



РАНХиГС

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Э М И Т
И Н С Т И Т У Т
ЭКОНОМИКИ, МАТЕМАТИКИ
И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И
ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Анализ влияния информационного фона на стоимость финансовых активов

магистерская программа «Экономика и Финансы»

2 июля 2018 г.

Студент: Ульяновкин Филипп Валерьевич

filfonul@gmail.com

Научный руководитель: к.э.н. Синельникова-Мурылёва Е.В.

- ▶ Каждый год объёмы неструктурированной текстовой информации растут;
- ▶ В последнее время интенсивно развиваются алгоритмы машинного обучения, позволяющие обрабатывать её;
- ▶ Понимание общего фона новостей и записей в социальных сетях может помочь в объяснении динамики различных экономических переменных.

Анализ предметной области

Авторы	Выборка, период	Метод исследования	Результат
(Utaka, 2003)	Япония, 04/1983 - 09/1998	ИФН на основе соц. опросов	Индекс поиска грейнджер причина для потребления. AIC, BIC и RMSE выбирают расширенную модель.
(Varian and Choi, 2012)	США, 01/2004 - 12/2011	Индекс поиска	Индекс поиска улучшает прогнозы для рынка недвижимости, безработицы, позволяет прогнозировать индекс доверия потребителей.
(Bloom, 2009)	США, 01/1965 - 12/2009	Индекс новостей	Индекс принимает высокие значения перед различными знаковыми событиями.
(Curme et all, 2014)	США, 04/01/2004 - 16/01/2012	Тематическое моделирования, индекс поиска	Увеличение объема поиска по политическим и экономическим темам предшествует падению рынка. Торговля на основе поиска даёт значимый доход. Прочность связи уменьшается.
(Yakovleva, 2017)	Россия, 01/2014 - 01/2017	Тематическое моделирование	Встречаемость тем в СМИ помогает предсказать PMI. AIC, BIC и RMSE выбирают расширенную модель.
(Frisbee, 2010)	США, 10/07/2006 - 21/12/2009	Анализ тональности мнений в финансовых блогах	Объем торговли в большей степени зависит от общих рыночных условий (ставка по федеральным фондам), чем от настроений инвесторов.
(Bollen, Mao and Zeng, 2011)	28/02/2008 - 19/12/2008	Анализ тональностей Twitter	Отношение позитивных твиттов к негативным, спокойные посты, счастливые посты грейнджер-причина для динамики индекса Доу-Джонса. Учёт эмоций улучшает прогнозы нейросети.

Цели и задачи

Цель:

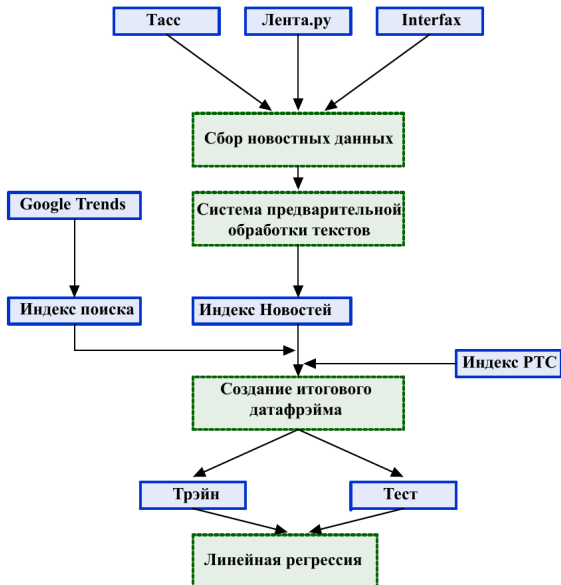
- ▶ Проверить оказывает ли изменение информационного фона влияние на формирование цен финансовых активов в России.

Задачи:

1. Обобщить актуальные исследования по вопросу взаимосвязи информационного фона с котировками.
2. Проанализировать методологии, используемые для анализа информационного фона.
3. Выбрать методы машинного обучения и эконометрические методы для дальнейшего анализа.
4. Собрать и провести предварительную обработку необходимых данных.
5. Проверить работоспособность выбранных способов учёта информационного фона на российских данных. Сравнить их между собой.

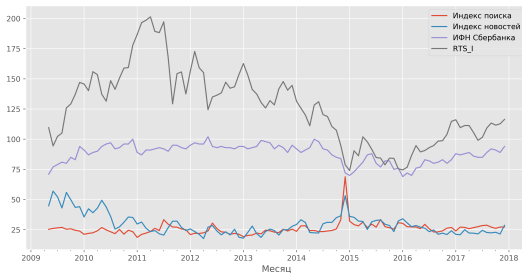
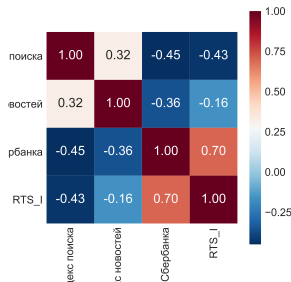
1. Индекс РТС (дневная и месячная динамика);
 2. Индекс финансовых настроений сбербанка (месячная динамика);
 3. Данные о динамике поисковых запросов из Google-trends (месячная динамика);
 4. Новостные тексты из ряда интернет СМИ.
- Новости собраны за период с 1 января 2009 года по 31 декабря 2017 года (больше 1 млн. новостей, 200 тыс. экономических новостей).

Первый этап исследовательской работы



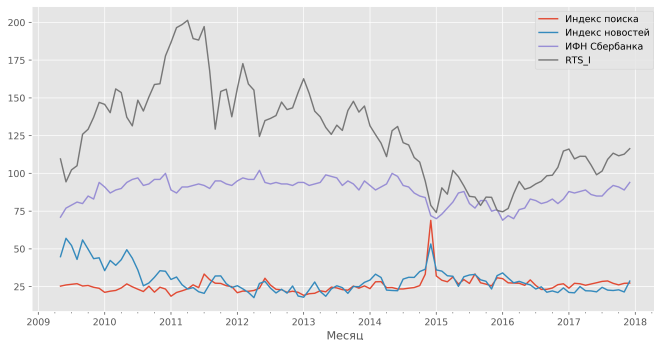
Частотные индексы

целевая переменная	причина	p-значение	результат теста (5%)
RTS d	ind IFN	0.336	не является причиной по Грейнджеру
ind IFN	RTS d	0.663	не является причиной по Грейнджеру
RTS d	ind poisk	0.005	причина по Грейнджеру
ind poisk	RTS d	0.114	не является причиной по Грейнджеру
RTS d	ind news	0.008	причина по Грейнджеру
ind news	RTS d	0.038	причина по Грейнджеру

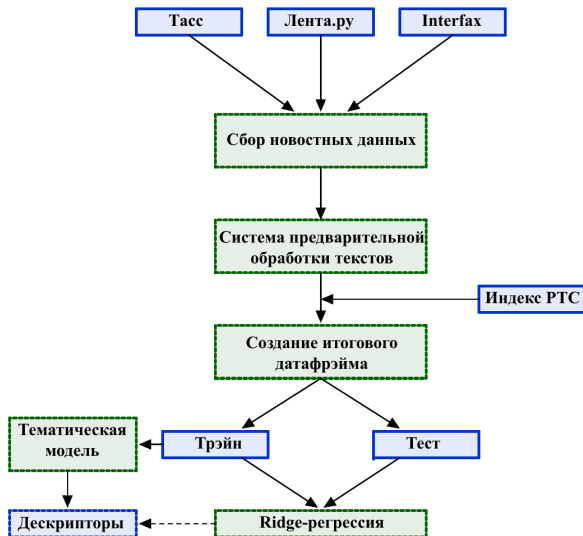


Частотные индексы

	модель 1	модель 2	модель 3	модель 4
$\Delta \text{ind_poisk}$	-0.005 (1.5e-3) *			-0.0042 (1e-3)*
$\Delta \text{ind_news}$		-0.004 (1.7e-3) *		-0.0008 (2e-3)
$\Delta \text{ind_IFN}$			0.003 (2e-3)	0.0019 (2e-3)
AIC	-172	-166	-164	-172
BIC	-167	-161	-159	-167



Второй этап исследовательской работы



Тематическое моделирование

- ▶ Только экономические новости;
- ▶ LDA-модель с 50 априорно заданными темами;
- ▶ Выборка: 01.01.2011 - 01.01.2017;
- ▶ Объясняемые переменные: приросты тем за день.

	Estimate	Std. Error	t value	Pr(> t)
X14	0.52	0.22	2.40	0.016 *
X19	-0.55	0.21	-2.60	0.009**
X20	0.37	0.11	3.18	0.001**
X37	-0.43	0.15	-2.80	0.005**

Интерпретация тем

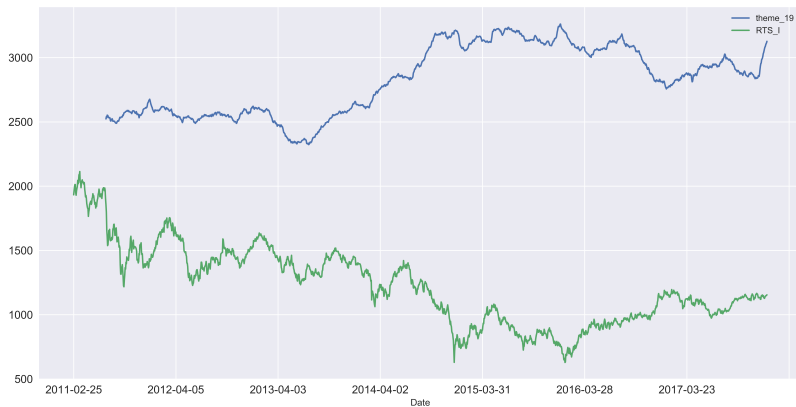
Тема 14: Западные страны,

Тема 19: Суды, следствия,

Тема 20: Поставки газа, газпром,

Тема 37: Страны ближнего востока и развивающиеся страны.

Динамика индекса РТС и темы 19: суды, следствия



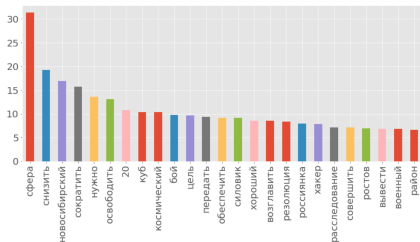
Ridge-регрессия

- ▶ Экономические новости очищены от стоп-слов, токенизированы, лемматизированы, tf-idf;
- ▶ Логистическая регрессия с l_2 регуляризатором;
- ▶ Обучение модели: SGD;
- ▶ Train: 01.01.2014 - 01.01.2017;
- ▶ Test: 01.01.2017 - 01.01.2018.

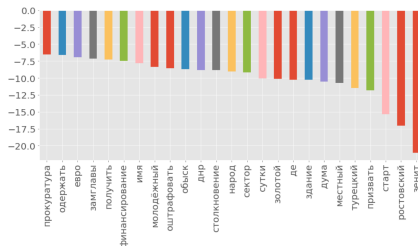
Гиперпараметр C	Доходность	SR	Просадка	5% VaR	Accuracy
10^{-8}	-1.4%	-1.3	18%	1.6%	47.4%
10^{-5}	0.7%	0.6	20.4%	1.7%	51.4
10^{-1}	24.1%	22.7	11.5%	1.6%	51.4%
10^1	30.1%	28.4	10.6%	1.6%	51.8%
10^5	24.3%	23	11.1%	1.6%	50.6%
10^8	27.5%	25.9	11.3%	1.6%	51.4%

Слова, формирующие фон

Слова, формирующий положительный фон:



Слова, формирующий отрицательный фон:



- Попытка использовать эти слова как дескрипторы для индекса новостей. Получена незначимая модель.

Теоретические результаты исследования

Месячные данные:

Модель	AIC	SIC	MAPE	RMSE
Наивный прогноз	—	—	3.77	0.052
Индекс поиска	-172	-167	1.17	0.051
Индекс новостей	-166	-161	2.32	0.052
Индексы на основе Ridge	-134	-127	3.98	0.062

Дневные данные:

Модель	AIC	SIC	MAPE	RMSE
Наивный прогноз	—	—	4.67	0.016
Индекс новостей	-5914	-5904	1.02	0.011
Тематическая модель	-6017	-5991	1.78	0.012
Индексы на основе Ridge	-5760	-5744	—	—
Ridge-регрессия	—	—	1.22	0.010

- ▶ Разработаны индексы поиска, новостей, тематические индексы, позволяющие отслеживать динамику информационного фона;
- ▶ Установлено, что новости, публикуемые в СМИ формируют информационный фон, оказывающий значимое влияние на индекс РТС;
- ▶ Выявлено, что значимое влияние на индекс РТС оказывают темы новостей, связанные с Западом, судом, поставками газа и странами ближнего Востока.
- ▶ Найдены конкретные слова, появление которых в СМИ формирует положительный/отрицательный информационный фон.

Спасибо за внимание

Анализ влияния информационного фона на
стоимость финансовых активов

Студент: Ульянов Филипп Валерьевич

filfonul@gmail.com

2 июля 2018 г.