

Т.В. Теплова

Центр финансовых исследований и анализа данных (ЦФИАНД) факультета экономических наук (ФЭН), Национальный исследовательский университет «Высшая школа экономики», Москва

Т.В. Соколова

ЦФИАНД ФЭН, Национальный исследовательский университет «Высшая школа экономики», Москва

А.Ф. Томтосов

ЦФИАНД ФЭН, Национальный исследовательский университет «Высшая школа экономики», Москва

Д.В. Бучко

ЦФИАНД ФЭН, Национальный исследовательский университет «Высшая школа экономики», Москва

Д.Д. Никулин

ЦФИАНД ФЭН, Национальный исследовательский университет «Высшая школа экономики», Москва

Сентимент частных инвесторов в объяснении различий в биржевых характеристиках акций российского рынка¹

Аннотация. В нашей работе впервые исследовано влияние настроений инвесторов в социальных сетях на биржевые характеристики акций российского рынка и показаны нелинейные эффекты. Месячная доходность и объемы торгов анализируются под контролем финансовых показателей и индикаторов качества корпоративного управления эмитентов акций, а также меняющейся внешней среды на отрезке с 2013 по 2020 г. Выборка для разнообразных метрик сентимента построена на уникальных данных – сообщениях на платформах Telegram и mfd.ru. Тональность сообщений диагностируется по авторской методике с использованием искусственного интеллекта (нейросети). Главный вывод – сентимент может рассматриваться как объясняющий фактор в ценообразовании и торговой активности. Предложен авторский NURE-индикатор настроений, который сопоставлен по объясняющей способности на биржевые характеристики с широким спектром прокси-переменных. Объясняющая способность выявления различий реализуется через регрессионные построения на панельных данных. Показано, что биржевые характеристики более чувствительны к нарастанию негативных сообщений, что согласуется с постулатами поведенческих финансов. Рост числа сообщений и позитивной, и негативной тональности способствует росту торговой активности. Важный практический вывод – следование за толпой при высокой активности обсуждений компании не позволяет инвестору получить высокую доходность.

Ключевые слова: настроения инвесторов; сентимент инвесторов; поведенческие финансы; обработка текста; тональность сообщений; нейросети; доходность акций; торговая активность.

Классификация JEL: G11, G12, G41.

Цитирование: Теплова Т.В., Соколова Т.В., Томтосов А.Ф., Бучко Д.В., Никулин Д.Д. (2022). Сентимент частных инвесторов в объяснении различий в биржевых характеристиках акций российского рынка // *Журнал Новой экономической ассоциации*. № 1 (53). С. 53–84. DOI: 10.31737/2221-2264-2022-53-1-3

1. Введение

Поведенческие финансы являются признанным направлением в экономических исследованиях, постулаты строятся на опытах и наблюдениях. Однако

¹ Работа подготовлена за счет средств гранта на поддержку исследовательских центров в сфере искусственного интеллекта, в том числе в области «сильного» искусственного интеллекта, систем доверенного искусственного интеллекта и этических аспектов применения искусственного интеллекта, предоставленного АНО «Аналитический центр при Правительстве Российской Федерации» в соответствии с соглашением о предоставлении субсидии (идентификатор соглашения о предоставлении субсидии 000000D730321P5Q0002) и договором с ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» от 2 ноября 2021 г. № 70-2021-00139.

существует относительно немного эмпирических исследований, которые бы охватывали разные группы инвесторов и учитывали бы тренды в источниках получения информации. В последние годы четко прослеживается рост числа частных инвесторов в биржевых торгах и в РФ, и в США, и в Китае. Яркая характеристика процесса инвестирования со стороны розничных (частных) инвесторов – использование рекомендаций других инвесторов, которые делятся своими соображениями и инвест-решениями на различных биржевых платформах и в социальных сетях. В 2021 г. регуляторы и в РФ, и в США признали роль частных инвесторов во влиянии на биржевые характеристики активов (для США – дело GameStop, для РФ – дело Россетей-Юг с блокировкой брокерских счетов активных участников Телеграм-чатов).

В ранее проведенных исследованиях анализа настроений инвесторов (далее – сентимента) на фондовом рынке преобладают исследования заголовков и текстов ведущих деловых СМИ или новостных агрегаторов. Например, (Tetlock, 2007) показал, что высокий пессимизм аналитических колонок в «Wall Street Journal» прогнозирует будущий разворот на падающем рынке акций США, а экстремальные значения пессимизма и оптимизма приводят к росту объема торгов. У такого подхода есть ряд недостатков. Во-первых, авторы не раскрывают вопроса о том, являются ли все подписчики указанных изданий инвесторами на фондовом рынке США и какова доля среди них частных инвесторов. Во-вторых, крайне затруднительно проверить реакцию подписчиков на опубликованные тексты. Остается открытым вопрос о том, были ли в действительности прочитаны тексты и заголовки, совершали ли подписчики сделки с акциями за период исследования в соответствии с тональностью статьи.

С 2010-х годов появились работы, посвященные анализу сентимента инвесторов на базе социальных сетей (Stock Twits, Twitter). В этом направлении уже четко видна роль частных инвесторов, так как участники прямо выражают свое отношение к перспективам роста и падения акций, делятся своими решениями. Базы позволяют отделить сообщения частных инвесторов от публикаций профессиональных участников. Но подавляющее большинство работ фокусируется на рынке США (Oliveira, Cortez, Areal, 2013, 2017; Al-Nasseri, Ali, 2018).

Актуальность нашего исследования обусловлена, во-первых, недостатком работ, анализирующих влияние сентимента инвесторов на биржевые характеристики активов на развивающихся рынках. В нашей работе впервые исследовано влияние нескольких метрик сентимента частных инвесторов на российском рынке акций. Во-вторых, на российском рынке активно развиваются платформы типа social-trading (платформы инвестиционных сообществ), в том числе независимые; стремительно растет число частных (розничных) инвесторов, которые не обладают профессиональными знаниями и совершают торговые сделки «на слухах», следуя за толпой. В Приложении на рис. А1 показано, что число активных счетов¹ частных инвесторов с января 2014 г. по ноябрь 2019 г. выросло в 7,5 раз. На январь 2021 г. число физических лиц на Мосбирже превысило 10 млн человек. Отчасти приток физических лиц на биржу способствовал росту фондовых индексов. Заметим, что рост рынка не привел к открытию новых счетов юридическими лицами (Приложение, рис. А2). Число счетов нерезидентов выросло, но в значительно меньшей степени в сравнении с физическими лицами. Возможная

¹ По методике МосБиржи активными считаются счета с минимум одной сделкой с ценными бумагами за календарный месяц.

причина – в антироссийских санкциях, введенных с 2014 г., и в общем снижении интереса к развивающимся рынкам.

В классическом варианте платформы social trading – это приложения и сервисы брокеров для своих клиентов, но такие приложения появились сравнительно недавно либо имеют закрытый характер и не хранят историю сообщений (например, чат в торговом терминале). Поэтому в нашей работе мы рассматриваем не зависящие от брокеров площадки с открытым (бесплатным) доступом, сохраняющие историю сообщений: относительно новый популярный мессенджер «Telegram», включающий каналы и чаты с инвестиционной тематикой, и старейший российский профессиональный биржевой форум МФД (mfd.ru). Охват данных площадок сопоставим с классическими СМИ. Так, число индивидуальных физических подписчиков бумажной версии газеты «Ведомости» на 2020 г. составляет 13–30 тыс. человек², число активных пользователей на форуме mfd.ru – 59 тыс. человек на май 2020 г., а число подписчиков двух самых популярных каналов на инвестиционную тематику в Telegram – «Биткоган» и «РынкиДеньгиВласть» – 107,7 и 102,9 тыс. человек соответственно. Отметим, что платформа Telegram была официально запрещена на российском рынке до июня 2020 г., хотя активно использовалась российскими инвесторами, несмотря на запрет. В целом число пользователей платформы составляет 15 млн человек.

Выбранные нами для анализа платформы Telegram и mfd.ru охватывают разные возрастные группы инвесторов, что позволяет сопоставить влияние их сентимента на биржевые характеристики акций и выявить отличия. Telegram можно рассматривать как молодежную площадку со средним возрастом пользователей до 25 лет³. Можно предположить, что средний счет у этих инвесторов ниже, но их доля будет расти со временем за счет развития финансовых технологий и популяризации инвестиций и сбережений в России. Открытых данных о среднем возрасте участников форума mfd нет, но по размеченным нами вручную сообщениям можно заключить, что средний возраст участников старше 30 лет.

В отличие от каналов профессиональных участников («Пulsь» Тинькова и др.) независимые каналы намного более популярны. Сравнительный анализ числа подписчиков самых популярных Telegram-каналов и официальных каналов ведущих брокеров представлен в табл. 1.

Важное нововведение нашей работы – уникальные данные и высокий охват большей части сентимента частных инвесторов – в сумме порядка 50% активных счетов на Московской бирже. Заметим, что во многих работах, посвященных рынку США, анализ не охватывает большей части сентимента частных инвесторов (Tetlock, 2007; Oliveira, Cortez, Areal, 2017; Renault, 2017; Huang, Rojas, Convery, 2019).

Помимо решения проблем анализа газетных СМИ, представленных в ранее опубликованных исследованиях, используемые нами новые источники данных добавляют свои проблемы. Если текст в газетах относится к деловому письму и сообщения можно разметить на позитивные, нейтральные и негативные группы с помощью готовых словарей и пакетов, то на площадках свободного общения в основном используется неформальный стиль, много сленговых

² См. <https://www.sostav.ru/publication/dozhd-obognal-vedomosti-po-platnoj-podpiske-9672.html>. Данные 2019 г. (менее 200 тыс. читателей) по данным издания Tadviseer со ссылкой на исследование Mediascope (<http://www.tadviseer.ru/index.php/Компания:Ведомости#.D0.90.D1.83.D0.B4.D0.B8.D1.82.D0.BE.D1.80.D0.B8.D1.8F>).

³ По данным издания «Взгляд» от февраля 2018 г. с отсылкой на исследовательский холдинг Romir (<https://vz.ru/news/2018/2/13/908109.html>).

Таблица 1

Число подписчиков тематических Telegram-каналов и официальных каналов брокеров, декабрь 2019 г.

Telegram-каналы инвестиционных компаний и брокеров			Независимые каналы по инвестициям и финансовой грамотности		
Позиция по клиентским счетам	Компания	Число подписчиков, человек	Позиция в рейтинге	Канал	Число подписчиков, человек
1	Сбербанк	Нет канала	1	РынкиДеньгиВласть (РДВ)	66 243
2	Тинькофф	13 896	2	МММ	60 239
3	ВТБ	9376	3	Бондовик	43 778
4	БКС	8031	4	Финансовый монстр	42 036
5	Открытие	1194	5	Дневник трейдера	35 742

Источник: составлено авторами.

выражений (например, ГК Норникель называют «норкой» или «гамаком»). Дополнительно усложняет задачу и тот факт, что сленг со временем меняется, в одном сообщении могут быть как негативные слова, так и позитивные (например, «рыночек падает, а мы еще порастем»). Нам пришлось отказаться от готовых словарей в пользу самостоятельно обученной нейросети⁴. Нами вручную были размечены более 30 тыс. сообщений и составлен словарь, отражающий биржевой сленг российского рынка.

Для дальнейшей классификации сообщений по тональности на уже распознанной выборке в русле работ (Oliveira, Cortez, Areal, 2017; Renault, 2017) мы применяем модель на основе CNN.

Точность на основе полносвязной нейронной сети и TF-IDF с корректировками под решаемую задачу составила 55%. Модель на основе трансформеров и CNN, предложенная (Yang et al., 2019), показала 60,1% точности.

В отличие от предыдущих работ (Oliveira, Cortez, Areal, 2013; Renault, 2017; Al-Nasseri, Ali, 2018), где предполагалось одинаковое влияние сентимента на динамику цен акций всех компаний, мы тестируем гипотезу о том, что влияние сентимента на биржевые характеристики (доходность по цене акций и полную доходность с учетом дивидендов, торговую активность) будет различным для «голубых фишек», где высока доля нерезидентов и профессиональных участников, средних по капитализации компаний и для акций компаний третьего эшелона, и где высока доля розничных инвесторов (физических лиц).

Оригинальность нашей работы – сопоставление нескольких конкурирующих метрик сентимента в отношении отдельных акций и учет эффектов нелинейного влияния сентимента на биржевые характеристики. Нелинейность объясняется нами возможностью, особенно на сфокусированных по отдельным акциям ветках платформ (MFD), реализовать массивный вброс информации и соответствующего настроения – слишком сильного оптимизма или негатива, т.е. фактически с помощью навязчивой рекламы или антирекламы провести манипуляцию рынком. Мы показываем, что заработать инвестору в такой ситуации невозможно; следуя за такими однонаправленными настроениями,

⁴ В работе (Jegadeesh, Wu, 2013) подчеркивается проблема эффективности стандартных словарей в пояснениях к финансовой отчетности. Авторы ссылаются на использование «HarvardIV-4 Psychosociological Dictionary» в работах (Tetlock, 2007). Более продвинутые методы, но с использованием равных весов для всех слов с последующим подсчетом позитивных и негативных, например (Loughran, McDonald, 2011).

инвестор скорее проиграет. Напротив, на платформе Telegram нет профильных тематических веток по компаниям, и вброс с целью манипуляции на одном из каналов нивелируется общим большим числом сообщений. В предыдущих работах эффекты нелинейности не рассматривались (Oliveira, Cortez, Areal, 2013; Al-Nasseri, Ali, 2018).

В разд. 2 приведен краткий обзор предыдущих работ по теме исследования. В разд. 3 выдвигаются гипотезы нашего исследования. В разд. 4 представлена методика разметки текстовых сообщений с применением методов машинного обучения и нейросети, описано построение индикаторов на основе тональности сообщений и предложены эконометрические модели для проверки гипотез. В разд. 5 представлена описательная статистика по выборке. В разд. 6 приводятся и интерпретируются результаты эконометрических расчетов влияния различных прокси-переменных сентимента и делаются выводы.

2. Обзор предыдущих работ

В Приложении в табл. А1 представлен краткий обзор работ по теме исследования. Подавляющее большинство работ посвящено рынку США (Tetlock, 2007; Oliveira, Cortez, Areal, 2013; Al-Nasseri, Ali, 2018; Huang, Rojas, Convery, 2020), тогда как акции компаний развивающихся рынков гораздо реже являются объектом исследования (Renault, 2017). В нашей работе впервые исследовано влияние сентимента соцсетей на российском рынке акций.

Во многих работах тестируются только модели влияния сентимента на доходность акций (Baker, Wurgler, 2006; Renault, 2017; Huang, Rojas, Convery, 2020). Большинство авторов рассматривают только один источник сообщений для построения прокси-переменных сентимента, не деля источники на классы – форумы и мессенджеры (Tetlock, 2007; Oliveira, Cortez, Areal, 2013; Renault, 2017; Al-Nasseri, Ali, 2018). В отличие от них мы рассматриваем и сопоставляем объясняющую силу индексов сентимента по двум источникам: новому популярному мессенджеру Telegram и одному из старейших биржевых форумов mfd.ru.

В ряде предыдущих работ объясняющая сила предложенных индексов сентимента не сравнивается с такими более простыми метриками, как число сообщений (Tetlock, 2007; Renault, 2017; Al-Nasseri, Ali, 2018), тогда как мы не только предлагаем оригинальный авторский индекс (*HYPE Indicator*), но и рассматриваем широкий спектр различных метрик и сопоставляем их объясняющую способность.

Как правило, предыдущие работы фокусируются на фондовых индексах широкого рынка, ETF, высоколиквидных акциях (Tetlock, 2007; Oliveira, Cortez, Areal, 2013; Renault, 2017; Al-Nasseri, Ali, 2018), и лишь в некоторых работах (Oliveira, Cortez, Areal, 2017) выделяются компании по размеру или степени ликвидности акций. Оригинальность нашей работы – выделение трех групп компаний: высокой, средней и малой капитализации (для которых характерна низкая ликвидность).

Преимущество анализа русскоязычных социальных сетей заключается в возможности выявить концентрированное влияние сентимента на биржевые характеристики российских акций. Так, в нашей работе охват сентимента частных инвесторов составляет более 50% (общего числа частных инвесто-

ров в России). Из-за популярности акций американского рынка их обсуждение ведется на различных национальных языках и платформах, площадки рассеяны, и каждая площадка составляет малую долю в общем объеме обсуждений (Oliveira, Cortez, Areal, 2017; Renault, 2017; Huang, Rojas, Convery, 2020).

В русле современных работ (Renault, 2017; Al-Nasseri, Ali, 2018) мы используем методы машинного обучения и искусственного интеллекта для классификации сообщений по тональности. Для проверки гипотез о влиянии сентимента инвесторов на биржевые характеристики мы тестируем регрессионные модели на панельных данных с учетом эффектов нелинейности. В большинстве предыдущих работ для тестирования также применялись регрессионные модели, но эффекты нелинейности не учитывались (Приложение, табл. А.1).

3. Гипотезы

В нашей работе мы исследуем влияние сентимента на следующие ежемесячные биржевые характеристики акций: а) чистая (без дивидендов) и полная (с учетом дивидендов) доходность акций; б) относительный объем торгов акциями.

Гипотеза 1. *Сентимент частных инвесторов значимо нелинейно влияет на доходность акций.*

В работах (Renault, 2017; Sun, Najand, Shen, 2016; Huang, Rojas, Convery, 2020) по развитым рынкам капитала показано, что более оптимистичные настроения инвесторов относительно перспектив компании-эмитента приводят к росту доходности акций, что объясняется повышением спроса. Мы рассматриваем две платформы для обмена мнениями инвесторов: 1) широкую, не включающую каналы обсуждений по отдельным компаниям, – мессенджер Telegram; 2) узко сфокусированный на отдельных компаниях форум MFD. Мы принимаем во внимание эффекты нелинейности, так как во втором случае мы допускаем возможность наличия манипуляций за счет вбросов позитивных или негативных настроений. Мы предполагаем, что с ростом положительных настроений доходность растет. Но инвесторы распознают навязчивую рекламу, что не приводит к росту доходности: при экстремально высоких значениях оптимизма большая часть инвесторов фиксирует прибыль.

Гипотеза 2. *Сентимент частных инвесторов значимо влияет на относительный объем торгов акциями (торговую активность).*

Мы предполагаем, что число сообщений (любой тональности) на профильных social trading платформах отражает интерес к акциям той или иной компании и мотивирует покупать-продавать. Впервые влияние сентимента на рост объема торгов было выявлено в работе (Tetlock, 2007), аналогичный эффект был показан в работе (Oliveira, Cortez, Areal, 2013) на примере рынков США на базе платформы Stock Twits.

Мы также предполагаем, что по платформе MFD, где ведутся узко сфокусированные обсуждения акций отдельных компаний, значимы нелинейные эффекты. При слишком высоком уровне оптимизма или пессимизма (скорее вброс информации) рост объема торгов прекращается из-за настороженности инвесторов.

Гипотеза 3. *Пессимистические настроения сильнее влияют на биржевые характеристики активов, чем оптимистические.*

Классическая теория финансов постулирует: новостная лента влияет на цену акций. Различный вклад позитивных и негативных новостей лежит в основе поведенческих финансов. Различные влияния на доходность акций на развитых рынках подтверждены в работе (Caporale, Spagnolo F., Spagnolo N., 2016).

Гипотеза 4. *Влияние сентимента частных инвесторов на биржевые характеристики акций малых и средних компаний более значимо, чем для крупных компаний.*

Мы предполагаем, что сентимент частных инвесторов больше влияет на акции малых и средних компаний, у которых доля физических лиц в объеме торгов выше. Для примера, доля нерезидентов в торгах обыкновенными акциями ПАО «Сбербанк» составляет 45%, а среднечасовой объем торгов за 2019 г. – 2 млрд руб. Напротив, в акциях ПАО «Тантал» в объеме торгов акциями нерезиденты отсутствуют, крайне мала доля институциональных инвесторов, объем акций в свободном обращении составляет всего 5%, а среднечасовой объем торгов за тот же период – 80 тыс. руб. в день. Можно предположить, что влияние сентимента физических лиц будет больше в случае ПАО «Тантал». Авторы (Baker, Wurgler, 2006) на данных рынка США получили вывод о том, что сентимент значимо влияет на доходность акций новых и небольших компаний.

4. Методология

4.1. Применение модифицированной нейросети, подготовка и разметка данных

Для построения индекса сентимента требовалось обрабатывать большие массивы сообщений (высказываний) в профильных социальных сетях, посвященных акциям и инвестированию в них на российском рынке. Мы отмечаем, что участники социальных сетей в переписке выражают свои мысли самыми разными способами, переделывая названия компаний и глаголы, используя жаргонизмы и искаженные английские термины, эмодзи. В связи с этим широко применяемый метод выявления тональности слов мы сочли для нашей работы неприемлемым. Оригинальность нашего исследования состоит в том, что единицей обработки в нейронной сети является не слово, а сообщение в целом.

Наш прием связан с тем, что, в отличие от статей в газетах и журналах или в официальных сообщениях, нельзя классифицировать сообщения как негативные, если они содержат слово «вниз». Отчасти это – проблема русского языка, который передает большое разнообразие тонких эмоций, а отчасти – проблема молодежной аудитории. Структурирование неструктурированной информации (в нашем случае, текста) – серьезная проблема на пути к решению нашей исследовательской задачи. Мы подошли к ее решению современными методами интеллектуального анализа данных.

На этапе предварительной обработки данных мы применили процедуры лемматизации и стемминга (stemming) для приведения слов к нормальной форме, удаления окончаний и стоп-слов, знаков препинания. В результате такой обработки распределение числа повторов слов стало более равномерным, уменьшилась доля выбросов. Далее нами был составлен оригинальный русскоязычный словарь с официальными (включая тикеры) и неформальными названиями компаний. В качестве последних на форумах часто используются жаргонизмы, фамилии CEO и основных владельцев, названия материнских компаний, названия основных продуктов и услуг.

Следующим этапом в построении базовых моделей машинного обучения стало преобразование сообщений в числовые векторы. Для этого применены алгоритмы One Hot Encoding и TF-IDF Encoding. One Hot Encoding создает бинарные переменные в числе, равном числу слов в выборке данных, что позволяет каждое сообщение представить в виде вектора из нулей и единиц. TF-IDF Encoding сопоставляет сообщению вектор, элементы которого характеризуют частоту употребления каждого слова в данном сообщении и во всей выборке сообщений (Sparck, 1972). Это позволило нам ослабить значимость слов, не несущих практического смысла.

Метод нейронных сетей использовался для присвоения сообщению вероятности принадлежности соответствующему классу сентимента. Мы различаем три класса эмоциональных окрасок: негативные (для выделения медвежьего тренда и рекомендаций продавать), нейтральные (для выделения сообщений, из которых непонятно, как поведут себя участники обсуждения) и позитивные (для выделения прогноза участников на рост цены и рекомендации, зачастую завуалированной, открывать позиции на покупку).

Для решения проблемы классификации через «обучение с учителем» мы сначала выбрали модель (Wang et al., 2018), представляющую собой нейронную сеть, которая находит закономерности между словами в сообщениях, имея информацию об априорном семантическом взаимоотношении слов.

Чтобы оценить преимущества нейронной сети (модели (Wang et al., 2018)), мы реализовали на нашей выборке сообщений стандартные методы машинного обучения в качестве эталонов. Цель такого сопоставления – оценить, насколько собранные и вручную размеченные данные информативны (табл. 2).

В табл. 2 показаны результаты тестирования моделей на обучающей и тестовой выборках (train и test соответственно) по двум типам метрик: Accuracy и F_1 -мера. Если Accuracy показывает процент правильно классифицированных ответов, то невзвешенная макропеременная F_1 -мера учитывает качественные особенности набора данных, влияющие на процесс обучения, включая про-

Таблица 2

Сопоставление результатов обработки текста в нейронной сети и методами машинного обучения

Метод	OneHotEncoding						TF-IDF					
	Acc (Train)	Acc (Test)	F1 (Train)	F1 (Test)	F1 weigh (train)	F1 weigh (test)	Acc (Train)	Acc (Test)	F1 (Train)	F1 (Test)	F1 weigh (train)	F1 weigh (test)
SGD	0,812	0,728	0,710	0,564	0,811	0,732	0,769	0,729	0,625	0,563	0,775	0,734
Случайный лес	0,984	0,736	0,977	0,524	0,984	0,756	0,989	0,744	0,983	0,507	0,989	0,782
Дерево решений	0,785	0,550	0,742	0,455	0,771	0,511	0,890	0,614	0,856	0,472	0,885	0,591
Константная модель	0,751	0,609	0,286	0,252	0,644	0,461						

блему несбалансированности классов. Поэтому лучшие модели мы отбираем по метрике F_1 (таблицы матриц ошибок для обучающей и тестовой выборок для моделей SGD, случайный лес и дерево решений представлены в Приложении, табл. А2, панели 1–12).

Из данных табл. 2 следует, что особенности наших данных позволяют обучить модель, предсказывающую верный класс сообщения, в 72,8% случаях. F_1 -мера определила нижнюю границу эффективности для обучаемой модели (SGD + ONE). Как следствие, был сделан следующий вывод: тестирующий набор данных может позволить моделировать связь между сообщениями и их сентиментом с точностью не меньше, чем 72,8% и F_1 -мерой в 0,564.

Далее для анализа данных мы применили модель на основе трансформеров и CNN, предложенную (Yang et al., 2019). Так как One-Hot Encoding может привести к переобучению и не очень хорошо справляется с неформальной, часто меняющейся лексикой, в итоговой модели мы используем процедуру CNN Embedding⁵.

На втором шаге после CNN Embedding мы применили полносвязную нейронную сеть с тремя слоями. В первом слое – 128 нейронов, во втором – 64. В обоих слоях в качестве функции активации используется гиперболический тангенс. Также используется dropout (исключение определенного процента случайных нейронов) с параметром 0,5 после каждого слоя для предотвращения переобучения. Последний слой состоит из трех нейронов и имеет функцию активации softmax. Модель обучалась 20 эпох, с коэффициентом скорости обучения (learning rate) 0,0015. CNN-based-модель на тестовой выборке показала F_1 -меру 57,7% и точность (Accuracy), равную 69,4% (Приложение, табл. А2, панели 13, 14).

4.2. Построение индикаторов на основе сентимента частных инвесторов

В нашей работе анализируется широкий спектр метрик сентимента частных инвесторов (табл. 3). Оригинальные метрики тональности, рассчитанные нами с применением методов машинного обучения для классификации сообщений, сопоставляются с дополнительными, относительно простыми прокси-переменными сентимента.

Для эконометрической проверки гипотез мы использовали только классы позитивных и негативных сообщений, так как нейтральные сообщения в основ-

Таблица 3

Прокси-переменные сентимента частных инвесторов

Прокси сентимента	Описание	Ожидаемое направление влияния на:	
		доходность	объем торгов
Авторские индикаторы тональности сообщений частных инвесторов, выявленной с применением нейросети			
<i>Mes_pos</i>	Число сообщений позитивной тональности в отношении акций данной компании	+	+
<i>Mes_neutr</i>	Число сообщений нейтральной тональности в отношении акций данной компании	+ / –	+
<i>Mes_neg</i>	Число сообщений негативной тональности в отношении акций данной компании	-	+

⁵ <https://tfhub.dev/google/universal-sentence-encoder-multilingual-large/3>

Окончание таблицы 3

Прокси сентимента	Описание	Ожидаемое направление влияния на:	
		доходность	объем торгов
<i>HYPE</i>	Авторский индикатор одновременно учитывает и рост числа сообщений позитивной тональности, и рост совокупного числа сообщений. Рассчитывается по формуле (1)	+	+
Дополнительные прокси сентимента			
<i>Yandex_abs</i>	Абсолютное число запросов в Яндекс Метрике рассчитывается для каждой компании отдельно на основе запросов в виде «Название компании» + «акции»	+ / -	+
<i>Yandex_rel</i>	Относительное число запросов в Яндекс Метрике как отношение абсолютного числа запросов по компании к совокупному числу поисковых запросов	+ / -	+
<i>Num_mes</i>	Число сообщений о каждой компании в MFD и Telegram без разметки тональности сообщений	+ / -	+
<i>TR_Indicator</i>	Индикатор Refinitiv Credit Text Mining News Component оценивает риск компании на основе публикаций, брокерских отчетов, записей конференц-звонков с аналитиками о компании. Измеряется по шкале от 1 (максимальный риск) до 100 (минимальный риск)	+	+

ном представляют собой текст, не относящийся напрямую к обсуждению ситуации на фондовом рынке. Разбив сообщения пользователей на классы с позитивной и негативной тональностью, мы вводим авторский индикатор *HYPE* ((1) и табл. 3). Рассчитывается по формуле:

$$HYPE_T = \frac{MSG_T^{POSITIVE}}{MSG_T^{ALL}} \left(\frac{MSG_T^{ALL}}{MSG_{T-1}^{ALL}} - 1 \right), \quad (1)$$

где $MSG_T^{POSITIVE}$ – число позитивных сообщений по каждой акции; MSG_T^{ALL} – сумма позитивных и негативных сообщений по каждой акции; $MSG_T^{ALL} / MSG_{T-1}^{ALL} - 1$ – изменение числа позитивных и негативных сообщений к предыдущему месяцу; не может превышать 10.

Предложенный нами *HYPE*-индикатор принимает значения от 0 (все сообщения негативные или число сообщений снизилось / не изменилось) до 100 (все сообщения позитивные и число сообщений выросло в 10 и более раз).

В качестве дополнительных переменных анализа выбраны прокси-метрики сентимента, часто применяемые в других работах (см. табл. 3). Это – число запросов в Яндекс Метрике, суммарное число всех сообщений в Telegram и MFD и индикатор Refinitiv Credit Text Mining News Component, оценивающий риск компании по публикациям в медиа.

4.3. Выбор контрольных переменных

Мы рассматриваем следующие контрольные переменные.

Финансовые показатели компаний на базе квартальной отчетности. Рентабельность активов (отдача на активы, ROA – отношение суммы чистой прибыли и процентных платежей по долгу с учетом налогового щита к величине активов) и долговая нагрузка, фиксируемая по показателям коэффициента чистого долга (Net Debt к EBITDA) и коэффициента покрытия процентов (соотношение EBIT и процентных платежей). В работе (Анкудинов, Лебедев, 2014) предполагается, что ROA отражает способность компании создавать конкурентные преимущества, и по выборке 107 крупнейших компаний российского рынка эмпирически выявляется, что ROA значимо положительно влияет на доходность акций. Авторы (Nurazi., Kananlua, Usman, 2015) по рынку Индонезии выявили, что с ростом ROI (Net Profit After Tax к активам) растет объем торгов.

Мы предполагаем значимое положительное влияние ROA на торговую активность. (Анкудинов, Лебедев, 2014) выявили, что высокий уровень долговой нагрузки, измеряемый на основе финансового левериджа (D/E), значимо положительно влияет на доходность акций российских компаний, авторы объясняют данный эффект тем, что привлечение большего долгового финансирования увеличивает контроль со стороны кредиторов за деятельностью менеджеров компании. Мы учитываем базовую гипотезу современных корпоративных финансов о наличии оптимального уровня долга, когда слишком большой долг негативно влияет на рост цены акции, так же, как и субоптимальный низкий долг. Таким образом, в нашу модель закладывается нелинейное влияние долговой нагрузки на доходность акций. (Nurazi, Kananlua, Usman, 2015) на данных рынка Индонезии выявили, что чем выше леверидж, тем ниже объем торгов. В русле классической теории финансов для несовершенных рынков капитала мы предполагаем нелинейное влияние показателей долговой нагрузки (парабола ветвями вниз) на все рассматриваемые нами зависимые переменные: существует оптимум, при котором достигается максимальная доходность и максимальный объем торгов.

Показатели корпоративного управления (КУ): размер совета директоров и доля независимых директоров. В русле работы (Andreou et al., 2016) мы предполагаем, что для российского рынка с наличием проблем защиты прав собственности инвесторы распознают защиту в виде различных механизмов корпоративного управления. Увеличение числа членов совета директоров (СД) и повышение доли независимых директоров (НД) ограничивают нежелательное оппортунистическое поведение менеджеров, что способствует более эффективному функционированию компании и росту цен акций. Авторы (Andreou et al., 2016) по выборке из 8 тыс. американских компаний выявили, что чем больше число членов совета директоров, тем меньше риск снижения цен акций компаний. В работе (Rostami S., Rostami Z., Kohansala, 2016) на данных развивающегося рынка Ирана показано, что доля независимых директоров значимо положительно влияет на доходность акций.

Мы также предполагаем, что высокое качество КУ повышает уверенность инвесторов в перспективах будущего роста компании, что способствует росту объема торгов. В работе (Sidhu, Kaur, 2019) на данных рынка Индии показано, что при увеличении размера СД и росте доли НД снижается мера неликвидности Амихуда. В работе (Ali, Liu, Su, 2017) по рынку Австралии показано, что повышение качества КУ (измеряемого по интегральному индексу, учитывающему в том

числе численность и состав совета директоров) значимо положительно влияет на торговый оборот, уменьшает число дней с нулевым объемом торгов, снижает меру неликвидности Амихуда.

Размер компании. На развитых рынках имеет место премия за размер, когда акции компаний малой капитализации демонстрируют более высокую доходность. На развивающихся рынках такая четкая зависимость не выполняется. Мы предполагаем, что для крупных компаний доходность акций выше, чем для малых – в соответствии с работой (Анкудинов, Лебедев, 2014) на данных российского рынка.

Акции более крупных компаний («голубых фишек») демонстрируют более высокую торговую активность. (Ali, Liu, Su, 2019) на данных рынка Австралии выявили, что чем больше рыночная капитализация, тем значимо выше торговый оборот, ниже число дней с ненулевым объемом торгов и ниже мера неликвидности Амихуда.

Биржевые характеристики акций: волатильность доходности и показатели ликвидности (относительный объем торгов, число дней с ненулевым объемом торгов, доля акций в свободном обращении – Free Float). Фактор систематического риска на российском рынке мало значим по сравнению с развитыми: если на рынке США систематический риск объясняет около 30–40% доходности акций, то на развивающихся – около 5–10%. Волатильность дневной доходности рассматривается нами как прокси-показатель общего риска, позволяющий провести самостоятельный анализ влияния фундаментальных факторов. Мы предполагаем, что чем выше волатильность, тем выше доходность.

Мы предполагаем, что чем ниже показатели ликвидности, тем выше доходность акций, так как инвесторы требуют премию за риск неликвидности. (Brown, Crocker, Foerster, 2009) на данных рынка США получили вывод об отрицательном влиянии объема торгов на доходность акций.

Мы предполагаем, что чем выше волатильность доходности акций, тем выше объем торгов, что согласуется с гипотезой и выводом (Tayeh, 2016) на данных рынка акций Иордании. В работе (Ali, Liu, Su, 2019) на данных рынка Австралии также показано, что волатильность доходности отдельных акций играет значимую роль в объяснении их ликвидности: чем выше волатильность, тем выше торговый оборот и ниже число дней с ненулевым объемом торгов.

Макроэкономические показатели: темп роста ВВП России, доходность к погашению государственных облигаций США, изменение индекса волатильности RVI на российском рынке, индекс деловой активности в США. Мы предполагаем, что рост российского ВВП и индекса деловой активности в США сигнализирует о благополучной ситуации на национальном и глобальном рынке соответственно. С ростом ВВП и деловой активности увеличиваются денежные потоки инвесторов на фондовый рынок, растет рыночная капитализация и доходность акций. В работе (Celebi, Honig, 2019) на данных рынка Германии выявлено значимое положительное влияние темпа роста ВВП на индекс фондового рынка с лагом в 1 квартал и незначимое влияние в текущем квартале. Мы также предполагаем, что в случае благоприятной макроэкономической ситуации повышается торговая активность на фондовом рынке. В работе (Salamat, 2016) на данных рынка Иордании показано, что чем выше темп роста ВВП, тем выше торговая активность.

Мы предполагаем, что чем ниже доходность государственных облигаций США (Treasuries), тем больше стимулов появляется у инвесторов вкладывать деньги на рынках акций развивающихся стран (погоня за доходностью при нарастании аппетита к риску). В работе (Celebi, Honig, 2019) показано, что доходность государственных облигаций значимо отрицательно влияет на доходность фондового индекса Германии. Авторы объясняют это эффектом замещения между рынками акций и облигаций (инвесторы ищут большую доходность), а также тем, что доходность государственных облигаций является прокси безрисковой доходности. Мы ожидаем, что при снижении доходности US Treasuries растут объемы торгов акциями на развивающихся рынках, включая российский. Аналогичный эффект повышения ликвидности при снижении уровня процентных ставок был выявлен и на рынке Индии (Debata, Mahakud, 2018).

«Индекс страха» RVI диагностирует панику и порождает падение цен акций, затухает торговая активность (на месячных данных), что подтверждено в работе (Mai, Chuan, Fang, 2016) на данных рынка Австралии (индекс VIX) и рынка Пакистана в кризисные периоды (Khan, Rehman, 2020).

Список рассматриваемых нами контрольных переменных приведен в табл. 4.

Таблица 4

Комментарии по контрольным переменным

Контрольная переменная	Описание	Ожидаемое направление влияния на:	
		доходность	объем торгов
Показатели волатильности и ликвидности акций			
<i>Volatility</i>	Волатильность дневной доходности акций за месяц	+	+
<i>TradeVolFF</i>	Объем торгов с учетом free-float, %	–	Не рассматривается
<i>TradeVolFF_2</i>	Квадрат переменной TradeVolFF	–	
<i>FreeFloat</i>	Доля акций в свободном обращении, %	–	
<i>NzTradeDaysShare</i>	Доля торговых дней с ненулевым объемом торгов за месяц, %	–	
Финансовые показатели эмитента			
<i>ROA</i>	Отдача на активы эмитента	+	+
<i>NetDebt_EBITDA</i>	Коэффициент чистого долга, рассчитывается как отношение Net Debt к EBITDA	+	+
<i>NetDebt_EBITDA_2</i>	Квадрат переменной NetDebt_EBITDA	–	–
<i>EBIT_Interest</i>	Коэффициент покрытия EBIT процентных расходов	+	+
<i>EBIT_Interest_2</i>	Квадрат переменной EBIT_Interest	–	–
Показатели корпоративного управления			
<i>BoardSize</i>	Численность Совета директоров	+	+
<i>IndepDirPercent</i>	Доля независимых директоров в Совете директоров	+	+

Окончание таблицы 4

Контрольная переменная	Описание	Ожидаемое направление влияния на:	
		доходность	объем торгов
Макроэкономические показатели			
<i>GDPgrowth</i>	Темп роста ВВП в РФ, квартал к кварталу, %	+	+
<i>RVI_change</i>	Изменение индекса волатильности RVI на российском рынке за месяц, %	–	–
<i>Gov10Y</i>	Ставка доходности к погашению 10-летних ОФЗ, %	–	–
<i>PMI_US</i>	Индекс деловой активности PMI в сфере услуг в США на начало месяца	+	+
<i>d14, d15, ..., d18</i>	Дамми-переменные временных периодов: 2014, 2015, ..., 2018 г. соответственно	+ / –	
Дамми-переменные размера компаний			
<i>dummy_large</i>	Дамми-переменная крупных компаний: 1 – если компания входит в 20 крупнейших в выборке, 0 – иначе	+	+
<i>dummy_small</i>	Дамми-переменная малых компаний: 1 – если компания входит в 20 малых в выборке, 0 – иначе	–	–

4.4. Тестирование эконометрических моделей

Мы тестируем на панельных данных следующие эконометрические модели со случайными эффектами:

$$y_{it} = \beta_1 \text{Sentiment}_{it} + \beta_2 \text{Sentiment}_{it}^2 + \sum \beta_{3,k} x_{i,t,k} + u_{it}, \quad (2)$$

где i – номер компании; t – месяц наблюдения; y_{it} – зависимая переменная; Sentiment_{it} – одна из прокси-переменных сентимента (переменные сентимента включаются в модели по одной, см. табл. 3); $x_{i,t,k}$ – контрольные переменные (см. табл. 4); $\beta_1, \beta_2, \beta_{3,k}$ – коэффициенты регрессии, u_{it} – индивидуальные эффекты.

Зависимые переменные и показатели сентимента берутся в помесечной разбивке, как и ряд макроэкономических показателей (например, доходность к погашению по 10-летним ОФЗ (среднее значение по месяцу) или индекс деловой активности), показатели корпоративного управления (изменения фиксируются на дату). Поквартально взяты некоторые контрольные переменные: финансовые показатели компании–эмитента акций и годовой темп роста ВВП. Для разнесения квартальных значений по месяцам мы используем прием дублирования (присваивания неизменного значения каждому из трех месяцев), аналогичный подход использовался в ряде предыдущих работ, например (Ahmad, Han, Hutson, 2015; Chen, Shangguan, Liu, 2020). Переменные включаются в модели с учетом проблемы мультиколлинеарности, вводятся дамми-переменные отдельных лет.

5. Данные. Описательная статистика по выборке

Выборка нашего исследования состоит из 60 компаний, временной период – с 1 января 2013 г. по 31 марта 2020 г. Выборка была разделена на три равные группы:

- 20 компаний – высоколиквидных «голубых фишек», входящих в индекс Мосбиржи и имеющих кросс-листинг на биржах LSE/ NYSE/ Франкфурт. Среднедневной объем торгов выше 100 млн руб.;
- 20 компаний с акциями средней ликвидности: среднедневной объем торгов от 10 до 100 млн руб. Все компании находятся в отраслевых индексах и хоть один раз были включены в индекс Мосбиржи;
- 20 компаний с низколиквидными акциями, которые не входят в индекс Мосбиржи и не имеют кросс-листинга на других биржах.

Были собраны данные о доходности акций по всем компаниям, волатильности доходности и объеме торгов, прокси-переменные сентимента, финансовые и нефундаментальные показатели. Основной источник данных о биржевых характеристиках, финансовых и нефундаментальных показателях, макроэкономических переменных: Eikon Refinitiv. В ряде случаев (при отсутствии данных в Eikon Refinitiv) показатели корпоративного управления были взяты из ежеквартальных отчетов эмитентов.

Данные для расчета индекса сентимента были собраны из биржевых форумов с 20 тыс. и более зарегистрированных пользователей, телеграм-каналов с 5 тыс. и более пользователей и чатов с 500 и более пользователей. Общий охват – 50% активных счетов на Мосбирже на март 2020 г.

Только вручную нами было размечено свыше 30 тыс. текстовых сообщений (не включая сообщений со смайлами и т.п.). Всего полученных токенов (слов, знаков препинания, смайлов) 727 326. Сформированный корпус был исходно несбалансированным (4061 негативное, 24 249 нейтральных и 4266 позитивных сообщений), что препятствовало обучению. Поэтому мы использовали веса. Впоследствии несбалансированность классов учитывалась при обучении классификатора следующим образом: оптимизируемая функция потерь учитывала сообщения из малых классов сильнее, чем сообщения из нейтрального класса. В итоге, в обучающей выборке ручной обработки у нас осталось 28 546 текстов, в тестовой – 4030.

В табл. 5 представлена описательная статистика по выборке.

Таблица 5
Описательная статистика по выборке

Переменная	Число наблюдений	Среднее	Стандартное отклонение	Минимум	Максимум
Зависимые переменные					
<i>total_return</i>	5185	0,99	10,734	-25	40
<i>TradeVolFF</i>	5185	6,682	1,858	0	12,843
Объясняющие переменные сентимента					
<i>ln_tele_hype</i>	2339	1,389	1,311	0	4,805
<i>ln_mfd_hype</i>	4320	1,799	0,877	0	5,0
<i>ln_HYPE</i>	4499	1,816	0,857	0	5,0
<i>ln_Yandex_abs</i>	1440	8,276	1,965	3,829	13,511
<i>ln_Yandex_relat</i>	1440	3,846	1,956	-0,728	9,038

Окончание таблицы 5

Переменная	Число наблюдений	Среднее	Стандартное отклонение	Минимум	Максимум
<i>ratingTR</i>	4859	6,693	1,093	0,526	10,0
<i>ln_all_mes</i>	4575	4,915	2,355	0	10,643
<i>ln_all_neg_mes</i>	4575	3,427	2,192	0	8,888
<i>ln_all_ntr_mes</i>	4575	4,450	2,281	0	10,211
<i>ln_all_pos_mes</i>	4572	3,408	2,074	0	8,38
Контрольные переменные					
<i>volatility</i>	5187	0,021	0,014	0,005	0,08
<i>FreeFloat</i>	5039	36,478	25,564	0	100
<i>NzTradeDays-Share</i>	5187	96,848	12,92	0	100
<i>ROA</i>	5082	1,24	2,576	-7,421	7,58
<i>BoardSize</i>	4995	9,858	3,012	0	19
<i>IndepDirPercent</i>	4389	29,443	20,736	0	100
<i>GDPgrowth</i>	5307	0,069	0,283	-0,704	0,61
<i>RVI_change</i>	4636	3,124	20,219	-32,74	86,48
<i>Gov10Y</i>	5307	8,515	1,620	6,136	14,38
<i>PMI_US</i>	5307	54,415	2,802	39,8	61

6. Результаты эконометрических расчетов

В табл. 6–7 показаны результаты расчетов для зависимых переменных доходности и объема торгов соответственно. Как следует из табл. 6, наш авторский индикатор NYRE, который учитывает и количество сообщений, и их позитивный настрой, оказывает положительное и значимое на уровне 10% влияние на доходность акций по всем каналам. Чем больше количество обсуждений компании в социальных сетях по рассматриваемому месяцу и чем более позитивен тон сообщений, тем выше месячная доходность акции. Данный вывод согласуется с поставленной гипотезой 1 и с результатами работ (Tetlock, 2007; Renault, 2017; Sun, Najand, Shen, 2016) по зарубежным рынкам. Положительно и влияние компонент NYRE индикатора, рассчитанное по платформам MFD и Telegram в отдельности, но статистическая значимость данных факторов невысока. Для сравнения отметим, что чем больше количество запросов по компании в поисковой системе Yandex, тем выше доходность акций, но данный фактор статистически незначим.

Из данных в табл. 6 можно сделать вывод о том, что наибольшей значимостью обладают негативные сообщения: и переменная числа негативных сообщений, и квадрат данной переменной значимы на уровне 1%. Мы объясняем это тем, что инвесторы сильнее реагируют на негативные сообщения, что согласуется с выводами (Caronale, Spagnolo F., Spagnolo N., 2016) на данных рынка восьми развитых стран. Направление влияния можно описать параболой с вет-

виями вниз. Мы объясняем это тем, что экстремально низкое число негативных сообщений тождественно сильному позитиву, что воспринимается инвесторами с недоверием и порождает снижение спроса. Если число негативных сообщений относительно велико, с ростом их числа доходность уменьшается. Зависимость доходности от числа нейтральных по тональности сообщений также описывается параболой ветвями вниз, но значимость числа сообщений и его квадрата снижается по сравнению с переменными числа сообщений негативной тональности (факторы значимы на уровне 10 и 5% соответственно). Для позитивных сообщений сохраняется зависимость в виде параболы ветвями вниз, но на уровне 10% значим только квадрат переменной. Мы делаем вывод, что инвесторы не доверяют слишком позитивному обсуждению компании: доходность ее акций начинает снижаться. При существенном нарастании настроений той или иной тональности доходность акций снижается, т.е. следование за толпой (консенсусному настроению) не позволяет получить высокую доходность.

Все контрольные переменные показали ожидаемое направление влияния. Мы отмечаем значимость общего риска (волатильности) на уровне 1% (табл. 6). Низколиквидные акции демонстрируют большую доходность, что согласуется с положениями финансовой экономики и эмпирическими выводами (Brown, Crocke, Foerster, 2009), полученными на данных рынка США. Подробный комментарий относительно контрольных переменных дан в Приложении.

Таблица 6

Влияние прокси сентимента на полную доходность

Переменные и характеристики моделей	Номера регрессионных моделей							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Wald</i>	168,2	171,2	172,46	243,73	242,89	250,07	220,99	185,61
<i>Num obs</i>	2138	3836	3836	1239	1239	3869	4182	3869
<i>ln_YandexAbs</i>				0,2071 (0,1777)				
<i>ln_YandexRel</i>					0,219 (0,1787)			
<i>ln_HYPE</i>			0,5286* (0,3147)					
<i>ln_mfd_hype</i>		0,3171 (0,3024)						
<i>ln_tele_hype</i>	0,2063 (0,2009)							
<i>ln_all_neg_mes</i>						0,6423*** (0,1028)		
<i>ln_all_neg_mes_2</i>						-0,4073*** (0,1203)		
<i>ln_all_neutr_mes</i>							0,6029* (0,3281)	
<i>ln_all_neutr_mes_2</i>							-0,0762** (0,0335)	
<i>ln_all_pos_mes</i>								0,4808 (0,3438)
<i>ln_all_pos_mes_2</i>								-0,0662* (0,0399)

Окончание таблицы 6

Переменные и характери- стики моделей	Номера регрессионных моделей							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Volatility</i>	212,6217*** (24,0821)	168,1124*** (21,2447)	166,433*** (21,0436)	281,7558*** (23,6214)	281,982*** (23,666)	172,4047*** (22,2873)	203,521*** (22,0071)	173,9585*** (20,8977)
<i>TradeVolFF</i>	-0,4161*** (0,1535)	-0,0166 (0,145)	-0,0158 (0,1448)	-0,5974*** (0,1757)	-0,6038*** (0,1742)	0,02 (0,4519)		-0,0086 (0,1399)
<i>FreeFloat</i>							-0,0026 (0,0082)	
<i>NzTradeDaysShare</i>							0,0223 (0,0143)	
<i>ROA_lag</i>	0,4775*** (0,0944)	0,3584*** (0,0746)	0,355*** (0,0743)	0,4034*** (0,0948)	0,4015*** (0,0947)	0,3369*** (0,076)		0,3593*** (0,078)
<i>ROA</i>							0,3356*** (0,0592)	
<i>BoardSize</i>	0,301*** (0,0943)	0,2374*** (0,0873)	0,2359*** (0,0873)	0,351*** (0,0873)	0,3485*** (0,0874)	0,2547*** (0,0887)	0,2472*** (0,0965)	0,2475*** (0,0931)
<i>IndepDirPercent</i>	0,0403*** (0,0136)	0,0243* (0,0134)	0,0237* (0,0134)	0,0265 (0,0169)	0,0263 (0,0169)	0,0218* (0,0127)		0,0247** (0,0124)
<i>GDP growth</i>	0,4625 (1,2479)	0,73 (0,7514)	0,7368 (0,7536)	-0,2761 (1,5361)	-0,3374 (1,532)	0,8848 (0,7524)		0,7386 (0,7496)
<i>RVI_change</i>							-0,1273*** (0,0149)	
<i>Gov10Y</i>	-0,3457 (0,3826)	-0,0196 (0,2765)	-0,0121 (0,2773)	-0,1558 (0,3897)	-0,1583 (0,3897)	-2,3371** (1,1164)		-0,017 (0,2751)
<i>Gov10Y_2</i>						0,1173** (0,0568)		
<i>PMI_US</i>							-0,0647 (0,0921)	
<i>_cons</i>	-1,1589 (3,4687)	-4,6976* (2,5784)	-5,127** (2,6207)	-3,4909 (3,4479)	-2,5151 (3,4298)	4,9016 (5,5622)	-3,7883 (5,222)	-5,0605** (2,4101)
<i>d 14</i>		-3,9705*** (0,6671)	-3,9169*** (0,6664)			-2,3627*** (0,8211)	-1,6713** (0,7902)	-3,9526*** (0,7284)
<i>d 15</i>		-0,1431 (0,8521)	-0,1056 (0,8503)			1,3489 (0,8755)	-0,4866 (0,6802)	-0,1236 (0,8551)
<i>d 16</i>		0,9506* (0,5694)	0,9943* (0,5672)			2,7677*** (0,7356)	0,599 (0,5273)	1,0289* (0,5586)
<i>d 17</i>	-2,6781*** (0,5902)	-2,6772*** (0,4757)	-2,6431*** (0,4775)			-1,375** (0,5454)	-2,8261*** (0,4641)	-2,5855*** (0,4811)
<i>d 18</i>	-2,743*** (0,5388)	-2,6458*** (0,4602)	-2,6311*** (0,459)	-3,4301*** (0,5894)	-3,4412*** (0,5882)	-2,3342*** (0,5031)	-1,8187*** (0,4523)	-2,6847*** (0,4613)

В табл. 7 приведено влияние оцененных факторов на торговую активность (относительный объем торгов). Авторский индикатор сентимента NYFE положительно и значимо (на уровне 1%) влияет на объем торгов. Иными словами, рост числа сообщений и их позитивная тональность приводят к повыше-

нию торговой активности. Данный вывод согласуется с результатами (Oliveira, Cortez, Areal, 2013) по рынку США.

Влияние компоненты индикатора HYPE на платформе MFD описывается параболой ветвями вниз. При низком и высоком значении индикатора сентимента MFD объем торгов акциями снижается по сравнению со средним уровнем. Низкое число сообщений и их негативный окрас отражают низкий интерес инвесторов к акциям. Слишком активные обсуждения и слишком сильный позитивный настрой вызывают недоверие инвесторов, многие профессиональные инвесторы боятся идти за толпой. В отличие от платформы MFD индикатор HYPE по платформе Telegram положительно влияет на торговую активность, но его статистическая значимость невысока.

Зависимость объема торгов от числа поисковых запросов по компании в Yandex также описывается параболой ветвями вниз, причем и число запросов, и его квадрат значимы на уровне 1%. Если число запросов относительно мало или, напротив, велико, то торговая активность снижается (табл. 7). Мы объясняем это тем, что после появления новостей о компании события начинают обсуждаться на форумах и параллельно растет число поисковых запросов с целью получения более детальной информации. Подтверждение новости приводит к формированию определенного мнения у инвесторов и, как следствие, к сделке – торговый объем растет. В дальнейшем рост запросов уже не обязательно связан с конкретной новостью: либо инвесторы осуществляют поиск показателей компании для более долгосрочного инвестирования, либо, если новость вышла достаточно давно, складывается понимание, что информация уже учтена в ценах, и сделки не совершаются.

Тональность обсуждений компании положительно значима на уровне 1% (табл. 7). Увеличение числа позитивных и числа негативных сообщений способ-

Таблица 7

Влияние прокси сентимента на объем торгов

Переменные и характеристики моделей	Номера регрессионных моделей							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Wald</i>	132,84	316,5	251,7	368,78	417,16	470,92	453,90	491,32
<i>Num obs</i>	2138	3833	3833	1239	1239	3869	3869	3869
<i>ln_YandexAbs</i>				1,648*** (0,4131)				
<i>ln_Yandex Abs_2</i>				-0,0513** (0,0207)				
<i>ln_YandexRel</i>					1,175*** (0,2468)			
<i>ln_Yandex Rel_2</i>					-0,0498** (0,0208)			
<i>ln_HYPE</i>			0,1403*** (0,0516)					
<i>ln_mfd_hype</i>		0,5654*** (0,1585)						
<i>ln_mfd_hype_2</i>		-0,1039*** (0,0299)						

Окончание таблицы 7

Переменные и характеристики моделей	Номера регрессионных моделей							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>ln_tele_hype</i>	0,0195 (0,0241)							
<i>ln_all_mes</i>						0,2593*** (0,0478)		
<i>ln_all_neg_mes</i>							0,2873*** (0,0463)	
<i>ln_all_pos_mes</i>								0,2924*** (0,0483)
<i>Volatility</i>	39,7148*** (4,137)	30,2739*** (2,9964)	29,4468*** (3,3557)	19,1056*** (2,898)	19,5567*** (2,9047)	23,8892*** (2,8868)	23,2249*** (2,8807)	23,5296*** (2,8919)
<i>ROA_lag</i>	-0,0491 (0,0526)	-0,09*** (0,0314)	-0,0832*** (0,0309)	-0,085 (0,0571)	-0,0847 (0,0572)	-0,0856** (0,0354)	-0,083*** (0,0324)	-0,0863*** (0,0329)
<i>BoardSize</i>	-0,0354 (0,0401)	0,0017 (0,0595)	-0,0032 (0,0622)	-0,0579 (0,0432)	-0,0566 (0,0447)	0,0177 (0,053)	0,0163 (0,0535)	0,0207 (0,0544)
<i>IndepDirPercent</i>	0,0055 (0,0066)	0,0055 (0,0036)	0,0052 (0,0038)	0,0013 (0,0057)	0,0012 (0,0056)	0,0037 (0,0038)	0,0035 (0,0037)	0,0036 (0,0036)
<i>GDP_growth</i>	-0,1027 (0,0975)	-0,0654 (0,0644)	-0,0625 (0,0619)	0,0727 (0,0954)	-0,0268 (0,0994)	-0,0289 (0,0538)	-0,0482 (0,0532)	-0,0548 (0,0532)
<i>Gov10Y</i>	-0,1334*** (0,041)	-0,078*** (0,0218)	-0,0743*** (0,0216)	-0,0062 (0,0329)	-1,8403*** (0,4912)	-0,3002*** (0,0844)	-0,3*** (0,0815)	-0,3073*** (0,0815)
<i>Gov10Y_2</i>					0,1204*** (0,0319)	0,0121*** (0,0041)	0,012*** (0,004)	0,0123*** (0,004)
<i>_cons</i>	8,2201*** (0,6193)	7,0475*** (0,6876)	7,4474*** (0,6815)	-2,0198 (1,9429)	11,2377*** (2,0385)	6,5548*** (0,7914)	6,8987*** (0,7403)	6,8742*** (0,7568)
<i>d14</i>		-0,5207** (0,2285)	-0,5608** (0,2429)			-0,0789 (0,1664)	-0,0866 (0,165)	-0,085 (0,1675)
<i>d15</i>		-0,5261*** (0,2042)	-0,56*** (0,2108)			-0,2533* (0,1466)	-0,2711* (0,1421)	-0,2635* (0,1451)
<i>d16</i>		-0,4695*** (0,1596)	-0,4792*** (0,1671)			-0,2389** (0,1197)	-0,2597** (0,1199)	-0,2641** (0,1217)
<i>d17</i>	-0,4333*** (0,1681)	-0,4182*** (0,1469)	-0,4198*** (0,1547)			-0,2905*** (0,1109)	-0,3024*** (0,1124)	-0,2988*** (0,1129)
<i>d18</i>	-0,376*** (0,108)	-0,3635*** (0,0995)	-0,3687*** (0,1033)	-0,0247 (0,0728)	-0,0878 (0,0726)	-0,2627*** (0,0699)	-0,2605*** (0,0695)	-0,2492*** (0,0684)

стует росту торговой активности, что согласуется с выводами (Tetlock, 2007). Влияние контрольных переменных прокомментировано в Приложении.

Поскольку в нашей выборке данные по платформе MFD доступны с 2014 г., а по платформе Telegram – только с 2017 г., мы проверили устойчивость полученных результатов на подвыборке с 2017 по I квартал 2020 г. Большинство выявленных ранее зависимостей сохраняется. Авторский индикатор HYPE значимо положительно на уровне 10% влияет на доходность акций. Влияние индикатора HYPE и его компоненты по платформе MFD на торговую активность можно представить в виде параболы ветвями вниз.

Рентабельность активов, численность совета директоров и процент независимых директоров положительно влияют на доходность. С ростом доходности

по государственным облигациям США уменьшается доходность акций российского рынка и снижается объем торгов.

В табл. 8 показаны результаты расчетов с учетом размера компаний. Для всех компаний сохраняется устойчивое положительное влияние сентимента на доходность акций, при этом для крупных компаний влияние сентимента на доходность акций ослабевает, а для малых компаний – усиливается, что соответствует нашей гипотезе 4. Ранее (Baker, Wurgler, 2006) на данных рынка США получили аналогичный вывод о значимом влиянии сентимента на доходность акций новых и небольших компаний.

Мы отмечаем, что инвестирование в акции крупных компаний обеспечивает большую доходность по сравнению с малыми компаниями (направление влияния дамми крупных компаний на доходность положительно, дамми малых компаний – отрицательно). Ранее аналогичный вывод по российскому рынку был получен в работе (Анкудинов, Лебедев, 2014).

Мы выявили нелинейное и значимое на уровне 1% влияние показателя сентимента на относительный объем торгов (парабола ветвями вниз): сначала при увеличении показателя сентимента торговая активность усиливается, но при экстремально высоком значении сентимента, напротив, снижается (табл. 8). Мы объясняем это тем, что инвесторы не доверяют чрезмерно позитивному фону, складывающемуся вокруг компании.

Значимы дамми-переменные размера компаний (табл. 8): относительный объем торгов выше для крупных компаний и ниже для малых компаний. Такой результат согласуется с выводами (Ali et al., 2019) на данных рынка Австралии.

Таблица 8
Результаты расчетов с учетом размера компаний

Переменные и характеристики моделей	Номера регрессионных моделей					
	(1)	(2)	(3)	(4)	(5)	(6)
	Зависимая переменная – полная доходность				Зависимая переменная – относительный объем торгов	
<i>Wald</i>	204,56	204,56	202,04	202,04	549,59	549,59
<i>Num obs</i>	3836	3836	3527	3527	3833	3833
<i>ln_HYPE</i>	0,3313 (0,5475)	0,3313 (0,5475)	0,4451 (0,5287)	0,4451 (0,5287)	0,5482*** (0,1663)	0,5482*** (0,1663)
<i>ln_HYPE_2</i>					-0,0989*** (0,0314)	-0,0989*** (0,0314)
<i>dummy_large</i>	1,8714 (2,0079)	1,8714 (2,0079)	1,649 (2,0149)	1,649 (2,0149)	1,9106*** (0,429)	1,9106*** (0,429)
<i>dummy_small</i>	-1,9043 (1,8003)	-1,9043 (1,8003)	-2,167 (1,8603)	-2,167 (1,8603)	-1,0618** (0,5228)	-1,0618** (0,5228)
<i>ln_HYPE* dummy_large</i>	-0,428 (1,1228)	-0,428 (1,1228)	-0,4661 (1,1249)	-0,4661 (1,1249)	-0,0495 (0,0611)	-0,0495 (0,0611)
<i>ln_HYPE* dummy_small</i>	0,4527 (0,7132)	0,4527 (0,7132)	0,7049 (0,7055)	0,7049 (0,7055)	-0,0001 (0,0934)	-0,0001 (0,0934)
<i>NetDebt_Ebitda</i>	0,0244 (0,0407)	0,0244 (0,0407)			-0,0006 (0,0165)	-0,0006 (0,0165)

Окончание таблицы 8

Переменные и характеристики моделей	Номера регрессионных моделей					
	(1)	(2)	(3)	(4)	(5)	(6)
	Зависимая переменная – полная доходность				Зависимая переменная – относительный объем торгов	
<i>NetDebt_Ebitda_2</i>	-0,0017 (0,0011)	-0,0017 (0,0011)			-0,0001 (0,0005)	-0,0001 (0,0005)
<i>Ebit_Interest</i>			0,1343*** (0,0514)	0,1343*** (0,0514)		
<i>Ebit_Interest_2</i>			-0,0025*** (0,0009)	-0,0025*** (0,0009)		
<i>Volatility</i>	182,048*** (23,8833)	182,048*** (23,8833)	178,6723*** (25,5803)	178,6723*** (25,5803)	30,3336*** (3,0055)	30,3336*** (3,0055)
<i>TradeVolFF</i>	-0,2468 (0,1978)	-0,2468 (0,1978)	-0,2294 (0,2267)	-0,2294 (0,2267)		
<i>ROA</i>	0,1999* (0,1096)	0,1999* (0,1096)	0,1225 (0,1051)	0,1225 (0,1051)	-0,0945*** (0,0331)	-0,0945*** (0,0331)
<i>BoardSize</i>	0,1401* (0,076)	0,1401* (0,076)	0,1183 (0,0765)	0,1183 (0,0765)	-0,0182 (0,0569)	-0,0182 (0,0569)
<i>IndepDirPercent</i>	0,0183 (0,0128)	0,0183 (0,0128)	0,0187 (0,0132)	0,0187 (0,0132)	0,0051 (0,0038)	0,0051 (0,0038)
<i>GDP_growth</i>	0,7316 (0,7537)	0,7316 (0,7537)	0,6617 (0,7944)	0,6617 (0,7944)	-0,0647 (0,0668)	-0,0647 (0,0668)
<i>Gov10Y</i>	-0,0484 (0,273)	-0,0484 (0,273)	-0,0489 (0,281)	-0,0489 (0,281)	-0,0745*** (0,0218)	-0,0745*** (0,0218)
<i>_cons</i>	-5,8367* (3,1529)	-5,8367* (3,1529)	-6,6199** (3,3169)	-6,6199** (3,3169)	6,4722*** (0,8011)	6,4722*** (0,8011)
<i>d15</i>	3,7541*** (0,8887)	3,7541*** (0,8887)	3,7152*** (0,871)	3,7152*** (0,871)	-0,0206 (0,1487)	-0,0206 (0,1487)
<i>d16</i>	4,9182*** (0,7669)	4,9182*** (0,7669)	5,0839*** (0,8522)	5,0839*** (0,8522)	0,041 (0,1678)	0,041 (0,1678)
<i>d17</i>	1,4233** (0,6511)	1,4233** (0,6511)	1,646** (0,7284)	1,646** (0,7284)	0,0911 (0,1869)	0,0911 (0,1869)
<i>d18</i>	1,465** (0,5728)	1,465** (0,5728)	1,6723** (0,6647)	1,6723** (0,6647)	0,1329 (0,1732)	0,1329 (0,1732)
<i>d19</i>	4,1779*** (0,723)	4,1779*** (0,723)	4,2295*** (0,8094)	4,2295*** (0,8094)	0,4904** (0,2218)	0,4904** (0,2218)

Примечание. В табл. 6–8 символами «*», «**», «***» отмечены оценки, значимые на уровне 10, 5 и 1% соответственно.

7. Комментарии к оценке контрольных переменных

Данные в табл. 7 оценки показывают, что чем выше рентабельность активов, тем выше доходность акций: фактор ROA как по прошлому кварталу, так и по текущему значим на уровне 1% во всех моделях, что совпадает с выводами (Анкудинов, Лебедев, 2014) по российскому рынку. Из двух рассматриваемых нами показателей долговой нагрузки наиболее значимым является покрытие

ЕВИТ процентных выплат. По обоим показателям (коэффициент чистого долга и покрытие процентов) фиксируется нелинейность (парабола ветвями вниз): существует оптимум долговой нагрузки, при котором доходность акций компании максимальная.

Значимы факторы корпоративного управления. Чем больше численность совета директоров и чем больше доля независимых директоров, тем выше доходность акций компании (см. табл. 7). Этот вывод совпадает с результатами (Andreou, Antoniou, Horton et al., 2015) на данных рынка США и (Rostami S., Rostami Z., Kohansala, 2016) на данных рынка Ирана.

По макрофакторам выявлено следующее: в кризисные периоды рост индекса страха RVI приводит к снижению доходности акций (фактор значим на уровне 1%). Данный вывод соответствует результатам работы (Mai, Chuan, Fang, 2016) на данных рынка Австралии. Чем ниже доходность американских Treasuries, тем больше инвесторы вынуждены искать альтернативы на развивающихся рынках, включая российский. Данный вывод соответствует выводам работы (Celebiand, Honig, 2019) на данных рынка Германии – доходность акций и торговая активность растут.

Квартальный темп роста ВВП в большинстве моделей оказывает положительное влияние на доходность акций, но статистическая значимость данного показателя не очень высока, так как показатель относительно статичен (см. табл. 7). Положительное направление влияния темпа роста ВВП ранее было выявлено в работе (Celebi, Honig, 2019) на данных рынка Германии. Значимы дамми временных периодов: в 2014, 2017 и 2018 г. фиксировалось снижение доходности акций российского рынка (см. табл. 7).

Влияние волатильности доходности на торговую активность положительно и значимо на уровне 1% (см. табл. 8), что согласуется с результатами (Tayeh, 2016) на данных рынка акций Иордании, (Ali, Liu, Su, 2019) – на данных рынка Австралии.

Чем ниже рентабельность активов в прошлом квартале, тем выше торговая активность (см. табл. 8). Показатель ROA привязан к раскрытию квартальной отчетности. Мы делаем вывод о том, что реакция инвесторов на отчетность избыточна: раскрытие отчетности с низкой ROA порождает большую торговую активность. Часть инвесторов продает акции на плохих новостях, часть покупает с надеждой на рост в долгосрочной перспективе. Высокая отдача на капитал не мотивирует инвесторов совершать сделки, они предпочитают сохранить позиции, а приток денег новых инвесторов в 2014–2019 гг. на российский фондовый рынок оказался не очень высоким. Аналогичную зависимость выявили (Nurazi, Kananlua, Usman, 2015; Asle et al., 2013).

Численность совета директоров не оказывает статистически значимого влияния на объем торгов (в отличие от зависимой переменной доходности), направление влияния неустойчивое (см. табл. 8). Доля независимых директоров устойчиво и положительно влияет на объем торгов, что согласуется с результатами (Sidhuand, Kaur, 2019) на данных рынка Индии и (Ali, Liu, Su, 2017) на данных рынка Австралии.

Мы выявили значимое нелинейное влияние доходности десятилетних US Treasuries, которое можно описать параболой с ветвями вверх (см. табл. 8).

Близкие к нулю и высокие ставки на рынке государственных облигаций США мотивируют торговую активность, так как есть возможность для одних участников дешево купить, для других – выйти в кэш. В предыдущих работах исследовалось только линейное влияние уровня процентных ставок на торговую активность: было выявлено обратное влияние (Debata, Mahakud, 2018).

Темп роста ВВП незначимо влияет на объем торгов (см. табл. 8). В 2017 и 2018 г. фиксировалось снижение объема торгов, дамки этих временных периодов значимы во всех рассмотренных моделях.

8. Заключение

В нашей работе впервые раскрывается влияние настроений частных инвесторов в социальных сетях на показатели доходности и торговой активности на российском рынке акций. Выборка сообщений построена на уникальных данных порядка 50% владельцев активных брокерских счетов за 2013–2020 гг. Проверка устойчивости реализована на периоде с 2017 г. по I квартал 2020 г.

В работе предложен оригинальный авторский *HYPE*-индикатор, построенный на основе обработки более 30 тыс. сообщений с использованием адаптированной нейронной сети. Индикатор построен по сообщениям двух платформ – широко используемому, несмотря на официальный запрет (действовавший до июня 2020 г.), мессенджеру Telegram и самому популярному биржевому форуму в России mfd.ru. Тестировалась как объединенная оценка тональности сообщений, так и *HYPE*-индикатор для каждой базы в отдельности. Мы выявляем отличия влияния *HYPE* по платформам Telegram и MFD. Наряду с индикатором *HYPE* мы рассматриваем широкий спектр прокси-переменных сентимента: число позитивных, нейтральных и негативных сообщений, число поисковых запросов в Yandex, предложенный Refinitiv индикатор кредитного риска, построенный по сентименту брокеров и аналитиков.

Тестируя гипотезы о влиянии настроений инвесторов по панельным данным, мы на основе регрессионных построений выявили взаимосвязи между факторами сентимента частных инвесторов и биржевыми характеристиками акций с учетом эффектов нелинейности.

Результаты эмпирического тестирования на месячных данных показали, что предложенный нами оригинальный индикатор *HYPE* положительно и значимо влияет на доходность акций на уровне 10%. По акциям с большим числом позитивных обсуждений доходность при прочих равных выше. Для сравнения отметим, что число поисковых запросов по компании в Yandex и оценка риска компании на основе публикаций, брокерских отчетов, настроений аналитиков, предложенная Refinitiv, статистически незначимы, хотя показывают ожидаемые знаки влияния.

Большую роль в объяснении различий доходности акций играет тональность обсуждений: доходность сильнее всего реагирует на негативные сообщения, их число значимо влияет уже на уровне 1%. Направление влияния отдельных по тональности сообщений можно описать параболой с ветвями вниз: мы делаем вывод о том, что следование за толпой при высокой активности обсуждений компании не позволяет получить высокой доходности.

Еще более выражено положительное влияние нашего индикатора сентимента *HYPE* (на уровне 1%) на объем торгов. По сфокусированной платформе

MFD зависимость описывается параболой ветвями вниз: небольшое число сообщений и их негативный окрас отражают низкий интерес инвесторов к акциям и объем торгов по таким акциям ниже. Слишком активные обсуждения и слишком сильный позитивный настрой вызывают недоверие инвесторов, что также блокирует рост объема торгов. В отличие от платформы MFD индикатор *HYPE* на платформе Telegram положительно влияет на торговую активность, но его статистическая значимость невысока. Мы отмечаем, что выводы по платформе Telegram могут быть неокончательными. Это связано, во-первых, с большей сложностью обработки сообщений по сравнению с форумом MFD, так как на платформе Telegram нет столь однозначного разделения сообщений по компаниям (по веткам форума). Из-за большого числа сленговых выражений мы допускаем возможность ошибок по диагностированию «акция–тональность». Во-вторых, Telegram приобрел массовую популярность лишь в последние годы (2019–2020 гг.), до этого он долгое время находился в «серой» зоне в связи с законодательными ограничениями.

Увеличение числа как позитивных сообщений, так и негативных способствует росту торговой активности. Для малых компаний эффект сентимента более значимый, чем для крупных, что подтверждает постулаты поведенческих финансов для просроченных инвесторов.

ПРИЛОЖЕНИЕ

Таблица П1

Обзор предыдущих работ по теме исследования

Работа	Прокси-переменные сентимента инвесторов	Рынок, объект исследования, период тестирования	Методы анализа тональности и построения индекса сентимента	Методы тестирования гипотез	Влияние сентимента на доходность	Влияние сентимента на объем торгов
Baker, Wurgler, 2006	Разность между СЧА закрытых фондов и рыночной стоимостью паев, торговый оборот акций на бирже NYSE, доля акций в новых размещениях акций и облигаций, дивидендная премия, число IPO и их доходность в первый день торгов	США, акции, 1962–2001 гг.	Метод главных компонент	Регрессионный анализ	Чем выше сентимент, тем ниже доходность акций новых, небольших или финансово неустойчивых компаний, а также доходность волатильных акций	–

Продолжение таблицы П1

Работа	Прокси-переменные сентимента инвесторов	Рынок, объект исследования, период тестирования	Методы анализа тональности и построения индекса сентимента	Методы тестирования гипотез	Влияние сентимента на доходность	Влияние сентимента на объем торгов
Tetlock, 2007	Тональность текста в рубрике «Abreast of the Market» в газете «Wall Street Journal»	США, индекс Dow-Jones, 1984–1999 гг.	Метод главных компонент, применение словарей для анализа тональности текста	Регрессионный анализ (VAR-модели)	Чем выше индекс пессимизма инвесторов, тем ниже доходность акций на следующий день. С 5-дневным лагом наблюдается обратный эффект: высокий пессимизм приводит к развороту на падающем рынке	Экстремальные значения пессимизма и оптимизма приводят к росту объема торгов
Oliveira, Cortez, Areal, 2013	Число сообщений и его скользящее среднее, индекс позитивной тональности, индекс сентимента и его прирост на основе данных с платформы StockTwits	США, 5 высоколиквидных компаний и индекс S&P 500, 2010–2012 гг.	Анализ ключевых слов, расчет соотношений между числом позитивных и негативных сообщений	Регрессионный анализ (многофакторные линейные регрессии)	Взаимосвязь не обнаружена	Сентимент частных инвесторов объясняет различия в будущем объеме торгов большинства активов
Oliveira, Cortez, Areal, 2017	Тональность и число сообщений в Twitter, индикаторы тональности обзоров в Datastream	США, индексы широкого рынка и отдельных секторов, портфели акций, факторы модели Fama-French, 2012–2015 гг.	Применение словарей для анализа тональности текста, фильтр Калмана	Регрессионный анализ и методы машинного обучения	Сентимент частных инвесторов позволяет спрогнозировать доходность S&P 500, ряда отраслевых индексов, портфелей акций компаний малой капитализации	Сентимент частных инвесторов позволяет спрогнозировать объем торгов акций в составе индексов S&P 500 и DJIA
Renault, 2017	Тональность сообщений на платформе StockTwits	США и развивающиеся страны, 11 ETF акций и облигаций, 2012–2016 гг.	Использование словарей и методов машинного обучения для классификации сообщений по тональности	Регрессионный анализ (линейные многофакторные регрессионные модели с включением лагов переменных)	Сентимент частных инвесторов позволяет спрогнозировать внутридневную доходность большинства ETF акций	–

Окончание таблицы П1

Работа	Прокси-переменные сентимента инвесторов	Рынок, объект исследования, период тестирования	Методы анализа тональности и построения индекса сентимента	Методы тестирования гипотез	Влияние сентимента на доходность	Влияние сентимента на объем торгов
Al-Nasseri, Ali, 2018	Дивергенция мнений на платформе StockTwits	США, 30 высоколиквидных акций, входивших в индекс Dow-Jones, 2012–2013 гг.	Методы машинного обучения для классификации сообщений по тональности	Регрессионный анализ (линейные многофакторные регрессионные модели и модели с включением мультипликативных переменных)	Высокая дивергенция мнений негативно влияет на доходность на растущем рынке и положительно на падающем	Высокая дивергенция предшествует повышенному объему торгов и на растущем, и на падающем рынке
Huang, Rojasand, Convery, 2019	Объем поисковых запросов в GoogleTrends	США, индекс S&P 500, 2004–2017 гг.	Значение единственной прокси-переменной	Регрессионный анализ (тесты Грейнджера, SVM, Ridge regressions и др.)	Сентимент в GoogleTrends позволяет спрогнозировать доходность S&P 500	–

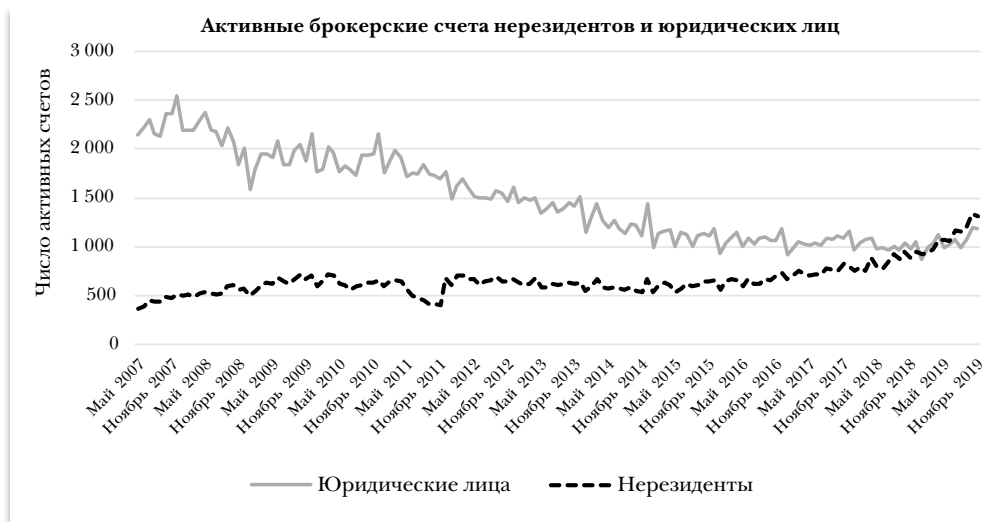
Источник: составлено авторами на основе обзора работ.



Рис. П1

Динамика индекса Мосбиржи и числа активных счетов физических лиц

Источник: МосБиржа.

**Рис. П2**

Динамика активных счетов нерезидентов и юридических лиц

Источник: данные Мосбиржи.

Таблица П2

Матрицы ошибок для обучающей и тестовой выборок

Класс сообщений		Спрогнозированные (Predicted)		
		негативные	нейтральные	позитивные
1. OneHot Encoding. SGD. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	2328	920	310
	нейтральные	937	16 067	1184
	позитивные	421	983	2157
2. OneHot Encoding. SGD. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	386	384	120
	нейтральные	303	3878	366
	позитивные	143	402	345
3. OneHot Encoding. Дерево решений. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	3409	71	78
	нейтральные	2421	13 285	2482
	позитивные	334	57	3170
4. OneHot Encoding. Дерево решений. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	481	216	193
	нейтральные	934	2633	980
	позитивные	269	253	368
5. OneHot Encoding. Случайный лес. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	3449	103	6
	нейтральные	68	18 033	87
	позитивные	25	107	3429

Окончание таблицы П2

Класс сообщений		Спрогнозированные (Predicted)		
		негативные	нейтральные	позитивные
6. OneHot Encoding. Случайный лес. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	333	496	61
	нейтральные	187	4108	252
	позитивные	97	578	215
7. TF-IDF. SGD. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	1729	1448	381
	нейтральные	786	16 091	1311
	позитивные	442	1466	1653
8. TF-IDF. SGD. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	360	412	118
	нейтральные	248	3896	403
	позитивные	128	405	357
9. TF-IDF. Дерево решений. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	3519	28	11
	нейтральные	1291	15 652	1245
	позитивные	185	17	3359
10. TF-IDF. Дерево решений. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	391	320	179
	нейтральные	639	3182	726
	позитивные	229	350	311
11. TF-IDF. Случайный лес. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	3465	88	5
	нейтральные	38	18 105	45
	позитивные	14	99	3448
12. TF-IDF. Случайный лес. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	297	548	45
	нейтральные	150	4252	145
	позитивные	91	643	156
13. CNN. Матрица ошибок для обучающей выборки				
Действительные (actual)	негативные	2459	819	724
	нейтральные	2245	13 921	2772
	позитивные	1218	960	1882
14. CNN. Матрица ошибок для тестовой выборки				
Действительные (actual)	негативные	222	88	70
	нейтральные	130	1620	450
	позитивные	159	20	241

ЛИТЕРАТУРА / REFERENCES

- Анкудинов А.Б., Лебедев О.В.** (2014). Детерминанты создания акционерной стоимости российскими компаниями // *Корпоративные финансы*. № 29 (1). Р. 32–45. [Ankudinov A.B., Lebedev O.V. (2014). Determinants of creating shareholder value by Russian companies. *Corporate Finance*, 29 (1), 32–45 (in Russian).]
- Ahmad K., Han J.G., Hutson E., Kearney C., Liu S.** (2015). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, 37 (C), 152–172.
- Ali S., Liu B., Su J.J.** (2017). Corporate governance and stock liquidity dimensions: Panevidence from pure order-driven Australian market. *International Review of Economics and Finance*, 50, 275–304.
- Al-Nasseri A., Ali F.M.** (2018). What does investors' online divergence of opinion tell us about stock returns and trading volume? *Journal of Business Research*, 86, 166–178.
- Andreou P.C., Antoniou C., Horton J., Louca C.** (2016). Corporate governance and firm-specific stock price crashes. *European Financial Management*, 22 (5), 916–956.
- Asle H.G., Valahzaghari M., Ahranjani B.** (2013). A survey on the relationship between stock liquidity with firm performance: A case study of Tehran stock exchange. *Management Science Letters*, 3 (2), 635–640.
- Baker M., Wurgler J.** (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, LXI (4), 1645–1680.
- Brown J.H., Crocker D.K., Foerster S.R.** (2009). Trading volume and stock investments. *Financial Analysts Journal*, 65 (2), 67–84.
- Caporale G.M., Spagnolo F., Spagnolo N.** (2016). Macro news and stock returns in the Euro area: A VAR-GARCH-in-mean analysis. *Journal of Behavioral and Experimental Finance*, 45, 180–188.
- Celebi K., Honig M.** (2019). The impact of macroeconomic factors on the German stock market: Evidence for the crisis, pre- and post-crisis periods. *International Journal of Financial Studies*, 7 (18), 1–13.
- Chen X., Shangguan W., Liu Y., Wang S.** (2020). Can network structure predict cross-sectional stock returns? Evidence from co-attention networks in China. *Finance Research Letters*. In Press.
- Debata B., Mahakud J.** (2018). Interdependence between monetary policy and stock liquidity: A panel VAR approach. *Margin*, 12 (4), 387–413.
- Huang M.Y., Rojas R.R., Convery P.D.** (2020). Forecasting stock market movements using Google trend searches. *Empirical Economics*, 59, 2821–2839.
- Jegadeesh N., Wu D.** (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*, 110 (3), 712–729.
- Khan J., Rehman S.U.** (2020). Impact of corporate governance compliance and financial crisis on stock liquidity: Evidence from Pakistan. *Journal of Accounting and Finance in Emerging Economies*, 6 (2), 475–486.
- Loughran T., McDonald B.** (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66 (1), 35–65.
- Mai V.A.V., Chuan A.T., Fang V.** (2016). Aggregate volatility risk and the cross-section of stock returns: Australian evidence. *Pacific-Basin Finance Journal*, 36, 134–149.
- Nurazi R., Kananlua P.S., Usman B.** (2015). The effect of Google trend as determinant of return and liquidity in Indonesia stock exchange. *Jurnal Pengurusan*, 45, 131–142.
- Oliveira N., Cortez P., Areal N.** (2013). On the predictability of stock market behavior using

- StockTwits sentiment and posting volume. *EPIA 2013: Progress in Artificial Intelligence*, 355–365.
- Oliveira N., Cortez P., Areal N.** (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, 73, 125–144.
- Renault T.** (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking and Finance*, 84, 25–40.
- Rostami S., Rostami Z., Kohansala S.** (2016). The effect of corporate governance components on return on assets and stock return of companies listed in Tehran stock exchange. *Procedia Economics and Finance*, 36, 137–146.
- Salamat W.** (2016). Macroeconomics, firm-specific factors and stock liquidity: An empirical evidence from Jordan. *International Journal of Financial Research*, 7 (5), 110–123.
- Sidhu M.K., Kaur P.** (2019). Effect of corporate governance on stock market liquidity: Empirical evidence from Indian companies. *Decision*, 46 (3), 197–218.
- Sparck J.K.** (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28, 11–21.
- Sun L., Najand M., Shen J.** (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking and Finance*, 73, 147–164.
- Tayeh M.** (2016). Determinants of market liquidity: Evidence from the Jordanian stock market. *International Journal of Economics and Finance*, 8 (10), 48–59.
- Tetlock P.C.** (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, LXII (3), 1139–1168.
- Wang J.-H., Liu T.-W., Luo X., Wang L.** (2018). An LSTM approach to short text sentiment classification with word embeddings. *Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018)*, 214–223. Available at: <https://aclanthology.org/O18-1021>
- Yang Y., Cer D., Ahmad A., Guo M., Law J., Constant N., Abrego G., Yuan S., Tar C., Sung Y., Strophe B., Kurzweil R.** (2019). Multilingual universal sentence encoder for semantic retrieval. arXiv: 1907.04307v1 [cs.CL]. 9 Jul 2019. Available at: <https://arxiv.org/pdf/1907.04307.pdf>

Поступила в редакцию 15.07.2020

Received 15.07.2020

T.V. Teplova
HSE University, Moscow, Russia

T.V. Sokolova
HSE University, Moscow, Russia

A.F. Tomtosov
HSE University, Moscow, Russia

D.V. Buchko
HSE University, Moscow, Russia

D.D. Nikulin
HSE University, Moscow, Russia

The sentiment of private investors in explaining the differences in the trade characteristics of the Russian market stocks⁶

Abstract. In our paper, for the first time, we examine the influence of the sentiment of private investors in social networks on the trade characteristics of stocks in the Russian market. Monthly return rates and trading volumes are analyzed under the control of financial indicators and indicators of the quality of corporate governance of stock issuers, as well as the changing external environment in the period from 2013 to 2020. The sample for various sentiment metrics is based on unique data: messages in the Telegram and mfd.ru platforms. The tonality of messages is diagnosed according to the authors' method using artificial intelligence (neural network). The main conclusion is: the sentiment can be seen as an explanatory factor in pricing and trading activity. The influence of sentiment is non-linear. The author's HYPE indicator of sentiment is proposed and compared in terms of explanatory ability of the trade characteristics with a wide range of proxy variables. The explanatory ability to identify differences is realized through regression constructions on panel data. It is shown that trade characteristics are more sensitive to the growth of negative messages, which is consistent with the postulates of behavioral finance. An increase in messages' number of both positive and negative sentiment contributes to the growth of trading activity. An important practical conclusion is: following the crowd when the company is most intensely discussed will not result in high returns to an investor.

Keywords: *mood; text processing; investor sentiment; tone of messages; neural networks; stock returns; trading activity.*

JEL Classification: G11, G12, G41.

For reference: **Teplova T.V., Sokolova T.V., Tomtosov A.F., Buchko D.V., Nikulin D.D.** (2022). The sentiment of private investors in explaining the differences in the trade characteristics of the Russian market stocks. *Journal of the New Economic Association*, 1 (53), 53–84. DOI: 10.31737/2221-2264-2022-53-1-3

⁶ The publication was supported by the grant for research centers in the field of AI provided by the Analytical Center for the Government of the Russian Federation (ACRF) in accordance with the agreement on the provision of subsidies (identifier of the agreement 000000D730321P5Q0002) and the agreement with HSE University 70-2021-00139.