

Транскрипция к презентации

Гишин Андрей Юрьевич

6 апреля 2023 г.

1 Слайд (Начало)

Здравствуйте! Меня зовут Гришин Андрей. Я студент группы э408. И сегодня расскажу о моем исследовании, связанном с изучением способности как эконометрических, так и нейросетевых моделей к прогнозированию доходностей акций в зависимости от типа анализируемого рынка. Основным объектом исследования является цена открытия китайских и американских компаний.

2 Слайд (О плане)

На данном слайде для ознакомления представлен план текущей презентации.

3 Слайд (О «сегодня»)

В современном мире вокруг почти каждого человека каждый день вращается множество разнообразных слов. Однако слово «деньги» встречается, пожалуй чаще всего. Таким образом, предполагая рациональность современного человека, получаем его желание делать то, что выгодно и не делать то, что не выгодно. В формальных терминах это можно назвать — максимизацией индивидуальной полезности или максимизацией дохода. То есть человека хочет получать больше, чтобы мочь больше. Но, рассматривая частный случай этого, а именно — биржу и фондовые рынки, становится понятно, что просто словами или логикой не получится точно предугадать поведение как цен, так и доходностей акций той или иной компании. Как раз для этого и придумано очень много математических подходов, однако как выбрать нужный? Вопрос сложный.

На данный момент уже количество имеющихся данных в сети превышает огромные значения, то есть просто «данные» стали «большими данными», для обработки которых необходимы «большие» модели. Следовательно, область Статистики расширилась и стала областями Искусственного Интеллекта и Эконометрики. Но там, где много подходов, высока неопределенность.

4 Слайд (Актуальность — зачем)

Настоящая работа призвана выбрать наилучшую среди всех рассмотренных моделей, проведя их последовательный сравнительный анализ. Выбор такой модели в будущем позволит развивать лучшее из имеющегося, а не блуждать в поисках одного из множества подходов. На практике станет проще, а значит и быстрее принимать достаточно неоднозначное биржевое решение по вопросу «Buy, Hold or Sell».

5 Слайд (Цели — что хотим получить)

Целями настоящего исследования является помощь как трейдерам, так и кому бы то ни было связанному с биржей, в оперативном формировании точного ответа на вопрос: «Buy, Hold or Sell». Что в итоге сделает биржевые сделки менее рискованными из чего получается планирование на более длительный промежуток времени в силу финансовой определенности. В итоге людям, например, не имеющим прямого отношения к бирже (или являющимся неопытными в этом вопросе) станет проще адаптироваться к фондовым рынкам, так как они (люди) получают возможность достаточно точно прогнозировать необходимые для принятия решения показатели.

6 Слайд (Задачи)

Задачами же моего исследования является проведение последовательного сравнительного анализа рассматриваемых эконометрических, статистических и нейросетевых моделей в задаче прогнозирования доходностей и — как будет показано далее — цен акций 15 американских и 15 китайских компаний сроком на 1 рабочий день биржи. Для наиболее точного результата без смещения на особенности рассматриваемой индустрии, обсуждается как наиболее оптимальное число, с точки зрения системы китайского фондового рынка, именно 15 компаний Соединенных штатов и 15 компаний Китая, взятых из разных сфер производства и услуг.

7 Слайд (Гипотезы — что предполагаем)

Гипотезами, на которых строится все исследование являются: Гипотеза Рыночная Эффективность Юджина Фамы, статья о которой выпущена в 1970 году, а также Гипотеза Рыночной Фрактальности, сформулированная Бенуа Мандельбротом и опубликованная им в 2006 году.

В противоположность ко всем вышесказанному идет гипотеза «рыночной неэффективности» Мартина Севелла, опубликованная в 2011 году, по которой ГРЭ является лучшим, что может быть на сегодняшний день, однако ложным в силу наличия большого количества исследований это показывающих. В своей работе Мартин Севелл анализирует всю доступную информацию, посвященную становлению ГРЭ в том виде, в котором она нам известна.

8 Слайд (Больше о данных)

Более подробно об использованных данных. Рассматриваются компании развитого и развивающегося рынков. Но так как административный район Китая — Гонконг — является развитым, с точки зрения финансовой составляющей, его в анализе нет. Вместо него идет Шанхайская биржа. А для анализа американских компаний использована Нью-Йоркская биржа. Границами исследуемого временного ряда является выход компании на IPO и 13 декабря 2022 года. Так как после этой даны финансовые рынки анализируемых стран стали в полной мере подвергаться воздействию политических решений.

Индустрии, анализируемые в работе включают в себя самые разнообразные отрасли производства и услуг. Некоторые из них представлены на слайде. Это IT, медиа, продажи, банки, страхование и так далее.

9 Слайд (Визуальный анализ данных)

Все графики, представленные в работе, выполнены автором при использовании таких программных пакетов как Python, MATLAB, Excel на основании собранных с сайта «Yahoo! Finance» данных, описанных ранее.

Для лучшего понимания рассматриваемых данных, анализируем пример двух компаний: Coca Cola и Kweichow Moutai (один из крупнейших производителей алкогольной продукции Китая). Глядя на представленные графики видим 2 вещи. Первая — на обоих верхних графиках (представлены цены) наблюдается тенденция роста с увеличением амплитуды колебаний с течением времени. Вторая — на нижних двух графиках волатильность доходностей Coca Cola более «плотно» находится в пределах от -10% to 10%, чем аналогичный показатель для Kweichow Moutai. Отсюда вывод, что, скорее всего, потребуются разные подходы для моделирования данных показателей. Аналогично возникает чувство, что предсказывать цены проще, так как в них четко выделяется тренд и некоторый шум. Однако это пока что бездоказательный вывод. Чтобы сделать их более обоснованными, обращаемся к принципу Wavelet анализа, известного в области квантовой физики. Целью данного перехода является поиск доминирующих «частот» в имеющемся сигнале, которые впоследствии можно использовать в прогнозировании.

10 Слайд (Скалограмма Coca Cola)

Опираясь на представленный график скалограммы, имеющий по оХ — время, по оУ — частоту, а по оZ — амплитуду (силу) частоты, имеем отсутствие какой-либо значимой частотной информации в ценах акций компании Coca Cola. Однако в доходностях — наоборот содержится статистически значимые показатели. Говорит ли это о целесообразности работы исключительно с доходностями?

11 Слайд (Скалограмма Kweichow Moutai)

Нет! Так как на примере скалограммы Kweichow Moutai безоговорочно видны частотные паттерны, в то время как подобные паттерны доходности менее значимы. Таким образом, необходимо анализировать не что-то одно, а оба показателя. То есть цены и доходности вместе.

12 Слайд (Примененные методы)

В настоящем исследовании применяются модели как эконометрического характера, основанные на корреляции, статистические модели, вообще ничего не «думающие» об имеющихся данных, а также — нейросетевые модели, держащие в своей основе сложную самообучающуюся функцию, и называющуюся нейронной сетью.

В эконометрической области выбраны наиболее популярные ARIMA и ее фрактальное дополнение ARFIMA, GARCH и аналогично FIGARCH.

Из статистики тут EWMA (экспоненциальная средняя), а также SSA (в народе более известный как метод «Гусеница»), представляющий из себя сингулярное спектральное разложение имеющегося ряда.

От нейросетей взят наиболее простой набор моделей (полносвязные, рекуррентные, а также wavelet сеть, заимствующая подход у Wavelet анализа), чтобы показать, как ошибка предсказаний достаточно сложных Эконометрических и Статистических моделей отличается от Нейросетевых.

В качестве очистителя от шума имеющихся данных используется алгоритм MSSA.

Наиболее популярные на данный момент в сфере обработки естественного языка «трансформеры» не используются. Причина — отсутствие контекста для анализа.

13 Слайд (Результат для цен)

Переходя к результатам, полученным для цен, видим безоговорочное превосходство нейронных сетей над Эконометрическими методами.

Однако Статистические подходы тоже дали о себе знать: EWMA делит второе место с MLP, в то время как на первом идет комбинация MLP и MSSA — нейронная сеть на очищенных данных. Частотный анализ, к сожалению, не дал ожидаемого результата, так как находится на последнем месте в рейтинге нейросетевых моделей.

Среди Эконометрических моделей, комбинация ARIMA и FIGARCH занимают первое место, что однако никак не приближает ее к состязанию с нейронными сетями или статистическими методами.

Также интересно, что, основываясь на топологической функции, Эконометрическим моделям лучше удается прогнозировать значения для развитых рынков, чем для развивающихся. Это еще одно подтверждение неполной корректности ГРЭ. Однако для нейронных сетей так сходу сделать вывод не получится в силу наличия противоположных показаний.

14 Слайд (Результат для доходностей)

Переходя к доходностям, видим проигрыш почти всех методов — отклонение от реальности свыше 100% у большинства моделей. Однако рекуррентные нейронные сети в сочетании с алгоритмом удаления шума показывают намного более хороший результат, что выводит их на первое место в списке всех моделей для доходностей.

Столбец MLP + EWMA не вычислялся, так как проведение экспоненциального сглаживания для доходностей акций, на взгляд автора, не имеет как логического, так и экономического смысла.

15 Слайд (Обсуждение выводов)

В заключение презентации устно привожу наиболее важные выводы, полученные в результате проведения исследования, все остальные (в том числе и озвученные) представлены на слайде.

Видим, что лучшей модели (из рассмотренного круга) для прогнозирования цен акций — это комбинация MSSA и MLP, в то время как для доходностей — MSSA и RNN.

Алгоритм очистки данных от шума дает хороший прирост к точности прогноза, однако WN — в силу своей архитектуры — является исключением.

Эконометрические методы плохо подходят как для прогнозирования цен, так и доходностей, а вот статистические (EWMA и SSA) — показывают сравнимый с нейросетевым результат и при этом не требуют таких больших вычислительных мощностей, как нейронные сети при обучении. Получаем компромисс: если нужно более точно — нейронные сети, иначе — статистические (EWMA и SSA).

Более того, алгоритм бустинга (обучение одной модели на остатках предыдущей) для Эконометрических моделей — плохая идея. Лучше — в случае Эконометрики — максимизировать совокупную функцию правдоподобия.