

Data + Method

▼ АРХИВ

▼ texts - собрать тексты

- ☒ убрать ссылки на автора
- ☒ почистить знаки препинания (пока не критично)
 - ☒ кавычки
 - ☒ повторяющиеся точки тире
- ☒ убрать nap
- ☒ проверить хедеры с маленькой длиной строки (норм обычно техническая инфо или выдача лицензий)
- ☒ проверить тексты с маленькой длиной строки
 - ☒ если что убрать их или поставить флаг, что не релевантные — добрал данные
- ☒ чистить выступления наби
 - ☒ добавить норм текст
 - ☒ разбить на её спич и вопросы

▼ sentiment - получить dep var

positive/ negative

- ☐ найти хорошую обученную русскую модель для позитив и негатив
 - ☐ <https://www.kaggle.com/ziliwang/baseline-upsampling-balanced-softmax-regression>
- ☐ напредиктить на своих данных
- ☐ смотреть вариацию по выступлениям и на сабсемплах

clarity by Alina

- ☐ нагенерить её признаков
- ☐ обучить её трансформер
- ☐ напредиктить для своих данных

▼ пайплайн по методе (old)

- попробовать с dictionary based method / Dostoevsky

- ✓ найти норм словарь
- ✓ нафитить
- ✓ обзор вариации
- ✓ собрать данные для пропозала
 - прогнать регрессии
- попробовать обучить модель не руберт
 - написать мужику
 - научиться и разобраться
 - нафитить прогнать регрессии
- попробовать обучить трансформер от алины
 - научиться и разобраться
 - нафитить прогнать регрессии
- переложить

▼ problems

- несколько пресс-релизов в один день - смотреть только выступления по ставке?
 - да, только выступления по ставке
- спред очень шумный
 - используем только спотовые показатели - как городниченко
- что за шоки в качестве контроля у городниченко
 - разные шоки феда - из другой статьи
 - цб юзает только стави и валютные интервенции, другие шоки плохо квантифицируются

▼ что можно сделать

▼ добавить интеракшн между dep var и рейт шоком

- несильно зависит, но можно попробовать

▼ добавить ещё шоки

- пока что только ставка
- добавить в реги уровень доверия к цб

▼ добавить трансперенси индекс

- смотреть статью чела
- взаимосвязь между сентиментом и трансом
- добавить в основные регрессии пресс релизы в тот же день

дополнительно:

- странные резы - плохие контроли/ переменные?
 - test b/w rate shock & sentiment?
- sentiment:
 - добавить clarity и pos. нег продвинутые
 - посмотреть как связана тональность пресс-релизов и выступлений (tests b/w units)
 - посмотреть как между собой связаны (tests b/w properties)
- data
 - разбить наблюдения выступлений на Q&A session и основную часть
- структура
 - добавить динамику на несколько дней
 - rolling window regression
 - мб перейти на минутки
- outcomes
 - infl expectations - как fxgd, мб брать месячные из отчетов
 - добавить ещё outcome переменные (бонды разной maturity, отраслевые рынки)
- controls
 - смотреть шоки в изаков / отчет мвф

later:

- обучить модель предсказывать ставку по тексту гыгыгы
- анализ метрик - как доверие зависит от transperncy
 1. смотреть городниченко комувмент
- нет ли reverse causality? что тональность зависит от шока рынка?
 - хотя и не зависит от шока ставки

▼ вопросы к стырину по тех частб

- как правильно кластеризовать? robust или достаточно обычной hc

- нужно ли добавлять интеракшн в полной модели?
- как тестить co-movement
- как правильно репортить - только максимально полную таблу или **треугольник**

▼ доп ссылки

- отсюда берем ставки (необяз): <https://normativ.kontur.ru/rates/cbrkey>
- процентные шоки: <https://dlib-eastview-com.proxylibrary.hse.ru/browse/doc/58304216>
- изаков и мвф: смотреть шоки в изаков / отчет мвф
- al-mashat <https://www.imf.org/en/Publications/WP/Issues/2018/09/28/An-Index-for-Transparency-for-Inflation-Targeting-Central-Banks-Application-to-the-Czech-46192>
- инфл ожидания http://cbr.ru/eng/analytics/dkp/inflation_expectations/

Структура:

▼ Introduction

- **описывающее основной исследовательский вопрос**
- **объясняющее его значимость для экономической науки,**
- **также примеры из реальной жизни, где полученные результаты могли бы быть применены**
- **введение также должно давать ясное представление об основном вкладе автора работы в исследуемый вопрос.**

что можно рассказать:

▼ городниченко:

- много статей про эффективность влияния текстов
- он использует тональность голоса
- press conferences have been commonly used as an important communication tool
- vast literature on policy communication and high-freq analys
- анализ вокал работ
- что вносят нового

- повышается точность
- может быть дообучена
- почему такие outcomes?

▼ алина

- говорит про важность текстовой коммуникации - it режим, история
- izakov - summary imf analys of monetary surprises - смотреть есть ли данные о шоках
- график- уровень доверия! - можно вставить как **контроль или интеракшн**
- readability and transparency
- получает ридабилити индекс с помощью своей модели
- ещё есть al mashat transparency index - **можно достать**
- есть апппендикс
- про forward guidance, informational shock - найти подходящие статьи из городниченко в т.ч.
- event studies
- https://www.cbr.ru/dkp/objective_and_principles/
рассказать про период тишины
и про меры цб
following tradition of the previous works - каналы информации такие-то - в своей работе сконцентрируюсь на этих двух
(https://www.cbr.ru/dkp/information_policy/)
- сказать почему рассматриваю несколько outcomes
- почему новое

the rest of the paper organised as follows

▼ Related Literature

на городниченко – мол у него аудио хорошие, а у меня тексты хорошо записаны на алину

▼ Data description

1. все тексты с сайта цб
 - а. На сайте есть разные разделы
Новость выступления и пресс релизы
Новости не релевантны поэтому мы их не рассматриваем как источник шока

- b. здесь основные статистики (длина, ширина текста, мб топик моделинг)
 - i. кол-во слов, средняя длина предложений и всякое такое

2. Компоненты коммуникации

- a. сделать ограничение до заявлений по ставке
- b. рассказать, что много нерелевантных + непонятно, как оценить влияние технической инфы + влияние совокупности новостей
- c. график с динамикой ставки + пунктир - моменты заявлений
- d. на статью сметанина - Сказать про его сентимент подход + про руберн
- e. Рассказ про модели

Надо сослаться На проект достоевский – сказать про классы

<https://github.com/bureaucratic-labs/dostoevsky>

данные от Roger <https://aclanthology.org/C18-1064/>

- f. рассказ про трансперенси

3. данные outcome:

- a. про рынок, бонды разной доходности и инфляционные ожидания

overall, такие-то датасеты

▼ Empirical Model

переменные:

outcome

rate shock - как городниченко и как контроль и как интеракшн

уровень доверия - и как контроль и как интеракшн

tone - рассказать про подходы - сказать что пока обращаюсь к достоевскому

transparency - рассказать про подходы ал машат+ сказать что беру оценку алины

- выписать все конфигурации
 - указать что для всех будут использовать разные outcome
- Детальное описание выбранной для анализа эмпирической стратегии, а также оцениваемое уравнение модели
- Ожидаемые результаты статистической оценки (например, какие знаки оценок коэффициентов ожидается получить);

▼ Results

▼ Robustness Checks (later)

1. bootstrap methods

▼ Conclusion

Поговорили со Стыриным (19.03.2022)

▼ 11 days later...

- forward guidance - прояснение повестки, видения ЦБ о будущем, сигнал рынку
 - нужно дать норм ссылку с определением
- подробнее рассказать про методы NLP
 - еще 1 страница про методы какие есть
 - объяснить почему я выбрал именно эти
- (low) добавить dictionary-based method как benchmark и интерпретируемый результат
 - найти словарь positive/ negative или hawkish/ dovish
 - константин предложил составить самому (если будут идеи и время)
- поменять forward guidance shock на shock from tone sentiment - p.9
- вопросы на подумать
 - как контролировать на ожидания относительно будущего
- добавить динамику индексов ОФЗ
 - композитный по всем maturities
- добавить индекс корпоративных облигаций
- заменить rateshock на rate change - p.9
- сделать rolling window regression
 - 1-35, 2-36 ... 36-70
 - посмотреть как эффект накапливается во времени
- dynamic effects
 - у не за сегодня, а за завтра
- в части про model написать что у меня
 - event study approach, event-analysis
 - нужны controls для того-то и того-то

- сделать ссылку на городниченко
- добавить interaction между tone и trust, tone и transp
- чтобы таблички были адекватнее
 - индекс → разделить на std
 - returns → return * 100
- bootstrap пока не надо

▼ Плюшки от Стырина (напомнить):

- скинул: статья от Нелюбиной
- скинул: статья от ранхигс
- **скип?** контакт с какими-то новостными челами
- **скип?** просить у Стырина источник с норм описанием forward guidance

▼ Done Дальнейшие шаги (01.04.2022)

- **calculations:** чтобы таблички были адекватнее: индекс → разделить на std, returns → return * 100
- **calculations:** robust se - сделать везде
- **формулировки:** добавить в notes
- **формулировки:**
 - в части про model написать что у меня
 - event study approach, event-analysis
 - нужны controls для того-то - что-то про них написать
 - сделать ссылку на городниченко
 - **заменить** rateshock на rate change - p.9
 - поменять forward guidance shock на shock from tone sentiment - p.9
- **дописать results, conclusion**
- **отправить** corrected стырину + напомнить о плюшках

▼ ГЛАВНАЯ ИДЕЯ

- есть индекс ал машат - трансперенси

- есть индекс сентимент - разные подходы, например достоевский
- есть индекс алины - из трансформера
- хочется оценить влияние их комбинации на рынок и инфляционные ожидания
 - как аффецитит точно
 - как влияет на волатильность - отвечали

то что можно нейронкой - нейронками разностороннее влияние на россию

- плюсы нейроподхода?
 - лучше понимаем метрику
- меньше контролей лучше
 - степеней свободы меньше!
 - проблема большой размерности

▼ TECH STRUCTURE

- outcomes (12 штук):
 - IMOEX, RTSI
 - FXGD
 - RGBITR, RUCBITR
 - RUABITR, RUBITR1Y, RUBITR3Y, RUBITR5Y, RUBITR10Y
 - кредитный спред
 - как считать, тоже будет багованный
 - валюта - daily return?
- список всех конфигураций
 - 1 тип: 5 наращивание контролей
 - 2 тип: 2 interactions между tones 1/2
 - rate change - использую
 - trust & transparency (незначимо) - не использую
 - 3 тип: 4 dynamic fixed effects (single, no controls & both, no controls & both, controls)
 - делать в статье
 - в текст график (мб только наиболее значимый)
 - в аппендикс таблицу

Вопросы ко встрече (16.05.22):

▼ ответы

▼ проблема с бондами разных матуритис и композитным индексом

- 1 раз в день считается - орен багованный - норм
- нужно ли поместить коммент в таблицы?
 - **в тексте сноской и в таблице**
- писать что корпоративный более чувствительный
 - видимо, недостаточно наблюдений
 - писать что у бондов знаки в одном направлении (если это правда)

▼ что с обменным курсом?

- считать так же: разность логов = ретерн в течение дня
 - попробовать **тудэй туморой**
- нет ли здесь reverse causality, ведь цб может делать прямые интервенции в день изменения ставки
- **если легко достать, то давайте**
- или все равно сделать, но возможно есть лишний конторль на интервенций которые я не включаю и сказать об этом

▼ что с кредитным спредом?

- если ужесточение, то реальная цена активов пдаает, долг сниажеся
- нетфорт сокращается (как доля активов) \Rightarrow более высокая премия за риск
- moral hazard
- **отказался, так как не понимаю мотивации и интуиции за ним**
- тут что-то на умном, ск всего скипаю

▼ как структурировать модели, куда уравнения, что рассказывать в приоритете

- никто не в приоритете
 - уравнения в основную часть с эмпирикал моделс
 - baseline model, model c interaction и dynamic effects model
- делать ли dynamic effects
 - не включать прошлый y_{T-1} ?
 - считать как n регрессий

- $\log(\text{Close price завтра}) - \log(\text{Open price сегодня})$
 - если верна гипотеза об эффективности финрынков эффект должен произойти и остановиться (не дб прироста)
 - взять ряд остатков и посчитать автокорреляцию
 - ск всего ее не будет
 - Newy-west command in stata
- ▼ как репортить?
 - таблицы с наращиванием контролей (instant effect)
 - просто список, не много ли регрессий?
 - не много, норм
 - добавить F-stats для суммарного эффекта
 - отдельной строкой во все таблы
 - interactions
 - норм ли вынести таблицу с summary без интеракшн? - норм
 - норм ли две панели? - норм
 - **отказался:** добавлять ли интеракшн для trust и trans?
 - dynamic fixed effects графики для tone sentiment 1/2 с/без контролей?
- ▼ смотреть траектории и выбросы во всех переменных (ноябрь 2014)
 - чекнуть если он вообще входит в выборку из 40 наблюдений
 - строить для всех переменных гистограммы
 - если есть заметные выбросы, то
 - winsorize 5-95 (заполнение хвостов крайними константами)
 - truncation 5-95
- ▼ какую таблицу кросс-корреляции регрессоров использовать?
 - питон - симметричная
 - сделать таблицу для регрессоров нижнетреугольной
 - посмотреть команду rwsorr (есть значимость!)
 - комментировать интересные моменты типа не коррелированности сентиментов
 - экспериментальный метод, нет лучших практик
 - что-то упущено

- большая корреляция (multi-collinearity) раздувает стандартные ошибки
 - test на совместную значимость (???)
 - построить PCA первая главная компонента
- или stata-based с ошибками как регрессии
 - это вообще какая-то хрень и лучше её не комментировать
- (skipped) есть команда в стате rwscore
- ссылка на мосбиржа?
 - да в текст ссылку на сайт
- ▼ по поводу евстигнеевой
 - у меня другой подход и другая метрика, видимо менее интерпретабл
 - сложно показать, но поищу способы
 - могу ли сослаться на петрову? а почему нет
 - хорошо бы сопоставить, но надо проверить одинаковая ли нормировка

Final Plans

▼ papers

- смотреть статью евстигнеевой (нелюбиной)
 - понять как описать эффект от конкретных слов (положительный вклад в Tone S)
 - LDA: расчет частности слов по темам - получается матрица - нормируется
 - затем рассчитывается разность между весами по темам (между повесткой цб и прогнозами аналитиков)
 - регрессия - сюрприз на нарратив гэп
 - рез-ты: сюрприз частично объясняется гэпом (больше гэп больше неожиданность)
- смотреть материалы с двух конференций который он прислал (конфа 1, конфа 2)
 - телегин: регулярные коммуникации повышают волатильность
 - петрова: похоже на евстигнееву
 - Lda расчет + мешок слов
 - смягчение, ужесточение денежной политики
 - рез-ты для у:

- прирост однодневной ставки денежного рынка МИАКР, прирост спреда 10- и 2-летних государственных облигаций, прирост спреда 10-летних и 3-месячных государственных облигаций, прирост спреда эффективной доходности корпоративных облигаций (индекс CBONDS) и 10-летних государственных облигаций;
- Ожидаемое в будущем смягчение денежно-кредитной политики будет сопряжено со снижением однодневной ставки МИАКР (таблица 1) на 0,275 п.п. при векторизации «мешок слов» и на 0,16 п.п. при векторизации tf-idf.
- картаев: переход к ИТ снижает зав-ть цен от валютного курса
 - <https://dlib-eastview-com.proxylibrary.hse.ru/search/simple/doc?pager.offset=0&id=52179494&hl=Влияние+инфляционного+таргетирования>
- божечкова: теор модель
 - медленное распространение при принятии ценовых решений экономическими агентами (жесткая информация)
- (if) найти работы MAF и других

▼ data

- индексы гособлигация для разных диапазонов сроков до погашения (МосБиржа)
 - динамику индексов ОФЗ, композитный, по всем maturities
- индекс корпоративных облигаций
- реакция обменных курсов (RUB/USD, RUB/EUR)
- (if) индексы отраслевых рынков
- (if) найти консенсус прогнозы Bloomberg по инфл ожиданиям

▼ calculations

- наstackать датафреймы с новыми outcomes
- рассчитать кредитный спред (разница между доходностью к погашению корпоративных облигаций и гособлигаций, рассчитать на основе индексов)

▼ model (summary):

- 2 тип: эксперимент с interactions (попарными произведениями изменения ключевой ставки и Tone Sentiment 1/2)
 - добавить interaction между tone и trust, tone и transp
- 3 тип: dynamic effects - у не за сегодня, а за завтра (вместе с минутками огонь)

- (if) rolling window regression 1-35, 2-36 ... 36-70
- ▼ считать новые регрессии (1 тип)
 - индексы гособлигация разных диапазонов
 - индекс корпоративных облигаций
 - кредитный спред
 - обменных курсов
- ▼ считать новые регрессии (2 тип)
 - сделать таблицу со всеми переменными
- ▼ считать новые регрессии (3 тип)
 - рассчитать датасеты
 -
- добавить t-статистику для суммарного эффекта в каждую таблицу
 - F statistics & p-value
- ▼ пересчитать функ зав-ть между tone variables (robust errors, w constants)
 - это не корреляции!!!
 - без констант появляется значимость, но по-моему, на это надо забыть
- (if) добавить график ликвидности FXGD
- (if) сопоставить с графиком FXGD, считать корреляцию
- ▼ robustness calculations
 - ▼ сделать таблицу с кросс-корреляциями между регрессорами и outcomes
 - график как одна с другой
 - кросс-корреляция между регрессорами
- ▼ writing
 - ▼ введение
 - чек адекватность
 - конкретнее описать forward guidance - прояснение повестки, видения ЦБ о будущем, сигнал рынку (нужно дать норм ссылку с определением)
 - ссылаться на mishkin handbook of banking
 - ▼ литра
 - видимо здесь сказать про нлп
 - ~~ссылаться на петрову~~

- подробнее рассказать про методы NLP: еще 1 страница про методы какие есть, объяснить почему я выбрал именно эти
 - не страница, уонечно но ок
- data
 - ▼ NLP
 - сослаться на качество достоевского мол тоже ничего + описать модель
 - писать что считаю сепаратели для конф и релизов
 - подробнее описать поведение тональности
 - сказать почему может быть низкая корреляция
 - мол неточные индексы и все такое - здесь бы ~~про петрову добавить...~~
 - экспериментальный метод, нет лучших практик
 - что-то упущено
 - уточнить, что именно вносить негативный вклад (евстигнеева и ко)
 - пояснить почему нельзя добавить иллюстрации, примеры с сильно позитив/ негатив текстом
 - ~~описать методы петровой...~~
 - сказать, что при расчете регрессий нормирую на стандартное отклонение
 - ▼ секция outcomes
 - дать в текст ссылку на сайт мосбиржа
 - сказать как считаю ретерны и что нормирую
 - разбить на несколько параграфов
 - добавить описание бондов
 - сказать что для них некоторые ретрены как log close/close
 - поместить коммент в таблицы
 - чек что использую percentage points или basis points
 - писать что умножаю на 100
 - и во всех таблицах менять
 - ▼ дать саммари
 - сказать что rate change важно добавить!
 - описать кросс-корреляции - в отдельную секцию тесты?
 - поместить в текст?

- no perfect multicollinearity

▼ эмпирикал модель

- Проверить написанное
 - проверить, что не пишу второй раз про rate change
 - убрать индекс i у outcomes
 - добавить в notes табличек что эта спецификация
 - сказать что считаю robust ошибки так как мало наблюдений
 - в help reg stat почитать
 - мб сюда и добавить про корр
- добавить items и новые уравнения
 - с интеракшнс
 - можно себе представить, что реакция outcome variable на заданное изменение процентной ставки может зависеть от его вербального сопровождения в ходе пресс-конференции и/или в тексте пресс-релиза.
 - добавление интеракшн с траст не дало эффекта видимо из-за корреляций
 - dynamic effects
 - ссылаться на городниченко и jorda - мол прогоняю отдельные регрессии
 - считал как $\log \text{close}_h - \log \text{open}_0$
 - сказать что тут как бы панель и возможны ошибки among events и ошибки among time - первую adress with robust std. второе проверил тестом
 - ссылки в инете
 - прогнав регрессии вулдридж тест
 - dynamic точно называется dynamic?

▼ результаты

- вообще везде писать одновременно про interaction и dynamic
 - ~~если верна гипотеза об эффективности финрынков эффект должен произойти и остановиться (не дб прироста)~~
- подробнее описать для рынка
- золота

- переписать
- добавить таблу
- добавить секцию про бонды новых регрессий
 - можно сопоставить с петровой — если одинаковая нормировка
 - абзац про госбонды
 - писать что у бондов знаки в одном направлении (если это правда)
 - про разные матъюритис
 - одинаковая динамика
 - про корпоративный и композитный
 - корпоративный более чувствительный
 - знаки в одну сторону с остальными бондами
- описать почему разные коэффициенты для tone sentiment
- ▼ добавить robustness section (или discussions & extensions):
 - OVB
 - третья переменная влияет
 - гвоорим, что цб проводит независимую полтику
 - написать нет ли reverse causality? что тональность зависит от шока рынка? хотя и не зависит от шока ставки
 - можно улучшить используя другую модель
 - вообще мало наблюдений - всего 40!
 - boot strap - более точный расчет значимости
 - добавить больше текстовых индикаторов
 - перейти на минутки, чтобы корректнее разделить эффекты от релизов и конференций
 - можно добавить новые метрики - обменный курс
 - нужно добавлять новые контроли - бодяга про валютные интервенции
- обновить conclusions
- написать abstract
- еще раз перечитать и проверить грамматику
- чек формат references
- (if) описать что FXGD sufficiently liquid for our purposes (+ график)

▼ Ultra-far-future plans & Robustness checks

▼ планировалось

- **ml part:** добавить clarity и pos. neg продвинутые
 - обучить bert по сметанину на kaggle russian news dataset
 - добавить clarity by alina
 - посмотреть как связана тональность пресс-релизов и выступлений (tests b/w units)
 - посмотреть как между собой связаны (tests b/w properties)
- **ml part:** добавить dictionary-based method как benchmark и интерпрет. результат
 - найти словарь positive/ negative или hawkish/ dovish
 - константин предложил составить самому (если будут идеи и время)
- **data:** разбить наблюдения выступлений на Q&A session и основную часть
- **data:** мб перейти на минутки
- **window rolling regression**
- bootstrap test (Gorodnichenko)
- как контролировать на ожидания относительно будущего?
- обучить модель предсказывать ставку по тексту гыгыгы
- анализ метрик - как доверие зависит от transperncy

DOOOOOOOOOOOOOOOOOONEEEE!!!!!!!