

ФИНАНСОВАЯ ЭКОНОМИКА

Е. А. Федорова¹,

Финансовый университет

при Правительстве РФ /

Национальный исследовательский университет

«Высшая школа экономики» (Москва, Россия)

О. Ю. Рогов²,

Финансовый университет

при Правительстве РФ (Москва, Россия)

В. А. Ключников³,

Национальный исследовательский университет

«Высшая школа экономики» (Москва, Россия)

ВЛИЯНИЕ НОВОСТЕЙ НА ИНДЕКС НЕФТЕГАЗОВОЙ ОТРАСЛИ ММВБ: ТЕКСТОВЫЙ АНАЛИЗ⁴

В данном исследовании выявлена зависимость между тональностью новостей и откликом фондового индекса нефтегазовой отрасли Российской Федерации. Эмпирическая база исследования включала 8,5 млн новостей из зарубежных источников. Методология исследования: нечеткие множества, наивный байесовский классификатор, коэффициент корреляции Пирсона. В результате исследования было выявлено: 1) негативные новости влияют сильнее позитивных на фондовый индекс; 2) новости по компаниям влияют на значение индекса, а новости по отрасли влияют на объем торгов; 3) санкции не повлияли существенно на освещение российских нефтегазовых компаний.

Ключевые слова: прогнозирование цен акций, анализ тональности текста, поведенческие финансы.

THE IMPACT OF NEWS ON THE MICEX OIL & GAS INDEX: TEXTUAL ANALYSIS

In this study, a relationship between the mood of news and the response of the oil and gas industry index of the Russian Federation was revealed. The empirical base of the study included 8.5 million news from foreign sources. Research methodology: fuzzy sets, naive Bayesian classifier, Pearson correlation coefficient. As a result of the research, it was discovered

¹ Федорова Елена Анатольевна, д.э.н., профессор; e-mail: ecolena@mail.ru

² Рогов Олег Юрьевич, соискатель; e-mail: fintech@gmx.ch

³ Ключников Валерий Андреевич, студент; e-mail: vaklyuchnikov@edu.hse.ru

⁴ Статья подготовлена по Госзаданию Правительства РФ 2018 г.

that: 1) negative news affects the stronger than the positive on the stock index; 2) news on companies affect the value of the index, and news on the industry affect the volume of trading; 3) the sanctions did not significantly affect the coverage of Russian oil and gas companies.

Key words: stock price forecasting, textual analysis, behavioral finance.

Введение

Наиболее сложным фактором с точки зрения прогнозирования движения рынка была и остается психология его участников. Количество и скорость поступления информации, доступной инвестору, увеличиваются с каждым днем, и даже профессионалам становится все сложнее реагировать на нее быстро и рационально. Помимо этого наблюдаются рост доли непрофессиональных инвесторов на российском фондовом рынке, кризисные явления в российской экономике и напряженность на международной политической арене. Все это делает фондовый рынок еще более восприимчивым к эмоциональному, а не фундаментальному содержанию новостей. По этой причине на первый план выходят исследования, посвященные изучению факторов, влияющих на настроение инвестора и его решения.

Одним из решений этой проблемы может стать сентимент-анализ, или анализ тональности новостей (sentiment analysis). В обзорной статье по сентимент-анализу¹ под сентимент-анализом понимается раздел компьютерной лингвистики, в котором рассматриваются задачи определения численного значения эмоциональной окраски текстов. В данной работе мы сужаем область сентимент-анализа до оценки эмоциональной окраски новостей из информационной системы Thomson Reuters на английском языке. Как будет показано ниже, подобная практика часто используется в работах западных исследователей, так как Thomson Reuters — один из основных поставщиков информации для инвесторов на финансовом рынке.

Целью исследования является оценка влияния позитивной или негативной тональности новостей мировой нефтегазовой отрасли и новостей по нефтегазовым компаниям на характеристики фондового индекса нефтегазовой отрасли ММВБ (цена, волатильность и объем торгов). Практическая направленность включает использование полученных результатов инвесторами, которые вкладывают средства в ценные бумаги на краткосрочные и среднесрочные периоды.

В нашем исследовании затрагивается еще один аспект, связанный с санкциями и освещением этих санкций в новостях, в том числе и для компаний нефтегазового сектора. Взаимосвязь влияния экономических санкций одних государств по отношению к другим достаточно подробно

¹ Boudad N. et al. Sentiment analysis in Arabic: A review of the literature // Ain Shams Engineering Journal xxx (2017) xxx–xxx

изучена в экономической литературе. Тем не менее можно выявить некоторые направления, которые в настоящий момент недостаточно проработаны, в том числе и на основе применения текстового анализа. Существует немного работ, которые направлены на изучение влияния санкций в контексте текстового анализа. Например, санкции в отношении Северной Кореи рассматриваются в работе Y. J. Kim, который исследует их влияние на характер высказываний руководства страны. Авторы устанавливают взаимосвязь между содержанием и эмоциональной окраской официальных заявлений руководства и ужесточением санкций¹. Исследование проводится на широкой выборке, охватывающей период с 1949 по 2010 г. М. Hoffmann и М. Neuenkirch анализируют влияние конфликта между Россией и Украиной на биржевую прибыль российского и украинского фондового рынка. Авторами разрабатывается новый индикатор эскалации (деэскалации) конфликта, основанный на степени напряженности текста новостей, относящихся к данной тематике. При этом в качестве одной из переменных выступает частота упоминания санкций в новостях. Авторы приходят к выводу, что усиление конфликта сокращает прибыль, получаемую на российской и украинской фондовых биржах. В нашем исследовании мы оцениваем влияние позитивной или негативной тональности новостей мировой нефтегазовой отрасли и новостей по нефтегазовым компаниям на характеристики фондового индекса нефтегазовой отрасли ММВБ (цена, волатильность и объем торгов) до введения санкций и после введения санкций.

Наше исследование состоит из нескольких частей: обзор литературы и гипотезы исследования, методология исследования, анализа данных, результаты и дискуссия.

Обзор литературы и гипотезы исследования

Для достижения цели исследования на основании обзора литературы были сформулированы следующие гипотезы.

Гипотеза 1. *Существует зависимость между тональностью новостей о компании и реакцией рынка (индекса). Рынок сильнее реагирует на негативные новости, чем на позитивные.*

К настоящему моменту поведение целого ряда финансовых индикаторов: цен акций и биржевых индексов, их доходностей, объема торгов и рыночной волатильности — объясняется с использованием анализа тональности текстовой информации. Источники информации о настроении бывают самые различные, например, помимо обычных новостей это могут

¹ Kim Y.J. Economic sanctions and the rhetorical responses of totalitarian regimes: Examining North Korean rhetorical strategies, 1949–2010 // Communist and Post-Communist Studies. — 2014. — Т. 47. — No. 2. — S. 159–169.

быть: социальные сети, поисковые системы, блоги. Наиболее широко используемым источником информации о рынке и наиболее релевантным источником сентиментов для компаний нефтегазовой отрасли являются новости [Qing Li et al., 2014]. На многих развитых рынках капитала доказано, что зависимость между тональностью новостей и поведением акций нефтегазовых компаний существует. Гипотеза строится по аналогии с гипотезами в работах [Tetlock, 2007; Gilbert, Karahalios, 2009; Oliveira et al., 2013; Porshnev et al., 2016].

В ряде работ было показано, что пессимизм новостей влияет на рынок дольше и сильнее, чем оптимизм, вследствие неприятия риска инвестором [Veronesi, 1999; Moat et al., 2013]. Различия в реакции инвесторов на новости отмечаются и в работе Чекли [Checkley, 2017]. В данном исследовании оценивался лаг между настроением инвесторов и характеристиками фондового рынка. Важным выводом из этой работы является тот факт, что рынок быстрее учитывает в цене негативную информацию, чем позитивную. По мнению авторов, это также является следствием неприятия риска инвестором. Также в ряде исследований оценивается влияние новостей на волатильность фондового индекса. В исследовании А. Atkins et al.¹ было выявлено, что движение волатильности более предсказуемо, чем движение цен на активы при использовании финансовых новостей.

***Гипотеза 2.** На акции компании влияют не только новости по компании, но и отраслевая специфика. Мы считаем, что существует зависимость между тональностью новостей по характеристике отрасли (нефть) и реакцией рынка.*

Российская нефтегазовая отрасль сильно зависит от экспорта и мировых цен на сырье, а значит — от хода переговоров и заключения сделок о поставках нефти и газа за рубеж и решений международных организаций, определяющих структуру отраслевых рынков. В частности, в работе Лутиа было показано, насколько сильно решения ОПЕК влияют на цену нефти, а значит, и на цены акций компаний отрасли [Loutia et al., 2016]. Чтобы учесть «глобальные» настроения, предлагается проверка гипотезы для более широкой выборки новостей. Ожидается, что новости с мировых отраслевых рынков и глобальная рыночная конъюнктура влияют на цены акций на российском рынке наравне с новостями о компаниях.

***Гипотеза 3.** Политические шоки влияют на зависимость между тональностью новостей и реакцией рынка.*

Часть исследований посвящена влиянию политической неопределенности (революции, восстания) на поведение финансовых рынков, в большинстве из них предполагается, что политическая неопределенность влияет на волатильность, доходность фондовых рынков и финансовую

¹ Atkins A. et al. Financial news predicts stock market volatility better than close price // The Journal of Finance and Data Science. — 2018. — 4. — 120–137.

интеграцию. В статье Трабелси рассматривается влияние политической неопределенности, вызванной гражданскими восстаниями (тунисская революция), на поведение и характеристики тунисского фондового рынка на различных временных периодах [Trabelsi, 2017]. В период после гражданских восстаний действительно нарастала амплитуда и волатильность циклов тунисского фондового рынка. Но в долгосрочной перспективе амплитуда и изменчивость биржевых циклов амортизируются до более низкого уровня. Исследование Шо посвящено оценке влияния политической неопределенности (вызванной гражданскими восстаниями в арабском мире, т.е. «арабской весной») на волатильность фондовых рынков стран Ближнего Востока и Северной Африки [Chau et al., 2014]. В этой же работе было выявлено, что политическая неопределенность способствует финансовой нестабильности. В одном исследовании также рассматриваются последствия демократии и политического риска для 49 фондовых рынков [Lehkonen, Neimonen, 2015]. Было выявлено, что демократия и политический риск оказывают влияние на доходность фондового рынка.

В качестве исследуемого периода со значительной политической напряженностью был выбран крымский кризис. В середине марта 2014 г., после того как Россия признала итоги крымского референдума, поддержала одностороннее провозглашение независимости и Республика Крым приняла ее предложение о вхождении в состав Российской Федерации, США, Евросоюз, Австралия, Новая Зеландия и Канада ввели в действие первый пакет санкций. В соответствии с выводами Лотт и Хассетт, ожидается, что введение санкций повлияло и на тон финансовых новостей [Lott, Hassett, 2012]. Мы также предполагаем, что после введения санкций новостной фон стал сильнее влиять на колебания фондового рынка. Несмотря на заявленную объективность бизнес-новостей, СМИ склонны представлять искаженную картину событий. Появление более полярных оценок должно было сказаться на большей волатильности российского рынка после введения санкций.

Также нельзя не обратить внимания на работу Анкутдинова и др., непосредственно посвященную этим событиям [Ankudinov et al., 2017]. Авторы статьи отмечают, что российский фондовый рынок является достаточно волатильным даже в отсутствие политических шоков, что характеризуется «широкими» хвостами в функциях распределения доходностей акций, обращаемых на Московской бирже. Так же как и в настоящей статье, в работе затрагивается влияние санкций на российский фондовый рынок и отмечается, что, как и ожидалось, большинство отраслевых индексов стали более волатильными после описанных выше событий. Однако авторы заявляют, что отделить политическую напряженность, оказывавшую давление на рынок, от общего спада на сырьевых рынках невозможно. В данном случае, пишут они, нельзя однозначно заявить о силе влияния политики на рынок. Ожидается, что использование сентимент-

анализа позволить разделить влияние новостей, описывающих ситуацию на сырьевом рынке, и новостей, касающихся непосредственно самих российских компаний.

Методология исследования

Методология исследования основана на последовательном применении двух этапов: сбор и предварительная обработка данных текстов новостей и применение программных методов оценки эмоциональной окраски с помощью специальных словарей финансовых терминов. Блок-схема, иллюстрирующая работу алгоритма, представлена на рис. 1.

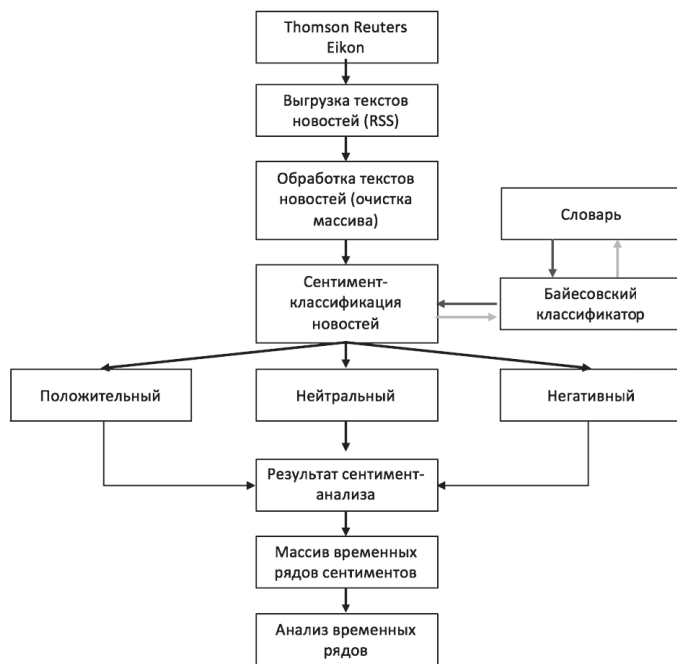


Рис. 1. Схема проведенного исследования с использованием словаря сентиментов

Первый этап заключается в загрузке данных новостей с сайтов информационных агентств Thomson Reuters. Особенностью данного источника является полнота информации по различным тематикам: база данных ежегодно пополняется более чем миллионом текстовых новостей из 200 различных авторизованных системой Thomson Reuters источников. Thomson Reuters Connect обеспечивает бесперебойный доступ к массиву данных через протоколы RSS и FTP. Новости размещаются в базе данных в едином

формате (дата публикации, время публикации, текст новости), что существенно упрощает работу с большими объемами текстовой информации.

Предварительная обработка включает в себя очистку массива от нерелевантных новостей: удаляются тексты, не содержащие информацию о российских компаниях из заданного списка. Затем каждая новость разбивается на предложения. Каждое предложение может быть проанализировано как группа слов или словосочетаний, для которых может быть найдено соответствие числовому значению по сентимент-шкале из выбранного словаря.

На втором этапе для анализа эмоциональной окраски применяется так называемый наивный байесовский классификатор. Алгоритм обращается к словарю, сформированному с использованием алгоритмов машинного обучения по правилам, установленным составителями. Каждому слову классификатор присваивает известное значение по шкале соответствия, которое в дальнейшем используется для формирования временных рядов сентиментов. Построенные временные ряды могут быть проанализированы для поиска причинности между величиной спроса на акции отдельных компаний и эмоциональной окраской публикуемых новостей¹.

Новизна используемых методов накладывает ряд ограничений на исследование, что также подробно рассказано в соответствующем разделе.

Ограничения исследования. Главное ограничение данного исследования — предположение о том, что новости на русском языке от российских поставщиков финансовой информации содержательно не отличаются от используемых в данной работе и выходят в течение того же торгового дня. Это важное допущение необходимо в виду того, что необходимые новости и инструменты сентимент-анализа доступны только на английском языке. Также важно отметить, что используемый в работе индекс нефтегазовой отрасли ММВБ не является самостоятельным торговым инструментом, однако, так как механизм его расчета известен, его можно легко воспроизвести в каждый момент времени, используя акции, входящие в него. Другим ограничением является использование «мешка слов» — в дальнейшем предполагается использовать более продвинутые методы анализа тональности текста, учитывающие контекст, а не просто эмоциональную окраску отдельных слов.

Рассматриваемые периоды времени отличаются по продолжительности и соответственно содержат различное количество новостей. Поэтому при вычислении частоты использования терминов показатели нормируются, а рассматриваемые «окна» для временных рядов при анализе вли-

¹ Manav M., Poornalatha G. Twitter Sentiment Analysis Using a Modified Naïve Bayes Algorithm, ISAT 2017: Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology. — ISAT, 2017. — P. 171–181.

яния по продолжительности не превышают месяца и несравнимы с временными рамками каждого из подпериодов, выбранных для анализа, составляющих семь и три года соответственно.

Хотелось бы отметить, что в нашем исследовании мы оцениваем взаимосвязь не только полученного ряда негативной и позитивной окраски новостей и фондового индекса, но и волатильности фондового индекса.

Условная волатильность рассчитывается из общего уравнения GARCH-модели. Спецификация модели рассчитывалась по следующей формуле:

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

где q и p — значения временных лагов для GARCH и ARCH-компонент соответственно.

Спецификация модели GARCH определяется на основе критериев Акаике и Шварца. Мы использовали GARCH-модели с $p = 1$ и $q = 1$: GARCH (1,1). Волатильность рынка рассчитывалась как условная волатильность, аналогично исследованиям P. Singh et al.¹, F. M. Jaleel, L. P. Samarakoon² и др.

Текстовый анализ

Общая схема исследования в сфере сентимент-анализа, как правило, имеет следующий вид: первый этап — определение вида данных и их получение; второй этап — предварительная обработка «сырых» данных и вычисление сентимент-фактора (безразмерной величины, рассчитанную с помощью «мешка слов» и отражающей эмоциональную окраску новостей в выбранный день); третий этап — использование сентимент-фактора для ответа на вопросы, перечисленные в цели исследования. По этой же логике и описывается методология данной работы: в начале данного раздела обосновывается выбор новостей как источника сентиментов, после чего раскрывается специфика применяемого в данной работе текстового анализа, и затем описываются экономико-математические инструменты, задействованные для проверки гипотез исследования, перечисленных выше.

Из всех потенциальных источников текстовой информации, влияющих на настроение рынка, были выбраны именно новости, так как анализ их влияния наиболее широко изучен и удобен для сентимент-

¹ Singh P. et al. Price and volatility spillovers across North American, European and Asian stock markets // International Review of Financial Analysis. — 2010. — 19. — 55–64.

² Jaleel F. M., Samarakoon L. P. Journal of Multinational Financial Management. — 2009. — 19. — 409–423.

анализа цен акций нефтегазовой отрасли. В качестве поставщика информации был выбран Thomson Reuters, так как новости из Thomson Reuters уже использовались для оценки рыночных настроений и их содержание было признано отвечающим требованиям sentiment-анализа для изучения влияния новостного фона на фондовый рынок в работе Хестон и Синха [Heston, Sinha, 2016]; они представляли наиболее полную и актуальную информацию, доступную о рынке в каждый момент времени.

В основе алгоритма отбора новостей в соответствии с их тональностью применяется кластерный анализ с использованием векторных моделей. В качестве алгоритма классификации задействован наивный байесовский классификатор — вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости. Под классификацией текста понимается задача определения категории, к которой принадлежит ранее не встречавшийся пример, для элементов задаваемого обучающего множества эти категории известны.

Для дальнейшей проверки гипотез исследования также необходим корпусный анализ текста новостей на основе статистики их появления в тексте (частота появления слова). Для вычисления частоты слов используется отношение числа вхождений некоторого термина к общему числу слов новости. Таким образом, оценивается весовое положение слова в пределах отдельной новости или всех текстов новостей. За значение «прямой частоты» TF (Term Frequency) принимают:

$$f_1(t, D) = \frac{N_{w_k}}{\sum_{w_k} N_{w_i}}.$$

Обратная частота — инверсия частоты f , с которой некоторое слово встречается в массиве новостей [Jones, 2004].

$$f_2(t, D) = \log \frac{|D|}{|D_i \in D | t \in D_i|},$$

где $|D|$ — количество текстов (новостей); $|D_i \in D | t \in D_i|$ — общее число слов число документов из D , в которых найден t (при $n_w \neq 0$).

Учет f_2 позволяет снизить влияние слов широкого употребления в текстах. Для каждого уникального слова в пределах конкретного набора новостей существует только одно значение IDF (Inverted Document Frequency).

На следующем этапе проводился анализ тональности новостей. На сегодняшний день уже существуют готовые решения в области sentiment-анализа, разработанные и распространяемые крупнейшими поставщиками финансовой информации, например: Thomson Reuters «MarketPsych»

и Bloomberg. Также доступны несколько более узкоспециализированных и бюджетных программ для, например, отслеживания эмоционального фона в Твиттере [«Pulse of the Nation»] или для продвинутого анализа финансовой отчетности от компании LIWC. Описанное выше программное обеспечение пользуется спросом со стороны профессиональных трейдеров и исследователей в области финансов. Однако для данной работы было принято решение не использовать описанные выше продукты, во-первых, ввиду их высокой стоимости и, во-вторых, по причине отсутствия у них открытого исходного кода и технической документации, подробно описывающей процесс получения сентиментов — что сделало бы процесс исследования «черным ящиком».

Наиболее широко распространенными инструментами определения тональности текстов являются «мешок слов» (bag-of-words), «именные группы» (noun phrases) и ряд других. Специфика применения «мешка слов» заключается в необходимости использования специального словаря, в котором экспертным мнением и/или с использованием машинного обучения каждому слову присваивается некоторая тональность. МакДональд и Лоугрэн в своей работе показали существенное преимущество комбинированного метода обучения с использованием специализированного словаря перед другими инструментами текстового анализа [Loughran, McDonald, 2011]. В ряде работ была показана высокая эффективность метода сентимент-анализа новостей, базирующегося на статистике использованных слов в применении к обработке большого объема (не менее 10 тыс.) финансовых новостных блоков [Loughran et al., 2015; Kearney, Liu, 2014]. Авторами применяемой в данном исследовании методики был составлен словарь из более 300 «положительных», 1200 «отрицательных» новостей на основе линейной шкалы измерения тональности анализируемых терминов — словам, обычно связанным с отрицательными, нейтральными или позитивными тональностями, ставились в соответствие числа по шкале от -10 до 10 (от самого отрицательного к самому положительному). Сначала фрагмент неструктурированного текста исследовался с помощью инструментов и алгоритмов обработки естественного языка, а затем выделенные из этого текста объекты и термины анализировались с целью понимания значения этих слов. Для наиболее точного сопоставления каждого слова в тексте выполняются корпусный и синтаксический анализ. На настоящий момент словарь постоянно пополняется и меняет свою структуру, реагируя таким образом на динамично развивающийся язык — что делает его одним из лучших решений для проведения исследования.

Процесс работы с «мешком слов» в данном исследовании: сначала удаляются так называемые «стоп-слова» (stop-words: например, артикли (в английском языке), цифры и знаки препинания), все буквы в тексте приводятся к единому регистру, текст раскладывается на отдельные слова,

а синтаксис и сведения о связях между словами — игнорируются. В итоге новость представляется в виде множества слов, и далее оцениваются свойства отдельно взятых слов, а не всего текста в целом. Подобным образом уже проводилось исследование финансовых текстов, например, в статье [Федорова, 2016]. В данном исследовании сначала были выбраны слова, относящиеся к инвестиционному процессу (слова «инвестиции», «инвестор» и «страны-инвесторы» (Китай, Германия, Франция, Италия и США). Потом были рассчитаны коэффициенты корреляции между выбранным словом и блоком слов с «положительной» или «отрицательной» окраской для каждой новости.

После того как каждая предварительно отобранная новость преобразовалась в вектор слов («мешок слов»), ей присваивается значение, соответствующее тональности по каждому эмоциональному оттенку, представленному в словаре. В фокусе данной работы находятся наиболее важные с точки зрения силы влияния на рынок настроения — оптимизм и пессимизм инвесторов, поэтому тональность текста оценивается по признакам слов: «негативное» и «позитивное» слово (и для российских, и для мировых новостей). В работе используются ежедневные новости: тональность новостей за день рассчитывается как сумма значений тональностей по всем новостям за каждый торговый день.

В результате описанных манипуляций с текстами новостей получается сентимент-фактор — временной ряд ежедневных значений новостей по выбранной теме (отечественные или мировые новости рынка) и тональности (положительной или негативной). Этот временной ряд отражает «настроение новостей» и в дальнейшем используется в качестве объясняющей переменной.

Корреляция Пирсона

Для двух стационарных временных рядов u и v , элементы которых имеют временной лаг:

$$l = (k - 1)\Delta t,$$

кросс-корреляционная функция находится по формуле:

$$C_k = \sum_i u_i v_{k+i-1}.$$

И максимальное значение соответственно:

$$C'_k = \max_{0 \leq N \leq l} \sum_{i=0}^{N-1} u_i v_{k+i-1}.$$

Расчет этого показателя подразумевает запаздывание реакции рынка по отношению к новостям.

Корреляция Пирсона применяется для оценки взаимосвязи полученных рядов с позитивной и негативной информацией о состоянии рынка нефти (основными компаниями нефтяной отрасли) с характеристиками фондового индекса нефтегазовой отрасли ММВБ (цена, волатильность и объем торгов). Новости по мировому рынку формировались по новостям с упоминанием слов: «ОПЕК», «crude oil», «Brent» — что соответствует нефтяному картелю ОПЕК, нефти (вообще) и самому распространенному сорту нефти. Также новости формировались по следующим нефтяным компаниям — «Rosneft», «ROSN», «Gazprom», «GAZP», «Lukoil», «LKOH», «Surgutneftegaz», «SNGS».

Описание данных

Эмпирическая база исследования включает нефтегазовый индекс ММВБ, цену открытия, закрытия, волатильность и объем торгов. В работе использовались ежедневные котировки по следующим периодам: 1-й подпериод с 01.01.2007 по 13.02.2014 и 2-й подпериод с 14.02.2014 по 17.08.2016.

Таблица 1

Описательная статистика по переменным

	VOLRUB	POSOIL	POSCOM	NEGOIL	NEGCOM	CLOSE
Mean	1,83E+10	5,298541	4,403135	8,024728	5,471226	3067,264
Median	1,61E+10	5,080000	1,030000	7,570000	0,450000	3034,270
Maximum	8,09E+10	32,04000	33,42000	27,54000	34,49000	5290,090
Minimum	0,000000	1,030000	1,030000	0,450000	0,450000	883,8000
Std.Dev.	9,51E+09	2,773639	5,117621	3,845862	6,223331	876,5878
Skewness	1,545104	1,228485	1,455377	1,054457	1,157663	0,290625
Kurtosis	7,122959	8,907992	4,915889	5,557823	4,094809	3,144915
Jarque-Bera	3556,350	5484,397	1626,672	1472,199	878,6783	48,07122

Тональность новостей за день считается как средняя по всем новостям за этот день. Расчеты по этому показателю приводятся на рис. 1.

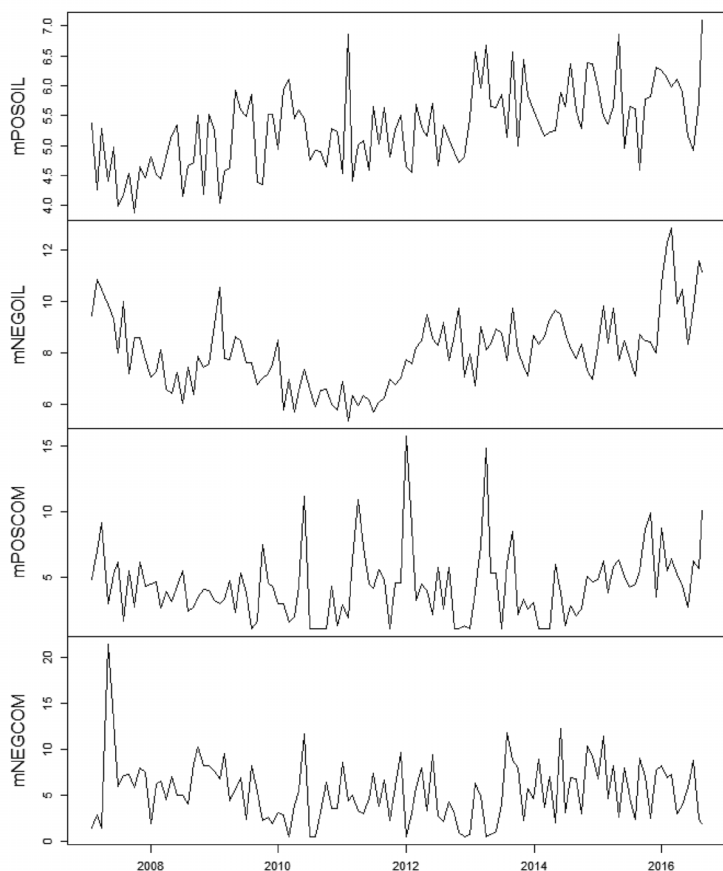


Рис. 1. Ежемесячные значения тональностей новостей о ситуации на рынке нефти (mPOSIL, mNEGOIL) и по wybranым российским компаниям (mPOSCOM, mNEGCOM)

Эмпирические результаты исследования

Результаты исследования представлены в табл. 2.

Таблица 2

Корреляция по Пирсону между исследуемыми показателями

	POSIL	NEGOIL	POSCOM	NEGCOM	CLOSE	VOLRUB	VLTY	RETS
POSIL	1,000	−0,115	0,026	0,003	0,109	−0,072	−0,015	0,023
NEGOIL	−0,115	1,000	0,044	0,022	0,129	−0,122	−0,031	−0,009

	POSOIL	NEGOIL	POSCOM	NEGCOM	CLOSE	VOLRUB	VLTY	RETS
POSCOM	0,026	0,044	1,000	−0,070	0,115	−0,005	−0,033	−0,009
NEGCOM	0,003	0,022	−0,070	1,000	−0,034	−0,016	0,085	0,021
CLOSE	0,109	0,129	0,115	−0,034	1,000	−0,289	−0,493	0,018
VOLRUB	−0,072	−0,122	−0,005	−0,016	−0,289	1,000	0,167	0,021
VLTY	−0,015	−0,031	−0,033	0,085	−0,493	0,167	1,000	−0,041
RETS	0,023	−0,009	−0,009	0,021	0,018	0,021	−0,041	1,000

POSOIL,NEGOIL,POSCOM,NEGCOM — временные ряды сентиментов новостей.

CLOSE — величина на закрытие в день торгов (ММВБ).

VOLRUB — объем торгов.

VLTY — волатильность вида Close-to-Close (σ_{cl}) из модели GARCH (1, 1).

RETS — логарифмическая доходность индекса.

$$r_i = \log \left(\frac{C_i}{C_{i-1}} \right).$$

Рассчитанные корреляции Пирсона при уровне значимости 95%

Величина k — аргумент, используемый при обработке временных рядов, позволяет задать шаг приращения последующих дат. Например, при $k = 12$ год, содержащий наблюдения, разбивается на 12 промежутков таким образом, что шаг приращения составит один месяц.

Действительно, на протяжении всего периода исследования существует незначительная статистически значимая связь между тональностью новостей и реакцией рынка. На нулевом лаге позитивная и негативная информация о состоянии отрасли больше влияет на индекс, чем информация по компаниям. Позитивная информация по компаниям влияет на уровне 0,115 коэффициента корреляции, а негативная информация влияет незначительно на уровне −0,034.

Негативная и позитивная информация и по мировой отрасли, и по российским компаниям фактически не влияет на волатильность доходности нефтегазового индекса ММВБ и объема торгов. Для получения более точных выводов по гипотезам 1 и 2 требуется дальнейший анализ.

Возможно, проходит слишком мало времени для реакции рынка, поэтому рассмотрим взаимосвязь тональности новостей по компаниям и реакции рынка на разных лагах. В табл. 2 представлена максимальная корреляция между тональностью новостей и нефтегазовым индексом ММВБ, его волатильностью и объемом торгов.

**Корреляция на разных лагах (г-Пирсон для результатов,
значимых на уровне 5%) с 01.01.2007 по 17.08.2016**

Индексы		k = 52 (недели)		k = 12 (месяцы)	
		cor	lag, дни	cor	lag, дни
POSOIL	CLOSE	0,1090	0,9	0,1090	17,5
NEGOIL	CLOSE	0,1348	4,3	0,1348	80,0
NEGCOM	CLOSE	−0,0270	3,6	−0,0270	67,5
POSCOM	CLOSE	0,1231	3,6	0,1231	67,5
POSCOM	vClose	−0,0086	1,2	−0,0086	22,5
NEGCOM	vClose	0,0868	1,8	0,0868	32,5
POSOIL	vClose	−0,0147	0,0	−0,0147	0,0
NEGOIL	vClose	0,0133	3,1	0,0133	57,5
POSOIL	VOLRUB	−0,0619	0,8	−0,0619	15,0
NEGOIL	VOLRUB	−0,0685	2,0	−0,0685	37,5
NEGCOM	VOLRUB	0,0243	1,8	0,0243	32,5
POSCOM	VOLRUB	0,0275	3,8	0,0275	70,0
POSOIL	RETS	0,0320	3,1	0,0320	57,5
NEGOIL	RETS	0,0360	4,0	0,0360	75,0
NEGCOM	RETS	0,0264	4,0	0,0264	75,0
POSCOM	RETS	0,0253	1,6	0,0253	30,0

Проведенные расчеты корреляционной зависимости на разных лагах (г-Пирсон для результатов, значимых на уровне 5%) между каждым показателем из группы индикаторов тональности новостей и реакцией рынка являются более высокими. Влияние отраслевых новостей осталось на прежнем уровне на лагах 1–4 дня, при этом влияние негативных новостей более длительно (до 80 дней). Аналогичная ситуация и для новостей по компаниям, позитивная информация по компаниям влияет на нефтегазовый индекс ММВБ уже на 4-й день. Любопытно, что негативные новости, касающиеся мировой отрасли, влияют более чем в 4 раза дольше позитивных, в то время как российские новости имеют эффект одинаковой длительности на цену закрытия. То есть гипотеза 1 не отвергается для мировых новостей. Возможно, это связано с меньшей значимостью внутренних новостей относительно мировых и/или с большей рациональностью отечественных инвесторов и потому — меньшей степенью риск-неприятя. Похожие результаты были получены Чекли [Checkley, 2017].

Также хотелось бы отметить, что в случае мировых новостей утверждение о большей значимости новостей на цены закрытия справедливо, а в случае российских компаний — нет. Также коэффициенты при оптимистичных российских новостях больше аналогичных при мировых, в то время как для пессимистичных верно обратное. При этом влияние негативных мировых новостей значительно превышает влияние российских. Вероятно, это тоже объясняется тем, что мировая рыночная конъюнктура — более важный фактор, влияющий на российские компании нефтегазового сектора, нежели внутренние новости, касающиеся, как правило, операционной деятельности бизнесов.

Что касается лагов, то повысилось значение корреляции Пирсона: как и ожидалось, зачастую влияние новостей носит отложенный характер, поэтому наиболее тесная взаимосвязь между тональностью новостей и движением рынка, в первую очередь изменениями в ценах закрытия, проявилась на более продолжительном временном горизонте. Как и предполагалось теорией и работами западных авторов на схожую тематику, было обнаружено заметное влияние настроения новостей на фондовый рынок, исключение составили только незначимые пессимистичные новости по компаниям. В остальном можно заключить, что гипотезы 1 и 2 не отвергаются. Следовательно, эмпирический анализ не противоречит работам Тетлока и Лутиа, исследовавших влияние новостей, касающихся компании и мирового отраслевого рынка [Tetlock, 2007; Loutia et al., 2016].

Для тестирования 3-й гипотезы рассмотрим влияние политических событий на взаимосвязь движения индекса и тональности новостей.

Результаты корреляции на двух подпериодах представлены ниже в табл. 4.

Таблица 4

Корреляция на разных лагах
(г-Пирсон для результатов, значимых на уровне 5%)

BEFORE								
	POSOIL	NEGOIL	POSCOM	NEGCOM	CLOSE	VOLRUB	VLTY	RETS
POSOIL	1,000	−0,168	0,024	0,002	0,073	−0,046	0,001	0,028
NEGOIL	−0,168	1,000	0,008	0,015	−0,046	−0,087	0,002	−0,018
POSCOM	0,024	0,008	1,000	−0,070	0,080	0,032	−0,039	−0,026
NEGCOM	0,002	0,015	−0,070	1,000	−0,126	0,006	0,104	0,017
CLOSE	0,073	−0,046	0,080	−0,126	1,000	−0,094	−0,669	0,022
VOLRUB	−0,046	−0,087	0,032	0,006	−0,094	1,000	0,118	0,027
VLTY	0,001	0,002	−0,039	0,104	−0,669	0,118	1,000	0,037
RETS	0,028	−0,018	−0,026	0,017	0,022	0,027	0,037	1,000

AFTER								
	POSOIL	NEGOIL	POSCOM	NEGCOM	CLOSE	VOLRUB	VLTY	RETS
POSOIL	1,000	−0,048	0,007	−0,021	0,007	−0,047	−0,003	0,009
NEGOIL	−0,048	1,000	0,101	0,005	0,141	−0,039	−0,043	0,020
POSCOM	0,007	0,101	1,000	−0,089	0,176	−0,081	0,029	0,043
NEGCOM	−0,021	0,005	−0,089	1,000	−0,121	−0,004	0,067	0,038
CLOSE	0,007	0,141	0,176	−0,121	1,000	−0,132	−0,195	0,076
VOLRUB	−0,047	−0,039	−0,081	−0,004	−0,132	1,000	0,211	−0,023
VLTY	−0,003	−0,043	0,029	0,067	−0,195	0,211	1,000	−0,544
RETS	0,009	0,020	0,043	0,038	0,076	−0,023	−0,544	1,000

Проведенные расчеты корреляционной зависимости (γ -Пирсон для результатов, значимых на уровне 5%) между каждым показателем из группы индикаторов тональности новостей и величиной CLOSE индекса различаются в период до начала введения санкций с 01.01.2007 по 14.02.2014 и после, что соответствует периоду с 14.02.2014 по 31.12.2016. Так, после введения санкций увеличилось влияние позитивных новостей по компаниям более чем в 2 раза, также выросло влияние негативной информации по отрасли. В то же время из-за падения сырьевых цен снизилась тональность новостей, связанных с нефтью, и стала практически незначима. Укрепление важности внутренних новостей после политического шока также подтверждается изменением главного фактора, влияющего на объем торгов акциями российских компаний нефтегазового сектора: если раньше это были мировые новости, то после вхождения Крыма в состав Российской Федерации — уже внутренние. Это позволяет говорить о том, что политическая нестабильность действительно повлияла на зависимость поведения рынка и новостей, однако не слишком сильно. Несмотря на различия в методологии исследования в целом, получен тот же результат, однако более статистически значимый [Ankutdinov et al., 2016].

Для определения продолжительности влияния новостного фона были получены значения максимальной корреляции, представленные ниже в табл. 5.

Таблица 5

**Корреляция на разных лагах на двух подпериодах
(r-Пирсон для результатов, значимых на уровне 5%)**

		с 01.01.2007 по 14.02.2014				с 14.02.2014 по 17.08.2016			
		k = 52		k = 12		k = 52		k = 12	
		CORR	Lag, дни	CORR	Lag, дни	CORR	Lag, дни	CORR	Lag, дни
POSOIL	CLOSE	−0,02	3,23	0,08	57,50	0,01	0,67	0,08	20,00
NEGOIL	CLOSE	−0,21	3,50	−0,03	75,00	0,15	3,37	0,15	62,50
NEGCOM	CLOSE	−0,12	2,56	−0,12	47,50	−0,10	3,50	−0,10	65,00
POSCOM	CLOSE	0,10	3,37	0,10	62,50	0,18	0,00	0,18	0,00
POSCOM	VLTY	−0,01	1,35	−0,01	25,00	0,08	0,54	0,08	10,00
NEGCOM	VLTY	0,12	4,04	0,12	75,00	0,07	0,00	0,07	0,00
POSOIL	VLTY	0,04	1,08	0,00	0,00	0,02	2,56	0,01	37,50
NEGOIL	VLTY	0,22	2,15	0,05	57,50	0,04	1,62	0,04	30,00
POSOIL	VOLRUB	0,03	1,21	−0,04	15,00	0,03	3,23	0,07	27,50
NEGOIL	VOLRUB	−0,06	3,50	−0,03	25,00	0,04	2,29	0,04	42,50
NEGCOM	VOLRUB	0,06	1,75	0,06	32,50	0,07	3,50	0,07	65,00
POSCOM	VOLRUB	0,06	2,56	0,06	47,50	0,00	3,50	0,00	65,00
POSOIL	RETS	0,07	1,48	0,04	60,00	0,04	1,62	0,05	55,00
NEGOIL	RETS	0,09	1,21	0,02	5,00	0,03	0,67	0,03	12,50
NEGCOM	RETS	0,03	4,04	0,02	55,00	0,04	0,00	0,04	0,00
POSCOM	RETS	0,03	1,62	0,03	32,50	0,06	0,40	0,06	7,50

Для определения продолжительности влияния новостного фона были получены значения максимальной корреляции. Представленные в табл. 5 данные расчетов позволяют сделать вывод об увеличении зависимости между сентимент-коэффициентами и индексом. Агрегация проводилась по неделям и месяцам. В целом можно отметить влияние политической неопределенности на взаимосвязь новостного фона и реакции рынка. В краткосрочном периоде увеличилась значимость позитивной информации по компании и негативной информации по отрасли. Также выросло влияние новостного фона на движение фондового индекса, любая (позитивная и негативная) информация по компании приводит к увеличению волатильности фондового индекса. Можно отметить, что гипотезы 1 и 2 по-прежнему не отвергаются и на при разбиении исследуемого временного отрезка на подпериоды.

Что касается долгосрочного периода, то влияние новостного фона на реакцию рынка снизилось, после санкций рынок уменьшил период реагирования на новость, при этом влияние позитивной информации как по компании, так и по отрасли остается на высоком уровне и в долгосрочном периоде. Это также подтверждает влияние политической напряженности на фондовый рынок, предполагаемое гипотезой 3.

Выводы

В ходе работы удалось решить все поставленные исследовательские задачи и достичь поставленной цели. В результате эмпирического анализа была выявлена взаимосвязь между тональностью новостей и отраслевым фондовым индексом ММВБ для российских нефтегазовых компаний. Большинство результатов западных исследователей, тестирующих гипотезы на развитых рынках, получили подтверждение и в данной работе.

Во-первых, положительные новости про российские компании действительно предшествуют росту фондового индекса, а отрицательные — его падению, одновременно с этим объяснить тональность новостной повестки на основании рыночных данных невозможно. Обратное наблюдается в случае с мировыми новостями: возможно, причина в том, что Россия — экспортер нефти, в то время как использованные в исследовании мировые новости отрасли представлены с позиций ее стран-импортеров. Это означает, что некоторые положительные новости мирового рынка могут отрицательным образом влиять на цены российских акций. При этом новости по компаниям влияют на значение индекса, а новости по отрасли влияют на объем торгов.

Во-вторых, как и ожидалось, российский фондовый рынок сильнее и медленнее реагирует на негативные новости, чем на позитивные. Также было подтверждено существование асимметричной волатильности. Все это дает основания полагать о существовании у российского инвестора высокой степени неприятия риска.

В-третьих, тональность новостей и их роль во влиянии на фондовый рынок изменились после введения санкций в отношении России: российские компании стали освещаться немного более негативно, и несколько возросла рыночная волатильность, однако существенного влияния санкций на освещение российских нефтегазовых компаний обнаружено не было.

Можно с уверенностью заключить, что сентимент-анализ помог лучше объяснить психологию российских инвесторов и роль финансовых новостей во влиянии на них и на движение российского фондового рынка.

Данные выводы могут быть полезны компаниям и менеджерам, управляющим активами на российском фондовом рынке.

Список литературы

1. Федорова Е. А., Федоров Ф. Ю., Толкачев А. В. Взаимосвязь новостного фона и притока прямых иностранных инвестиций в регионы России // Пространственная экономика. — 2016. — № 4–5 (48). — С. 75–92.
2. Ankudinov A., Ibragimov R., Lebedev O. Sanctions and the Russian stock market // Research in International Business and Finance. — 2017. — Vol. 40. — P. 150–162.
3. Bodnaruk A., Loughran T., McDonald B. Using 10-K Text to Gauge Financial Constraints // Journal of Financial and Quantitative Analysis. — 2015. — 50:4.
4. Chau F. et al. Political uncertainty and stock market volatility in the Middle East and North African (MENA) countries // Int. Fin. Markets, Inst. and Money. — 2014. — Vol. 28. — P. 1–19.
5. Checkley M. S., Añón Higón D., Alles H. The Hasty Wisdom of the Mob: How Market Sentiment Predicts Stock Market Behavior // Expert Systems with Applications. — 2017. — Vol. 77.
6. Coulter A. Slander: Liberal lies about the American right. — N.Y.: Crown Publishers, 2002.
7. Gilbert E., Karahalios K. Widespread worry and the stock market // Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 2010. — P. 1–8.
8. Goldberg B. Bias: A CBS insider exposes how the media distort the news. — Washington, DC: Regnery Publishing, 2001.
9. Heston Steven L. and Nitish R. Sinha. News versus Sentiment: Predicting Stock Returns from News Stories // Finance and Economics Discussion Series. — 2016. — P. 1–35.
10. Jones K. S. A statistical interpretation of term specificity and its application in retrieval // Journal of Documentation. — Vol. 60. — No. 5. — P. 493–502.
11. Kearney C., Liu S. Textual sentiment in finance: A survey of methods and models // International Review of Financial Analysis. — 2014. — Vol. 33.
12. Lehkonen H., Heimonen K. Democracy, political risks and stock market performance // Journal of International Money and Finance. — 2015. — No. 59. — P. 77–99.
13. Lott J. R., Hassett K. A. Is Newspaper Coverage of Economic Events Politically Biased? // SSRN Electronic Journal. — 2012.
14. Loughran T., McDonald B. When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks // Journal of Finance. — 2011. — No. 66:1. — P. 35–65.
15. Loutia A., Mellios C., Andriosopoulos K. Do OPEC Announcements Influence Oil Prices? // Energy Policy. — 2016. — No. 90. — P. 262–272.
16. Moat H. S., Curme C., Avakian A., Kenett D. Y., Stanley H. E., Preis T. Quantifying Wikipedia Usage Patterns Before Stock Market Moves // Scientific Reports. — 2013. — 3:1.
17. Oliveira N., Cortez P., Areal N. On the predictability of stock market behavior using stock with sentiment and posting volume // Progress in Artificial Intelligence: Lecture Notes in Computer Science. — 2013. — 8154. — P. 355–365.
18. Porshnev A., Lakshina V., Redkin I. Could Emotional Markers in Twitter Posts Add Information to the Stock Market ARMAX-GARCH Model // Higher School of Economics Research Paper No. WP BRP 54/FE/2016.

19. *Qing L., Wang T., Li P., Liu L., Gong Q., Chen Y.* The Effect of News and Public Mood on Stock Movements // *Information Sciences*. — 2014. — No. 278. — P. 826–840.
20. *Tetlock P. C.* Giving content to investor sentiment: the role of media in the stock market // *J. Finance*. — 2007. — No. 62. — P. 1139–1168.
21. *Trabelsi Mnif A.* Political uncertainty and behavior of Tunisian stock market cycles: Structural unobserved components time series models // *Research in International Business and Finance*. — 2017. — No. 39. — P. 206–214.
22. *Veronesi P.* Stock market overreactions to bad news in good times: a rational expectations equilibrium model // *Rev. Finance. Stud.* — 1999. — 12 975.

The List of References in Cyrillic Transliterated into Latin Alphabet

1. *Fedorova E. A., Fedorov F. Ju., Tolkachev A. V.* Vzaimosvjaz' novostnogo fona i pritoka prjamyh inostrannyh investicij v regiony Rossii // *Prostranstvennaja jekonomika*. — 2016. — № 4–5 (48). — S. 75–92.