



Research proposal

Статьи из литературе ревью

- **Евстигнеева**

- с помощью всяких библиотек и методов собрала датасет - 10к текстов разной сложности (экспертная оценка) x 44 разные языковые метрики
- прогнала 10 ML моделей, получила, что у трансформерной самая высокая точность
- какие машинные подходы есть
- какие были рез-ты до этого:

1. **Ошима**

- a. публикация влияет на волатильность в день-два после публикации
- b. какие-то темы особенно сильно влияют на Unexpected Volatility
- c. *to exclude other factors - use tick-by-tick data + narrower event window*
- d. market sentiment directionally from the words appearing in statement
 - i. **Lougharn** - считает сентимент как доля негативных слов из словаря и смотрит влияние доли на цену акции

2. **Dincer and Eichengreen**

- a. сделали трансперенси индекс - экспертная оценка
- b. показали что больше трансперенси - меньше волатильность инфляции

3. **Bulíř et al. (2014)**

- a. Flesch-Kincaid index - волатильность фин рынков

4. **Hernández-Murillo and Shell**

- a. Просто анализ FOMC - показали что длина текстов и сложность изменяется
- 5. **Bulíř et al. (2013)**
 - a. низкая ридабилити - повышается волатильность инфляции
- 6. **Bruno (2017)**
 - a. LSA + ARI + Flesch-Kincaid index
 - b. для италии отчеты понятны только для людей с высшим образованием
- 7. **Tumala (2019)**
 - a. sentiment analysis
 - b. positive - negative words коррелирует с экономической обстановкой
- 8. **Gorodnichenko**
 - a. DL model with controlling for the Fed's actions and the sentiment in policy texts - какая непонятно, по-моему, просто из словаря доли посчитали
 - b. positive tone in the voices of Fed Chair - выше цены акций, бондов и инфляционные ожидания, и влияет на валютный курс
- 9. **Головко**
 - a. влияние публикаций на доходности
 - b. + предложение трейдинг страты
- 10. **Рыбински**
 - a. считает индекс из словаря ястребиной голубиной политик
 - b. посчитал корреляции внутри тем публикаций со статьями до и после (несколько дней недель)
- 11. **Luangram**
 - a. Оценивают читаемость - готовое решение
 - b. LDA - топики + сентимент анализ - позитив негатив слова
 - c. просто следят по странам 20 лет

d. смотрят периоды снижения ставок

12. Lee

- a. Посчитали индекс стиля - функция от нграм и кол-ва слов (dovish hawkish)
- b. публикация хорошо объясняет изменение долгосрочных ставок
- c. кол-во статей в газетах растёт значимо (1.5 раза) в день публикации

посмотреть как можно посмотреть влияние одного текста на другое
смотреть как смотрят влияние текста и публикации на акции

накидать план что нужно писать

- сам план
 - research question
 - motivation & novelty
 - мотивация - взять у Алины
 - новизна - methodological twist - мол такую текстовую машинку в экономике не юзали
 - hypotheses (if causal relationship?)
 - пока хз
 - data
 - чек как описывает евстигнеева свой датасет
 - моменты публикации от Стырина
 - индексы рынка
 - газетки???

- methodology

Смотреть пропозал к вальсекки

План статьи:

важно взять именно публичные данные или сравнить поддасеты

- описание данных - как у алины?(
 - морфо и лингво метрики
 - наложить фреймворки других статей
 - сделать топик моделлинг и динамика его - булщит, но можно
 - реакция рынка (лонгтерм + шорттерм) на публикации, но не просто, а с учетом лингво метрик
 - просто уже делал головко, нужно именно завязаться на текст
 - и в другой статье
 - + коммуникация банка Англии - первая статья
- интеракшн панелька?*
- реакция журналистов
 - оценить корреляции текстов на основе каких-то индексов

какая dependent variable

какие independent

Какие регрессии

какие fixed effects

какие контроли

какие гипотезы:

- акции реагируют на одни топики сильнее чем на другие
- мб посмотреть влияние на инфляционные ожидания?
- влияние на волатильность рынков?
 - Евстигнеева:

Jansen (2011a, 2011b) and Bulíř et al. (2014) propose the use of the classical Flesch-Kincaid index to assess the clarity of communication and demonstrate the existence of an inverse relationship between the readability of central bank texts and the volatility of financial markets.

- правые левые газеты реагируют по-разному
- дамми ан спикера
доходность на конкретное слово/ topic
- дифиндиф переключения режима
ауткамы на тексты
- измерить эффект на сабсетах экспорт-импорт ориентированных

мб разделить месседж о повышении/ понижении (бинарная, дамми переменная) на две переменных

но вообще она становится континуум

например берем сложность текста, она может быть высокой как для повышения так и для понижения, поэтому надо брать интеракшн сложности текста с самим дамми

дамми + индекс - будет ли вообще индекс влиять на рост, типа чем сложнее текст, меньше растет - не поняли

дамми + индекс * дамми + индекс - а тут даем воз-ть сильнее влиять сложности при повышении

мб диф ин диф до и после публикаций? + добавить все прошлые