



Факультет экономических наук

Школа финансов

Москва 2022

ВЛИЯНИЕ ИНФОРМАЦИИ О ПАНДЕМИИ КОРОНОВИРУСА НА ФИНАНСОВЫЕ РЫНКИ

Выполнили:

Королева Софья Андреевна, БЭК1811

Разуваев Никита Сергеевич, БСТ181

Научный руководитель:

д.э.н, профессор Федорова Елена Анатольевна



Содержание

- Введение
- Обзор литературы
- Гипотезы исследования
- Основные этапы исследования
- Методология
- Данные
- Эмпирические результаты
- Выводы



Актуальность данного исследования заключается в отсутствии достоверной оценки влияния информации о вакцинации против Covid-19 на доходность основных и отраслевых индексов финансовых рынков:

1

Высокая значимость неофициальной информации для инвесторов в последнее десятилетие

2

Особое внимание к программам вакцинации за последние несколько лет

3

Отсутствие исследований на данную тему, использующих текстовый анализ

Цель исследования: выявление наличия влияния информации о разработке вакцин и программах вакцинации против Covid-19 на доходность основных фондовых индексов, а также индексов потребительского и транспортного секторов, как наиболее пострадавших в период пандемии, для США, Китая и России.

Задачи исследования:

- Изучить и описать в обзоре литературы последние научные результаты по рассматриваемой теме;
- Выдвинуть гипотезы исследования;
- На основе анализа соответствующей литературы выбрать оптимальную методологию и описать данные;
- Построить модель, отвечающую всем требованиям и провести эмпирический анализ влияния тональности информации о вакцинах на доходность индексов финансовых рынков рассматриваемых стран.

Объект исследования: основные фондовые индексы на финансовых рынках США, Китая и России, а именно S&P500, CSI300 и IMOEX соответственно, а также отраслевые индексы Wholesale & Retail (SSE: W&S) и Transportation (SSE: T) для Китая, MOEXCN и MOEXTN для России и Dow Jones Consumer Goods (DJCG) и Transportation (DJT) indices для США.

Предмет исследования: влияние информации о вакцинах из официальных и неофициальных источников на доходность индексов финансовых рынков рассматриваемых стран.



Нарративы и
финансовые рынки

Covid-19 и
финансовые рынки

Тема	Публикации	Основные выводы
Влияние информации из официальных источников на финансовые рынки	Zhang и др. (2020), Rouatbi и др. (2021), Apergis и др. (2021), Zorgati и др. (2021)	<ul style="list-style-type: none">• Наблюдается корреляция между поведением инвесторов и публикуемыми официальными данными;• От особенностей рассматриваемой страны и от уровня доверия населения к государству зависит величина эффекта влияния.
Влияние информации из неофициальных источников на финансовые рынки	Jiao & Walther (2016), Audrino и др. (2020), Ульянов (2020)	<ul style="list-style-type: none">• Интенсивное обсуждение новостей в социальных сетях коррелирует с повышенной волатильностью и объемом торгов;• Переменные тональности и настроения способны значительно улучшить прогнозы волатильности;
Информация о Covid-19 и динамика фондовых рынков	Long и др. (2021), eFerguson и др. (2020), Zhang и др. (2020), Malik, Sharma (2021)	<ul style="list-style-type: none">• Новости положительно коррелируют с волатильностью ценных бумаг в краткосрочном периоде;• Наблюдается корреляция в поведении финансовых рынков территориально близко расположенных стран.
Информация о программах вакцинации и динамика фондовых рынков	Rouatbi и др. (2021), Apergis и др. (2021), Chan и др. (2021), Bakry и др. (2021), Zorgati и др. (2021)	<ul style="list-style-type: none">• Программы вакцинации стабилизируют финансовые рынки;• Влияние программ вакцинации оказывается более значительным в развитых странах, чем в странах с переходной экономикой.



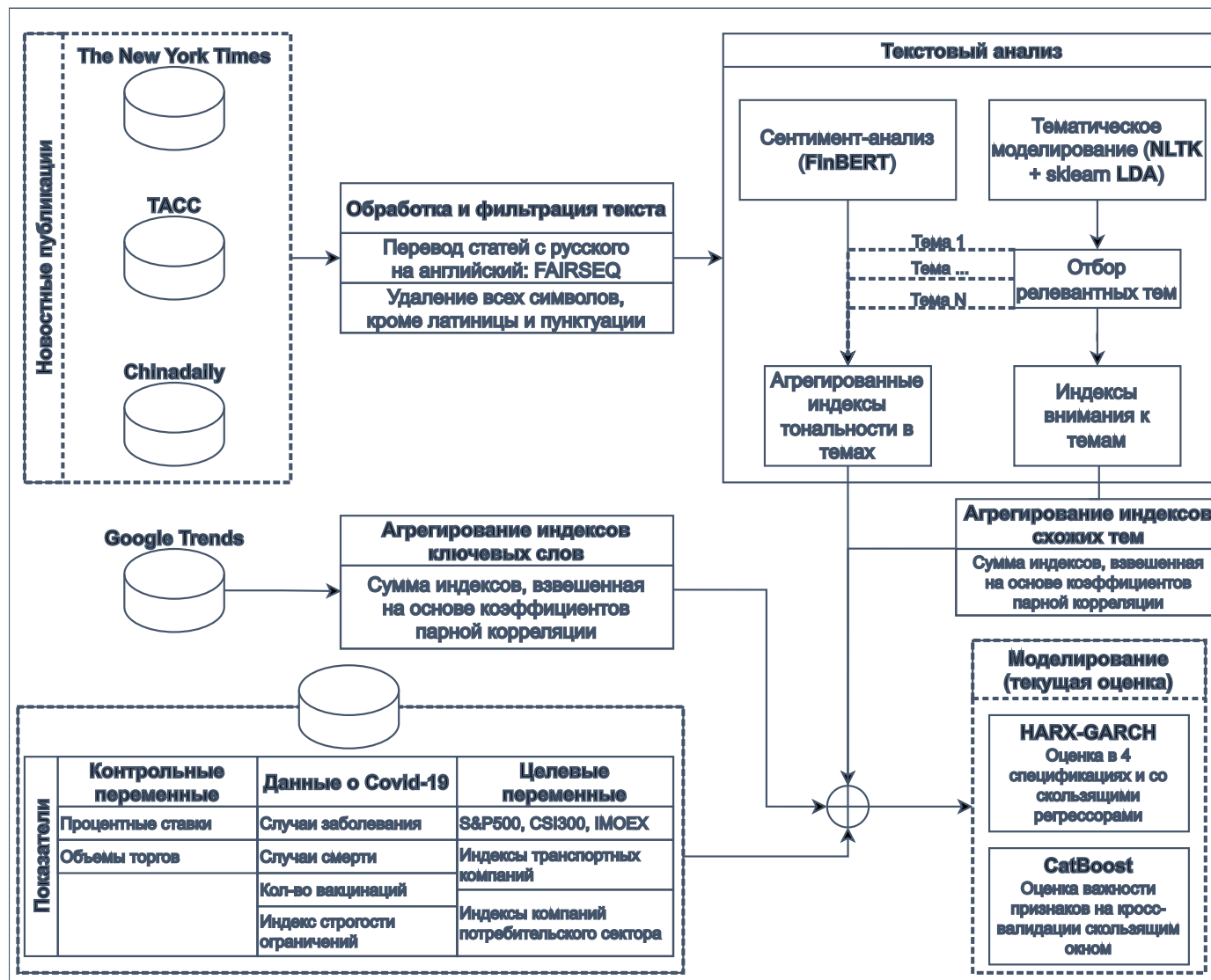
Гипотезы исследования

Гипотеза 1	Гипотеза 1а
<p>Информация о разработке вакцин и программах вакцинации против Covid-19 влияет на доходность индексов фондового рынка.</p> <p><i>Chan и др. (2021), Rouatbi и др. (2021)</i></p>	<p>Любые новости о вакцинах приводят к росту доходности индексов транспортного и потребительского секторов.</p> <p><i>Ранее не тестировалась</i></p>
Гипотеза 1б	Гипотеза 1в
<p>Новости о вакцинах с отрицательной тональностью имеют более существенное влияние на доходность индексов фондового рынка, чем новости с положительной тональностью.</p> <p><i>Jiao & Walther (2016)</i></p>	<p>Информация из разных источников о разработке вакцинах и программах вакцинации по-разному влияет на поведение индексов фондовых рынков.</p> <p><i>Ranco (2016)</i></p>



Гипотезы исследования

Гипотеза 2	Гипотеза 3
<p>Фондовые рынки реагируют на информацию о вакцинах не только в своей стране, но и в других странах.</p> <p><i>Consoli и др. (2021)</i></p>	<p>Переменные, которые описывают внимание, настроения и темпы вакцинации, по-разному влияют на фондовые индексы.</p> <p><i>Piantadosi и др. (2014), Colladon и др. (2020)</i></p>
Гипотеза 4	Гипотеза 5
<p>Реакция инвесторов на информацию о вакцинах и программах вакцинации не является немедленной.</p> <p><i>Consoli и др. (2021)</i></p>	<p>Информация о вакцинации больше повлияла на доходность индексов транспортного сектора, чем на индексы потребительского сектора.</p> <p><i>Sahoo & Ashwani (2020)</i></p>





Временные ряды

- Дневные данные с 1 января 2020 по 31 декабря 2021.

Короткое название	Полное название	Полнота (из 731 дня)	Источник	Короткое название	Полное название	Полнота (из 731 дня)	Источник
Целевые переменные (индексы)				Данные о COVID-19			
us_ind_comp_close_price	S&P 500, цена закрытия, USD	504	finance.yahoo.com	{us, ru, cn}_new_cases	Новые зарегистрированные случаи заражения COVID-19	729	covid19.who.int
us_ind_wr_close_price	Dow Jones U.S. Consumer Goods Index, цена закрытия, USD	505	finance.yahoo.com	{us, ru, cn}_new_death	Новые зарегистрированные случаи смерти от COVID-19		
us_ind_tr_close_price	Dow Jones Transportation Average, цена закрытия, USD	504	finance.yahoo.com	{us, ru, cn}_total_vaccinations_per_hundred	Общее число вакцинаций, приходящееся на 100 млн человек	384 (us), 281 (ru) и 297 (cn)	ourworldindata.org
ru_ind_comp_close_price	Индекс IMOEX, цена закрытия, RUB	505	finam.ru	{us, ru, cn}_stringency_index	Индекс строгости ограничений против COVID-19	710 (us), 701 (ru), 710 (cn)	
ru_ind_wr_close_price	Индекс MOEXCN, цена закрытия, RUB		finam.ru	Контрольные переменные			
ru_ind_tr_close_price	Индекс MOEXTN, цена закрытия, RUB		finam.ru	{us, ru, cn}_ind_{comp, wr, tr}_vol	Объемы торгов соответствующими индексами	То же, что и у целевых переменных	
cn_ind_comp_close_price	Shanghai Shenzhen CSI 300, цена закрытия, CNY	486	finam.ru	us_overnight_{ind, 1m, 3m, 6m}	USD LIBOR interest rate	494	iborate.com
cn_ind_wr_close_price	Shenzhen Stock Exchange: Wholesale & Retail, цена закрытия, CNY		szse.cn	ru_overnight_{ind, 1m, 3m, 6m}	Срочная версия ставки RUONIA	731	cbr.ru
cn_ind_tr_close_price	Shenzhen Stock Exchange: Transportation, цена закрытия, CNY			cn_overnight_{ind, 1m, 3m, 6m}	Shanghai Interbank Offered Rate	499	shibor.org

Текстовый анализ

LDA

Приведение к нижнему
регистру, очистка от
пунктуации, стоп-слов и
шума

Лемматизация
(приведение слова к
нормальной
(словарной) форме)

Матрица частоты терминов (1
тыс. слов, встречающихся в
менее, чем 95% статьях и более
2 раз)

Применение
алгоритма **LDA**
(50 тем)

В результате тематического моделирования LDA были отобраны темы, связанные с пандемией Covid-19 и разделены на 3 смысловых блока – **наука и здравоохранение, общественные институты и экономика**.

FinBERT

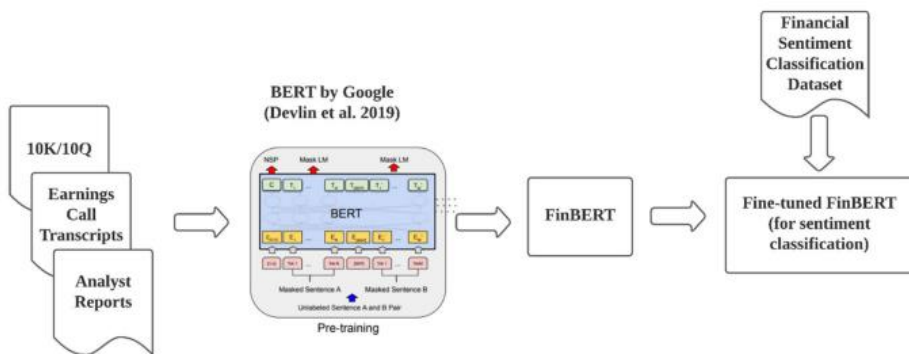


Рисунок 1. Предобучение и дообучение модели
BERT (Источник: Huang A. и др., 2020)

Преимущества

- ✓ Отсутствие необходимости в предобработке текста (лемматизации, токенизации), end-to-end подход;
- ✓ Учитывается **порядок, вес и связи** слов (токенов) в тексте.
- ✓ При прочих равных **превосходит по качеству** оценку тональности с помощью словарей.

Недостатки

- ✗ Ограничение входного слоя модели в **512 токенов** (слов, пунктуации или частей слов).
Решение: разбиение больших текстов новостных статей на части (их не более 1-2% в каждом издании) и усреднение для получения тональности всего текста
- ✗ **Смещение данных** (как распределения слов (токенов), так и длины текстов (статей и заголовков)) относительно обучающей выборки, что может привести к **зашумленным или ложным выводам**.

Агрегирование индексов Google Trends и тем LDA

$$I = w_1 I_1 + w_2 I_2 + \dots + w_n I_n$$

$$w_i = \frac{\sum_{j=1}^n \rho_{ij}}{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij}}$$

Темы LDA

- **Наука и здравоохранение.** В блоке часто встречаются упоминания вакцинации, активно публикуются новости о производстве вакцин и ходе программ вакцинации.
- **Общественные институты.** Новости о ковидных ограничениях, включающие закрытия школ, перевод сотрудников на удалённый режим работы, а также, в целом, об изменении жизненного уклада.
- **Экономика.** Упоминание пандемии в контексте экономических новостей.
- Итоговая выборка новостей включала **12451** публикаций в США, в Китае – **19826**, в России – **5195**

Таблица 1. Веса ключевых слов в итоговых индексах
(Источник: составлено автором)

Россия		США		Китай	
Ключевое слово	Вес	Ключевое слово	Вес	Ключевое слово	Вес
вакцинация	0.085	vaccination	0.169	疫苗接种 (вакцинация)	0.181
вакцина	0.077	vaccine	0.157	疫苗 (вакцина)	0.244
тест на антитела	0.104	test covid	0.124	covid test	0.111
тест на ковид	0.099	test antibody	0.089	coronavac	0.119
qr код	0.117	pfizer	0.147	sinovac	0.188
спутник v	0.165	moderna	0.179	sinopharm	0.157
эпиваккорона	0.110	novavax	0.136		
спутник лайт	0.125				
ковивак	0.118				

Моделирование с HARX-GARCH (1, 1), текущая оценка

- Модель состоит из двух компонент, описывающих динамики среднего значения μ_t и шоков ϵ_t :

$$r_t = \mu_t + \epsilon_t$$

$$\epsilon_t = \sigma_t e_t$$

$$r_t = \frac{y_t}{y_{t-1}} - 1; e_t \sim N(0, 1)$$

$$\text{GARCH}(1, 1): \sigma_t^2 = \omega + \sum_{i=1}^1 \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^1 \beta_j \sigma_{t-j}^2$$

$$\text{HARX}(1, 5, 22): \mu_t = \mu + \sum_{i=1}^p \phi_i \bar{\mu}_{t-L_i:t-1} + \gamma_1^T \text{controls}_t + \gamma_2^T \text{official_info}_t + \gamma_3^T \text{gtrends}_t + \gamma_4^T \text{news}_t$$

HAR(1, 5, 22):
лаговый многочлен
со скользящими
средними за 1, 5 и
22 дня

Контрольные
переменные:
ставки
overnight и
объемы торгов

Официальные данные о
Covid-19:
случаи заболевания,
смерти, новые
вакцинации, индекс
строгости ограничений

Индекс
внимания к
вакцинации
Google Trends

Новостные
переменные:
тональности,
темы Covid,
кол-во
новостей

$\bar{\mu}_{t-L_i:t-1} = \frac{1}{L_i} \sum_{j=1}^{L_i} \mu_{t-j}$ – среднее значение переменной между моментами времени $t - L_i$ и $t - 1$

HARX(1, 5, 22) со скользящими регрессорами: $\mu_t + \theta_1^T \overline{\text{controls}}_{t-L_i:t-1} + \theta_2^T \overline{\text{official_info}}_{t-L_i:t-1} + \theta_3^T \overline{\text{gtrends}}_{t-L_i:t-1} + \theta_4^T \overline{\text{news}}_{t-L_i:t-1}$

Оценка важности признаков в модели CatBoost

- Модель CatBoost представляет собой разновидность градиентного бустинга, построенного на деревьях решений. Каждое следующее дерево строится на подвыборках наблюдений и признаков так, чтобы минимизировать ошибку предыдущего дерева. Данные свойства позволяют снизить степень переобучения под конкретную выборку.
- Важность признаков оценивалась путем обучения модели CatBoost с максимальной глубиной дерева 2 и максимальным числом деревьев равным 500. Для обучения модели использовалась функция потерь MAE:

$$MAE = \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Кросс-валидация осуществлялась скользящим расширяющимся окном с максимальным размером выборки 264 дня. В результате модель оценивалась на 400 выборках для каждого индекса.



Рисунок 2. Механизм скользящего окна.
(Источник: составлено автором)

Динамика индексов внимания к темам

- Наблюдается схожая динамика освещения тематических блоков между странами.



Рисунок 3. Частота попадания новостей в выделенные тематические группы в странах. (Источник: составлено автором)

Тональности внутри тематических блоков

- Анализ тональностей внутри тематических блоков показывает, что экономические темы наиболее ярко окрашены с точки зрения модели FinBERT, что валидирует использование указанных подходов.
- Для России и США характерны ярко выраженные негативные заголовки, в то время как в Китае заголовки чаще всего позитивные.

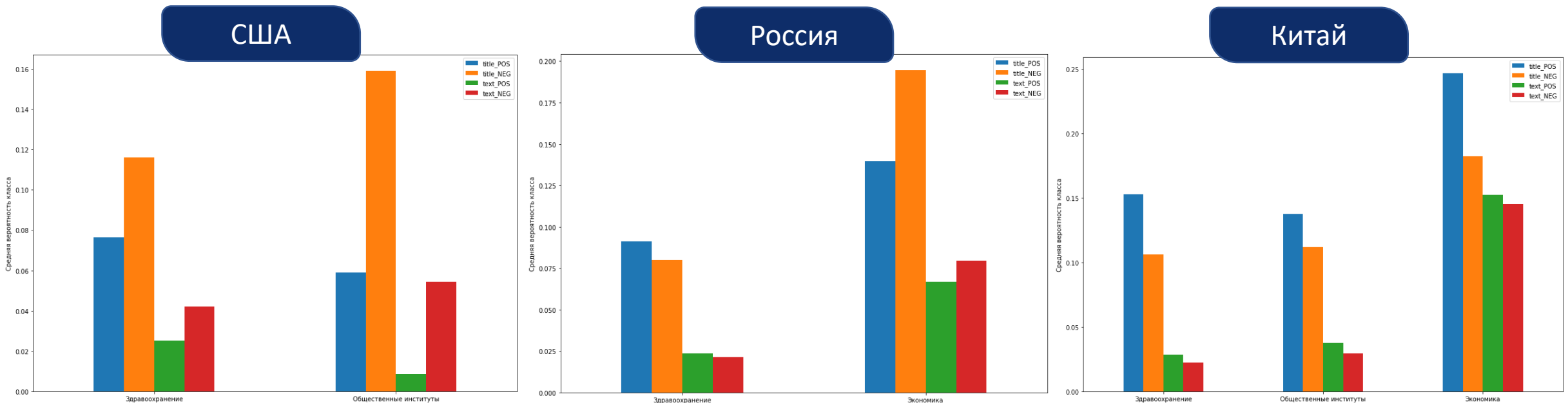


Рисунок 4. Тональности внутри тематических блоков. (Источник: составлено автором)

Индексы тональностей новостей на базе LDA



Рисунок 5. Динамика изменения сентимента новостей в странах. (Источник: составлено автором)

Результаты оценивания CatBoost на скользящем окне

- Новостные переменные являются наиболее важными признаками для объяснения доходности индексов в России и 2-ми по важности для США и Китая
- Контрольные переменные практически всегда относительно важны для объяснения доходности индексов
- Наименьшую ценность для оценки текущей доходности представляют официальные данные о Covid-19 во всех 3-х странах.
- Внимание к вакцинации по индексу Google Trends имеет большую значимость только в Китае

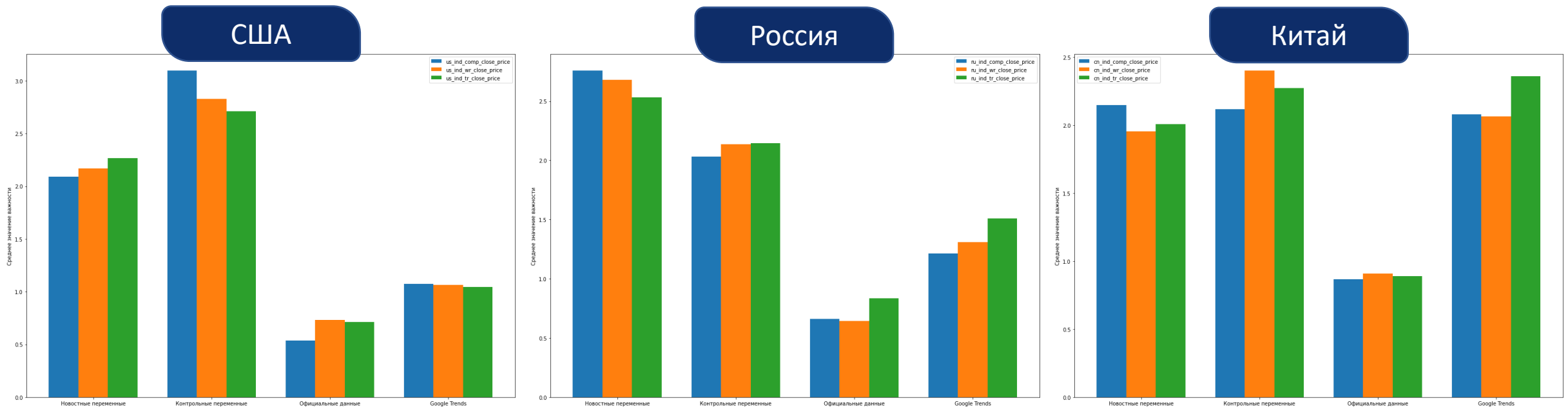


Рисунок 6. Важность признаков в модели CatBoost, микроусреднение. (Источник: составлено автором)



Гипотеза 1

Принимается

Информация о разработке вакцин и программах вакцинации против Covid-19 **влияет** на доходность индексов фондового рынка.

Гипотеза 16

Отвергается

Новости о вакцинах с отрицательной тональностью **не во всех случаях** имеют более существенное влияние на доходность индексов фондового рынка, чем новости с положительной тональностью.

Variable	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.0617	0.1009	0.1006	0.1156
R2_adj	0.0458	0.0779	0.0755	0.0772
Const	0.364 (1.749)	-0.343 (1.726)	-0.309 (1.722)	-1.129** (1.627)
i...ice[0:1]	-0.095*** (0.051)	-0.103*** (0.052)	-0.103*** (0.052)	-0.086*** (0.049)
i...ice[0:5]	-0.188** (0.140)	-0.221** (0.147)	-0.225** (0.149)	-0.304*** (0.151)
i...ice[0:22]	-0.295** (0.319)	-0.584*** (0.323)	-0.599*** (0.321)	-0.575*** (0.299)
ind_comp_vol	-0.010*** (0.004)	-0.017*** (0.005)	-0.017*** (0.005)	-0.020*** (0.005)
overnight_ind	-0.008 (0.021)	-0.008 (0.020)	-0.008 (0.020)	0.002 (0.019)
overnight_1m	0.002 (0.005)	0.001 (0.005)	0.001 (0.005)	0.001 (0.005)
overnight_3m	0.007** (0.005)	0.005** (0.005)	0.005** (0.005)	0.007** (0.004)
overnight_6m	0.002 (0.007)	0.001 (0.007)	0.001 (0.007)	-0.002 (0.006)
omega	0.069*** (0.033)	0.056*** (0.022)	0.056*** (0.022)	0.040*** (0.016)
alpha[1]	0.280*** (0.110)	0.230*** (0.064)	0.232*** (0.067)	0.258*** (0.073)
beta[1]	0.696*** (0.093)	0.740*** (0.048)	0.739*** (0.049)	0.734*** (0.048)
new_cases		-0.006*** (0.004)	-0.005** (0.005)	-0.004** (0.005)
new_death		0.003 (0.004)	0.002 (0.004)	0.002 (0.005)
total_vaccinations_per_hundred		0.003** (0.002)	0.004** (0.003)	0.003** (0.003)
stringency_index		0.015*** (0.006)	0.015*** (0.006)	0.015*** (0.005)
gtrend_corr			-0.002 (0.003)	-0.000 (0.003)
news_count				0.014** (0.008)
title_POS				0.004*** (0.002)
title_NEG				-0.003** (0.005)
text_POS				0.003** (0.003)
text_NEG				-0.009*** (0.002)
topic_1_corr				-0.006 (0.011)
topic_2_corr				0.007** (0.004)

Таблица 2. Результаты оценивания HARX-GARCH модели на композитном индексе **США**.
(Источник: составлено автором).



Гипотеза 1

Принимается

Информация о разработке вакцин и программах вакцинации против Covid-19 **влияет** на доходность индексов фондового рынка.

Гипотеза 16

Отвергается

Новости о вакцинах с отрицательной тональностью **не во всех случаях** имеют более существенное влияние на доходность индексов фондового рынка, чем новости с положительной тональностью.

Нельзя сделать однозначный вывод из-за разницы в результатах для каждой страны.

Таблица 3. Результаты оценивания HARX-GARCH модели на композитном индексе России. (Источник: составлено автором).

Variables	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.0149	0.0306	0.0313	0.1037
R2_adj	-0.0018	0.0058	0.0044	0.0648
Const	0.495*** (0.242)	0.073 (0.373)	0.057 (0.374)	0.289 (0.539)
i...ice[0:1]	0.010 (0.049)	0.011 (0.050)	0.010 (0.050)	0.002 (0.049)
i...ice[0:5]	0.034 (0.131)	0.037 (0.131)	0.032 (0.131)	0.027 (0.129)
i...ice[0:22]	-0.016 (0.287)	-0.121 (0.303)	-0.115 (0.302)	-0.144 (0.318)
ind_comp_vol	-0.008** (0.006)	-0.008** (0.006)	-0.008** (0.006)	-0.008** (0.007)
overnight_ind	-0.003** (0.005)	-0.004** (0.005)	-0.004** (0.005)	-0.004** (0.004)
overnight_1m	-0.006** (0.004)	-0.004** (0.004)	-0.004** (0.004)	-0.006** (0.004)
overnight_3m	0.007*** (0.004)	0.009*** (0.004)	0.009*** (0.004)	0.011*** (0.004)
overnight_6m	-0.002** (0.004)	-0.000 (0.004)	-0.000 (0.004)	-0.001 (0.004)
omega	0.026*** (0.015)	0.025*** (0.014)	0.026*** (0.014)	0.024** (0.016)
alpha[1]	0.080*** (0.032)	0.079*** (0.032)	0.078*** (0.032)	0.080*** (0.038)
beta[1]	0.899*** (0.031)	0.900*** (0.030)	0.901*** (0.030)	0.900*** (0.039)
new_cases		0.001 (0.005)	0.003 (0.006)	0.002 (0.006)
new_death		-0.005** (0.006)	-0.006** (0.006)	-0.006** (0.007)
total_vaccinations_per_hundred		-0.001 (0.004)	0.000 (0.004)	-0.003 (0.004)
stringency_index		0.005** (0.003)	0.005** (0.003)	0.009*** (0.004)
gtrend_corr			-0.003** (0.003)	-0.001 (0.003)
news_count				-0.001 (0.004)
title_POS				0.003 (0.008)
title_NEG				-0.028*** (0.010)
text_POS				0.013*** (0.007)
text_NEG				-0.009** (0.010)
topic_1_corr				0.011** (0.012)
topic_3_corr				-0.001 (0.007)



Гипотеза 1а

Принимается

Любые новости о вакцинах приводят к изменениям доходности индексов транспортного и потребительского секторов.

Гипотеза 5

Отвергается

Информация о вакцинации **по-разному** повлияла на доходность индексов транспортного сектора и потребительского сектора для каждой из стран.

Variables	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.031	0.0466	0.0453	0.0591
R2_adj	0.0146	0.0222	0.0188	0.0183
Const	0.641 (2.105)	0.307 (2.140)	0.390 (2.137)	-0.046 (2.113)
i...ice[0:1]	-0.040** (0.054)	-0.043** (0.053)	-0.040** (0.053)	-0.041** (0.052)
i...ice[0:5]	-0.015 (0.127)	-0.100** (0.138)	-0.120** (0.140)	-0.149** (0.143)
i...ice[0:22]	-0.236** (0.253)	-0.566*** (0.264)	-0.622*** (0.261)	-0.556*** (0.263)
overnight_ind	-0.013 (0.025)	-0.017** (0.025)	-0.017** (0.025)	-0.011 (0.024)
overnight_1m	0.003 (0.007)	0.003 (0.007)	0.003 (0.007)	0.004 (0.007)
overnight_3m	0.013*** (0.006)	0.012*** (0.006)	0.012*** (0.006)	0.013*** (0.006)
overnight_6m	-0.002 (0.008)	-0.003 (0.007)	-0.002 (0.007)	-0.003 (0.007)
omega	0.063** (0.053)	0.072*** (0.041)	0.068*** (0.039)	0.063** (0.038)
alpha[1]	0.149*** (0.073)	0.184*** (0.064)	0.185*** (0.064)	0.182*** (0.064)
beta[1]	0.811*** (0.095)	0.776*** (0.070)	0.778*** (0.069)	0.784*** (0.070)
ind_wr_vol	-0.010*** (0.004)	-0.010*** (0.005)	-0.011*** (0.005)	-0.012*** (0.005)
new_cases		0.001 (0.005)	0.005** (0.005)	0.002 (0.006)
new_death		-0.000 (0.006)	-0.001 (0.006)	-0.000 (0.006)
total_vaccinations_per_hundred		-0.002** (0.002)	0.001 (0.003)	0.001 (0.003)
stringency_index		0.012*** (0.006)	0.012*** (0.006)	0.010** (0.006)
gtrend_corr			-0.005*** (0.003)	-0.004** (0.003)
news_count				-0.000 (0.011)
title_POS				0.004** (0.003)
title_NEG				-0.003 (0.005)
text_POS				0.001 (0.003)
text_NEG				-0.007*** (0.003)
topic_1_corr				0.011** (0.011)
topic_2_corr				0.006** (0.005)

Таблица 4. Результаты оценивания HARX-GARCH модели на индексе **потребительского сектора США**.
(Источник: составлено автором).



Гипотеза 1а

Принимается

Любые новости о вакцинах приводят к изменениям доходности индексов транспортного и потребительского секторов.

Гипотеза 5

Отвергается

Информация о вакцинации **по-разному** повлияла на доходность индексов транспортного сектора и потребительского сектора для каждой из стран.

Variables	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.0408	0.06	0.06	0.0757
R2_adj	0.0246	0.0359	0.0338	0.0355
Const	-1.296 (3.941)	-2.204 (4.250)	-2.239 (4.279)	-4.023** (3.862)
i...ice[0:1]	-0.022 (0.067)	-0.016 (0.067)	-0.016 (0.069)	-0.003 (0.061)
i...ice[0:5]	-0.135** (0.139)	-0.167** (0.177)	-0.168** (0.184)	-0.190** (0.194)
i...ice[0:22]	0.153 (0.381)	-0.113 (0.405)	-0.130 (0.407)	-0.165 (0.571)
overnight_ind	-0.012 (0.046)	-0.010 (0.049)	-0.010 (0.049)	0.004 (0.049)
overnight_1m	0.013** (0.013)	0.014** (0.013)	0.014** (0.013)	0.017** (0.022)
overnight_3m	0.005 (0.010)	0.004 (0.012)	0.004 (0.012)	0.006 (0.017)
overnight_6m	0.017** (0.016)	0.015** (0.021)	0.014 (0.022)	0.012 (0.029)
omega	0.233** (0.292)	0.235** (0.315)	0.232** (0.337)	0.238** (0.340)
alpha[1]	0.184** (0.195)	0.191** (0.217)	0.191** (0.230)	0.227** (0.243)
beta[1]	0.735*** (0.271)	0.725*** (0.302)	0.726*** (0.323)	0.691*** (0.331)
ind_tr_vol	0.017** (0.017)	0.008 (0.021)	0.008 (0.022)	0.005 (0.025)
new_cases		-0.010 (0.016)	-0.009 (0.018)	-0.012 (0.018)
new_death		0.005 (0.018)	0.004 (0.019)	0.007 (0.019)
total_vaccinations_per_hundred		0.003** (0.003)	0.004** (0.005)	0.003 (0.007)
stringency_index		0.013** (0.016)	0.014** (0.017)	0.009 (0.016)
gtrend_corr			-0.002 (0.005)	0.002 (0.006)
news_count				0.017** (0.020)
title_POS				0.009*** (0.005)
title_NEG				0.005 (0.007)
text_POS				0.003 (0.006)
text_NEG				-0.014*** (0.004)
topic_1_corr				0.005 (0.024)
topic_2_corr				0.015** (0.014)

Таблица 5. Результаты оценивания HARX-GARCH модели на индексе **транспортного сектора США**.
(Источник: составлено автором).



Гипотеза 1в

Принимается

Информация из разных источников о разработке вакцинах и программах вакцинации по-разному влияет на поведение индексов фондовых рынков.

Гипотеза 3

Принимается

Переменные, которые описывают внимание, настроения и темпы вакцинации, по-разному влияют на фондовые индексы.

Variable	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.0617	0.1009	0.1006	0.1156
R2_adj	0.0458	0.0779	0.0755	0.0772
Const	0.364 (1.749)	-0.343 (1.726)	-0.309 (1.722)	-1.129** (1.627)
i...ice[0:1]	-0.095*** (0.051)	-0.103*** (0.052)	-0.103*** (0.052)	-0.086*** (0.049)
i...ice[0:5]	-0.188** (0.140)	-0.221** (0.147)	-0.225** (0.149)	-0.304*** (0.151)
i...ice[0:22]	-0.295** (0.319)	-0.584*** (0.323)	-0.599*** (0.321)	-0.575*** (0.299)
ind_comp_vol	-0.010*** (0.004)	-0.017*** (0.005)	-0.017*** (0.005)	-0.020*** (0.005)
overnight_ind	-0.008 (0.021)	-0.008 (0.020)	-0.008 (0.020)	0.002 (0.019)
overnight_1m	0.002 (0.005)	0.001 (0.005)	0.001 (0.005)	0.001 (0.005)
overnight_3m	0.007** (0.005)	0.005** (0.005)	0.005** (0.005)	0.007** (0.004)
overnight_6m	0.002 (0.007)	0.001 (0.007)	0.001 (0.007)	-0.002 (0.006)
omega	0.069*** (0.033)	0.056*** (0.022)	0.056*** (0.022)	0.040*** (0.016)
alpha[1]	0.280*** (0.110)	0.230*** (0.064)	0.232*** (0.067)	0.258*** (0.073)
beta[1]	0.696*** (0.093)	0.740*** (0.048)	0.739*** (0.049)	0.734*** (0.048)
new_cases		-0.006*** (0.004)	-0.005** (0.005)	-0.004** (0.005)
new_death		0.003 (0.004)	0.002 (0.004)	0.002 (0.005)
total_vaccinations_per_hundred		0.003** (0.002)	0.004** (0.003)	0.003** (0.003)
stringency_index		0.015*** (0.006)	0.015*** (0.006)	0.015*** (0.005)
gtrend_corr			-0.002 (0.003)	-0.000 (0.003)
news_count				0.014** (0.008)
title_POS				0.004*** (0.002)
title_NEG				-0.003** (0.005)
text_POS				0.003** (0.003)
text_NEG				-0.009*** (0.002)
topic_1_corr				-0.006 (0.011)
topic_2_corr				0.007** (0.004)

Таблица 2. Результаты оценивания HARX-GARCH модели на композитном индексе **США**.
(Источник: составлено автором).



Гипотеза 1в

Принимается

Информация из разных источников о разработке вакцинах и программах вакцинации по-разному влияет на поведение индексов фондовых рынков.

Гипотеза 3

Принимается

Переменные, которые описывают внимание, настроения и темпы вакцинации, по-разному влияют на фондовые индексы.

Variables	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.0149	0.0306	0.0313	0.1037
R2_adj	-0.0018	0.0058	0.0044	0.0648
Const	0.495*** (0.242)	0.073 (0.373)	0.057 (0.374)	0.289 (0.539)
i...ice[0:1]	0.010 (0.049)	0.011 (0.050)	0.010 (0.050)	0.002 (0.049)
i...ice[0:5]	0.034 (0.131)	0.037 (0.131)	0.032 (0.131)	0.027 (0.129)
i...ice[0:22]	-0.016 (0.287)	-0.121 (0.303)	-0.115 (0.302)	-0.144 (0.318)
ind_comp_vol	-0.008** (0.006)	-0.008** (0.006)	-0.008** (0.006)	-0.008** (0.007)
overnight_ind	-0.003** (0.005)	-0.004** (0.005)	-0.004** (0.005)	-0.004** (0.004)
overnight_1m	-0.006** (0.004)	-0.004** (0.004)	-0.004** (0.004)	-0.006** (0.004)
overnight_3m	0.007*** (0.004)	0.009*** (0.004)	0.009*** (0.004)	0.011*** (0.004)
overnight_6m	-0.002** (0.004)	-0.000 (0.004)	-0.000 (0.004)	-0.001 (0.004)
omega	0.026*** (0.015)	0.025*** (0.014)	0.026*** (0.014)	0.024** (0.016)
alpha[1]	0.080*** (0.032)	0.079*** (0.032)	0.078*** (0.032)	0.080*** (0.038)
beta[1]	0.899*** (0.031)	0.900*** (0.030)	0.901*** (0.030)	0.900*** (0.039)
new_cases		0.001 (0.005)	0.003 (0.006)	0.002 (0.006)
new_death		-0.005** (0.006)	-0.006** (0.006)	-0.006** (0.007)
total_vaccinations_per_hundred		-0.001 (0.004)	0.000 (0.004)	-0.003 (0.004)
stringency_index		0.005** (0.003)	0.005** (0.003)	0.009*** (0.004)
gtrend_corr			-0.003** (0.003)	-0.001 (0.003)
news_count				-0.001 (0.004)
title_POS				0.003 (0.008)
title_NEG				-0.028*** (0.010)
text_POS				0.013*** (0.007)
text_NEG				-0.009** (0.010)
topic_1_corr				0.011** (0.012)
topic_3_corr				-0.001 (0.007)

Таблица 3. Результаты оценивания HARX-GARCH модели на композитном индексе России. (Источник: составлено автором).



Гипотеза 1в

Принимается

Информация из разных источников о разработке вакцинах и программах вакцинации по-разному влияет на поведение индексов фондовых рынков.

Гипотеза 3

Принимается

Переменные, которые описывают внимание, настроения и темпы вакцинации, по-разному влияют на фондовые индексы.

Variables	HARX-GARCH	HARX-GARCH-O	HARX-GARCH-OG	HARX-GARCH-OGN
R2	0.0125	0.018	0.0238	0.0251
R2_adj	-0.0049	-0.0082	-0.0045	-0.0213
Const	0.215 (0.528)	-0.130 (0.571)	-0.000 (0.572)	0.362 (1.031)
i...ice[0:1]	0.026 (0.052)	0.023 (0.052)	0.021 (0.052)	0.016 (0.052)
i...ice[0:5]	-0.036 (0.124)	-0.038 (0.126)	-0.018 (0.127)	-0.024 (0.132)
i...ice[0:22]	-0.342** (0.319)	-0.500** (0.320)	-0.510** (0.317)	-0.518** (0.325)
ind_comp_vol	0.004** (0.004)	0.005** (0.004)	0.005** (0.004)	0.005** (0.005)
overnight_ind	0.000 (0.005)	-0.000 (0.005)	-0.000 (0.005)	-0.000 (0.005)
overnight_1m	-0.018** (0.014)	-0.020** (0.014)	-0.020** (0.014)	-0.020** (0.015)
overnight_3m	-0.008 (0.029)	-0.005 (0.029)	-0.006 (0.029)	-0.004 (0.030)
overnight_6m	0.018 (0.029)	0.016 (0.029)	0.016 (0.029)	0.015 (0.031)
omega	0.074*** (0.035)	0.073*** (0.037)	0.074*** (0.038)	0.075*** (0.042)
alpha[1]	0.094*** (0.037)	0.094*** (0.039)	0.095*** (0.040)	0.099*** (0.045)
beta[1]	0.857*** (0.042)	0.857*** (0.045)	0.856*** (0.047)	0.850*** (0.054)
new_cases		0.012 (0.035)	0.011 (0.036)	0.013 (0.038)
new_death		-0.027 (0.131)	-0.027 (0.133)	-0.041 (0.143)
total_vaccinations_per_hundred		-0.002** (0.002)	0.000 (0.003)	0.001 (0.003)
stringency_index		0.005** (0.003)	0.005** (0.003)	0.005** (0.003)
gtrend_corr			-0.005** (0.004)	-0.005** (0.004)
news_count				-0.002 (0.010)
title_POS				-0.001 (0.008)
title_NEG				-0.002 (0.008)
text_POS				-0.003 (0.006)
text_NEG				-0.006** (0.006)
topic_1_corr				0.001 (0.011)
topic_2_corr				-0.000 (0.014)

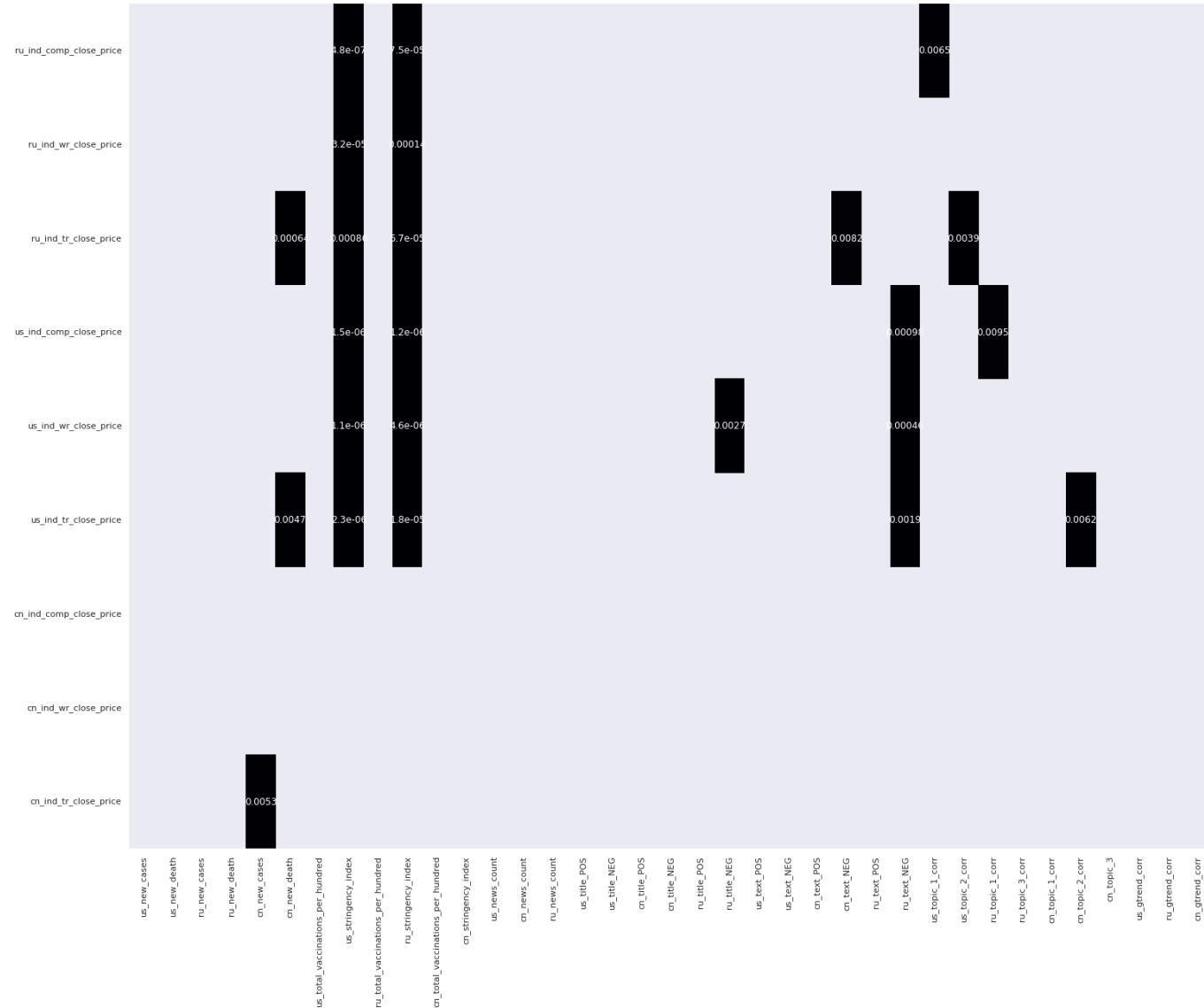
Таблица 6. Результаты оценивания HARX-GARCH модели на композитном индексе *Китая*. (Источник: составлено автором).



Отвергается

Исследование показало связь между новостями и фондовыми рынками **США и России.**

Рисунок 7. Тестирование на причинность по Грейнджеру до 5 лагов. Выделенные цветом ячейки означают, что признак в столбце j является причиной по Грейнджеру индекса в строке i на уровне значимости 0.01. В ячейках указано p -значение F -статистики (Источник: составлено автором).





Гипотеза 4

Отвергается

Реакция инвесторов на информацию о вакцинах и программах вакцинации действительно **не всегда** является **немедленной**.

Эффект от новости можно наблюдать как в тот же день, так и спустя время.

Таблица 7. Результаты оценивания HARX-GARCH модели с добавлением скользящих средних за предшествующие 5 дней на композитном индексе **России**
(Источник: составлено автором).

VARIABLES	HARX-GARCH-R5	HARX-GARCH-O-R5	HARX-GARCH-OG-R5	HARX-GARCH-OGN-R5
R2	0.0145	0.0493	0.0501	0.1275
R2_adj	-0.0129	0.0059	0.0024	0.0548
Const	0.600** (0.534)	-0.502** (0.682)	-0.447 (0.687)	-0.609 (0.965)
i...ice[0:1]	0.011 (0.049)	0.003 (0.050)	0.000 (0.050)	-0.014 (0.049)
i...ice[0:5]	0.029 (0.131)	0.044 (0.132)	0.043 (0.134)	0.021 (0.137)
i...ice[0:22]	-0.023 (0.288)	-0.153 (0.309)	-0.184 (0.314)	-0.174 (0.346)
new_cases		-0.024** (0.022)	-0.021** (0.024)	-0.029** (0.025)
new_death		-0.020** (0.020)	-0.018** (0.020)	-0.018** (0.019)
total_vaccinations_per_hundred		-0.001 (0.005)	-0.001 (0.005)	-0.003** (0.005)
stringency_index		-0.015** (0.014)	-0.018** (0.015)	-0.020** (0.016)
new_cases_rolling_5		0.028** (0.024)	0.027** (0.024)	0.032** (0.025)
new_death_rolling_5		0.011 (0.020)	0.007 (0.021)	0.009 (0.021)
total_vaccinations_per_hundred_rolling_5		-0.004 (0.013)	0.000 (0.014)	-0.004 (0.013)
stringency_index_rolling_5		0.022** (0.015)	0.025** (0.016)	0.030*** (0.017)
gtrend_corr			0.005 (0.011)	0.007** (0.010)
gtrend_corr_rolling_5			-0.007 (0.012)	-0.008** (0.011)
news_count				-0.003 (0.005)
title_POS				0.002 (0.008)
title_NEG				-0.031*** (0.011)
text_POS				0.016*** (0.008)
text_NEG				-0.010** (0.011)
news_count_rolling_5				-0.003 (0.009)
title_POS_rolling_5				0.008** (0.008)
title_NEG_rolling_5				0.004 (0.017)
text_POS_rolling_5				-0.003 (0.024)
text_NEG_rolling_5				-0.001 (0.017)



Результаты проверки гипотез

Гипотеза 1	Информация о разработке вакцин и программах вакцинации против Covid-19 влияет на доходность индексов фондового рынка.	<input checked="" type="checkbox"/>
<i>Гипотеза 1a</i>	Любые новости о вакцинах приводят к росту доходности индексов транспортного и потребительского секторов.	<input checked="" type="checkbox"/>
<i>Гипотеза 1б</i>	Новости о вакцинах с отрицательной тональностью имеют более существенное влияние на доходность индексов фондового рынка, чем новости с положительной тональностью.	<input checked="" type="checkbox"/>
<i>Гипотеза 1в</i>	Информация из разных источников о разработке вакцинах и программах вакцинации по-разному влияет на поведение индексов фондовых рынков.	<input checked="" type="checkbox"/>
Гипотеза 2	Фондовые рынки реагируют на информацию о вакцинах не только в своей стране, но и в других странах.	<input checked="" type="checkbox"/>
Гипотеза 3	Переменные, которые описывают внимание, настроения и темпы вакцинации, по-разному влияют на фондовые индексы.	<input checked="" type="checkbox"/>
Гипотеза 4	Реакция инвесторов на информацию о вакцинах и программах вакцинации не является немедленной.	<input checked="" type="checkbox"/>
Гипотеза 5	Информация о вакцинации больше повлияла на доходность индексов транспортного сектора, чем на индексы потребительского сектора.	<input checked="" type="checkbox"/>



Выводы

- Информация о вакцинации и программах вакцинации **оказывает влияние** на доходность композитных и отраслевых индексов финансовых рынков России, США и Китая.
- Влияние оказалось **неодинаковым** для каждой страны, разные источники информации по-разному влияют на доходность индексов для каждого рынка. Не представляет возможности сделать **универсальный вывод** для всех стран в целом – случай каждой страны является уникальным.
- Было выявлено, что внимание население к вакцинации влияет на доходность индекса **потребительского сектора** больше, чем **транспортного**.
- Влияние новостного сентимента в **России** на фондовые индексы оказалось самым значительным.

Новизна исследования

- Использован разнообразный набор данных, собранный из популярных СМИ и новостных источников в России, США и Китае и официальной коронавирусной статистики, оценивая влияние информации именно о вакцинации на финансовые рынки.
- В работе использованы самые современные методы текстового анализа и подходы к моделированию (LDA).
- В работу включена оценка влияния информации о вакцинации не только на основные фондовые индексы, но и на отраслевые индексы: индексы транспортного и потребительского секторов.

Дальнейшие перспективы

- Ожидается подготовка статьи к публикации в Quarterly Journal of Economics и American Economic Review.
- Планируются дальнейшие исследования, с включением в модель переменных внимания в социальных сетях и др.

Спасибо за внимание!



- Apergis N., Mustafa G., Malik S. Covid-19 pandemic, stock returns, and volatility: The role of the vaccination program in Canada // *Applied Economics*. 2022. С. 1–14.
- Audrino F., Sigrist F., Ballinari D. The impact of sentiment and attention measures on stock market volatility // *International Journal of Forecasting*. 2020. Т. 36. № 2. С. 334–357.
- Bakry W. и др. Response of stock market volatility to COVID-19 announcements and stringency measures: A comparison of developed and emerging markets // *Finance Research Letters*. 2022. Т. 46.
- Birjali M., Kasri M., Beni-Hssane A. A comprehensive survey on sentiment analysis: Approaches, challenges and Trends // *Knowledge-Based Systems*. 2021. Т. 226.
- Brown G., Cliff M.T. Investor sentiment and asset valuation // *SSRN Electronic Journal*. 2001.
- Chan K.F. и др. Covid-19 vaccines: Saving lives and the global stock markets // *SSRN Electronic Journal*. 2021.
- Chen M.-H., Jang S.C., Kim W.G. The impact of the SARS outbreak on Taiwanese Hotel Stock Performance: An event-study approach // *International Journal of Hospitality Management*. 2007. Т. 26. № 1. С. 200–212.
- Colladon A.F. и др. Forecasting financial markets with semantic network analysis in the COVID-19 crisis // 2020.
- Consoli S., Pezzoli L.T., Tosetti E. Emotions in macroeconomic news and their impact on the European Bond Market // *Journal of International Money and Finance*. 2021. Т. 118.
- Devlin J. и др. // *Proceedings of the 2019 Conference of the North*. 2019.
- Dorogush A. V., Ershov V., Gulin A. CatBoost: gradient boosting with categorical features support. 2018.
- Duan J.-C., Ritchken P., Sun Z. Approximating GARCH-jump models, jump-diffusion processes, and option pricing // *Mathematical Finance*. 2006. Т. 16. № 1. С. 21–52.
- Edelman. 2022 EDELMAN TRUST BAROMETER // *Global Report*. 2022.
- Ferguson N. Big Pandemics and Their Economic, Social and Political Consequences // *Working Paper*, Hoover Institution. Finance Studies. 2020.
- Hoffman M., Bach F., Blei D. Online learning for latent dirichlet allocation // *Advances in neural information processing systems*. 2010. Т. 23.
- Huang A., Wang H., Yang Y. Finbert—a deep learning approach to extracting textual information // *SSRN Electronic Journal*. 2020.
- ICAO. Effects of Novel Coronavirus (COVID-19) on Civil Aviation: Economic Impact Analysis // 2022.
- Jiao P., Walther A. Social Media, news media and the stock market // *SSRN Electronic Journal*. 2016.
- Long W. и др. Does the emotional tendency of covid-19 news affect financial markets? // *Procedia Computer Science*. 2022. Т. 199. С. 532–539.
- Malik K., Sharma S., Kaur M. Measuring contagion during COVID-19 through volatility spillovers of BRIC countries using diagonal bekk approach // *Journal of Economic Studies*. 2021. Т. 49. № 2. С. 227–242.
- Ng N. и др. Facebook fair’s WMT19 news translation task submission // *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*. 2019.
- Ozaydin O., Ulengin F. Impacts of covid-19 on the transport sector and measures as well as recommendations of policies and future research: A report on Turkey // *SSRN Electronic Journal*. 2020.
- Petryk V.L. An analysis of the Global Civil Aviation Market and a forecast of its development in the face of unstable demand for Air Transportation // *Business Inform*. 2020. Т. 3. № 506. С. 112–119.



- Piantadosi S.T. Zipf's word frequency law in natural language: A critical review and future directions // Psychonomic Bulletin & Review. 2014. Т. 21. № 5. С. 1112–1130.
- Rakshit B., Neog Y. Effects of the COVID-19 pandemic on stock market returns and volatilities: Evidence from selected emerging economies // Studies in Economics and Finance. 2021.
- Ranco G. и др. The effects of twitter sentiment on stock price returns // PLOS ONE. 2015. Т. 10. № 9.
- Rouatbi W. и др. Immunizing markets against the pandemic: Covid-19 vaccinations and stock volatility around the world // International Review of Financial Analysis. 2021. Т. 77.
- Sahoo P., Ashwani. Covid-19 and Indian Economy: Impact on growth, manufacturing, trade and MSME sector // Global Business Review. 2020. Т. 21. № 5. С. 1159–1183.
- Serrano F., Kazda A. The future of Airports Post covid-19 // Journal of Air Transport Management. 2020. Т. 89.
- Shiller R.J. Narrative economics: How stories go viral and drive major economic events // Princeton University Press. 2019.
- Stolbov M. Statistics of search queries in Google as an indicator of financial conditions // Voprosy Ekonomiki. 2011. № 11. С. 79–93.
- Tardivo A., Sánchez Martín C., Carrillo Zanuy A. Covid-19 impact in transport, an essay from the Railways' System Research Perspective // 2020.
- Ulyankin F. Forecasting Russian macroeconomic indicators based on information from news and search queries // Russian Journal of Money and Finance. 2020. Т. 79. № 4. С. 75–97.
- Zhang D., Hu M., Ji Q. Financial Markets under the global pandemic of covid-19 // Finance Research Letters. 2020. Т. 36. С. 101528.
- Zorgati I., Garfatta R. Spatial financial contagion during the COVID-19 outbreak: Local Correlation Approach // The Journal of Economic Asymmetries. 2021. Т. 24.
- Similarweb. Аналитика трафика и доля рынка nytimes.com. URL: <https://www.similarweb.com/ru/website/nytimes.com/#overview> (дата обращения: 30.04.2022).
- Similarweb. Аналитика трафика и доля рынка tass.ru. URL: <https://www.similarweb.com/ru/website/tass.ru/#overview> (дата обращения: 30.04.2022).
- Similarweb. Аналитика трафика и доля рынка chinadaily.com.cn. URL: <https://www.similarweb.com/ru/website/chinadaily.com.cn/#overview> (дата обращения: 30.04.2022).
- Johns Hopkins Coronavirus Resource Center. Understanding vaccination progress by country. URL: <https://coronavirus.jhu.edu/vaccines/international> (дата обращения: 30.04.2022).



Моделирование с HARX-GARCH (1, 1), текущая оценка

- Динамика среднего значения μ_t описывается процессом гетерогенной авторегрессионной моделью HARX (p, L), в которую включаются средние значения зависимой переменной за предыдущий период, 5 и 22 дня до текущего периода. Модель позволяет учитывать среднесрочные и долгосрочные частоты торговли индексами.

$$\mu_t = \mu + \sum_{i=1}^p \phi_i \bar{\mu}_{t-L_i:t-1} + \gamma_1^T \text{controls}_t + \gamma_2^T \text{official_info}_t + \gamma_3^T \text{gtrends}_t + \gamma_4^T \text{news}_t$$

- Где $\bar{\mu}_{t-L_i:t-1} = \frac{1}{L_i} \sum_{j=1}^{L_i} \mu_{t-j}$ – среднее значение переменной между моментами времени $t - L_i$ и $t - 1$, controls_t – контрольные переменные, official_info_t – вектор признаков официальной информации о Covid-19, gtrends_t – значение индекса Google Trends, news_t – вектор новостных признаков.
- В работе используются параметры $p = 3, L = [1, 5, 22]$. Модель оценивалась с использованием библиотеки arch для языка программирования Python.



Построение индексов Google Trends

- Многочисленные исследования показывают, что данные о популярности запросов в поисковой системе Google могут быть значимы для прогнозирования финансовых рынков и отражают общественное внимание к определенным темам и проблемам.
- В индексы Google Trends были включены ключевые слова, отражающие внимание к вакцинам и вакцинации, а также тестированию на Covid-19.
- Протестированы 2 метода агрегирования индексов популярности ключевых слов – РСА и взвешивание по коэффициентам парных корреляций (Столбов, 2011):

$$I = w_1 I_1 + w_2 I_2 + \dots + w_n I_n$$
$$w_i = \frac{\sum_{j=1}^n \rho_{ij}}{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij}}$$

Таблица #. Веса ключевых слов в итоговых индексах
(Источник: составлено автором)

Россия		США		Китай	
Ключевое слово	Вес	Ключевое слово	Вес	Ключевое слово	Вес
вакцинация	0.085	vaccination	0.169	疫苗接种 (вакцинация)	0.181
вакцина	0.077	vaccine	0.157	疫苗 (вакцина)	0.244
тест на антитела	0.104	test covid	0.124	covid test	0.111
тест на ковид	0.099	test antibody	0.089	coronavac	0.119
qr код	0.117	pfizer	0.147	sinovac	0.188
спутник v	0.165	moderna	0.179	sinopharm	0.157
эпиваккорона	0.110	novavax	0.136		
спутник лайт	0.125				
ковивак	0.118				

Модель для предсказания тональностей новостей

- Оценка тональности текста производилась моделью **FinBERT** (Huang A. и др., 2020) – трансформером с BERT архитектурой.
- Модель была предварительно обучена на выборке корпоративных годовых и квартальных отчетов, отчетов финансовых аналитиков и стенограмм конференц-звонков о прибылях и убытках, а затем доучена на классификацию предложений из аналитических финансовых отчетов объемом выборки 10 тыс. предложений.

Сравнение FinBERT с Bag-Of-Words (мешок слов), Word2Vec

Преимущества	Недостатки
<ul style="list-style-type: none"> ✓ Отсутствие необходимости в предобработке текста (лемматизации, токенизации), end-to-end подход; ✓ Учитывается порядок, вес и связи слов (токенов) в тексте. ✓ При прочих равных превосходит по качеству оценку тональности с помощью словарей. 	<ul style="list-style-type: none"> ✗ Ограничение входного слоя модели в 512 токенов (слов, пунктуации или частей слов). Решение: разбиение больших текстов новостных статей на части (их не более 1-2% в каждом издании) и усреднение для получения тональности всего текста ✗ Смещение данных (как распределения слов (токенов), так и длины текстов (статей и заголовков)) относительно обучающей выборки, что может привести к зашумленным или ложным выводам.

Построение индексов Google Trends



Рисунок #. Сравнение индексов Google Trends. (Источник: составлено автором)

- Индексы, взвешенные с помощью парных коэффициентов корреляции, практически совпали с 1-й компонентой метода PCA.
- 2-я компонента метода PCA по Google Trends Китая напоминает шум.
- Для сопоставимости индексы отмасштабированы в шкалу от 0 до 100:
$$I_t^* = 100 \cdot \frac{I_t - \min_t I_t}{\max_t I_t - \min_t I_t}$$
- В итоговой модели использовали индексы по Столбову как более простые и интерпретируемые.



Обработка текстов новостей

Таблица #. Источники новостных данных
(Источник: составлено автором, данные о посещаемости – портал
SimilarWeb, 2022)

Страна	Новостное издание	Язык	Кол-во новостей	Позиция по посещаемости (аудитория) на март 2022
Россия	TACC (tass.ru)	Русский	86407	13 (76,8 млн)
США	The New York Times (nytimes.com)	Английский	107313	4 (662,1 млн)
Китай	Chinadaily (chinadaily.com.cn)	Английский	117411	79 (6,9 млн)

- Новостной фон формирует настроения и ожидания инвесторов на рынке.
- Тексты новостей издания TASS были переведены на английский язык с помощью модели fairseq компании Facebook, обученной на задаче перевода новостей WMT19 News Translation (Ng N. и др., 2019.).
- Выборка была очищена путем удаления всех токенов, кроме пунктуации и латиницы, и отбрасывания статей на языках отличных от английского.
- Были исправлены критические ошибки перевода. Например, наименование вакцины «спутник» переводилось как «satellite», а не «sputnik»

Прогнозирование тональностей новостей

- Соотношения классов заголовков и текстов статей отличаются от распределения в обучающей выборке, где 36% текстов позитивны, 46% нейтральны и 18% негативны (Huang A. и др., 2020)
- Вероятно, модель адекватно оценивает тональность новостей, поскольку большая часть публикаций не имеет отношения к финансовому рынку.

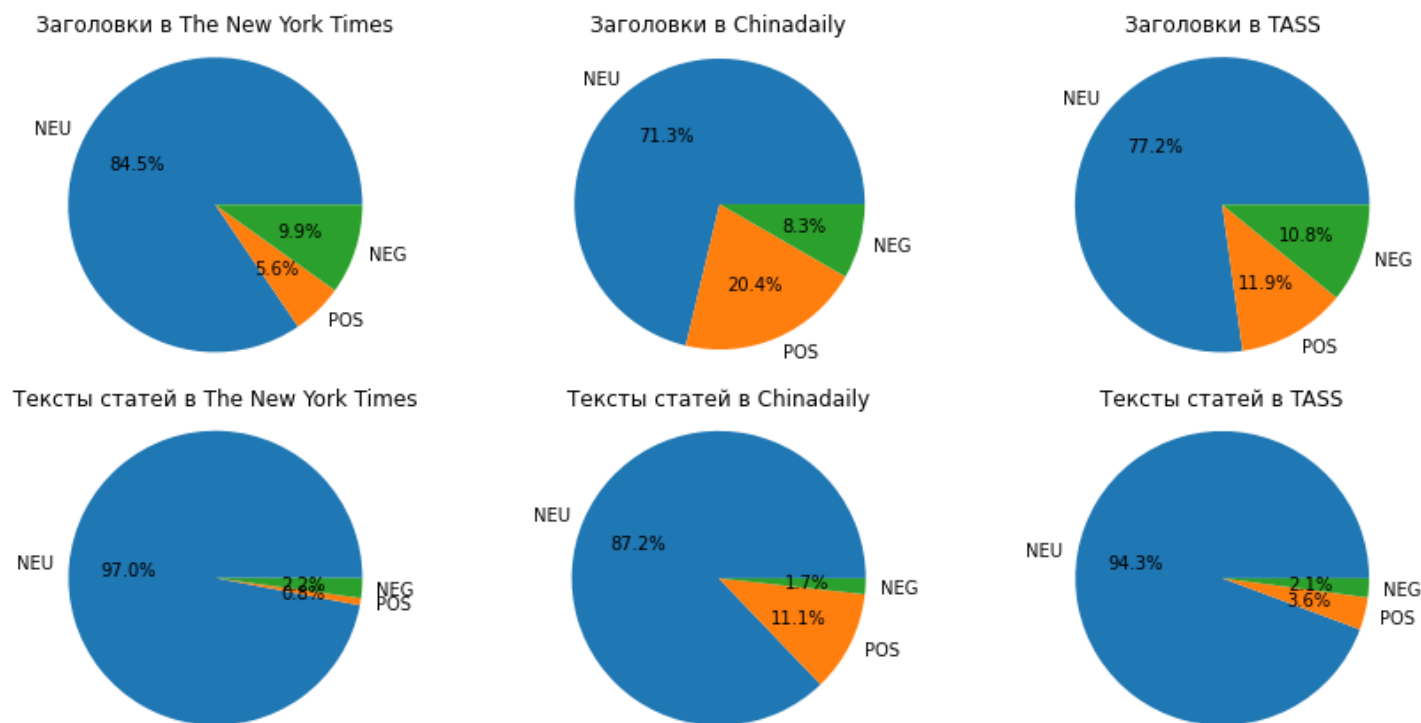


Рисунок #. Тональность публикаций новостных изданий
(Источник: составлено автором)

Тематическое моделирование новостей алгоритмом LDA



В результате тематического моделирования LDA были отобраны темы, связанные с пандемией Covid-19 и разделены на 3 смысловых блока – **наука и здравоохранение, общественные институты и экономика**.

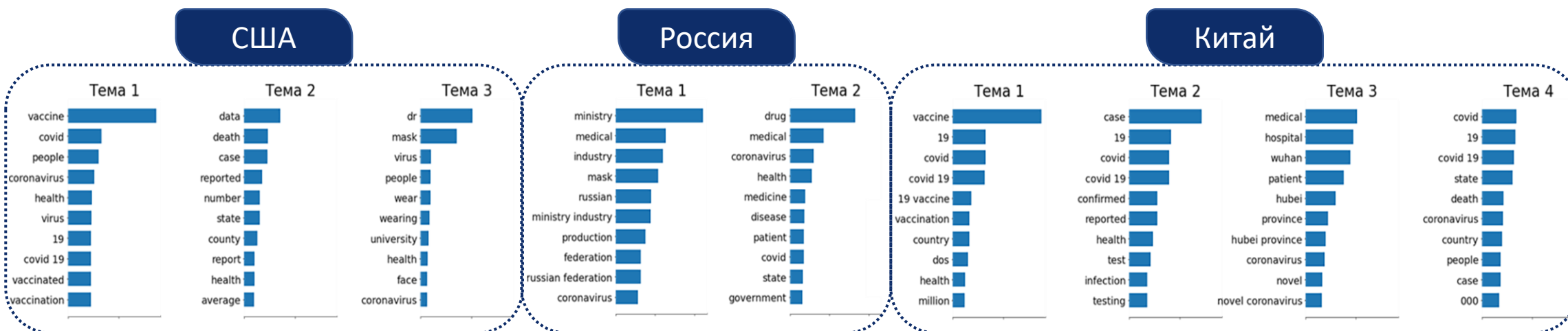


Рисунок #. Тематическое моделирование LDA. Тема **наука и здравоохранения**.
(Источник: составлено автором)

Тематические блоки

Наука и здравоохранение. В блоке часто встречаются упоминания вакцинации, активно публикуются новости о производстве вакцин и ходе программ вакцинации.

Общественные институты. Новости о ковидных ограничениях, включающие закрытия школ, перевод сотрудников на удалённый режим работы, а также, в целом, об изменении жизненного уклада.

Экономика. Упоминание пандемии в контексте экономических новостей.

* Из-за особенностей выборки и новостных изданий не все тематические блоки одинаково представлены в странах

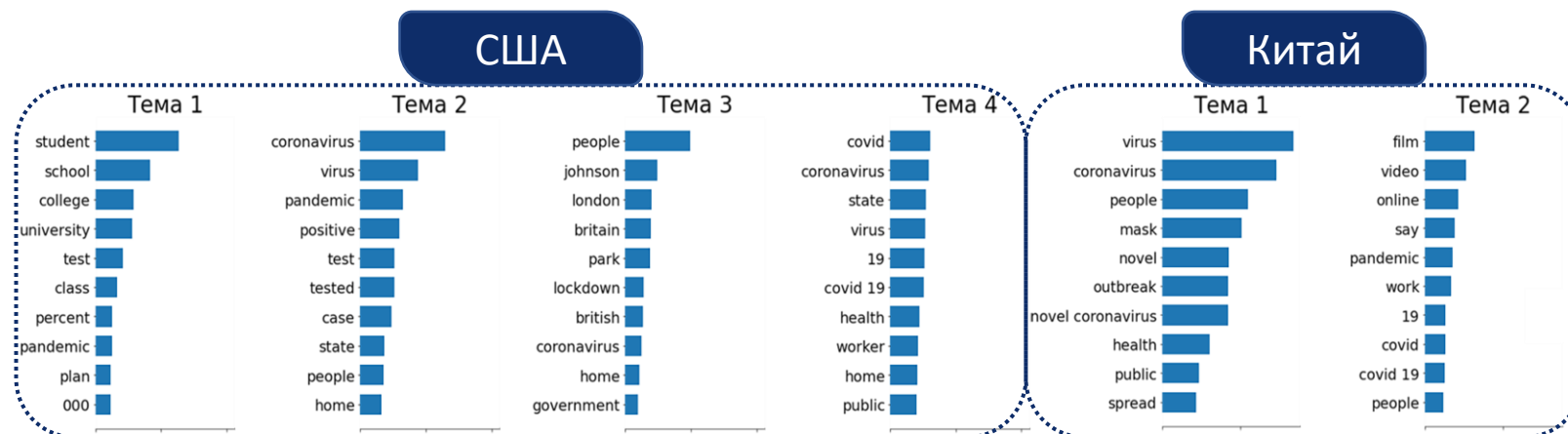


Рисунок #. Тематическое моделирование LDA. Тема **общественных институтов**.
(Источник: составлено автором)



Рисунок #. Тематическое моделирование LDA. **Экономические темы**.
(Источник: составлено автором)

Динамика индексов внимания к темам

- Формирование агрегированных индексов тем по блокам осуществлялось путем взвешенной суммы, где веса определяются на основе коэффициентов парной корреляции (по аналогии с индексами Google Trends). Затем индексы масштабируются в шкалу от 0 до 100.
- Наблюдается схожая динамика освещения тематических блоков между странами.

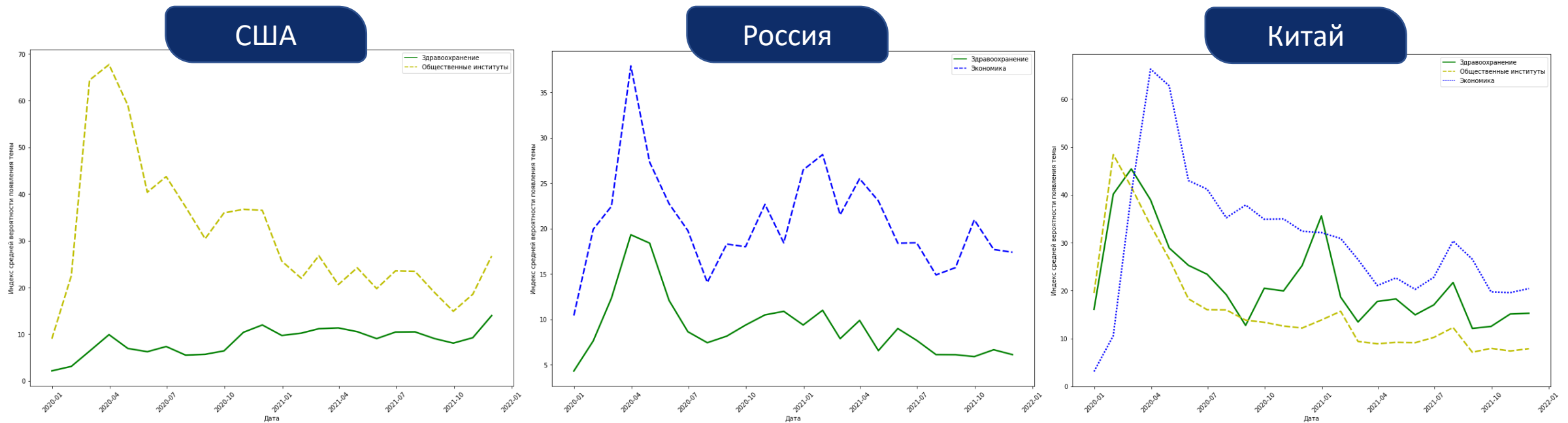


Рисунок #. Частота попадания новостей в выделенные тематические группы в странах. (Источник: составлено автором)

Тональности внутри тематических блоков

- Анализ тональностей внутри тематических блоков показывает, что экономические темы наиболее ярко окрашены с точки зрения модели FinBERT, что валидирует использование указанных подходов.
- Для России и США характерны ярко выраженные негативные заголовки, в то время как в Китае заголовки чаще всего позитивные.

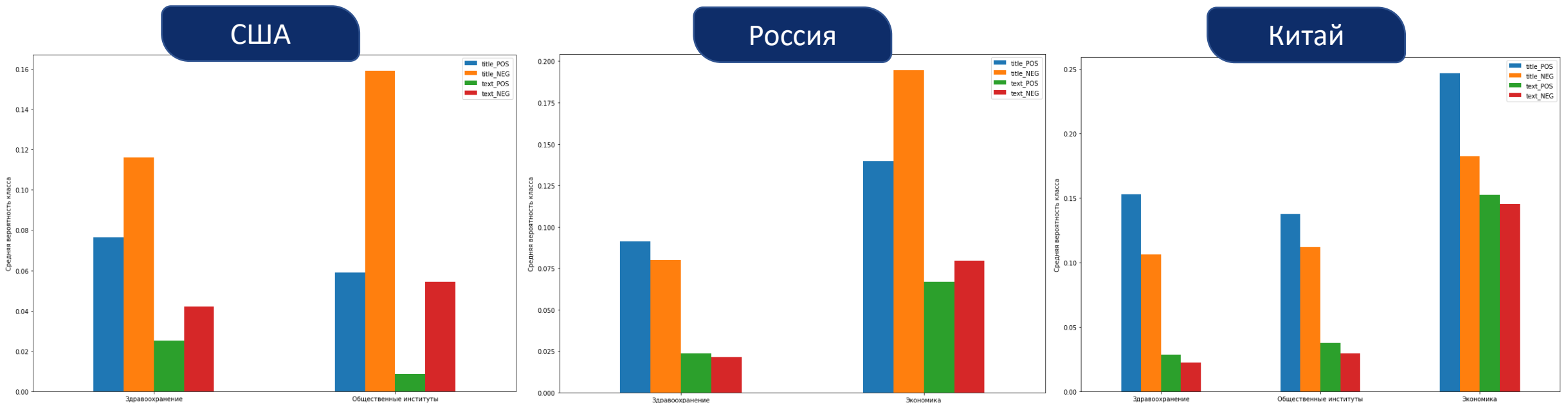


Рисунок #. Тональности внутри тематических блоков. (Источник: составлено автором)

Индексы тональностей новостей на базе LDA

- По спрогнозированным с помощью LDA темам, связанным с Covid-19, были сформированы общие индексы тональностей новостей (в некоторых тематических блоках наблюдалось малое количество новостей).
- Сентимент-индекс в США включает оценку 12451 новостей, в Китае – 19826, в России – 5195.



Рисунок #. Динамика изменения сентимента новостей в странах. (Источник: составлено автором)

Моделирование с HARX-GARCH (1, 1), текущая оценка

- Модель состоит из двух компонент, описывающих динамики среднего значения μ_t и шоков ϵ_t :

$$\begin{aligned}r_t &= \mu_t + \epsilon_t \\ \epsilon_t &= \sigma_t e_t\end{aligned}$$

- Где $r_t = \frac{y_t}{y_{t-1}} - 1$ – доходность индекса y в момент времени t . Предполагается, что $e_t \sim N(0,1)$.
- Динамика волатильности σ_t описывается процессом обобщенной авторегрессионной условной гетероскедастичности GARCH (p, q), что позволяет модели учитывать гетероскедастичность целевых переменных.

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

- Где $p = 1$ – порядок симметричных шоков, $q = 1$ – количество лагов условной дисперсии.

Результаты оценивания HARX-GARCH модели в США

Гипотеза 1

- Данные о новых вакцинациях, строгости ограничений и новостные факторы значимы для объяснения доходности индексов США

Гипотеза 1a

- Наблюдается позитивное влияние количества новостей как в общем, так и в темах общественных институтов и здравоохранения

Гипотеза 1б

- Влияние негативных и позитивных новостей на индексы США сонаправлено с доходностью, при этом значимы обе группы новостей

Гипотеза 3

- Внимание к вакцинации по индексу Google Trends негативно отражается на доходности компаний потребительского сектора, в то время как влияние новостей (особенно позитивных) скорее положительно

Таблица #. Значимые на уровне 5% факторы в модели для индексов США.
(Источник: составлено автором)

Переменные	S&P500	Dow Jones U.S. Consumer Goods Index	Dow Jones Transportation Average
HAR	-1, -5, -22	-1, -5, -22	-5
GARCH	$+\omega, +\alpha_1, +\beta_1$	$+\omega, +\alpha_1, +\beta_1$	$+\omega, +\alpha_1, +\beta_1$
Контрольные переменные	-Объем торгов, -LIBOR 6m	-Объем торгов, -LIBOR 3m	+LIBOR 1m
Официальные данные о Covid-19	+Случаи заболевания, +Новые вакцинации, +Индекс строгости ограничений	+Индекс строгости ограничений	+Новые вакцинации, +Индекс строгости ограничений
Индекс Google Trends		-Внимание к вакцинации	
Новостные переменные	+Общее кол-во новостей, +Позитивные заголовки, -Негативные заголовки, +Позитивные тексты, -Негативные тексты, +Общественные институты	+Позитивные заголовки, -Негативные тексты, +Здравоохранение, +Общественные институты	+Общее кол-во новостей, +Позитивные заголовки, -Негативные тексты, +Общественные институты

Результаты оценивания HARX-GARCH модели в России

Гипотеза 1

- Данные о вакцинации блоков Google Trends и новостей значимо влияют на доходности индексов России

Гипотеза 1а

- Общее кол-во новостей вносит положительный вклад в доходность только для индекса потребительского сектора в России. А тема здравоохранения положительна для доходности композитного и транспортного индексов

Гипотеза 1б

- Сентимент новостей в темах, связанных с Covid-19 значим только для композитного и транспортного индексов

Гипотеза 3

- Выявлено разнонаправленное влияние Google Trends и новостных переменных

Таблица #. Значимые на уровне 5% факторы в модели для индексов России.
(Источник: составлено автором)

Переменные	IMOEX	MOEXCN	MOEXTN
HAR		+1, -5	+5
GARCH	$+\omega, +\alpha_1, +\beta_1$	$+\alpha_1, +\beta_1$	$+\beta_1$
Контрольные переменные	-Объем торгов, -RUONIA, -1m, +3m	+RUONIA, -1m, +6m	+RUONIA 3m, +6m
Официальные данные о Covid-19	-Случаи смерти, +Индекс строгости ограничений	-Случаи заболевания	+Случаи заболевания, -Случаи смерти
Индекс Google Trends		-Внимание к вакцинации	-Внимание к вакцинации
Новостные переменные	-Негативные заголовки, +Позитивные тексты, -Негативные тексты, +Здравоохранение	+Общее кол-во новостей	+Позитивные заголовки, -Негативные заголовки, +Позитивные тексты, +Здравоохранение, -Экономика

Результаты оценивания HARX-GARCH модели в Китае

Гипотеза 1

- Для всех индексов Китая выявлена значимость внимания к вакцинации Google Trends и некоторых новостных факторов

Гипотеза 1a

- Фактор общего количества новостей положительно сказывается на доходностях потребительского и транспортного секторов. Влияние тем, связанных с пандемией незначимо.

Гипотеза 1б

- Тональность новостей незначима для объяснения доходности потребительского сектора Китая. Выявлено отрицательное влияние позитивных текстов на доходность

Гипотеза 3

- Индекс внимания Google Trends негативно влияет на доходности индексов Китая. С другой стороны, наблюдается разнонаправленное воздействие новостных переменных.

Таблица #. Значимые на уровне 5% факторы в модели для индексов Китая.
(Источник: составлено автором)

Переменные	CSI300	Shenzhen Stock Exchange: Wholesale & Retail	Shenzhen Stock Exchange: Transportation
HAR	-22	+1, -5	+1, -5, -22
GARCH	$+\omega, +\alpha_1, +\beta_1$	$+\alpha_1, +\beta_1$	$+\omega, +\alpha_1, +\beta_1$
Контрольные переменные	+Объем торгов, -SHIBOR 1m	+SHIBOR, -1m, +6m	-SHIBOR 1m
Официальные данные о Covid-19	+Индекс строгости ограничений	-Случаи заболевания	+Новые вакцинации, +Индекс строгости ограничений
Индекс Google Trends	-Внимание к вакцинации	-Внимание к вакцинации	-Внимание к вакцинации
Новостные переменные	-Негативные тексты	+Общее кол-во новостей	+Общее кол-во новостей, +Позитивные заголовки, -Позитивные тексты

Оценка важности признаков в модели CatBoost

- Модель CatBoost представляет собой разновидность градиентного бустинга, построенного на деревьях решений. Каждое следующее дерево строится на подвыборках наблюдений и признаков так, чтобы минимизировать ошибку предыдущего дерева. Данные свойства позволяют снизить степень переобучения под конкретную выборку.
- Важность признаков оценивалась путем обучения модели CatBoost с максимальной глубиной дерева 2 и максимальным числом деревьев равным 500. Для обучения модели использовалась функция потерь MAE:

$$MAE = \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Кросс-валидация осуществлялась скользящим расширяющимся окном с максимальным размером выборки 264 дня. В результате модель оценивалась на 400 выборках для каждого индекса.



Рисунок #. Механизм скользящего окна.
(Источник: составлено автором)

Результаты оценивания CatBoost на скользящем окне

- Новостные переменные являются наиболее важными признаками для объяснения доходности индексов в России и 2-ми по важности для США и Китая
- Контрольные переменные практически всегда относительно важны для объяснения доходности индексов
- Наименьшую ценность для оценки текущей доходности представляют официальные данные о Covid-19 во всех 3-х странах.
- Внимание к вакцинации по индексу Google Trends имеет большую значимость только в Китае

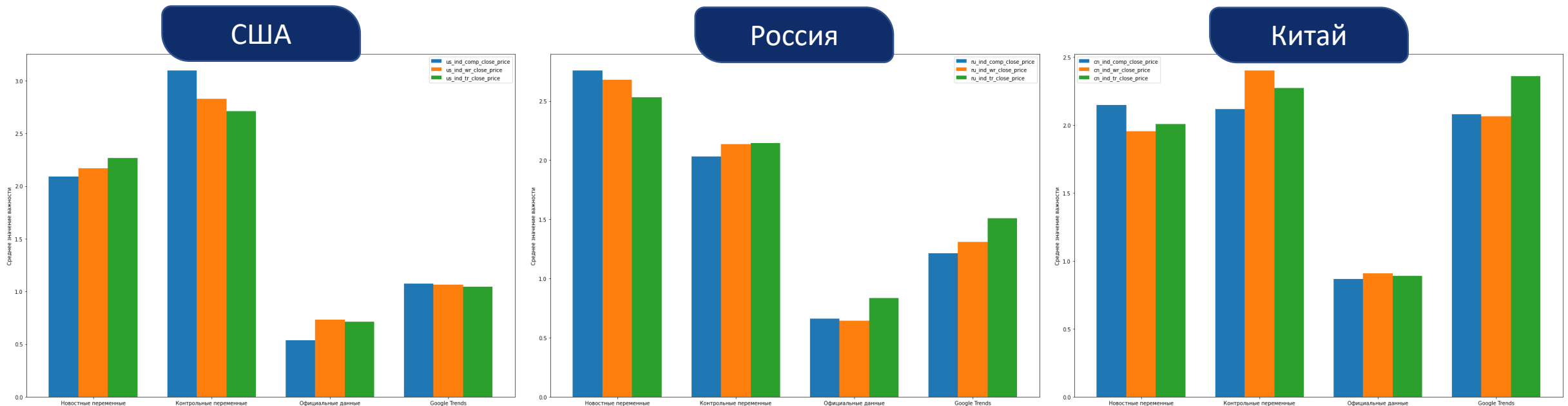


Рисунок #. Важность признаков в модели CatBoost, микроусреднение. (Источник: составлено автором)

Динамика важности признаков

- Новостные факторы, данные о Covid-19 и Google Trends на определенных промежутках времени замещают контрольные переменные при объяснении доходности индексов

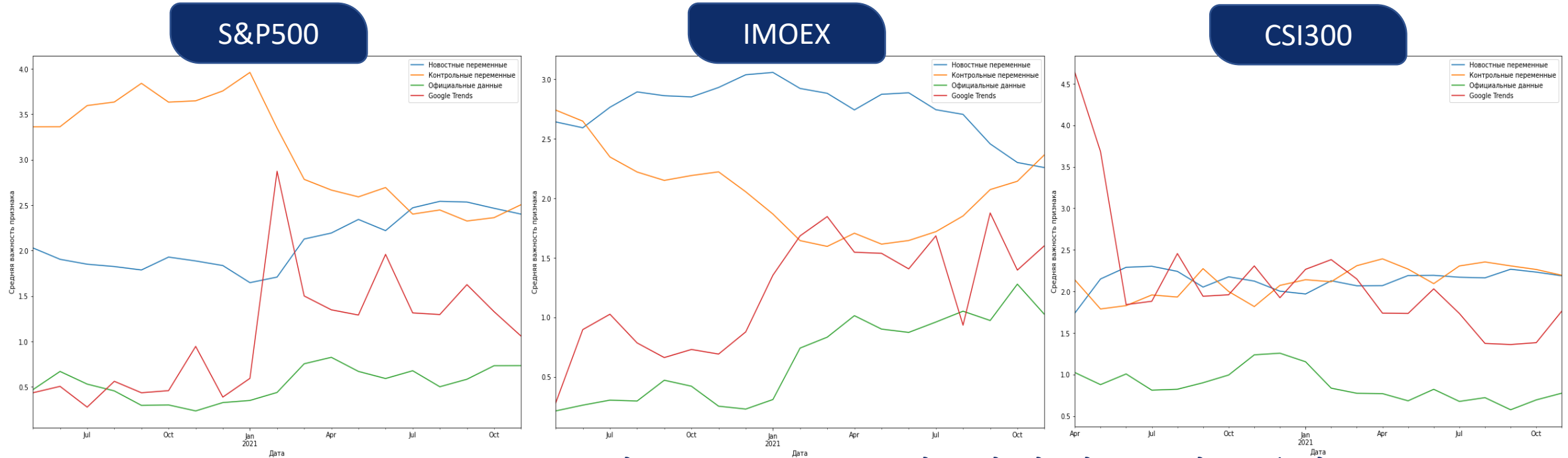


Рисунок #. Динамика важности признаков в модели CatBoost на кросс-валидации для доходности индексов фондового рынка, микроусреднение.

(Источник: составлено автором)