

ФИНАНСЫ

В. ГАВРИЛОВ

студент магистратуры экономического факультета
МГУ имени М.В. Ломоносова

М.А. ИВАНОВ

студент магистратуры экономического факультета
МГУ имени М.В. Ломоносова

О.А. КЛАЧКОВА

кандидат экономических наук, доцент кафедры
математических методов анализа экономики экономического факультета
МГУ имени М.В. Ломоносова

В.Ю. КОРОЛЕВ

доктор физико-математических наук, профессор, заведующий кафедрой
математической статистики факультета вычислительной математики
и кибернетики МГУ имени М.В. Ломоносова

Я.А. РОЩИНА

кандидат экономических наук, доцент кафедры
математических методов анализа экономики экономического факультета
МГУ имени М.В. Ломоносова

ВЛИЯНИЕ ТЕМАТИЧЕСКИХ НОВОСТНЫХ ПОТОКОВ НА КОМПОНЕНТЫ ВОЛАТИЛЬНОСТИ ФОНДОВОГО РЫНКА РОССИИ

В работе моделируется волатильность доходностей ценных бумаг на российском фондовом рынке с помощью моделей условной авторегрессионной гетероскедастичности, учитывающих поступающие на рынок тематические новостные потоки. Для учета новостного фона в модель в качестве независимого регрессора включается числовой показатель, характеризующий число новостей по каждой из ключевых тем. Выделение тем и построение такого показателя производятся методами обработки естественного языка. Для оценки влияния новостного фона не на волатильность доходности в целом, а на ее компоненты в стандартные GARCH модели вводятся предположения о том, что случайные ошибки являются смесью двух нормальных распределений. Было показано, что одна из компонент имеет существенно больший вес, но меньшую волатильность. Мы интерпретируем это как то, что «распространенные» темы составляют привычный новостной фон и слабо влияют на волатильность, тогда как более редко встречающиеся (и, следовательно, несущие больше информации) темы влияют на волатильность сильнее.

Ключевые слова: фондовый рынок, новостная аналитика, компоненты волатильности, обработка естественного языка.

JEL: C32, C53, G17.

DOI: 10.52180/2073-6487_2022_2_93_111.

Введение

* Компания Meta Platforms Inc., которой принадлежит социальная сеть Facebook, признана экстремистской организацией и ее деятельность запрещена на территории РФ.

В данной работе оценивается влияние тематических новостных потоков на фондовый рынок России: рассматривается зависимость изменчивости доходности некоторого актива (волатильности доходности) от тематики новостей о финансах и фондовом рынке РФ.

Информационный фон может оказывать значимое влияние на реальные экономические процессы. В частности, в ряде академических работ показано, что настроения СМИ могут влиять на репродуктивное поведение населения. Например, в статье [9] автор приходит к выводу, что в период с 1974 по 1981 г. во Франции общественное мнение в отношении демографических процессов коррелировало с позицией, выраженной в средствах массовой информации. Также выводы о влиянии СМИ на рождаемость подтверждаются и в статье [28].

Макроэкономическая динамика также может быть связана с информационным фоном. Например, в работе [2] строится индикатор инфляционных ожиданий населения на основе комментариев к новостным сообщениям об инфляции в сообществах Ведомости, Коммерсант, РБК и Эксперт Онлайн в социальных сетях Вконтакте и Facebook* в период 2014–2016 гг. Авторы показывают, что предложенный ими индикатор сильно коррелирует с классической мерой инфляционных ожиданий, построенной по опросам населения Центральным Банком России. Непосредственное влияние новостей на динамику инфляционных ожиданий в РФ обсуждается в работе [3].

Зависимость показателей фондовых рынков от новостей также исследовалась в академических работах. В основополагающей работе [32] доказано существование значимого влияния финансовых новостей из Wall Street Journal на фондовый рынок США (индекс Dow Jones), причем влияние является асимметричным: негативные новости оказывают большее влияние, чем позитивные. Аналогичные выводы получены и в работах [25; 5]. В книге [4] приведен обзор различных подходов к выделению факторов, определяющих волатильность, и к разложению волатильности на соответствующие компоненты.

В настоящей работе впервые на российских данных оценивается влияние новостного фона не на волатильность доходности финансовых индексов в целом, а на ее компоненты. Оценивание производится с помощью включения новостных характеристик в смешанные модели типа $MN(k)$ - $GARCH(p, q)$.

Расчеты проведены на данных 2021 г., в относительно спокойный период функционирования российского фондового рынка, поэтому содержательные выводы, к сожалению, нельзя напрямую экстраполировать на текущую ситуацию. Тем не менее предлагаемая в работе методология может быть повторена на данных нестабильного пери-

ода и является актуальной и для его анализа, поскольку позволяет в том числе моделировать воздействие большого числа новостей, посвященных конкретной теме, на высоковолатильную компоненту волатильности.

Работа организована следующим образом: в первой части обсуждается методология исследования, во второй части приводится описание данных и результаты эконометрического моделирования. Выводы и дальнейшие направления исследования в этой области приведены в заключении.

Методология исследования

Описание методологии исследования состоит из двух разделов. Сначала обсуждаются существующие модели волатильности и предлагается авторский вариант модели, в котором в качестве параметра выступает вектор $topics_t$, отвечающий за новостной фон в момент времени t . Во втором разделе описаны существующие подходы новостной аналитики и выбранный метод построения вектора $topics_t$.

Модели волатильности

Традиционная мера волатильности – скользящая дисперсия (среднее значение квадратичного отклонения доходности от ее среднего значения на отрезке времени из W наблюдений), то есть ряд вида:

$$\sigma_t^2 = \frac{1}{W} \cdot \sum_{i=t-W+1}^t (y_i - \bar{y})^2,$$

где: y_i – доходность некоторого актива в момент времени i .

Базовыми моделями волатильности являются модели ARCH [13] и GARCH [11].

Пусть ε_t – случайные ошибки в модели, описывающей динамику доходности некоторого актива y_t , тогда модель GARCH(p, q) имеет вид:

$$\begin{cases} \varepsilon_t = \sigma_t \cdot \eta_t, \quad \eta_t \sim iidN(0,1) \\ \sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \end{cases}$$

В свою очередь модель ARCH(q) является упрощенной версией, в которой коэффициенты $\beta_j = 0$.

Существует множество модификаций данных моделей. Например, в работе [15] описаны следующие варианты:

- *Threshold GARCH* (TGARCH), *Exponential GARCH* (EGARCH) и *Asymmetric Power gARCH* (APARCH), в которых учитывается, в отличие от оригинальной GARCH, наблюдаемая на финансовых рынках асимметрия;

- GARCH-X, в которых в уравнении волатильности используются независимые регрессоры;
- GARCH-MIDAS, предложенные в работе [14], в которых содержится дополнительная макроэкономическая информация.

В работе [17] на данных фондовых рынков была протестирована предсказательная способность большого числа (более трехсот) моделей типа ARCH (в этот класс входят все модели, описанные выше). Авторы приходят к выводу, что простейшая модель GARCH(1,1) почти не имеет себе равных (в смысле использованных в статье функций потерь) при предсказании волатильности обменных курсов на отложенной выборке.

Современное развитие моделей GARCH происходит в предположении о том, что ε_t имеют смешанный закон распределения, в частности, представляют из себя смесь нескольких нормальных распределений. В работе [16] были предложены модели $MN(k)$ – GARCH(p, q) (Mixed Normal GARCH):

$$\begin{cases} \varepsilon_t \sim MN(p_1, \dots, p_k, \mu_1, \dots, \mu_k, \sigma_{1,t}^2, \dots, \sigma_{k,t}^2), \mathbb{E}\varepsilon_t = 0 \\ \sigma_{k,t}^2 = \omega_k + \sum_{i=1}^q \alpha_{ki} \varepsilon_{k,t-i}^2 + \sum_{j=1}^p \beta_{kj} \sigma_{k,t-j}^2 \end{cases}$$

Здесь плотность распределения ошибок ε_t представляет собой линейную комбинацию (с весами p_1, \dots, p_k) плотностей нормальных распределений с параметрами $(\mu_k, \sigma_{k,t}^2)$, $k = 1, \dots, K$.

При оценке моделей этого класса на дневных доходностях индекса NASDAQ Haas ([16]) приходит к выводу, что однокомпонентная модель (т. е. обычная GARCH) описывает данные хуже смешанных моделей.

Таким образом, в настоящей работе оценивается модель $MN(2)$ –GARCH(1,1) – X, где в качестве независимого регрессора выступает переменная новостного фона $topics_t$:

$$\begin{cases} y_t = a + b \cdot y_{t-1} + \varepsilon_t \\ \varepsilon_t \sim MN(p, 1-p, \mu_1, \mu_2, \sigma_{1,t}^2, \sigma_{2,t}^2), \mathbb{E}\varepsilon_t = 0 \\ \sigma_{k,t}^2 = \omega_k + \alpha_k \cdot \varepsilon_{t-1}^2 + \beta_k \cdot \sigma_{k,t-1}^2 + T_k' \cdot topics_{t-1}, k = 1, 2 \end{cases}$$

где: y_t – доходность некоторого актива, моделируется авторегрессионным процессом первого порядка AR(1); ε_t – случайные шоки, имеющие распределение смеси двух нормальных распределений (с весами p и $1-p$), дисперсии которых описываются моделями GARCH(1,1).

Авторская новизна заключается в том, что дисперсия каждой компоненты волатильности зависит от новостного фона. Это дает возможность проверить гипотезу о различном влиянии новостей на разные компоненты волатильности.

Новостная аналитика

Согласно определению в статье [5], новостная аналитика – это измерение количественных и качественных характеристик текстовых новостей и исследование их влияния на финансовые показатели. Таким образом, применение новостной аналитики позволяет исследовать влияние новостного фона на волатильность, включая числовой показатель, характеризующий новостной фон, в качестве независимого регрессора в эконометрические модели, оценивающие волатильность. Существует два основных подхода к построению такого числового показателя.

Первый подход состоит в подборе подходящих замещающих переменных и косвенной оценке новостного фона с его помощью. Если выбор был удачным, то изменения волатильности должны (частично) объясняться такой замещающей переменной, а значит ее включение в модель должно приводить к ослаблению GARCH-эффекта (снижению модулей коэффициентов β). В эмпирических исследованиях встречаются следующие варианты замещающих переменных:

- индикатор (бинарный) выхода макроэкономических новостей, например, в работах [12; 21];
- число газетных заголовков и финансовых отчетов, например, в работе [10];
- число анонсов по заданным фирмам и/или ценным бумагам, например, в работах [24; 19];
- наиболее «популярный» вариант – объем торгов, например, в работах [22] для акций двадцати компаний США, в [29] и [18] для индексов Австралийской фондовой биржи, в [26] для индекса Токийской фондовой биржи, в [27] для рынка фьючерсов, в [5] для двадцати компаний FTSE100;
- отклонения фактического объема торгов от прогнозного уровня по ARMA модели, например, в работе [7]. Однако стоит отметить, что включение авторами такой переменной не снизило колебания условной волатильности для рыночных индексов девяти рассматриваемых в работе стран, т. е. данный вариант замещающей переменной, в отличие от предыдущих, требует дальнейшего тестирования применимости;
- новостная интенсивность (число новостных сообщений), например, в работах [18; 5] (в обеих работах используются два варианта замещающих переменных).

Второй подход более трудоемкий и состоит в выделении ключевых тем новостного потока и отнесении каждой исходной новости к одной из выделенных тем. Тогда новостной фон в каждый момент времени характеризуется числом новостей по каждой теме (представляет из себя r -мерный числовой вектор, где r – число выделенных тем). В рамках

данного подхода возможна как обработка новостного фона «вручную», как это сделано, например, в [3], где авторы считают количество выступлений официальных лиц и учитывают тематику выступления (она также определяется вручную), так и автоматизация этого процесса. Такая автоматизация привлекательна возможностью учесть значительные массивы данных, а также легкостью реплицирования результатов (например, для проверки их устойчивости или для их сравнения на спокойном и активном новостных фонах).

В данной работе мы пользуемся вторым подходом в его автоматизированном варианте. Для обработки текста новостей используется его векторное представление с помощью модели упрощенного представления текста «Мешок слов» («Bag-of-words»). После этого из полученного набора («мешка») слов удаляются стоп-слова (предлоги, местоимения и т. д.), а все оставшиеся слова методом лемматизации приводятся к их нормальной (базовой) форме: существительные – к именительному падежу единственного числа; прилагательные – к именительному падежу единственного числа мужского рода; глаголы, причастия и деепричастия – к глаголу в инфинитиве несовершенного вида. Каждой i -ой новости ставится в соответствие вектор из R_k^+ : {Новость $_i$ } \rightarrow (x_1^i, \dots, x_k^i), где k – выбранное число базовых слов из словаря соответствующей направленности (на практике выбирают k от 500 до 1000), по следующему правилу. В качестве j -ой компоненты x_j^i берется мера $TF-IDF$ значимости j -го базового слова в i -ой новости, характеризующая важность этого слова в i -ой новости, нормированная на его важность во всем массиве новостей. Данная мера вычисляется как произведение частоты встречаемости слова в новости (сомножитель TF , term frequency) на обратную частоту встречаемости слова во всем массиве новостей (сомножитель IDF , inverse document frequency).

$$TF \times IDF(j) = TF(j), j = 1, \dots, k.$$

Преимуществом (перед простым подсчетом частоты встречаемости) использования в качестве меры значимости слова меры $TF - IDF$ является то, что при этом большой вес получают слова с высокой частотой в пределах данной новости и одновременно с низкой частотой в остальных новостях. Таким образом, при равной встречаемости слова в новости большую значимость будет иметь более редкое «в целом» слово.

Результатом применения преобразования {Новость $_i$ } \rightarrow (x_1^i, \dots, x_k^i) ко всем n новостям является новостная матрица A ($n \times k$). Для выделения из нее ключевых тем новостного потока мы пользуемся методологией [31]. В рамках данной методологии исходная матрица A размера $n \times k$ (имеющая на практике очень большой ранг) аппроксимируется

некоторой матрицей A_r существенно (на порядки) меньшего заранее фиксированного ранга r (на практике не более 10) с помощью SVD-разложения, в котором у стоящей в центре матрицы оставляются r наибольших сингулярных чисел (остальные отбрасываются). Согласно теореме Экарта-Янга¹, полученная матрица A_r является наилучшей (в смысле L_2 нормы) среди матриц ранга r аппроксимацией исходной матрицы A . При этом норма разности $\|A - A_r\|_2$ (содержательно представляющая собой потери информации при такой аппроксимации) равна следующему по величине $(r + 1)$ -му сингулярному числу исходной матрицы. Таким образом, выбирая число r так, чтобы величина $(r + 1)$ -го сингулярного числа была приемлемо малой, мы получим аппроксимацию удовлетворяющей нас точности. При этом первым r сингулярным числам математически будут соответствовать вектора, по которым разложена исходная матрица, а содержательно – основные темы собранных новостей, другими словами, мы получим описание r основных тем новостного потока в виде набора описывающих каждую тему ключевых слов. Наконец, каждую исходную новость можно отнести к одной из выделенных тем, используя в качестве меры косинусное расстояние между векторами.

Описание данных и результаты моделирования

Мы используем данные о ежедневных значениях фондовых индексов, взятые с сайта Финам.ру² за период с 2021–03–06 до 2021–10–19. Для моделирования были выбраны ряды процентных доходностей индекса РТС, а также акций компаний трех разных отраслей (банки – СберБанк, SBER; телекоммуникации – МТС, MTSS; IT-технологии – Яндекс, YNDX). В качестве источника данных о новостях использовался сайт finanz.ru – один из крупнейших провайдеров новостей о финансах и фондовом рынке РФ³. Для сбора и обработки новостей использовался язык программирования Python, а для оценки модели – язык Julia.

Оценивание выбранной спецификации модели происходит методом максимального условного правдоподобия аналогично тому, как это сделано в работах [16; 6].

Первая сложность, которая возникает в процессе оценивания модели, – выбор начальных значений волатильности $\sigma_{1,1}^2$ и $\sigma_{2,1}^2$, ненаблюдаемых в данных. Простейший способ решить эту проблему – положить начальные дисперсии равными некоторой константе и рас-

¹ Eckart C. The approximation of one matrix by another of lower rank // Psychometrika. 1936. Vol. 1. № 3. P. 211–218.

² Экспорт котировок – Российские индексы – RTSI (2021). Финам.ру.

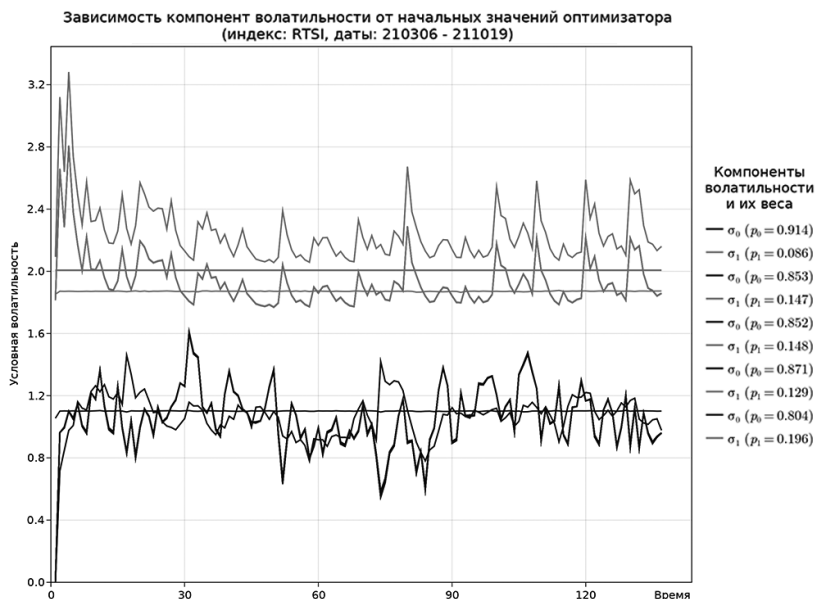
³ Экспорт финансовых новостей (2021). finanz.ru.

считывать, что коэффициенты при $\sigma_{1,t-1}^2$ и $\sigma_{2,t-1}^2$ (β_1 и β_2 соответственно) будут достаточно малы, чтобы воздействие произвольного первоначального выбора со временем сгладились. Второй способ, описанный в [15], состоит в том, чтобы положить $\sigma_{k,1}^2 = \omega_k$, иначе говоря, в предположении, что первое наблюдение распределено независимо от времени. Наконец, третий подход, также формализованный в [15], предполагает использование безусловных дисперсий, то есть $\mathbb{E}\sigma_{k,t}^2$. В данной работе используется третий подход.

Вторую сложность представляет непосредственно решение оптимизационной задачи в рамках реализации метода условного максимального правдоподобия – максимизация функции правдоподобия при заданных ограничениях на то, что веса смесей (p , $1 - p$) лежат от 0 до 1, на равенство нулю математического ожидания случайных ошибок модели ($\mathbb{E}\varepsilon_t = 0$), а также на стационарность процесса ($\alpha_k + \beta_k < 1$). В данной работе процесс оптимизации реализуется с помощью метода внутренней точки (библиотека Optim.jl), который частично инициализируется результатами работы EM-алгоритма, по аналогии с его применением в работе [8]. Стабильность оценок проверяется путем оптимизации правдоподобия с разных начальных точек допустимого множества задачи. На рис. 1 представлены оценки условной волатильности индекса RTSI в зависимости от начальных точек. Видно, что оценки зависят от начальных приближений – это типичная проблема смешанных моделей, вызванная тем, что их правдоподобие имеет множество локальных максимумов, к которым и сходятся оптимизационные алгоритмы. Однако для ряда RTSI во всех случаях компонента с большей волатильностью имеет намного меньший вес, а менее волатильная компонента во всех случаях имеет схожую динамику. Стабильность оценок также зависит от входных данных. Как показано на рис. 2, оценки для ряда YNDX практически не зависят от начальных значений.

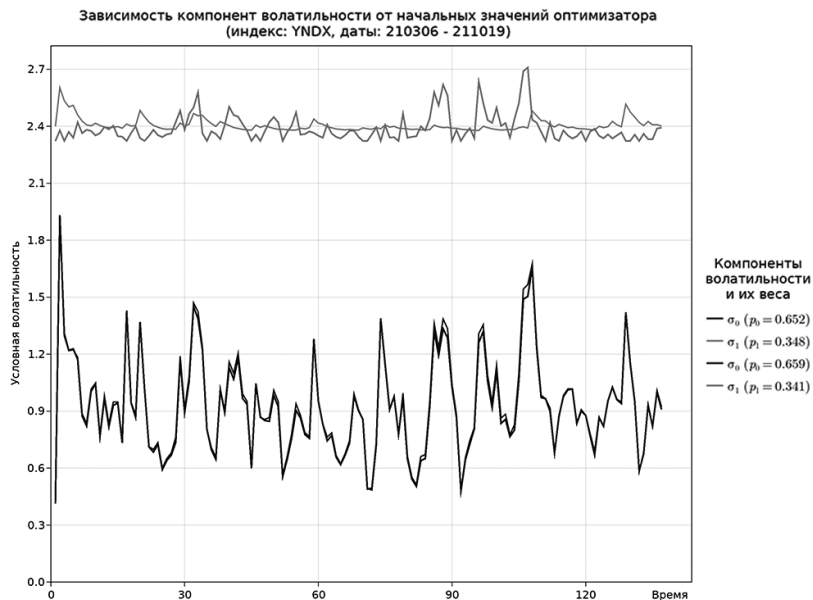
Всего было собрано 2380 новостей, выбрано $k = 1000$ базовых слов направленности экономика и финансы. Таким образом, исходная новостная матрица имела размер 23807×1000 . Построенная эмпирическая кривая отсортированных по убыванию сингулярных чисел SVD-разложения резко убывает до 6-го сингулярного числа (см. рис. 3).

Согласно описанной выше методологии, возьмем $r = 6$ и сформируем 6 основных тем новостей. Каждая тема представляет собой линейную комбинацию исходных 1000 слов словаря. Ключевые слова (было выбрано десять слов с наибольшим весом) выделенных тем приведены в табл. 1. Тема 1 скорее относится к «нефтяной» тематике, темы 2 и 6 описывают экономику РФ в целом, тема 3 близка к курсам валют и бирже, тема 4 касается геополитической обстановки, тема 5 относится к фондовому рынку.



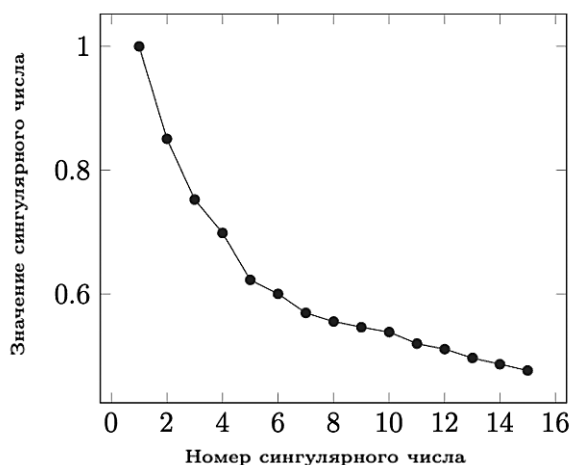
Источник: рассчитано авторами.

Рис. 1. Оценки условной волатильности индекса RTSI в зависимости от начальных точек – различных значений коэффициентов при лаге волатильностей компонент β_1 и β_2 при всех новостях T_1 и T_2 ; они варьируются по решетке от 10^{-1} до 10^{-4}



Источник: рассчитано авторами.

Рис. 2. Оценки условной волатильности YNDX в зависимости от начальных точек – различных значений коэффициентов при лаге волатильностей компонент β_1 и β_2 при всех новостях T_1 и T_2 ; они варьируются по решетке от 10^{-1} до 10^{-4}



Источник: рассчитано авторами.

Рис. 3. График убывания сингулярных чисел SVD-разложения новостной матрицы

Выделение в отдельную тему «нефтяной» темы 1 согласуется с результатами ряда работ, в которых доказывается влияние волатильности цен на нефть на волатильность фондовых индексов, в том числе для фондового рынка РФ (см., например, [1; 23]). Это выглядит вполне естественно, поскольку Россия является одним из крупнейших экспортеров нефти в мире. Аналогично, выделение в отдельную тему «геополитической» темы 4 согласуется, например, с исследованиями ([1; 30]), в которых авторам удалось показать существование воздействия геополитических изменений на российский фондовый рынок.

Ключевые слова тем приведены в таблице строчными (маленькими) буквами, поскольку они рассматриваются здесь просто как слова, полученные после применения модели упрощенного представления текста.

В табл. 2 приведены наиболее типичные новости для каждой темы.

Результаты оценивания⁴ приведены в табл. 3. Видно, что для каждого индекса первая компонента смеси имеет больший вес, чем вторая, причем наиболее весомая компонента имеет меньшую дисперсию. Это согласуется с традиционной интерпретацией смешанных моделей [20; 16]: компонента с относительно низкой дисперсией (и обычно с большим весом) может означать «обычное состояние экономики», а компоненты с большими дисперсиями (и более низкими весами) могут соответствовать всплескам волатильности, вызванным появлением важной новой информации.

⁴ Для начальных значений $\beta_1 = \beta_2 = T_1 = T_2 = 10^{-2}$.

Таблица 1

Ключевые слова основных тем

ТЕМА 1:	18704	ТЕМА 2:	752	ТЕМА 3:	670
СЛОВО	SCORE	СЛОВО	SCORE	СЛОВО	SCORE
год	0.3248	рф	0.5052	курс	0.3871
нефть	0.2226	цб	0.3042	рф	0.2967
цена	0.1875	банк	0.139	доллар	0.2612
рост	0.1665	млрд	0.1015	евро	0.2546
млн	0.1208	вырасти	0.0818	цб	0.2546
опек	0.0835	инфляция	0.0812	руб	0.2301
добыча	0.0825	ставка	0.0774	биржа	0.1485
баррель	0.0802	индекс	0.0666	московский	0.1413
ввп	0.0753	резерв	0.0585	понизить	0.0996
россия	0.0707	рубль	0.0576	повысить	0.0923
ТЕМА 4:	2164	ТЕМА 5:	1123	ТЕМА 6:	394
СЛОВО	SCORE	СЛОВО	SCORE	СЛОВО	SCORE
рф	0.2368	индекс	0.5697	руб	0.3834
год	0.2203	торг	0.3908	млрд	0.3619
нефть	0.1956	мосбиржа	0.366	вырасти	0.2719
россия	0.1916	ртс	0.3475	рф	0.246
цена	0.1879	открытие	0.2504	тасс	0.2215
рост	0.1814	дать	0.1505	год	0.2196
тасс	0.1808	снизиться	0.121	индекс	0.1679
сша	0.152	вырасти	0.1185	прибыль	0.159
газ	0.1357	тасс	0.0357	чистый	0.1477
северный	0.1148	пункт	0.0338	цб	0.1435

Источник: рассчитано авторами.

Таблица 2

Типичные новости по основным темам

Тема	Типичная новость
1	«Доктор апокалипсис» предсказал крах биткоина через 10 лет
2	ЦБ РФ выпускает памятную серебряную 2-рублевую монету, посвященную Н.А. Некрасову
3	ЦБ РФ повысил курс доллара на 27 июля до 74,1 руб., курс евро – до 87,32 руб.
4	ТАСС: РФ не исключает роста добычи нефти странами, не входящими в ОПЕК+, на фоне роста цен на нефть- Новак
5	Индекс Мосбиржи на открытии торгов вырос на 0,19%, индекс РТС – на 0,72%
6	ТАСС: Чистая прибыль «ЭНЕЛ РОССИИ» по МСФО за I полугодие выросла до 2 млрд руб. С 2,8 млрд руб. годом ранее – компания

Источник: рассчитано авторами.

Результаты оценивания моделей

Коэффициент	SBER	MTSS	YNDX	RTSI
AR(1): a	0,19937	0,12336	0,07454	0,20489
AR(1): b	-0,10956	-0,02262	0,03165	-0,00769
Смесь: p	0,55484	0,92112	0,65165	0,85209
Смесь: μ_1	0,05526	-0,08555	0,03315	-0,01823
Смесь: μ_2	-0,06888	0,99902	-0,06201	0,10501
GARCH: ω_1	0,00000	0,00000	0,00000	0,00000
GARCH: ω_2	0,60995	0,35728	5,39545	1,44413
GARCH: α_1	0,03584	0,03453	0,05953	0,00000
GARCH: α_2	0,37448	0,08016	0,00000	0,21836
GARCH: β_1	0,00000	0,27203	0,30624	0,31852
GARCH: β_1	0,62552	0,00000	0,00000	0,52839
Тема 1, режим 1: T_{11}	0,90001	0,00000	0,00000	0,12121
Тема 1, режим 2: T_{21}	0,00000	0,00000	0,00000	0,00000
Тема 2, режим 1: T_{12}	0,00000	1,98736	6,98004	0,00000
Тема 2, режим 2: T_{22}	0,00000	0,00000	9,35144	0,00000
Тема 3, режим 1: T_{13}	0,00000	0,00000	0,00000	16,15318
Тема 3, режим 2: T_{23}	0,00000	389,17110	0,00000	0,00000
Тема 4, режим 1: T_{14}	0,00000	0,13595	0,00000	0,00000
Тема 4, режим 2: T_{24}	0,00000	0,00000	0,00000	0,00000
Тема 5, режим 1: T_{15}	0,00000	0,45064	1,32596	4,78394
Тема 5, режим 2: T_{25}	0,00000	0,00000	0,00000	0,00000
Тема 6, режим 1: T_{16}	0,00000	1,63339	11,44925	0,00000
Тема 6, режим 2: T_{26}	0,00000	0,09182	0,00000	0,00000

Источник: рассчитано авторами.

Примечание: режим 1 соответствует первой компоненте волатильности в оцениваемой модели AR(1) – MN(2) – GARCH(1,1), режим 2 – второй компоненте.

Обратимся к коэффициентам при переменных новостей. Отметим, что по абсолютному значению наименьшими являются коэффициенты при теме 1, которая чаще всего встречается среди собранных новостей (18704 новости из 23807). Следовательно, более «редкие» темы 2–6 имеют более значительное влияние на волатильность компонент.

Для индекса РТС, характеризующего рынок в целом, от новостей зависит только волатильность первой компоненты, которая характери-

зует «обычное состояние экономики». При этом повышательное воздействие на нее оказывают новости по нефтяной тематике (тема 1), а также по темам 3 и 5, связанным с курсами валют и фондовым рынком. Для высокотехнологичной компании Яндекс на волатильность обеих компонент оказывает влияние новостной поток об экономике РФ в целом; кроме того, новости фондового рынка влияют на волатильность первой компоненты. Для компании МТС на обе компоненты влияют новости об экономике России; на первую компоненту повышательное воздействие оказывают геополитические новости, а также сообщения о фондовом рынке; вторая компонента зависит от новостей о курсах валют и бирже. Наконец, для компании банковского сектора СБЕР, новости влияют только на волатильность первой компоненты, причем влияние оказывают только новости, имеющие нефтяную тематику.

Если анализировать результаты, приведенные в табл. 3 «по горизонтали», то можно отметить, что на волатильность первой компоненты, которая имеет существенно больший вес, но меньшую среднюю дисперсию, влияют темы 2 и 6, соответствующие новостям по экономике России в целом, а также тема 5, характеризующая фондовый рынок. Данный результат согласуется с интуитивным представлением о том, что основная слабоволатильная компонента определяется общероссийскими новостями. Кроме того, можно отметить роль новостей фондового рынка – они сходны по важности с общероссийскими новостями. На волатильность меньшей по весу высоковолатильной компоненты также оказывают некоторое воздействие общие темы 2 и 6, но имеет влияние и тема 3, соответствующая изменениям курсов валют. Это также согласуется с интуитивным представлением о том, что меньшая высоковолатильная компонента определяется новостями динамично меняющейся сферы экономики.

Таким образом, в данной работе на российских данных еще раз подтверждается устоявшийся в англоязычной литературе вывод о том, что новостной фон оказывает влияние как на фондовый рынок в целом, так и на компании разных отраслей. Кроме того, нам удалось показать, что часть новостей имеет воздействие на «весомую», но низковолатильную компоненту, а некоторые специфические новости влияют на имеющую малый вес высоковолатильную компоненту.

Заключение

Любая попытка моделирования доходностей финансовых активов приводит к необходимости моделирования волатильности доходности. Одним из наиболее популярных современных направлений такого моделирования является применение GARCH моделей, причем в предположении о том, что случайные ошибки имеют смешан-

ный закон распределения (так называемых смешанных GARCH моделей). В данной работе мы следуем этому направлению, моделируя ошибки смесью двух нормальных распределений. При этом предметом нашего интереса служит возможность интерпретации новостей как внешних шоков, которые могут влиять на волатильность. Поэтому мы дополняем смешанные GARCH модели экзогенной переменной, отвечающей за тематику новостного фона, с целью оценки его влияния на волатильность фондового рынка России. Пользуясь тем, что применение смешанных моделей предполагает наличие двух компонент волатильности, мы оцениваем влияние тематики новостного фона на каждую из компонент. Было показано, что одна из них имеет значительно больший вес, но меньшую волатильность. Более часто встречающиеся темы имеют близкие к нулю коэффициенты, тогда как коэффициенты при реже встречающихся темах отделены от нуля. Это может говорить о том, что «распространенные» темы составляют привычный для участников рынка новостной фон и слабо влияют на волатильность, тогда как более редко встречающиеся (и, следовательно, несущие больше информации) темы влияют на волатильность сильнее. В частности, даже большой публикационный поток по «привычным» для участников рынка темам может не оказывать существенного влияния на волатильность доходностей активов. Поэтому инвесторам не стоит паниковать, видя большое число новостей, например, по экономике России в целом.

Однако нельзя полностью исключить возможности того, что обнуление некоторых коэффициентов в табл. 3 вызвано неограниченностью правдоподобия смеси нормальных распределений при нулевой дисперсии любой компоненты. В пользу первой версии говорит то, что в случае неограниченности правдоподобия обнулились бы и стандартные коэффициенты GARCH (ω , α , β), чего не наблюдается.

Стоит еще раз отметить, что результаты получены на относительно спокойном новостном фоне (на данных за 2021 г.). Повторение предложенной методологии на данных нестабильного периода является логичным направлением развития работы. Особый интерес представляет потенциальный рост относительного веса волатильной компоненты, а также проверка гипотезы о том, что «распространенные» темы слабее влияют на волатильность. Возможно, в условиях нестабильности, в том числе новостного фона, «распространенные» новости на «привычные» темы будут оказывать более значимое положительное «успокаивающее» воздействие.

Ограничением (и еще одним направлением дальнейшего развития) работы является определение темы по набору ее ключевых слов, проделанное нами вручную. Одним из возможных путей развития является переход к автоматизации такого определения, например, к при-

менению вероятностных тематических моделей⁵. Другим естественным продолжением работы является отказ от предпосылки о том, что количество компонент в гауссовской смеси равно двум, а также учет тональности новостей. Возможно, это поможет уловить более тонкое влияние новостного фона на фондовый рынок.

ЛИТЕРАТУРА

1. Аганин А.Д. Волатильность российского фондового индекса: нефть и санкции // Вопросы экономики. 2020. № 2. С. 86–100.
2. Голощапова И.О., Андреев М.Л. Оценка инфляционных ожиданий российского населения методами машинного обучения // Вопросы экономики. 2017. № 6. С. 71–93.
3. Жемков М.И., Кузнецова О.С. Вербальные интервенции как фактор формирования инфляционных ожиданий в России // Журнал Новой экономической ассоциации. 2019. № 2 (42). С. 49–69.
4. Королев В.Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. М.: Изд. МГУ, 2011.
5. Сидоров С., Дате П., Балаш В. Использование данных новостной аналитики в GARCH моделях // Прикладная эконометрика. 2013. № 29 (1). С. 82–96.
6. Alexander C., Lazar E. Normal mixture GARCH(1,1): Applications to exchange rate modelling // Journal of Applied Econometrics. 2006. 21(3). Pp. 307–336.
7. Arago V., Nieto L. Heteroskedasticity in the returns of the mainword stock exchange indices: Volume versus GARCH effects // International Financial Markets Institute and Money. 2005. 15. Pp. 271–284.
8. Ardia D., Bluteau K., Boudt K., Catania L., Trottier D.-A. Markov-Switching GARCH Models in R: The MSGARCH Package // Journal of Statistical Software. 2019. 91(4). Pp. 1–38.
9. Barrusse De Luca. Discussion on the return of «denatality» in France and its perception between 1974 and 1981 // Population and Economics. 2020. 4(3). P. 33–56.
10. Berry T. D., Howe K. M. Public information arrival // Journal of Finance. 1993. 49. Pp. 1331–1346.
11. Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity // Journal of Econometrics. 1986. Т. 31. № 3. Pp. 307–327.
12. Ederington L. H., Lee J. H. How markets process information: News releases and volatility // Journal of Finance. 1993. 48. Pp. 1161–1191.
13. Engle R. F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation // Econometrica. 1982. Vol. 50. No. 4. P. 987.
14. Engle R. F., Ghysels E., Sohn B. Stock Market Volatility and Macroeconomic Fundamentals // Review of Economics and Statistics. 2013. Т. 95. No 3. Pp. 776–797.

⁵ См., например, обзор вероятностных тематических моделей К.В. Воронцова в гл. 5 пособия «Автоматическая обработка текстов на естественном языке и анализ данных» (изд. НИУ ВШЭ, 2017 г.).

15. *Francq C., Zakoian J.-M.* GARCH Models: Structure, Statistical Inference and Financial Applications // Hoboken, NJ: John Wiley & Sons. 2019. 2 edition.
16. *Haas M.* Mixed Normal Conditional Heteroskedasticity // *Journal of Financial Econometrics*. 2004. Т. 2. No 2. Pp. 211–250.
17. *Hansen P. R., Lunde A.* A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? // *J. Appl. Econ.* 2005. Vol. 20. No 7. Pp. 873–889.
18. *Janssen G.* Public information arrival and volatility persistence in financial markets // *The European Journal of Finance*. 2004. 10. Pp. 177–197.
19. *Kalev P. S., Liu W.-M., Pham P. K., Jarnecic E.* Public information arrival and volatility of intraday stock returns // *Journal of Banking and Finance*. 2004. 28(6). P. 1447–1467.
20. *Kon, S. J.* Models of Stock Returns—A Comparison // *Journal of Finance*. 1984. 39(1). Pp. 147–165.
21. *Laakkonen H., Lanne M.* Asymmetric News Effects on Exchange Rate Volatility: Good vs. Bad News in Good vs. Bad Times // *Studies in Nonlinear Dynamics & Econometrics*. 2009. 14.
22. *Lamoureux C. G., Lastrapes W. D.* Heteroskedasticity in stock return data: Volume versus GARCH effects // *Journal of Business & Economic Statistics*. 1990. 2. Pp. 253–260.
23. *Lozinskaia A., Saltykova A.* Fundamental Factors Affecting the MOEX Russia Index: Retrospective Analysis // *CEUR Workshop Proceedings*. 2019. 2479. Pp. 32–45.
24. *Mitchell M. L., Mulherin J. H.* How markets process information: News releases and volatility // *Journal of Finance*. 1994. 49. Pp. 923–950.
25. *Mitra L., Mitra G.* Applications of news analytics in finance: A review // *The Handbook of News analytics in finance*. 2011. Pp. 1–36.
26. *Miyakoshi T.* ARCH versus information-based variances: Evidence from the Tokyo stock market // *Japan and the World Economy*. 2002. 14. Pp. 215–231.
27. *Najand M., Yung K.* A GARCH examination of the relationship between volume and variability in futures markets // *The Journal of Futures Markets*. 1991. 11. Pp. 613–621.
28. *Rabbi A. M. F.* Mass media exposure and its impact on fertility: Current scenario of Bangladesh // *Journal of Scientific Research*. 2012. Vol. 4. № 2. Pp. 383–383.
29. *Ragunathan V., Peker A.* Price variability, trading volume and market depth: Evidence from the Australian futures market // *Applied Financial Economics*. 1997. 7. Pp. 447–454.
30. *Rubtsov B., Annenskaya N.* Factor analysis of the Russian stock market // *Journal of Reviews on Global Economics*. 2018. Iss. 7 (Special Issue). Pp. 417–425.
31. *Sanjiv R.* News analytics: Framework, techniques, and metrics // *The Handbook of News analytics in finance*. 2011. Pp. 41–69.
32. *Tetlock Paul C.* Giving Content To Investor Sentiment: The Role Of Media In The Stock Market // *The Journal Of Finance*. 2007. Vol. Lxii. Pp. 1139–1168.

REFERENCES

1. *Aganin A.D.* Russian Stock Index volatility: Oil and sanctions. *Voprosy Ekonomiki*. 2020; (2): 86–100. (In Russ.).
2. *Goloshchapova I., Andreev M.* Measuring inflation expectations of the Russian population with the help of machine learning. *Voprosy Ekonomiki*. 2017; (6): 71–93. (In Russ.).
3. *Zhemkov M.I., Kuznetsova O.S.* Verbal Interventions as a Factor of Inflation Expectations in Russia // *The Journal of the New Economic Association*. 2019. No 2 (42). Pp. 49–69. (In Russ.).
4. *V. Yu. Korolev.* Probabilistic-statistical methods for the decomposition of the volatility of chaotic processes, 2011, Moscow: Moscow University Press. (In Russ.).
5. *S. P. Sidorov, P. Date, Balash V. A.* Using news analytics data in GARCH models // *Applied Econometrics*. 2013. V. 29, no. 1. Pp. 82–96. (In Russ.).
6. *Alexander C., Lazar E.* Normal mixture GARCH(1,1): Applications to exchange rate modelling // *Journal of Applied Econometrics*. 2006. 21(3). Pp. 307–336.
7. *Arago V., Nieto L.* Heteroskedasticity in the returns of the mainword stock exchange indices: Volume versus GARCH effects // *International Financial Markets Institute and Money*. 2005. 15. Pp. 271–284.
8. *Ardia D., Bluteau K., Boudt K., Catania L., Trottier D.-A.* Markov-Switching GARCH Models in R: The MSGARCH Package // *Journal of Statistical Software*. 2019. 91(4). Pp. 1–38.
9. *Barrusse De Luca.* Discussion on the return of «denatality» in France and its perception between 1974 and 1981 // *Population and Economics*. 2020. 4(3). P. 33–56.
10. *Berry T. D., Howe K. M.* Public information arrival // *Journal of Finance*. 1993. 49. Pp. 1331–1346.
11. *Bollerslev T.* Generalized Autoregressive Conditional Heteroskedasticity // *Journal of Econometrics*. 1986. T. 31. No 3. Pp. 307–327.
12. *Ederington L. H., Lee J. H.* How markets process information: News releases and volatility // *Journal of Finance*. 1993. 48. Pp. 1161–1191.
13. *Engle R. F.* Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation // *Econometrica*. 1982. T. 50. No 4. Pp. 987.
14. *Engle R. F., Ghysels E., Sohn B.* Stock Market Volatility and Macroeconomic Fundamentals // *Review of Economics and Statistics*. 2013. T. 95. No 3. Pp. 776–797.
15. *Francq C., Zakoian J.-M.* GARCH Models: Structure, Statistical Inference and Financial Applications // Hoboken, NJ: John Wiley & Sons. 2019. 2 edition. 492 c.
16. *Haas M.* Mixed Normal Conditional Heteroskedasticity // *Journal of Financial Econometrics*. 2004. T. 2. No 2. Pp. 211–250.
17. *Hansen P. R., Lunde A.* A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? // *J. Appl. Econ*. 2005. T. 20. No 7. Pp. 873–889.
18. *Janssen G.* Public information arrival and volatility persistence in financial markets // *The European Journal of Finance*. 2004. 10. Pp. 177–197.
19. *Kalev P. S., Liu W.-M., Pham P. K., Jarnecic E.* Public information arrival and volatility of intraday stock returns // *Journal of Banking and Finance*. 2004. 28(6). Pp. 1447–1467.
20. *Kon, S. J.* Models of Stock Returns—A Comparison // *Journal of Finance*. 1984. 39(1). Pp. 147–165.
21. *Laakkonen H., Lanne M.* Asymmetric News Effects on Exchange Rate Volatility: Good vs. Bad News in Good vs. Bad Times // *Studies in Nonlinear Dynamics & Econometrics*. 2009. 14.

22. *Lamoureux C. G., Lastrapes W. D.* Heteroskedasticity in stock return data: Volume versus GARCH effects // *Journal of Business & Economic Statistics*. 1990. 2. P. 253–260.
23. *Lozinskaia A., Saltykova A.* Fundamental Factors Affecting the MOEX Russia Index: Retrospective Analysis // *CEUR Workshop Proceedings*. 2019. 2479. P. 32–45.
24. *Mitchell M. L., Mulherin J. H.* How markets process information: News releases and volatility // *Journal of Finance*. 1994. 49. P. 923–950.
25. *Mitra L., Mitra G.* Applications of news analytics in finance: A review // *The Handbook of News analytics in finance*. 2011. P. 1–36.
26. *Miyakoshi T.* ARCH versus information-based variances: Evidence from the Tokyo stock market // *Japan and the World Economy*. 2002. 14. P. 215–231.
27. *Najand M., Yung K.* A GARCH examination of the relationship between volume and variability in futures markets // *The Journal of Futures Markets*. 1991. 11. P. 613–621.
28. *Rabbi A. M. F.* Mass media exposure and its impact on fertility: Current scenario of Bangladesh // *Journal of Scientific Research*. 2012. T. 4. №. 2. C. 383–383.
29. *Ragunathan V., Peker A.* Price variability, trading volume and market depth: Evidence from the Australian futures market // *Applied Financial Economics*. 1997. 7. P. 447–454.
30. *Rubtsov B., Annenskaya N.* Factor analysis of the Russian stock market // *Journal of Reviews on Global Economics*. 2018. Issue 7 (Special Issue). P. 417–425.
31. *Sanjiv R.* News analytics: Framework, techniques, and metrics // *The Handbook of News analytics in finance*. 2011. P. 41–69.
32. *Tetlock Paul C.* Giving Content To Investor Sentiment: The Role Of Media In The Stock Market // *The Journal Of Finance*. 2007. Vol. Lxii. P. 1139–1168

Дата поступления рукописи: 25.02.2022 г.

ABOUT THE AUTHORS

Gavrilov Vadim – graduate student of the Economics Faculty of the M.V. Lomonosov Moscow State University, Moscow, Russia
vad093@mail.ru

Ivanov Mikhail Aloisovich – graduate student of the Economics Faculty of the M.V. Lomonosov Moscow State University, Moscow, Russia
mia.m5@yandex.ru

Klachkova Olga Alexandrovna – Cand. Sci. (Econ.), Associate Professor at the Department of Mathematical Methods of Analysis of Economics of the Economics Faculty of the M.V. Lomonosov Moscow State University, Moscow, Russia
sparrow889@gmail.com

Korolev Viktor Yurievich – Dr. Sci. (Phys.-Math.), Professor, Head of the Statistics Department of the Faculty of Computational Mathematics and Cybernetics of the M.V. Lomonosov Moscow State University, Moscow, Russia
vkorolev@cs.msu.ru

Roshchina Yanina Alexandrovna – Cand. Sci. (Econ.), Associate Professor at the Department of Mathematical Methods of Analysis of Economics of the Economics Faculty of the M.V. Lomonosov Moscow State University, Moscow, Russia
janina-d@yandex.ru

IMPACT OF NEWS FLOWS ON COMPONENTS OF RUSSIAN STOCK MARKET VOLATILITY

The paper simulates the volatility of securities returns on the Russian stock market depending on thematic news flows, entering the market, applying conditional autoregressive heteroscedasticity models. To take into account the news background, the model includes a numerical indicator, characterizing the number of news on each of the key topics as an independent regressor. The selection of topics and the construction of such an indicator are carried out by natural language processing methods. To assess the impact of the news background not on the volatility of returns as a whole, but on its components, the standard GARCH models are subject to assumptions that random errors are a mixture of two normal distributions. It was shown that one of the components has a significantly larger weight but smaller volatility. Our interpretation is «common» themes form the usual news background and have a little effect on volatility, while more rare specific themes (and so more informative) have a stronger impact on volatility.

Keywords: *stock market, news analytics, volatility components, natural language processing.*

JEL: C32, C53, G17.