, and Michael Wolf. "Robust performance hypothesis testing with the Sharpe ratio." *Journal of Empirical Finance* 15.5 (2008): 850-859.
<https://doi.org/10.1016/j.jempfin.2008.03.002>
8. Balvers, Ronald J., and Yangru Wu. "Momentum and mean reversion across national equity markets." *Journal of Empirical Finance* 13.1 (2006): 24-48.
<https://doi.org/10.1016/j.jempfin.2005.05.001>
9. Avellaneda, Marco, and Jeong-Hyun Lee. "Statistical arbitrage in the US equities market." *Quantitative Finance* 10.7 (2010): 761-782.
<https://doi.org/10.1080/14697680903124632>
10. Poterba, James M., and Lawrence H. Summers. "Mean reversion in stock prices: Evidence and implications." *Journal of financial economics* 22.1 (1988): 27-59.
[https://doi.org/10.1016/0304-405X\(88\)90021-9](https://doi.org/10.1016/0304-405X(88)90021-9)
11. Chan, Ernie. *Algorithmic trading: winning strategies and their rationale*. Vol. 625. John Wiley & Sons, 2013.
<https://doi.org/10.1002/9781118676998>
12. Lakner, Peter. "Optimal trading strategy for an investor: the case of partial information." *Stochastic Processes and their Applications* 76.1 (1998): 77-97.
[https://doi.org/10.1016/S0304-4149\(98\)00032-5](https://doi.org/10.1016/S0304-4149(98)00032-5)
13. Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. "High-frequency trading and price discovery." *The Review of Financial Studies* 27.8 (2014): 2267-2306.

<https://doi.org/10.1093/rfs/hhu032>

14. Gârleanu, Nicolae, and Lasse Heje Pedersen. "Dynamic trading with predictable returns and transaction costs." *The Journal of Finance* 68.6 (2013): 2309-2340.
<https://doi.org/10.1111/jofi.12080>
15. Chan, Ernie. *Quantitative trading: how to build your own algorithmic trading business*. Vol. 430. John Wiley & Sons, 2009.
<https://doi.org/10.1002/9781119203377>
16. Chaudhuri, Kausik, and Yangru Wu. "Random walk versus breaking trend in stock prices: Evidence from emerging markets." *Journal of Banking & Finance* 27.4 (2003): 575-592.
[https://doi.org/10.1016/S0378-4266\(01\)00252-7](https://doi.org/10.1016/S0378-4266(01)00252-7)
17. Carrion, Allen. "Very fast money: High-frequency trading on the NASDAQ." *Journal of Financial Markets* 16.4 (2013): 680-711.
<https://doi.org/10.1016/j.finmar.2013.06.005>
18. Balvers, Ronald, Yangru Wu, and Erik Gilliland. "Mean reversion across national stock markets and parametric contrarian investment strategies." *The Journal of Finance* 55.2 (2000): 745-772.
<https://doi.org/10.1111/0022-1082.00225>
19. Taylor, Alan M. "Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price." *Econometrica* 69.2 (2001): 473-498.
<https://doi.org/10.1111/1468-0262.00199>
20. Thorp, Edward O. "The Kelly criterion in blackjack sports betting, and the stock market." *The Kelly Capital Growth Investment Criterion: Theory and Practice*. 2011. 789-832.
<https://doi.org/10.1016/B978-044453248-0.50015-0>
21. Carbone, Anna, Giuliano Castelli, and H. Eugene Stanley. "Time-dependent Hurst exponent in financial time series." *Physica A: Statistical Mechanics and its Applications* 344.1-2 (2004): 267-271.
<https://doi.org/10.1016/j.physa.2004.06.130>