Московский Государственный Университет
имени М.В. Ломоносова

Экономический Факультет

Кафедра Экономической Информатики

Выпускная квалификационная работа по теме:

«Применение нейронных сетей и эконометрических методов в прогнозировании доходностей акций»

Выполнил:
Студент 4 курса э408 группы

Гришин А. Ю.

Преподаватель:
Кандидат экономических наук,
Кандидат физико-математических наук,
Кандидат юридических наук,
Доцент кафедры Экономической Информатики

Сидоренко В.Н.

Москва, 2023

Оглавление 2

1 Введение

Современный мир нельзя представить без валюты и денег в принципе. Каждый день всякий человек на Земле имеет с ними дело: кто-то с большей суммой, кто-то с меньшей, однако все мы неразрывно связаны с деньгами. Апогеем валютного триумфа в истории человека можно назвать создание предприятий: объединений людей на основе некоторой общей цели, которой впоследствии стали считать - получение прибыли. Логика проста: один человек получает сумму N за один рабочий день $\Rightarrow$ 2 человек получат $2N$, однако, что даст им стимул к объединению усилий? Только факт того, что вместе они заработают $(2+\varepsilon)N$ денег, где ε - выигрыш от работы вместе. Но можно ли как-то получить сумму больше, чем указанная? Да, производить больше [1] , пытаться подстраиваться под клиента [2] : предугадывать потребности клиента и делать на конкретный товар скидку, пытаться угадать, когда спрос на тот или иной вид продукции будет больше в зависимости от времени года [3]. То есть можно подстраиваться с двух сторон: работая над собой и изменять что-то внутри организации, а можно смотреть на то, что необходимо клиентам и делать именно это, именно тогда, когда это нужно. Получаются 2 крайности, которые необходимо уметь комбинировать.

Развивая финансовый рынок, получаем закономерное действие компаний - выход на IPO [4] и становление их публичными. Таким образом, уже владелец компании (он же владелец акции) стремится увеличить свою выгоду от владения долей, иначе зачем ему вообще тратить деньги на покупку ценных бумаг компании. Ведь важно то, чтобы вложенная сумма "отбивалась" с некоторой надбавкой, то есть человек зарабатывает то, что вложил, а также что-то сверх. Это сверх и есть искомый ε, который человек стремится увеличить как можно сильнее. Справедливо задать вопрос: "А зачем ему все это? Зачем вкладываться куда-то, чтобы получить больше?". Иными словами, зачем иметь больше того, чем есть сейчас? Логический ответ на это пытались дать многие экономисты, начиная с Адама Смита (1729–1737) [13] и его идеей о рациональном человеке (homo economicus), однако большинство теорий, изложенных им же и позднее его последователями (членами классической школы) выдвигали это как предпосылку всего анализа. Следовательно, без выполнения данного условия о стремлении человека к максимизации собственной выгоды, большинство экономический теорий не являются работающими.

Однако, опуская этот философский вопрос и возвращаясь к методам моделирования человеческого поведения, нельзя не заметить, что сами классики не

[1] Делай больше, клиент все купит - это очень похоже на тезис, что каждый товар находит своего покупателя $\Rightarrow$ тезис о "базаре справедливого обмена" Роберта Оуэна. Подробнее о том, что именно предлагалось сделать рассказано в работе [19].

[2] Человека/компании, который пользуется услугами данной компании.

[3] Очевидно, что мороженое будут покупать заметно в меньших количествах зимой по сравнению с летом.

[4] IPO - Initial Public Offering - первый выход на биржу, когда компания продает свои акции неограниченному кругу лиц, делая их совладельцами.

углублялись в именно математическое моделирование и как следствие - описание экономической деятельности человека посредством формул, а делали это лишь через построения причинно-следственных связей, у истоков которых стояла проблема стоимости.

Намного позже Альфред Маршалл, английский экономист, живший в 1842-1924, являясь представителем неоклассической школы (кем его "обозвал" Торстейн Бунде Веблен), выпускает учебник "Принципы экономической науки" [8], в котором сводит воедино все труды и знания, полученные на пути развития и становления Экономики, а также собственные работы, посвященные применению наиболее полно описанного математического подхода к анализу экономической деятельности человека.

В настоящее время для анализа экономических показателей используются всевозможные методы будь то математические или нет. Однако Математика - точная наука, значит, если получится понять, когда и сколько продукции компании будет потреблять клиент (или получать дивидендов держатель акций), то можно будет без проблем еще сильнее уменьшить издержки на производство, что в свою очередь, приведет к получение еще большей экономической прибыли.

Так как в современной науке набирает популярность применение глубоких нейронных сетей к различным видам деятельности (медицина, металлургия, космическая промышленность, киноиндустрия, материаловедение, биология, биохимия, ...) [2], значит, не лишнее - проверить, а имеет ли смысл вообще применять данные методы к финансовым задачам, а конкретно именно к предсказанию доходностей акций.

Настоящая работа является попыткой произвести сравнительный анализ наиболее популярных на данный момент математических моделей предсказания доходностей акций, примененных к 30 компаниям развитого (США) и развивающегося (Китай: Шанхай) рынков соответственно. Сравнение производится по характеристике качества прогнозирования доходности на 1 рабочий день биржи.

1.1 Актуальность

Данное исследование актуально с двух позиций: научная - выявив наиболее удачную с точки зрения предсказания на 1 день модель, в дальнейших исследованиях можно стараться развивать только ее, чтобы получать более точные результаты, а не пытаться выбрать между тем, над какой именно моделью из их множества работать, практическая - трейдерам или просто акционерам будет намного легче воспринимать временные ряды доходностей, так как они смогут с определенной точностью предсказывать конкретное значение подобного ряда на момент времени $t + 1$, что возможно даже приумножит их доход.

1.2 Цели

Целью настоящего исследования является помощь трейдерам или акционерам в прогнозировании доходностей акций. А умение качественно (с определенной точностью) предсказывать доходность приводит к умению формировать ответ на вопрос вида Buy or Hold? [5] достаточно быстро и аккуратно. Ведь основная проблема финансовых временных рядов - непредсказуемость, таким образом, нужно постараться решить данный вопрос так, чтобы на выходе было наиболее прибыльно и наименее рискованно, следовательно - наиболее точно. Отсюда появляется большая надежность финансовых инструментов с точки зрения человека, который заключает опционы или просто старается приумножить свое благосостояние посредством формирования собственного портфеля. Далее следует успешность фирмы, а отсюда, ведя разговор о финансовых рынках, возможность страны развить их еще сильнее и, возможно - увеличить благосостояние своих резидентов, а значит, возможность стать из развивающейся развитой.

1.3 Задачи

Задачей текущего исследования является проведение сравнительного анализа между моделями машинного и глубокого обучения с целью выявления той, которая дает наиболее точный прогноз на период длиной 1 рабочий день биржи. Для этого: 1) Загружаем данные 2) Предобрабатываем данные 3) Проводим статистический анализ данных 4) Среди всех обученных алгоритмов выбираем тот, у кого лучший результат 5) Сформировать сравнительную таблицу между моделями по всем имеющимся данным. Несмотря на громоздкость поставленных задач, весь их комплект можно оформить в один алгоритм и, таким образом, применять его к любому набору входных данных (при условии, конечно, что входные данные - временной ряд). То есть появляется новая задача следующего вида: **на вход** программе подается временной ряд, **на выход** - показатели выбранной далее тестовой статистики для каждой из рассматриваемых моделей. [6]

2 Подготовительная часть

В чем заключается суть данной работы? С концептуальной стороны вопроса ответ очевиден: дан временной ряд, необходимо предсказать его следующее значение наиболее точным образом. Но прежде чем перейти к рассмотрению использованных моделей, формализуем поставленную задачу, ведь не всегда понятно, что имеется в виду под "наиболее точно" и "временной ряд".

[5] Buy of Hold - Покупаем акцию или продаем?

[6] Набор используемых моделей представлен в блоке ниже.

2.1 Формализация проблемы

Пусть на вход программе, назовем ее f, подается временной ряд вида: $y_t : t = \overline{1, N}$. Пока что никаких предпосылок относительно данного временного ряда нет. Тогда задача алгоритма f предсказать $\hat{y}_{t+1}$ так, чтобы значение $|\hat{y}_{t+1} - y_{t+1}|$ было минимальным. То есть $f : \mathbb{R}^{N \times 1} \to \mathbb{R} \Rightarrow f(y) = \hat{y}_{t+1},$ где $y = (y_1, \ldots, y_N)^T$. В терминах глубокого обучения, имеем задачу класса sequence to one [7] . **Q**: Но в принципе, что такое временной ряд? **A**: Временной ряд - это последовательность значений, относящихся к одному объекту в разные моменты времени. То есть, за тот период, когда за объектом наблюдали и снимали показатели. В нашем случае временной ряд - это набор доходностей акций за определенный период времени, то есть не ставится ограничение на знак данных величин, ведь она (доходность) может быть как отрицательной, так и положительной. **Q**: Почему именно доходности? **A**: Говоря о них, нет необходимости задумываться о том, сколько реально стоит та или иная акция/ценная бумага: 20 руб. или 2'000 руб. В анализе доходностей нас интересует: насколько они изменяются относительно некоторого момента времени. **Q**: Какого момента? **A**: Логично представлять доходность как прирост вида:

$$r_t = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 \tag{1}$$

То есть, как изменяется в процентах цена на некоторый актив (не обязательно акцию, хотя в нашем случае именно на нее) относительно предыдущего рабочего дня биржи. Однако внутри самой "акции" фигурирует несколько показателей, характеризующих ее в конкретный момент времени: цена открытия, максимальная и минимальная цены, цена закрытия, скорректированная цена закрытия и общая сумма сделок. Акцент делается на анализе цен открытия (так как с точки зрения автора работы важнее всего - хорошо начать рабочий день), хотя несомненно наличие взаимосвязи между ценой открытия и закрытия или иными доступными показателями. Когда есть возможность - другие показатели включаются в анализ, когда нет такой возможности (классическая модель не предназначена) - не включаются. Подводя итог: **Дано**: $y = (y_1, \ldots, y_n)^T$, **Найти**: $\hat{y}_{N+1} : |\hat{y}_{N+1} - y_{N+1}| \to \min$. Подставляя вышеупомянутое равенство, имеем задачу оптимизации:

$$|f(y) - y_{N+1}| \to \min_{\beta \in \mathbb{R}^k} \tag{2}$$

Где β - набор параметров модели, хотя некоторые из них имеют неоптимизируемые параметры (гиперпараметры), но в общем случае задача имеет подобный вид. Но проблема в том, что y_{N+1} неизвестно, а значит, невозможно подобрать алгоритм абсолютно точного прогнозирования, значит, необходимо на основе

[7] Sequence to one - задача получения одного значения, исходя из набора данных. К таким задачам относят: семантический анализ текста, классификацию картинок и так далее.

имеющейся информации сформировать алгоритм, который наиболее точным образом описывает предложенные ему данные, а далее делает предсказание, причем предсказание, как можно меньше отличающееся от реального значения. Только теперь, проведя подобные рассуждения, мы находимся в области машинного обучения и можем говорить о гипотезах рынка [8] . Ведь для каждого типа рынка характерны свои особенности, следовательно, закономерный вопрос: почему мы вообще имеем право пытаться предсказывать что-то для развитой или развивающейся экономик. Изложение двух нижестоящих теорий представлено в сжатом виде, что делает повествование о них поверхностным, но достаточным для понимания всех особенностей работы.

2.1.1 Гипотеза эффективного рынка

Это одна из самых неоднозначных в плане количества последователей инвестиционная теория, ставящая своей целью описать принципы движения цен на активы, первоначальная версия которой "представлена" Луи Башелье в 1900 году. В его работе показана независимость доходности акций от течения времени, таким образом, Башелье пришел к выводу: "Вероятность роста цены в любой момент времени равна вероятности ее падения, а математическое ожидание спекулянта равно нулю". Много раз менявшая свою формулировку, начиная с Пола Самуэльсона: "На конкурентных рынках на всякого продавца найдется покупатель. Если можно быть уверенным, что цена вырастет, значит, она уже выросла" , в 1960-ых годах она (гипотеза) приобрела формальный вид в труде Юджина Фамы, использовавшего в исследованиях модель случайного блуждания, выведенную Башелье. По итогам эксперимента, Юджин Фама [5] привел доказательство того, что вся доступная информация уже заложена в бумагах (позднее: "рынок полностью отражает всю доступную информацию"), то есть бесполезно пытаться предугадывать цены, любое предсказание не сбудется. Более того, единственный фактор, который способен повлиять на цену - это выходящие в будущем новости. А значит, перманентное доминирование над рынком не является возможным, а активное инвестирование не является состоятельным. Более подробная информация изложена в статье от 13 сентября 2022 года [6]. На данный момент существует 3 основных гипотезы эффективности рынка:

1. **Слабая гипотеза** - в цене содержится вся историческая информация об активе и только фундаментальный анализ иногда может обеспечить избыточную доходность.

2. **Полусильная гипотеза** - в ценах содержится вся публичная информация, таким образом, избыточную доходность может обеспечить только закрытая от широкой публики информация (инсайдерская).

[8] Гипотезы рынка - обоснование исследования на качественном уровне, то есть утверждения о возможности проводить какой-либо технический анализ.

3. **Сильная гипотеза** - в ценах содержится как общедоступная, так и закрытая информация. Таким образом, ничто не может дать инвесторам избыточную доходность по сравнению со среднерыночным показателем.

Тогда напрашивается вывод: если быть сторонником ГЭР, то настоящая работа не имеет смысла, ведь технический анализ не может дать дополнительной доходности для любой степени ее силы. Однако, по словам Мартина Свэлла, проанализировавшего историю данной гипотезы в работе [11]: "Строго говоря, гипотеза [прежде всего, в ее сильной форме] ложна, но по духу глубоко верна ... До тех пор, пока текущая гипотеза не будет заменена лучшей гипотезой, критика имеет ограниченную ценность". Отсюда все-таки следует обоснование, почему существует так много математических моделей, пытающихся прогнозировать доходность активов. Отсюда следует, что в современном мире количество методов предсказания настолько велико, что исследователю сложно выбрать нужный, это еще одно подтверждение, почему настоящая работа имеет смысл.

2.1.2 Гипотеза фрактального рынка

Часто в качестве доказательства ложности ГЭР приводятся в пример финансовые кризисы, так как по ГЭР вероятность возникновения подобного кризиса пренебрежимо мала или приблизительно ноль. Таким образом, появляется еще одна гипотеза: Гипотеза Фрактального Рынка (ГФР), чьим родоначальником является Бенуа Мандельброт [17], по которой можно объяснять кризисы. Ее основные характеристики: 1) график доходностей активов имеет фрактальную (всегда $1 < D < 2$) размерность 2) Различные окна (интервалы) исходного графика могут быть самоподобными 3) Каждому финансовому графику присуща своя уникальная структура и соответственно ее свойства 4) Финансовый график обладает памятью о своих исходных условиях (имеет долгосрочную память; формальный способ проверки данного утверждения вводится позже). Для выполнения данной гипотезы предполагается, что рынок является стабильным, если он включает в себя очень много инвесторов с различными горизонтами планирования (это гарантия ликвидности). Объяснение кризисов происходит следующим образом, описанным в статье Палювиной А.С. [20]: "Когда инвесторы меняют свои инвестиционные горизонты (например, фундаментальная информация становится ненадёжной, а долгосрочные инвесторы уходят с рынка или сокращают свои горизонты), баланс между краткосрочной и долгосрочной перспективами искажается, рынок становится менее ликвидным и возникает кризис". Таким образом, из данной гипотезы следует вывод, что информационный и инвестиционный горизонты оказывают влияние на поведение инвестора.

2.1.3 Проверяемая гипотеза настоящего исследования

Гипотеза, подтверждение которой настоящая работа ставит одной из своих ключевых задач, заключается в проверке суждения, что нейросетевой (далее NN) поход является наиболее эффективным применительно к исследуемой области финансовых рынков, а точнее к временным рядам цен/доходностей акций. То есть главный вопрос: нейросеть лучше справляется с прогнозированием доходностей акций по цене открытия на один рабочий день биржи по сравнению с другими использованными моделями или нет?

2.2 Наиболее популярные методы решения

В настоящем исследовании последовательно рассматриваются такие модели как: Exponentially Weighted Moving Average, Auto-regressive Integrated Moving Average, Generalized Auto-Regressive Conditional Heteroskedasticity, Autoregressive Fractionally Integrated Moving Average, Fractionally Integrated GARCH, Singular Spectrum Analysis, Fourier & Wavelet analysis, а также Transformers и моделирование сезонности. Далее рассказывается подробнейшим образом о математической подоплеке каждой из моделей. Важно понимать, что у данной работы нет задачи предоставить полноценное обоснование, почему та или иная модель однозначно работает, однако где-то все-таки приводится подробное описание, а где-то лишь качественные рассуждения и ссылки на более подробное обоснование. Сразу стоит отметить, обращаясь к последнему пункту плана (Transformers и моделирование сезонности), что он рассматривается с точки зрения: почему он тут не нужен [18], несмотря на революцию в Natural Language Processing, устроенную методом self-attention [16] и несмотря на важный аспект моделирования временных рядов - сезонность. Механизм self-attention создан для вычленения смысла из данных. То есть: дан набор слов, к ему в соответствие ставится вектор (она называется embedding), а после к данному набору чисел применяются известные алгоритмы ML или DL. Однако в текущей задаче в ряде доходстей или цен (одно можно получить из другого) нет контекста, то есть и сам механизм теряет свою значимость. Трансформеры оперируют понятием контекста, отвечающего за связь слова с окружающими его словами. Математически-интуитивно это можно рассматривать следующим образом: $\{w_j\}_{j=1}^n$ - все предложение, смотрим на конкретное слово (не в начале и не в конце: между словом и концом/началом должно быть как минимум еще одно слово) $w_i : i \in (1, n)$. Контекстом данного слова называется набор $w_k : k \in \{i-l, \ldots, i-1, i+1, \ldots, i+l\} : l \in \mathbb{N}$. тут пропущено само слово, так как для него окно одного размера влево и вправо является контекстом. Это значит, что при таком последовательном выборе контекста необходимо максимизировать вероятность получения конкретного слова w_i. Данная задача ставится в процессе создания качественного эмебддинга (сопоставления чисел словам; алгоритм Word2Vec [9]), но в текущей задаче нет необходимости рассматривать такую смысловую взаимосвязь между величинами ряда, так как, исходя из качественных рассуждений, на вход подается единственное число, которое никак, с точки зрения смысла, не связано с другими. Исходя из этого и делается вывод, что механизм трансформеров проигрывает более простым линейным моделям [18]. Что касается сезонности, то, несмотря на подтверждение ее наличия в финансовых временных рядах [10], все-таки рассматриваем модели без корректировки на данную составляющую. Подобное рассуждения приводится, опираясь на Гипотезу Эффективного Рынка, утверждающую, что финансовые временные ряды не зависят от времени, иначе рынок перестал бы быть эффективным [12]. Данное суждение весьма прозаично относительно показанного факта в блоке 2.1.1 о некорректности всей гипотезы в целом. В итоге получается, что из ГЭР рассматривается (принимается как верное) только суждение об отсутствии сезонности, а все остальные положения принимаются как ложные. Таким образом, в данной работе не рассматриваются как Трансформеры, так и сезонность ряда (даже если эта составляющая присутствует, то она все равно относится к шуму). Далее приступаем к описанию моделей, рассматриваемых в настоящей работе. Для удобства делим их на требующие стационарность и не требующие. **Q**: Что такое стационарность? **A**: Существует определение как для случая, когда 1) временной ряд воспринимается как последовательная реализация случайных величин как-то распределенных, так и для случая, когда 2) временной ряд - некоторый сигнал, например, от светодиода или от звукового динамика. Последовательно:

1. Временной ряд - реализация набора зависимых случайных величин с различным - изменяющимся во времени - распределением. Случайная выборка - простейший частный случай для временного ряда, так как все полученные данные являются следствием реализации не одной случайной величины с некоторым распределением, а некоторого набора. Данные распределения заранее неизвестны и их пока что не представляется физической возможности определить, так как для их определения необходимо было бы посетить параллельную вселенную, где данная величина реализовалась бы иначе. То есть для максимальной точности потребовалось бы посещения $n \to \infty$ параллельных вселенных. Отсюда и возможность оперировать только частными случаями временных рядов - случайной выборкой. Таким образом, под всеми последующими упоминаниями термина "временной ряд" понимается термин "случайная выборка" - частный случай реализации временного ряда. Для данного определения существуют 2 понятия стационарности: в узком (сильном) и в широком (слабом) смыслах. Взяты из работы Н. В. Артамонова, Е. А. Ивина [22], но переработаны для удобства восприятия.

Определение 1 ***Если***: *$\forall p, t_1, \ldots, t_p, l \in \mathbb{N} \Rightarrow$ распределение $y_{t_1}, y_{t_2}, \ldots y_{t_p}$ идентично распределению $y_{t_1+l}, y_{t_2+l}, \ldots, y_{t_p+l}$.* ***То***: *Временной ряд называется стационарным в узком смысле (строго стационарным).*

Определение 2 ***Если:*** *$\forall t \in \mathbb{N} \Rightarrow \mathbb{E}(y_t) = \mu \in \mathbb{R}, \mathbb{V}(y_t) = \gamma_0, cov(y_t, y_{t-k}) = \gamma_k : \mu, \gamma_0, \gamma_k$ не зависят от t.* ***То:*** *Временной ряд называется стационарным в широком смысле (слабо стационарным).*

2. Временной ряд - просто набор чисел, содержащий информацию об объекте и полученный в ходе эксперимента (наблюдения за объектом). В этом случае стационарности дается следующее определение:

   Определение 3 ***Если:*** *спектр амплитуд сигнала является постоянным во времени.* ***То:*** *Сигнал называется стационарным.*

   В данном случае имеет место деление: стационарный, значит, либо детерминированный (то есть: однозначно определяется по некоторому закону и является периодическим/квазипериодическим/непериодическим), либо случайный (например, модели финансового рынка, основанного на Броуновском Движении [14] или Белом шуме [7]); нестационарный, значит, непрерывный или переходный. Более подробно о данном понятии и формальном его осознании рассказывается в блоках 2.2.7 (Анализ Фурье) и 2.2.8 (Wavelet анализ).

Исходя из вышеперечисленных определений, выделяем 2 группы моделей: стационарные и нестационарные. Стационарные модели - те, для которых необходимо выполнение предпосылки о слабой (стр. 11 опр. 2) или сильной (стр. 10 опр. 1) стационарности. Нестационарные - те, для которых нет строгой необходимости в этом (они могут работать как со стационарными данными, так и с нестационарными).

2.2.1 Exponentially Weighted Moving Average

Предполагаем, что в наличии есть данные о некотором показателе $\theta_t : t = \overline{1, n}$. В данный момент перед нами не стоит задача предсказания следующего показателя, нужно только визуально выявить тренд, чтобы приблизительно понять направление движение показателя в осях (время, значение). Конечно, в этой формулировке отсутствует строгость, пока что останавливаемся на том, что есть. Экспоненциальная скользящая средняя занимается тем, что уже заложено в ее названии: сглаживает показатели посредством усреднения. То есть в текущем значении (t) используется как значение в предыдущий момент времени $(t-1)$, так и значение наблюдения в момент времени (t). Иными словами, получается выпуклая линейная комбинация текущего и предыдущего значений:

$$v_t = \beta \cdot v_{t-1} + (1-\beta) \cdot \theta_t : \ \beta \in (0, 1) \tag{3}$$

Где β - показатель, пропорциональный примерному количеству дней, по которым происходит усреднение, а $v_0 = 0$ - первое значение экспоненциальной скользящей средней. Для изучения данного рекуррентного соотношения "раскручиваем" его в обратном направлении:

$$\begin{aligned} v_t &= (1-\beta)\cdot\theta_t + \beta\cdot\overbrace{\{(1-\beta)\theta_{t-1} + \beta v_{t-2}\}}^{v_{t-1}} \\ &= (1-\beta)\cdot\theta_t + \beta\cdot\{(1-\beta)\cdot\theta_{t-1} + \beta\cdot\overbrace{\{(1-\beta)\cdot\theta_{t-2} + \beta\cdot v_{t-3}\}}^{v_{t-2}}\} \end{aligned} \tag{4}$$

Для лучшего понимания, рассматриваем частный случае:

$$\begin{aligned} v_4 &= \beta\cdot v_3 + (1-\beta)\cdot\theta_4 \\ v_3 &= \beta\cdot v_2 + (1-\beta)\cdot\theta_3 \\ v_2 &= \beta\cdot v_1 + (1-\beta)\cdot\theta_2 \\ v_1 &= \beta\cdot v_0 + (1-\beta)\cdot\theta_1 \\ v_0 &= 0 \end{aligned} \tag{5}$$

Тогда:

$$\begin{aligned} v_4 &= (1-\beta)\theta_4 + \beta((1-\beta)\theta_3 + \beta((1-\beta)\theta_2 + \beta((1-\beta)\theta_1 + \overbrace{\beta\cdot 0}^{v_0=0}))) \\ &= (1-\beta)\theta_4 + (1-\beta)\beta\cdot\theta_3 + (1-\beta)\beta^2\cdot\theta_2 + (1-\beta)\beta^3\cdot\theta_1 \\ &= (1-\beta)\sum_{t=1}^{4}\beta^{4-t}\cdot\theta_t \end{aligned} \tag{6}$$

Аналогично происходит раскрытие и для больших показателей v_n. А значит, получается формула вида:

$$v_n = (1-\beta)\sum_{t=0}^{n}\beta^t\cdot\theta_{n-t} \tag{7}$$

Отсюда качественно делаем вывод, что последним значениям наблюдаемого показателя (θ) соответствует большее значение коэффициента, то есть чем ближе θ_t к v_n с точки зрения индекса, тем больше при нем коэффициент. Теперь, возвращаясь к названию модели, отмечаем, что она должна усреднять по конкретному количеству временных единиц, следовательно, β (единственный гиперпараметр) должен регулировать указанный показатель (то есть количество дней, по которым происходит усреднение). Обычно, рассматривается следующее соотношение: $\beta = 1 - \alpha/(N+1)$, где N - количество дней, а α - фактор сглаживания. Опираясь на [15], полагаем $\alpha = 2$ как наиболее распространенное значение.

Далее рассматриваем подробнее первые значения EWMA, а конкретнее - для лучшего осознания - самое первое: $v_0 \Rightarrow v_1 = (1-\beta)\cdot\theta_1 \Rightarrow$ при $\beta \to 1-0$

получаем, что $(1-\beta)\cdot\theta_1 \ll \theta_1$, что крайне плохо отражается на самих значениях. Таким образом, вычленение тренда становится более трудным на начальных этапах из-за плохого "разогрева" модели. Данная проблема называет bias (смещение), а ее решение bias correction (коррекция смещения) соответственно. Основой подобной корректировки является множитель вида $(1-\beta^t)$. Очевидно, что при $\beta \in (0,1)$, а именно так и есть по построению, $\lim\limits_{t\to+0}(1-\beta^t) = 0$, а $\lim\limits_{t\to+\infty}(1-\beta^t) = 1$.

$$\begin{aligned} v_n &= (1-\beta^n)^{-1}\cdot(\beta\cdot v_{n-1} + (1-\beta)\cdot\theta_n) \\ v_n &= (1-\beta^n)^{-1}\cdot(1-\beta)\left(\sum_{t=0}^{n}\beta^t\cdot\theta_{n-t}\right) \end{aligned} \tag{8}$$

Тогда на начальных этапах коррекция выравнивает ранее сильно уменьшенные значения EWMA, а на последующих (при $n\to\infty$) ее влияние ослабеваем, постепенно сводясь к 0 (то есть - делению на 1). Данный способ занимает очень мало памяти компьютера, так как для вычисления ему всегда необходимо хранить только 2 переменные: v_t и β.

Иллюстрируем вышеизложенную теорию на примере реальных данных. Рассматривается цена открытия акций компании Apple за 2021 год. Первоначально сами значения, представленные в координатах (t - месяц года, y_t - цена в рублях), имеют вид:

Рис. 1: Цены открытия акций Apple (AAPL) 2021 (руб.)

Исходя из графика, получаем, что данные очень нестабильны, то есть с первого взгляда невозможно точно сказать, каков тренд. Однако в общих чертах графика однозначно видно, что от малых значений цен показатели переходят к бо́льшим. Применяем к данным модель EWMA.

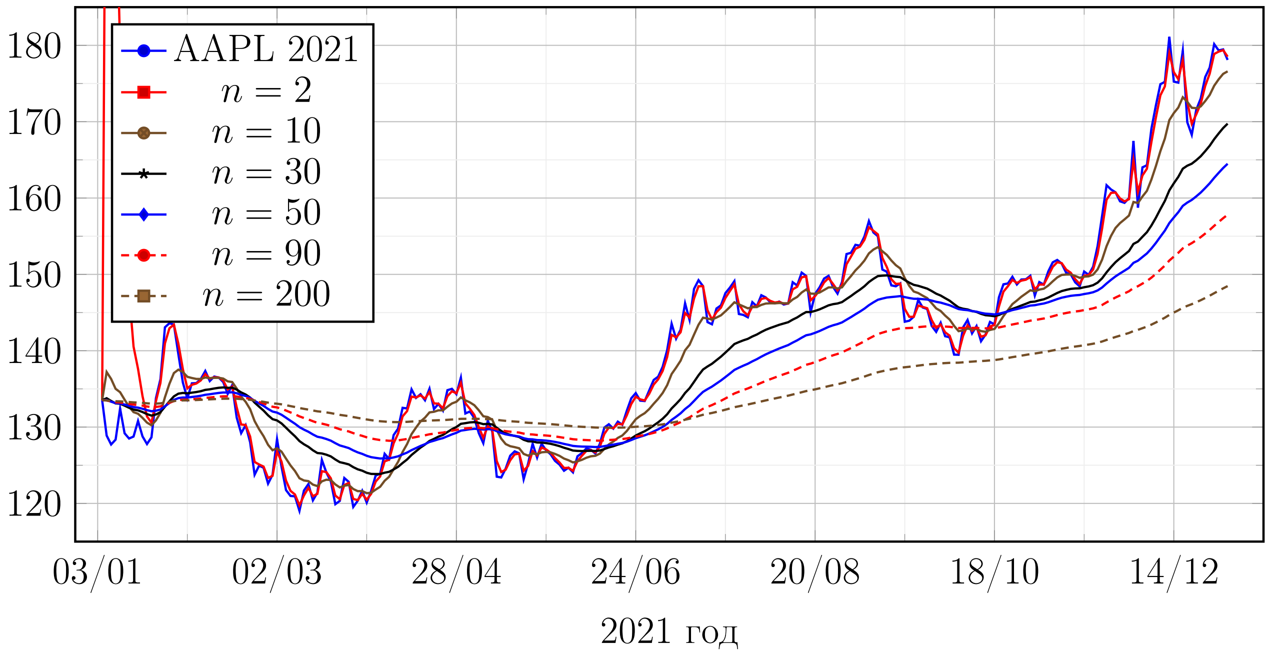

Рис. 2: EWMA, цены открытия AAPL 2021 (руб.)

Глядя на полученный результат сразу ясно, что тренд восходящий, однако это предположение делается только на основе визуального анализа. Пока что никаких алгоритмов нет. Однако нельзя недооценивать полученную информацию, так как более явный тренд позволяет моделям (нейронных сетей) более точно настраиваться на обучающую выборку, что часто приводит к улучшению результату предсказаний (если, конечно не довести до переобучения: данные термины и теория объясняется в блоке 2.2.9 стр. 19). Также стоит отметить всплеск, случившийся при $n = 2$. Это произошло именно именно из-за технических особенностей добавленного множителя, корректирующего смещение.

Последним пунктом необходимо рассказать более подробно о самом корректирующем множителе. **Q**: Почему у него именно такой вид, ведь можно просто сделать первое значение таким же, как в исходном ряде? **A**: 1) необходимо корректировать вычисления, исходя из одной формулы, так как цель - прийти именно к общности 2) при последующих наблюдения данный коэффициент становится чрезвычайно малым $\lim\limits_{t \to +\infty}(1-\beta^t) = 1$, а при первоначальных значениях не оказывает влияния $\lim\limits_{t \to +0}(1-\beta^t) = 0$. Однако при индексах (речь о непрерывных), стремящихся к 0, обратная величина стремится к ∞, что очень плохо с вычислительной точки зрения, так как у компьютера может произойти переполнение памяти и вместо числа получится NaN или Null, что не позволит далее осуществлять расчеты. Поэтому в начале графике наблюдается резкий всплеск, сходящийся впоследствии к исходному ряду. Далее приводится график значений корректирующих коэффициентов в зависимости от β и от t.

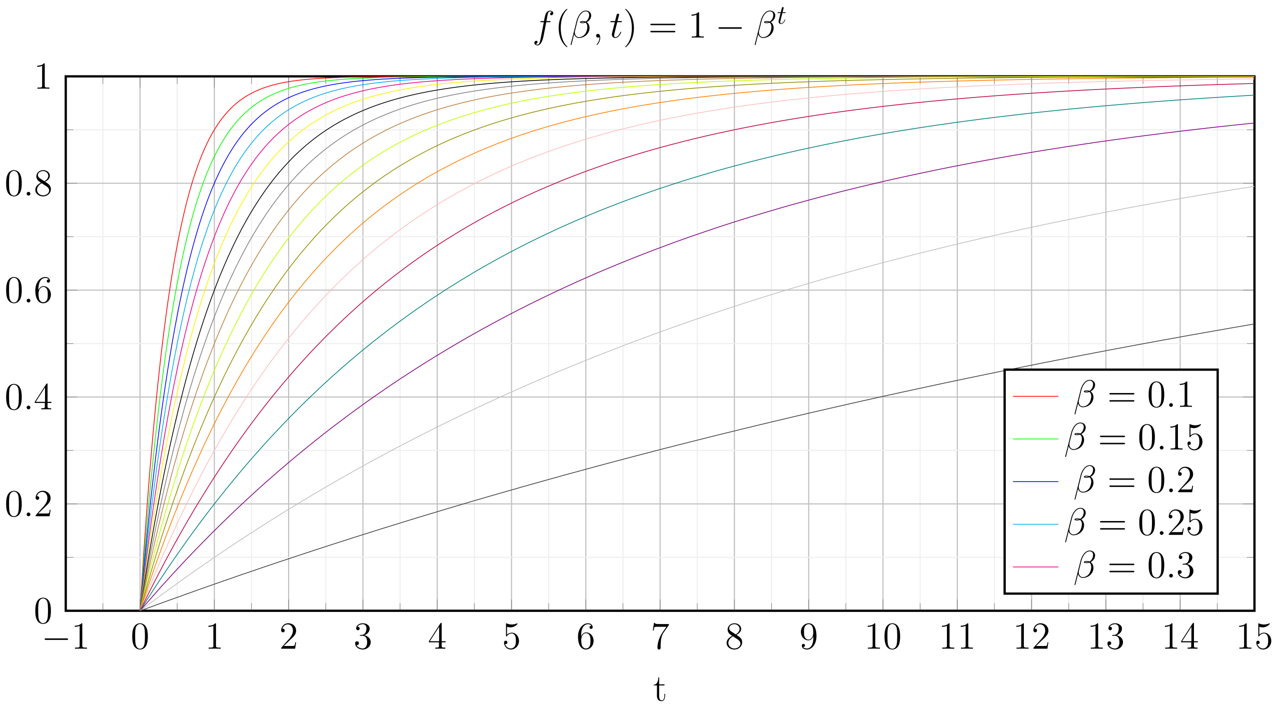

Рис. 3: График корректирующих коэффициентов

В трехмерном пространстве это выглядит так:

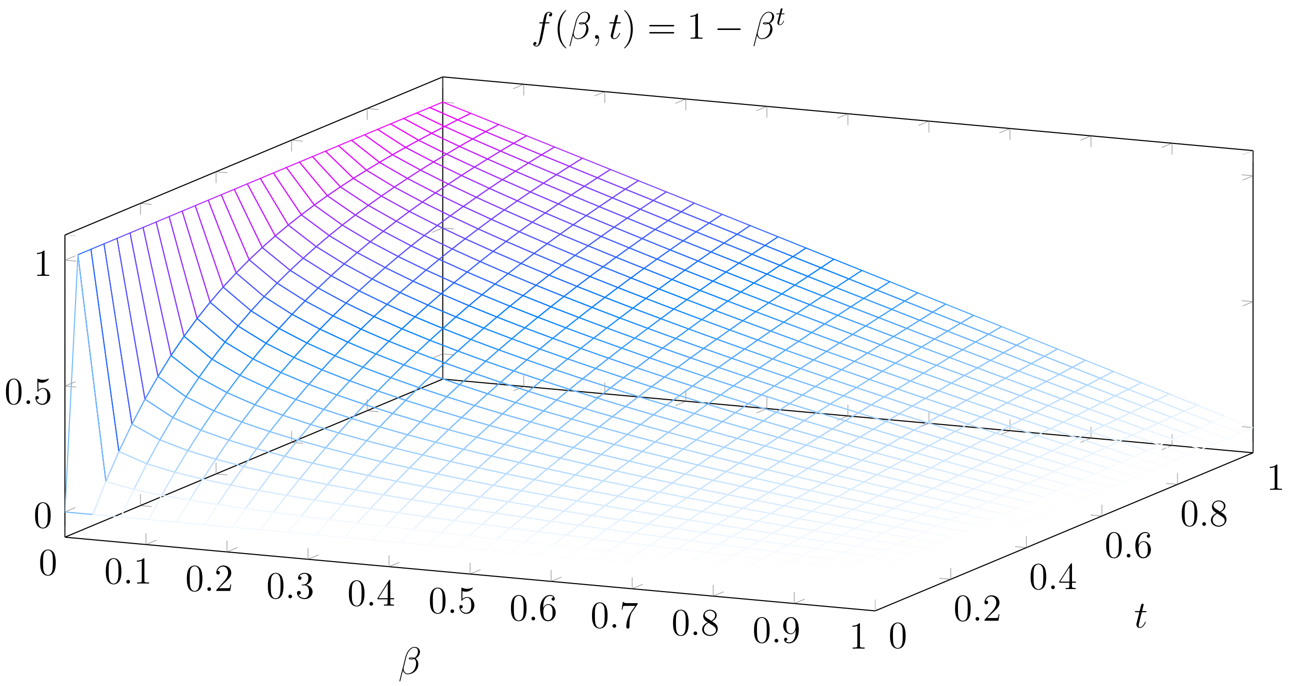

Рис. 4: Трехмерное изображение корректирующих коэффициентов

Данная модель является наиболее простой, однако она обладает возможностью предсказания.

$$v_{t+1} = k \cdot y_t + (1 - k) \cdot v_t : \ k \in (0, 1) \tag{9}$$

В данном случае v_t - это предсказанное значение на момент времени t. Модель не является устойчивой к наличию сезонности и тренду, однако факт того, что предсказания возможны есть. Значит, эта модель включается в список попадающих в сравнительную таблицу.

2.2.2 Auto-regressive Integrated Moving Average

Логичным предположением является тот факт, что значения конкретного временного ряда зависят от своих предыдущих значений, как было показано на примере EWMA (2.2.1 стр. 11). Обобщая данную мысль, текущее (t) значение временного ряда может зависеть от некоторого на k шагов отстающего от него значения $(t-k)$. Тогда получается формула вида:

$$y_t = \alpha_0 + \sum_{p=1}^{q} \alpha_p \cdot y_p + \varepsilon_t : \ \varepsilon_t \sim N(0, \sigma^2) \tag{10}$$

Пока что никаких ограничений на α_p не накладывается. **Q**: Но что такое переход от y_t элементу ряда к y_{t-1}? **A**: Это применение некоторого оператора лага (Lag/Backshift) $B : By_t = y_{t-1}, B^2 y_t = y_{t-2}$ и так далее. **Q**: Что должен представлять из себя данный оператор? **A**: 1) Он должен быть линеен, то есть: $B(x_t + k \cdot y_t) = Bx_t + k \cdot By_t = x_{t-1} + k \cdot y_{t-1} : k \in \mathbb{R}$. **Q**: Как грамотно задать данный оператор? **A**: Представляем ситуацию, в которой необходимо из вектора $Y_n = (y_1, \ldots, y_n)^T$ получить вектор $Y_{n-1} = (0, y_1, \ldots, y_{n-1})^T$, тогда матрица $B : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n \times 1}$ имеет вид.

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{pmatrix} \tag{11}$$

Во-первых, она квадратная, во-вторых интересно, что данный оператор является нильпотентным степени n. Из этого следует, что a) единственное собственное значение B это 0 b) $B^n = O$ (нулевая матрица), то есть степень нильпотентности не превосходит n. Более подробно в работе [21]. Но прежде, чем переписать полученную выше сумму в надлежащем виде, вспоминаем о возможной зависимости y_t от предыдущих значений скользящей средней (ε_{t-l}), где ε_{t-l} - значение, полученное из белого шума [9].

$$\begin{gathered} y_t = \alpha_0 + \sum_{k=1}^{p} \alpha_k \cdot B^k y_t + \sum_{j=1}^{q} \beta_j \cdot B^j \varepsilon_t + \varepsilon_t \\ \left(1 - \sum_{k=1}^{p} \alpha_k \cdot B^k\right) y_t = \alpha_0 + \left(\sum_{j=1}^{q} \beta_j \cdot B^j\right) \varepsilon_t \end{gathered} \tag{12}$$

[9] Белый шум - независимые, взятые из одного распределения величины - часто $N(0, \sigma^2)$.

При этом $B\alpha_0 = \alpha_0 : \alpha_0 \in \mathbb{R}$ по построению. Таким образом, получаем модель авторегрессионной скользящей средней ARMA(p,q). Но по первоначальной предпосылке о стационарности 2 стр. 11, есть желание обобщить модель на случай нестационарных рядов. То есть сделать ее более универсальной. Для этого формально выводим условие стационарности. Выражению $1-\sum_{k=1}^{p}\alpha_k B^k$ в соответствие ставим характеристическое уравнение $1-\sum_{k=1}^{p}\alpha_k z^k$. Тогда если все корни данного уравнения (в силу Основной теоремы Алгебры, полученный полином над комплексной плоскостью имеет не более, чем p различных корней) $|z_i| > 1$, то данный ряд стационарен. Подробнее для AR(1):

$$\begin{aligned} y_t = \alpha \cdot y_{t-1} + \varepsilon_t \Rightarrow 1 - \alpha \cdot z = 0 \Rightarrow z = \frac{1}{\alpha} \\ y_{t+1} - \alpha \cdot y_t \approx 0 \Rightarrow \lambda - \alpha = 0 \Rightarrow y_t \approx c \cdot \alpha^t : \ c \in \mathbb{R} \end{aligned} \tag{13}$$

Получается, чтобы числовые значения данного ряда не уходили в ∞ при $t \to \infty$ необходимо, чтобы в случае AR(1) $|z| > 1$, а $|\alpha| < 1$. Аналогично для случаев AR(p). Таким образом, вывод: необходим множитель, который помогал бы последовательности переходить к d - ому порядку интегрирования. **Q**: Как это сделать? **A**: $\nabla^1 y_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t, \nabla^2 y_t = \nabla^1 y_t - \nabla^1 y_{t-1} = (1-B)y_t - (1-B)y_{t-1} == (1-B)y_t - (1-B)By_t = (1-2B+B^2)y_t = (1-B)^2 y_t$ и так далее. Получается, что необходимый множитель: $\nabla^d = (1-B)^d$. Соответственно формула ARIMA(p,d,q) имеет вид:

$$\underbrace{\left(1 - \sum_{k=1}^{p} \alpha_k \cdot B^k\right) \overbrace{(1-B)^d}^{\mathrm{I}(d)}}_{\mathrm{AR}(p)} y_t = \overbrace{\alpha_0}^{\mathrm{const}} + \underbrace{\left(\sum_{j=1}^{q} \beta_j \cdot B^j\right) \varepsilon_t}_{\mathrm{MA}(q)} \tag{14}$$

Получившаяся модель включает в себя авторегрессионную составляющую порядка p, скользящую среднюю порядка q и интегрированность порядка d. Также стоит отметить, что любой стационарный процесс можно разложить в бесконечный (абсолютно сходящийся) процесс скользящего среднего. Более известен данный факт под названием - разложение Вольда [4]. Для примера предполагаем, что некоторый процесс AR(1) стационарен.

$$\begin{aligned} y_t &= \alpha_0 + \alpha \cdot y_{t-1} + \varepsilon_t \\ (1-\alpha B) \cdot y_t &= \alpha_0 + \varepsilon_t \\ (1-\alpha B)^{-1}(1-\alpha B) y_t &= (1-\alpha B)^{-1}(\alpha_0 + \varepsilon_t) \\ y_t &= (1-\alpha B)^{-1} \cdot \alpha_0 + (1-\alpha B)^{-1} \cdot \varepsilon_t \end{aligned} \tag{15}$$

$$\text{Но ряд стационарен} \Rightarrow y_t = \sum_{j=1}^{\infty} B^j \alpha_0 + \sum_{j=1}^{\infty} B^j \varepsilon_t = \frac{\alpha_0}{1-\alpha} + \sum_{j=1}^{\infty} \alpha^j \cdot B^j \varepsilon_t$$

Более подробно о данной модели написано в [3]. Процесс обучения (то есть подбора соответствующих коэффициентов модели) происходит посредством применения ММП (Метода Максимального Правдоподобия), максимизирующего вероятность наиболее точно описать входные данные. Глобальная задача, стоящая пред алгоритмом подбора параметров (читать далее - алгоритма обучения) это:

$$\sum_{p}^{n} e_p^2 \tag{16}$$

Где $e_p = y_p - \hat{y}_p$ - остаток (разница между предсказанным и исходным значением). Это очень похоже на Метода Наименьших Квадратов (МНК), только МНК дает аналитическую формулу оценок коэффициентов, а в общем случае это не всегда возможно, поэтому для подобных задач часто применяется ММП. Тут же встает вопрос: **Q**: Как понять, какая модель лучше описывает данные? **A**: 1) Критерий Акаике (AIC):

$$J_{\text{AIC}} = 2k + n \cdot \ln\left[\frac{1}{n}\sum_{p=1}^{n}(y_p - \hat{y}_p)^2\right] \tag{17}$$

Где k - количество параметров в обучаемой модели, n - общее количество наблюдений, $(y_p - \hat{y}_p)^2 = e_p^2$ - остаток от регрессии 2) Байесовский информационный критерий (BIC):

$$J_{\text{BIC}} = \frac{1}{n}\sum_{p=1}^{n}(y_p - \hat{y}_p)^2 + \frac{k\hat{\sigma}^2\ln(n)}{n} \tag{18}$$

Где $\hat{\sigma}^2$ - оценка (так как реальное значение неизвестно) дисперсии шума ряда, полученного по формуле $(y - \hat{y})$. [1] **Q**: Однако тут появляется еще один вопрос: как человеку понять, какой порядок AR($\cdot$) и MA($\cdot$) должен быть? **A**: Существует 2 способа, чтобы выяснить это. ACF(k) (Auto Correlation Function) и PACF(k) (Partial Auto Correlation Function). Как видно из названия - это некоторые функции от переменной k. **Q**: Что они характеризуют? **A**: 1) ACF - основанная на конкретном наборе данных дискретная функция, вычисляемая на основе формулы:

$$\text{ACF}(k) = \hat{\rho}_k = \widehat{\text{corr}}(y_t, y_{t-k}) \tag{19}$$

2) PACF - частная автокорреляционная функция. Вычисляется на основе оцененной модели AR(k). Строится график коэффициентов при лагах. В итоге формула функции приобретает вид:

$$\text{PACF}(k) = \hat{\alpha}_k \tag{20}$$

Как было сделано для предыдущей модели, иллюстрируем вышеизложенную теорию на примере. Рассматриваем также цены открытия акций компании Apple. Стоит заранее сказать, что прежде, чем подгонять модель, необходимо проверить сам ряд на стационарность, то есть - на наличие корней $|z_j| < 1$, что свидетельствовало бы о нестационарности исследуемого ряда.

Очевидно, что с помощью данной модели можно предсказывать последующие значения числовой последовательности, поэтому данная модель также включается в список сравнительной таблицы.

2.2.3 Generalized Auto-Regressive Conditional Heteroskedasticity

2.2.4 Auto-regressive Fractionally Integrated Moving Average

2.2.5 Fractionally Integrated GARCH

2.2.6 Singular Spectrum Analysis

2.2.7 Fourier analysis

2.2.8 Wavelet analysis

2.2.9 Neural Networks

2.2.9.1 Multilayer Perceptron

2.2.9.2 Recurrent Neural Network

2.2.9.3 Wavelet Network

2.2.9.4 Алгоритмы обучения: GD семейство

3 Описание данных и эксперимент

3.1 Развивающийся рынок (Китай)

3.2 Развитый рынок (США)

4 Обсуждение выводов

5 Заключение

Список литературы

[1] Antonios K Alexandridis and Achilleas D Zapranis. *Wavelet neural networks: with applications in financial engineering, chaos, and classification.* John Wiley & Sons, 2014. https://books.google.ru/books?hl=en&lr=&id=cYpgAwAAQBAJ&oi=fnd&pg=PR13&dq=wavelet+networks+and+applications&ots=PRTQtLGg33&sig=8S6--O6eRo8bYywFQpmCyT5rMME&redir_esc=y#v=onepage&q=wavelet%20networks%20and%20applications&f=false.

[2] Nathan Benaich and Ian Hogarth. *State of AI Report 2022.* Air Street Capital, Plural, 2022. https://www.stateof.ai/.

[3] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[4] Christiano. Wold representation theorem. *Northwestern University: FINC-520*, . https://faculty.wcas.northwestern.edu/lchrist/finc520/wold.pdf.

[5] Eugene Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970. https://www.jstor.org/stable/2325486.

[6] FinEX. *Гипотеза эффективного рынка: что это и как использовать.* FinEx, Статья от: 13 Сентября 2022. https://finex-etf.ru/university/news/gipoteza_effektivnogo_rynka_chto_eto_i_kak_ispolzovat.

[7] L.S. Lima and L.L.B. Miranda. Price dynamics of the financial markets using the stochastic differential equation for a potential double well. *Physica A: Statistical Mechanics and its Applications*, 490:828–833, 2018. https://www.sciencedirect.com/science/article/pii/S0378437117308324?casa_token=rzOzlZw26ZUAAAAA:TLYp15lOen3WJC65uWCpHNNGItqli5Iuqa5yb5LSZKPj5mznOHBbH3r_CwTyjDHAe2ZYaVGA.

[8] Alfred Marshall. *Elements of economics of industry: being the first volume of Elements of Economics*, volume 1. Macmillan, 1903.

[9] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. https://arxiv.org/abs/1301.3781.

[10] Mostafa Seif, Paul Docherty, and Abul Shamsuddin. Seasonal anomalies in advanced emerging stock markets. *The Quarterly Review of Economics*

and Finance, 66:169–181, 2017. https://www.sciencedirect.com/science/article/pii/S1062976917300832.

[11] Martin Sewell. History of the efficient market hypothesis. *Rn*, 11(04):04, 2011. http://www.cs.ucl.ac.uk/fileadmin/UCL-CS/images/Research_Student_Information/RN_11_04.pdf.

[12] Savva Shanaev and Binam Ghimire. A generalised seasonality test and applications for cryptocurrency and stock market seasonality. *The Quarterly Review of Economics and Finance*, 86:172–185, 2022. https://www.sciencedirect.com/science/article/pii/S1062976922000801.

[13] Adam Smith. *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*. Harriman House Limited, 2010.

[14] Pantelis Tassopoulos and Yorgos Protonotarios. Brownian motion & the stochastic behavior of stocks. *Journal of Mathematical Finance*, 12(1):138–149, 2021.

[15] CFI Team. Exponentially weighted moving average (ewma). *Corporative Finance Institute (CFI)*, 2023. url: https://corporatefinanceinstitute.com/about-cfi/.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[17] Constantinos E. Vorlow. The (mis)behavior of markets: A fractal view of risk, ruin and reward. *Journal of Economic Behavior and Organization*, 61(3):513–515, 2006. https://www.sciencedirect.com/science/article/pii/S0167268106000692.

[18] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. https://github.com/cure-lab/LTSF-Linear.

[19] Роберт Оуэн. *Спорная область между двумя мирами*. Strelbytskyy Multimedia Publishing, 2018.

[20] АС Палювина. Фрактальный анализ финансового рынка на основе коэффициента Хёрста. *Вектор экономики*, pages 30–30, 2019. http://www.vectoreconomy.ru/images/publications/2019/4/mathematicalmethods/Palyuvina.pdf.

[21] Тарас Евгеньевич Панов. Линейная алгебра и геометрия. *Механико-математический факультет МГУ*, 2021. http://higeom.math.msu.su/people/taras/teaching/slides-linalg/panov-linalg-lec10.pdf.

[22] Н. В. Артамонов; Е. А. Ивин; А. Н. Курбацкий; Д. Фантаццини. Введение в анализ временных рядов: учебное пособие для вузов. *Московский государственный университет имени М. В. Ломоносова, Московская школа экономики, Кафедра эконометрики и математических методов экономики. – Вологда : ВолНЦ РАН, 2021. – 134 с.: ил., табл.*, 2021. https://mse.msu.ru/wp-content/uploads/2021/03/%D0%92%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5-%D0%B2-%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7-%D0%B2%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D1%85-%D1%80%D1%8F%D0%B4%D0%BE%D0%B2-1.pdf.