

Экономический журнал ВШЭ. 2019. Т. 23. № 2. С. 264–289.
HSE Economic Journal, 2019, vol. 23, no 2, pp. 264–289.

О влиянии тональности новостей в международных СМИ на рыночный курс российского рубля: текстовый анализ¹

Афанасьев Д.О., Федорова Е.А., Рогов О.Ю.

В данной работе исследуется влияние тональности освещения России в международных средствах массовой информации (СМИ) на рыночный курс рубля к доллару США. В качестве источника новостей использовались данные международного новостного агентства Thomson Reuters в период с 10.01.2012 г. по 30.05.2018 г. (всего 162259 новостей о России), объединенные в пять тематических групп: бизнес, рынок, мировые события, политика и общая тематика. Для оценки тональности текстов использовался метод «мешка слов» и пяти различных словарей, а также учитывалась инерционность влияния новостного потока на поведение участников рынка. Эконометрическая методология исследования состояла из двух этапов: отбор потенциально значимых объясняющих переменных на базе метода эластичной сети и оценка параметров модели ARMAX-GARCH. В результате моделирования подтверждено, что динамика цены на нефть значимо влияет на рыночный курс рубля, в то время как косвенный инструмент денежно-кредитной политики Банка России – ставка межбанковского кредитования RUONIA – не была эффективна в изученный период. Тональность новостей, наряду с фундаментальными экономическими факторами, также оказывает систематическое влияние на курс рубля, которое тем не менее зависит от их тематики: наибольшее значение имеют новости бизнес-тематики, в то время как политические сообщения в СМИ не оказывают статистически значимого влияния на рыночный курс рубля. Полученные результаты потенциально применимы

¹ Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финансового университета на 2018 г.

Позиция Афанасьева Д.О., отраженная в данном исследовании, не является официальной позицией АО «Гринатом» и может не совпадать с ней.

Позиция Рогова О.Ю., отраженная в данном исследовании, не является официальной позицией Innoscripta GmbH и может не совпадать с ней.

Афанасьев Дмитрий Олегович – архитектор информационных систем АО «Гринатом». E-mail: dmafanasyev@gmail.com

Федорова Елена Анатольевна – д.э.н., профессор департамента корпоративных финансов и корпоративного управления Финансового университета при Правительстве РФ. E-mail: ecolena@mail.ru

Рогов Олег Юрьевич – м.н.с. Научно-исследовательского центра Учебно-научно-производственного комплекса Московского физико-технического института. E-mail: olegrg@yandex.ru

Статья поступила: 27.03.2019/Статья принята: 07.05.2019.

при прогнозировании динамики курса рубля инвесторами, Банком России, коммерческими банками, инвестиционными фондами и другими профессиональными участниками рынка.

Ключевые слова: текстовый анализ; курс рубля; влияние новостей; тональность новостей.

DOI: 10.17323/1813-8691-2019-23-2-264-289

1. Введение

Прогнозирование обменных курсов имеет особое значение как для инвесторов, так и политиков. Точные прогнозы обменных курсов позволяют инвесторам разрабатывать эффективные торговые стратегии и хеджировать валютные риски. Поведение обменных курсов также важно для центральных банков стран при формировании ключевых ставок. Курс обмена валюты играет значительную роль в международной торговле, инвестициях, управлении рисками в компаниях и экономической ситуации в стране за счет влияния на платежный баланс. Многие страны ввели свободно плавающий курс обмена валюты, опирающийся на механизмы рынка, позволяющий корректировать курс собственной валюты. В настоящее время вмешательство центральных банков стран на валютном рынке являются не очень частыми и одновременно редко успешными событиями. Неэффективность действий такого характера объясняется рыночными силами, влияющими на обменный курс. Система плавающего обменного курса сама по себе отражает экономические основы, но при этом флуктуация курса валют трудно интерпретируется из-за неполной информации о факторах, влияющих на них. Объяснение поведения обменного курса по-прежнему остается одной из самых важных областей в международной экономике.

В данном исследовании, наряду с фундаментальными экономическими детерминантами курса российского рубля, мы также рассматриваем информационный медиапоток и его эмоциональную окраску как фактор, потенциально влияющий на конъюнктуру валютного рынка. Когда аналитики, инвесторы и институциональные трейдеры оценивают текущие цены на финансовые активы (акции, курс валюты), новости играют важную роль в процессе оценки. Фактически новости содержат фундаментальную информацию, влияющую на ожидания участников рынка. Если финансовые новости содержат принципиально новую информацию, то они приводят к корректировке относительно ожидаемых будущих денежных потоков инвестиций или ставок дисконтирования, что влияет на доходность финансовых активов [Chatrath et al., 2014]. Аналогичное влияние потенциально существует и для обменного курса российского рубля по отношению к доллару. Учитывая это, целью данного исследования является оценка влияния тональности освещения Российской Федерации в международных средствах массовой информации (СМИ) на курс рубля.

Статья организована следующим образом. Второй раздел посвящен обзору литературы. В третьем разделе описывается методология исследования и проводится предварительный анализ данных. В четвертом разделе обсуждаются полученные результаты. Пятый раздел содержит ключевые выводы и завершает работу.

2. Обзор литературы

Временные ряды валютных курсов характеризуются высокой сложностью и нелинейностью, поскольку подвержены влиянию многочисленных нестабильных факторов, включая экономические условия и политические события. Разработка точных методов прогнозирования имеет большое значение и сможет предоставить инвесторам необходимые основы для разработки стратегий и хеджирования рисков. В настоящее время вопрос о том, как точно прогнозировать курсы валют, остается также открытым в отношении экономической и социальной организации современного общества. После ухода от Бреттон-Вудской системы фиксированных обменных курсов в начале 1970-х годов прогнозирование стоимости валюты стало критически важным для многих целей, таких как международные сопоставления доходов, доходов и стоимости жизни международных агентств, выравнивание обменных курсов правительствами и принятие корпоративных финансовых решений.

Можно выделить два основных направления повышения точности прогнозов курса валюты: 1) непосредственное развитие методологии прогнозирования обменных курсов; 2) анализ и учет внешних факторов, потенциально влияющих на динамику обменного курса.

В рамках *первого направления* в работе [Clements et al., 2010] были выделены два концептуальных подхода к построению методологии прогнозирования обменного курса.

1. Подход, основанный на опросах инвесторов. Иногда вместо опросов применяют некоторый «заменитель» валюты. Например, в основе исследования [Clements et al., 2010] лежит теория паритета покупательской способности и используется индекс Биг Мака, отражающий стоимость предельно однородной и типизированной продукции в двух выбранных странах. Авторы показали, что применение данного индекса приводит к более высокой прогностической способности, чем использование индекса потребительских цен.

2. Подход, основанный на линейных и нелинейных математических моделях временных рядов. Наиболее распространенной здесь является линейная модель ARIMA, которая была использована, например, в работе авторов [Ahmad et al., 2015] для прогнозирования валютного курса Индонезии. В то же время все большее применение находят и более сложные нелинейные техники. Так, для прогнозирования обменных курсов USD/INR, USD/EURO, YEN/INR и SGD/INR авторы в работе [Das et al., 2019] используют метод опорных векторов, дополненный процедурой отбора предикторов OS-ELM (On-Line Sequential Extreme Learning Machine). Данный подход сравним по затрачиваемым ресурсам с классическим методом опорных векторов и рекуррентными нейронными сетями, но показывает более высокую точность прогнозирования. В работе [Kawakami, 2013] используется попарное сравнение прогнозов ежемесячных курсов иены и доллара. Одним из перспективных методов прогнозирования является подход, основанный на разложении временных рядов на моды, характеризующие отдельные временные масштабы. Так, он используется в работе [Sun et al., 2017] совместно с методом опорных векторов и нейронными сетями как инструментами, позволяющими учесть высокую сложность и нелинейность поведения валютного курса. Также в исследовании [Mikova et al., 2019] доказано, что инвестиционная стратегия на валютном рынке является более эффективной при добавлении волатильности. Что касается отечественных исследований, то для прогнозирования курса валюты здесь используются, например, нейронные сети [Федорова, Линкова, 2013] и DSGE-модели с малым количеством уравнений [Крепцев, Селезнев, 2018].

Повысить точность прогноза валюты потенциально можно не только за счет непосредственного применения более сложной математической методологии, но и за счет выбора и учета различных внешних факторов, влияющих на курс валюты, что составляет *второе направление* из упомянутых выше. В исследовании [Morales-Arias et al., 2013] авторы принимают во внимание макроэкономические факторы, показатели экономического цикла, волатильность и доходность фондовых индексов. Совокупность этих факторов обеспечивает достаточно высокий уровень точности прогнозирования курса валюты. Авторы в работе [Sarno, Sojli, 2009] используют факторы на макро- и микроуровнях экономики и исследуют их применение для прогнозирования обменного курса валюты. Исследователи в работе [Yin, Li, 2014] предлагают безарбитражную интернациональную модель, которая показала сильную взаимосвязь между динамикой валютного курса (евро и доллар) и макроэкономическими показателями (в частности, доходность бескупонных облигаций).

Помимо имеющих экономическую природу факторов, влияние на обменные курсы способны оказывать информационные потоки, роль которых в последнее время исключительно возросла. Инвесторы могут использовать новости официальных агентств, сторонних источников – информационных посредников (например, финансовых аналитиков, финансовых консультантов, деловую прессу, публикации агентств кредитного рейтинга, оценки аудиторов) для получения своевременной и значимой информации о перспективах поведения цен финансовых активов. Однако в последнее десятилетие произошло появление новых источников информации, которые легко доступны для большинства участников рынка. С развитием интернета частные инвесторы все больше полагаются и учитывают мнение друг друга. Например, интернет-сервисы Yahoo! Finance, Silicon Investor Raging Bull и Twitter позволяют пользователям мгновенно публиковать свои взгляды и оценки для широкой аудитории. Такой канал передачи информации может оказать существенное влияние на ожидания участников рынка, и этот аспект отражен в довольно значительном количестве работ.

Авторы исследования [Beckmann, Czudaj, 2017] анализируют взаимосвязь публичных выступлений политиков и валютного курса, а основные их результаты говорят в пользу сильного влияния политической неопределенности и решений регуляторов (центральных банков) на курс иены к доллару. Исследование [Ederington et al., 2019] рассматривает влияние официальных отчетов об уровне безработицы в США на обменные курсы валют Pound/Dollar, Yen/Dollar, и Euro/Dollar. Авторами было показано, что реакция рынка опережает официальную дату публикации. В работе [Khadjeh Nassirtoussi et al., 2014] предлагается подход к прогнозированию дневного курса валют, основанный на анализе заголовков новостей с пометкой «срочно», изучается возможность сбора релевантной информации в новостном потоке и показывается преимущество предложенного подхода в смысле средней абсолютной ошибки (MAE). В исследовании [Lüdering, Tillmann, 2018] анализируется обсуждение монетарной политики США в Твиттере и отслеживается влияние таких обсуждений на показатели долговых инструментов страны. Особое внимание уделяется 2013 г., когда интенсивно велась дискуссия на тему ключевой ставки Федерального резерва, было собрано и проанализировано 90000 публикаций в Твиттере по данной тематике и с применением тематического моделирования (латентное размещение Дирихле), вычислены показатели частоты отдельных ключевых слов тематических групп внутри выборки. Отдельные выбранные темы далее анализируются численно с

помощью модели векторной авторегрессии (VAR) для оценки влияния таких групп на цены долговых бумаг. Обнаружено, что обсуждения в социальной сети по теме содержат количественные оценки, показано влияние текстовых численных параметров на цену долговых инструментов. В статье [Shen et al., 2019] рассматривается взаимосвязь между курсом биткойна, объемом торгов на криптобирже, волатильностью и вниманием инвесторов. «Текстовые» показатели рассчитываются методом сентимент-анализа на основе публикаций в Твиттере и сравниваются с показателями популярности тем в системе Google Trends. Авторами предложен подход оценки влияния числа публикаций в Твиттере на объем торгов следующего дня, а также волатильность. Взаимосвязь исследована с помощью линейных и нелинейных тестов причинности по Гренджеру.

В нашем исследовании мы также предполагаем, что существует систематическое влияние информационного (новостного) фона на курс валюты, но исследуем его воздействие на обменный курс именно российского рубля. Данная тема раньше практически не затрагивалась в контексте российской экономики, что, по нашему мнению, говорит об актуальности исследуемого вопроса. Кроме того, вклад нашей работы заключается в следующем.

1. В современной практике авторы обычно используют только один словарь для оценки сентиментальной окраски новостных текстов среди множества доступных на сегодняшний день. Это может приводить к определенным смещениям в получаемых результатах. В нашем исследовании мы применяем сразу пять наиболее популярных словарей, чтобы повысить объективность выводов, а также выполнить сравнение словарей.

2. Большинство работ охватывают только отдельный сегмент новостей, в то время как мы используем данные агрегатора Thomson Reuters, который включает новости таких изданий, как The New York Post, CNN, Breitbart, Reuters, Fox, Atlantic, The Washington Post и BuzzFeed. Мы также объединяем отдельные категории новостей в пять крупных тематических групп, чтобы исключить частную направленность отдельных статей.

3. В предыдущих работах оценивается влияние новостей только в момент их публикации. В данном исследовании мы предполагаем, что это влияние инерционно, а его убывание происходит линейным образом в течение определенного периода времени (рассмотрены пять вариантов продолжительности данного периода, включая отсутствие инерционности).

4. Принимая во внимание, что количество объясняющих переменных достаточно велико (за счет использования пяти словарей и рассмотрения пяти вариантов затухания влияния), мы применяем метод эластичной сети для отбора объясняющих переменных, влияние которых потенциально статистически отлично от нуля. Для дальнейшей проверки наличия систематического влияния мы также учитываем гетероскедастичность относительного изменения рыночного курса рубля.

3. Методология и данные

В данном исследовании мы рассмотрели период с 10.01.2012 г. по 30.05.2018 г. (всего 2333 календарных дней), который можно охарактеризовать достаточно высокой турбулентностью в области международных отношений России с иностранными государствами, введением с марта 2014 г. антироссийских санкций и последующим «санкционным бумом» (не прекращающимся до сегодняшнего дня), а также не самой простой ситуацией для российского рубля.

В качестве моделируемой переменной выступал рыночный курс российского рубля к доллару США. Для этого мы использовали спотовый курс по данным Банка Англии (база данных Quandl, код временного ряда BOE/XUDLBK69, <https://www.quandl.com/data/BOE/XUDLBK69-Spot-Exchange-Rate-Russian-Ruble-Into-Dollar>). В отличие от официального курса ЦБ РФ, данный показатель отражает характерную дневную стоимость рубля, которая сформировалась непосредственно на международном валютном рынке. Пропуски в данных были заполнены значениями предыдущего дня.

В свою очередь объясняющие переменные были разделены на два вида: фундаментальные факторы, которые имеют экономическую природу, и сентиментальные переменные, отражающие тональность новостей о России в международных СМИ. Остановимся более подробно на каждом из них, после чего обсудим итоговую эконометрическую спецификацию модели обменного курса.

3.1. Фундаментальные (экономические) факторы

Одним из общепризнанных факторов, существенно влияющих на курс рубля по отношению к доллару, является динамика цены на нефть. При этом наблюдается их отрицательная корреляция: при росте мировых цен на нефть курс рубля в долларах, в среднем, уменьшается, т.е. происходит укрепление российской валюты. Принимая это во внимание и основываясь на результатах предыдущих исследований, мы включили в нашу модель в качестве объясняющей переменной котировки нефти и ожидаем статистическую значимость коэффициента при ней, а также его отрицательный знак. В качестве цены на нефть мы использовали данные о стоимости нефтяной корзины ОПЕК (база данных Quandl, код временного ряда ОПЕК/ORB, <https://www.quandl.com/data/ОПЕК/ORB-ОПЕК-Crude-Oil-Price>), которая зачастую выступает как референсная цена и измеряется в долларах за баррель. По аналогии с обменным курсом пропущенные наблюдения были заполнены значениями цены предыдущего дня.

С ноября 2014 г. Банк России ввел свободно плавающий режим формирования обменного курса рубля, отказавшись от его ограничения валютным коридором, которое использовалось до указанного момента времени. Вместе с тем у Банка России существует ряд косвенных инструментов, которые применяются на практике и позволяют в определенной степени смягчить обесценивание рубля в моменты значительной его девальвации. В частности, таким инструментом выступает ставка межбанковского кредитования RUONIA (Ruble OverNight Index Average), представляющая собой индикативную взвешенную ставку однодневных рублевых кредитов и отражающая оценку стоимости необеспеченного рублевого заимствования на условиях «overnight» для заемщика из числа российских банков с минимальным кредитным риском. Учитывая это, мы также включили в нашу модель ставку RUONIA в качестве объясняющей переменной. Соответствующие данные были получены нами с официального сайта, содержащего основную информацию о данном косвенном инструменте денежно-кредитной политики Банка России, а также архив исторических значений (<http://www.ruonia.ru/eng/archive.html>). Опять же, пропуски во временном ряду были заменены на значение первого предшествующего доступного наблюдения.

Рисунок 1 показывает динамику упомянутых выше экономических показателей в рамках периода исследования. Визуальный анализ показывает, что временные ряды яв-

ляются нестационарными. Этот же вывод подтверждается и формальными статистическими тестами ADF и KPSS (с константой и трендом) для переменных в уровнях (см. табл. 1). Проверка коинтеграции обменного курса, цены нефти и ставки RUONIA с помощью теста Йохансена (с константой и трендом) [Johansen, 1995] не подтвердила наличие между ними долгосрочной взаимосвязи на 5-процентном уровне, так как ни одно значение статистики теста не превышает соответствующее критическое значение (см. табл. 2).

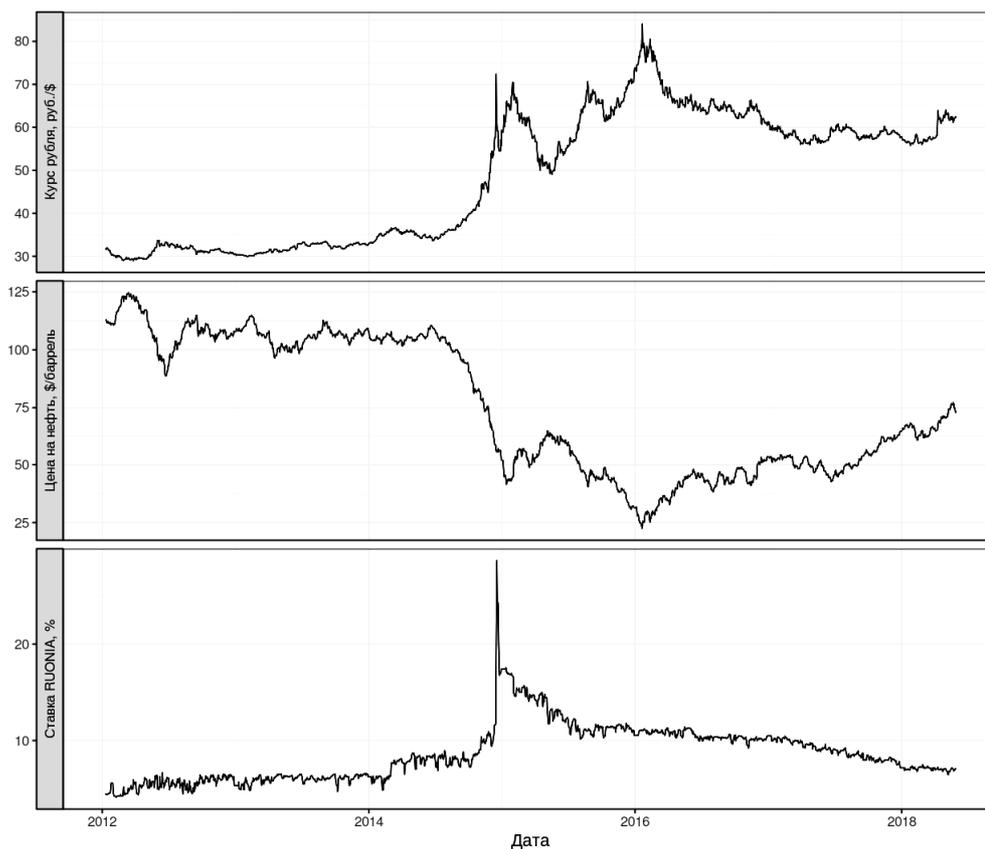


Рис. 1. Обменный курс рубля, стоимость нефтяной корзины ОПЕК и ставка межбанковского кредитования RUONIA в период с 10.01.2012 г. по 30.05.2018 г.

Принимая во внимание полученные результаты тестирования стационарности и коинтеграции временных рядов в уровнях, мы выполнили для них следующие преобразования: перешли к логарифмированным доходностям для обменного курса и цены на нефть (первые разности логарифмов) и приращениям ставки RUONIA. Тестирование указанных преобразованных переменных показало их стационарность на 1-процентном уровне в соответствии с ADF-тестом и 5-процентном уровне для KPSS-теста (см. табл. 1).

Таблица 1.

Проверка стационарности временных рядов экономических показателей, включаемых в модель обменного курса рубля, с использованием тестов ADF и KPSS (с константой и трендом)

Временной ряд	ADF			KPSS		
	статистика	p-значение	лаг	статистика	p-значение	лаг
Исходные данные в уровнях						
Курс рубля	-1,408	0,83	13	2,484	0,01	11
Цена на нефть	-0,751	0,97	13	2,564	0,01	11
Ставка RUONIA	-2,029	0,57	13	3,106	0,01	11
Преобразованные данные						
Курс рубля (лог-доходность)	-12,290	0,01	13	0,152	0,05	11
Цена на нефть (лог-доходность)	-11,395	0,01	13	0,151	0,05	11
Ставка RUONIA (приращение)	-16,647	0,01	13	0,019	0,10	11

Примечания. Нулевая гипотеза ADF-теста – наличие хотя бы одного единичного корня в модели с константой и трендом. Нулевая гипотеза KPSS-теста – временной ряд стационарен относительно тренда.

Таблица 2.

Результаты проверки коинтеграции обменного курса рубля к доллару США, цены нефтяной корзины ОПЕК и ставки межбанковского кредитования RUONIA с использованием теста Йохансена (с константой и трендом)

Число коинтеграционных векторов n	Статистика теста Йохансена	Критические значения		
		10%	5%	1%
2	2,46	10,49	12,25	16,26
1	13,13	22,76	25,32	30,45
0	25,84	39,06	42,44	48,45

Примечания. Нулевая гипотеза теста Йохансена – ранг коинтеграции r (число коинтеграционных векторов) не превышает заданное целое число n (где n меньше, чем число переменных). Критические значения теста были получены из работы [Osterwald-Lenum, 1992]. Количество учитываемых в тесте лагов составило 17 и было выбрано на основании информационных критериев Акаике (AIC), Шварца (SC) и Ханнана – Куинна (HC), рассчитанных для VAR-модели, включающей все экономические переменные.

Для дальнейшей корректной спецификации эконометрической модели мы дополнительно проверили наличие автокорреляции и ARCH-эффекта (т.е. присутствие гетероскедастичности) в логарифмированных доходностях обменного курса российского рубля к доллару США. Для проверки сериальной зависимости доходностей была рассчитана автокорреляционная функция и выполнен Q-тест Льюнга – Бокса с семью лагами². Анализ результатов показал, что автокорреляционные коэффициенты статистически незначимы на 5-процентном уровне, а гипотеза об отсутствии автокорреляции не отклоняется Q-тестом на 5-процентном уровне ($\chi^2 = 8,7863$, p -значение = 0,27). Вследствие этого, мы не стали включать в модель лаги объясняемой переменной. Для проверки наличия ARCH-эффекта мы выполнили Q-тест Льюнга – Бокса для квадратов и абсолютных значений доходностей с семью лагами. В обоих случаях гипотеза об одновременном равенстве нулю авторегрессионных коэффициентов не отклоняется на 5-процентном уровне значимости ($\chi^2 = 931,4$ и $\chi^2 = 884,3$ соответственно), что говорит о присутствии зависимости дисперсии доходности обменного курса от времени. Учитывая это, в дальнейшем мы включили в модель GARCH-компонент [Bollerslev, 1986].

3.2. Сентиментальные переменные

Для оценки сентиментальной окраски (эмоциональной тональности) новостей о России в международных СМИ мы применили широко распространенный подход на основе «мешка слов» или лексикона. Суть данного подхода заключается в том, что каждое слово в тексте классифицируется как позитивное, нейтральное или негативное на основе заранее сформированного перечня слов, который и образует лексикон («мешок слов», словарь, библиотеку). Данный словарь содержит как непосредственно сами униграммы (отдельные слова), так и коэффициенты, характеризующие их тональность. При этом можно выделить два основных вида применяемых на практике словарей: в первом, полностью дискретном, виде полярность слова выражена только значениями -1, 0, 1 (отрицательное, нейтральное и положительное слово соответственно; иногда значение 0 может отсутствовать), в то время как во втором, взвешенном, виде она принимает в том числе дробные значения, отражающие степень полярности сентимента, т.е. ее вес.

В данной работе мы использовали $L = 5$ словарей, которые достаточно часто встречаются в ранее выполненных исследованиях. Таблица 3 содержит их основные характеристики, а также источники, где можно получить указанные библиотеки. Оценка сентиментальной окраски выполнялась по всем указанным словарям, так как каждый из них состоит из различного набора слов, учитывает несколько отличающиеся тематики, а априорный выбор лишь одного из них несет в себе риск получения некорректных результатов.

² Значение в семь лагов было выбрано для учета возможного наличия недельной сезонности (периодичности) в доходностях обменного курса рубля. Учитывая природу данного временного ряда, выбор большего значения лагов видится экономически необоснованным.

Таблица 3.

**Основные характеристики словарей,
использованных для анализа эмоциональной окраски новостных текстов**

Словарь	Обозначение	Количество слов		
		негативные	нейтральные	позитивные
Harvard General Inquirer (Harvard IV-4 объединенный с Laswell) [Stone et al., 1966]*	GI	3695	0	2550
[Loughran, McDonald, 2011]**	LM	2350	0	352
[Mohammad, Turney, 2010]***	NRC	3241	0	2227
SentiWordNet 3.0 [Baccianella et al., 2010]****	SWN	11029	166	8898
Объединенный словарь [Jockers, 2017] и [Hu, Liu, 2004], дополненный [Rinker, 2018]*****	JR	7818	13	3878

Источники словарей: * – <http://www.wjh.harvard.edu/inquirer/homecat.htm>, ** – <https://sraf.nd.edu/textual-analysis/resources/#Master%20Dictionary>, *** – <http://www.purl.com/net/lexicons>, **** – <http://sentiwordnet.isti.cnr.it>, ***** – <http://github.com/trinker/lexicon>.

Оценка по тому или иному лексикону, который мы далее будем обозначать через l , тональности конкретного текста n , опубликованного в дату t , рассчитывалась нами как разница между числом положительных $N_{n,t,l}^+$ и отрицательных $N_{n,t,l}^-$ слов, нормированная на общее количество слов в тексте ($N_{n,t,l}^0$ – количество нейтральных слов):

$$(1) \quad S_t^{n,l} = \frac{N_{n,t,l}^+ - N_{n,t,l}^-}{N_{n,t,l}^+ + N_{n,t,l}^0 + N_{n,t,l}^-}.$$

Агрегирование значений тональности отдельных новостей в рамках дня мы выполнили усреднением с весами $w_{n,t}$, пропорциональными длине текста:

$$(2) \quad S_t^l = \sum_n w_{n,t} S_t^{n,l}.$$

Таким образом, эмоциональная окраска новости с большим количеством слов оказывает более существенное влияние на агрегированную в рамках дня эмоциональную тональность S_t^l .

В качестве наиболее простого случая мы использовали для моделирования обменного курса рубля непосредственно показатель S_t^l , что подразумевает воздействие агре-

гированной (усредненной) тональности новостей на поведение экономических агентов только непосредственно в день публикации сообщения. Однако, следуя работе [Ardia et al., 2019], мы также полагаем, что это влияние инерционно и может сохраняться в течение определенного периода времени³. Поясним это более подробно на частном примере. Предположим, что в день t_0 опубликована всего одна новость n , имеющая исключительно отрицательную окраску по лексикону l (т.е. $S_{t_0}^l = S_{t_0}^{n,l} = -1$). Если в последующие τ дней (где τ – период «затухания») публикации отсутствуют, то воздействие указанной новости на экономических агентов будет постепенно сходиться на нет ($S_{t_0+\tau}^l = 0$), так как она будет терять свою актуальность. Если в течение указанного периода τ в момент времени $t_1 \leq t_0 + \tau$ появляется еще одна новость m , например, с исключительно положительной окраской (т.е. $S_{t_1}^l = S_{t_1}^{m,l} = +1$), то ее тональность также полагается убывающей на промежутке времени с t_1 по $t_1 + \tau$. При этом на перекрывающемся периоде от момента t_1 до $t_0 + \tau$ учитывается влияние как новости n , так и новости m в виде усреднения соответствующих значений тональностей. Данный частный пример без труда обобщается на случай публикации более одной новости в течение каждого дня.

Мы полагаем, что убывание влияния сентиментальной окраски новостей происходит линейным образом в течение периода τ , т.е. вес каждого последующего наблюдения после момента t_0 определяется по формуле ($t = 1, \dots, \tau - 1$):

$$(3) \quad v_t = 1 - \frac{t}{\tau}.$$

При этом, так как величина периода затухания τ нам априори неизвестна, мы рассмотрели несколько его значений: три дня, одна неделя, две недели и один месяц. Временной ряд агрегированной в рамках дня t сентиментальной окраски новостей, полученный с использованием лексикона l и схемы затухания \mathcal{U} для периода τ , будем далее обозначать как $S_t^{l,\tau}$, опуская индекс \mathcal{U} , так как рассматривается только линейное убывание.

В качестве источника новостей в средствах массовой информации мы использовали данные одного из крупнейших международных новостных агентств – Thomson Reuters (<https://www.reuters.com>). Всего за исследуемый период (с 10.01.2012 г. по 30.05.2018 г.) было опубликовано 162259 новостей, которые так или иначе связаны с событиями в Российской Федерации⁴. При этом каждая из новостей имеет признак тематики, к которой она относится, а общее количество тематических категорий составляет 290. Мы исключили те из них, которые по своей сути маловероятно оказывают влияние на обменный курс рубля, в частности, новости спорта, кино, медиа, шоу-бизнеса, технологий, здоровья,

³ Отметим, что в указанной работе речь идет о влиянии тональности новостей СМИ на макроэкономические показатели, однако это не меняет сути обсуждаемого предположения.

⁴ Отметим, что для сопоставления временных шкал мы преобразовали время публикации новостей в московский часовой пояс.

окружающей среды, искусства, науки и др. Также были исключены «дублирующиеся» новости, возникающие в случае корректировок опечаток и т.п. и последующей их повторной публикации. В результате указанной предобработки мы сформировали набор из 101096 релевантных новостей, которые классифицируются по 185 темам.

Однако многие из оставшихся категорий новостей имели достаточно частную направленность, что не позволяло оценить их интегральное влияние на поведение экономических агентов. В целях обобщения мы объединили данные категории в пять крупных тематических групп:

- 1) Business – новости различных индустрий, банковские, финансовые, крупные сделки, слияния и поглощения, инвестиционные обзоры и форумы, корпоративный сектор, общеэкономические и макроэкономические новости;
- 2) Market – новости различных рынков, включая товарные, валютные, фондовые, энергетические, розничные и долговые, рыночные отчеты различных стран;
- 3) World – новости международных событий и отношений, а также конкретных регионов или стран;
- 4) Policy – новости политического характера, включая конкретных политиков, а также события (саммит G20, форум в Давосе и т.д.);
- 5) Common – новости общего характера, которые не удалось отнести по смысловой составляющей ни к одной из вышеперечисленных групп.

Таблица 4 показывает основные характеристики указанных групп, а также обобщенные значения для всего набора данных. Можно отметить некоторую неравномерность распределения количества новостей по группам. Так, в частности, новости группы Market публикуются наиболее часто: в среднем по 14,1 статьи в день. При этом новости групп Policy и Common появляются реже всего, а именно в среднем 4,3 и 4,5 статьи в день. Суммарно по всем группам средняя частота публикации составляет 43,3 новости в день.

Таблица 4.

**Основные показатели тематических групп новостей,
в рамках которых выполнялся расчет сентиментальной окраски**

Показатель	Тематические группы новостей					Итого
	Business	Market	World	Policy	Common	
Количество категорий новостей	46	48	33	25	33	185
Количество новостей	19663	32951	28104	9974	10404	101096
Среднее количество новостей в день	8,4	14,1	12,0	4,3	4,5	43,3

По каждой группе новостей мы рассчитали временные ряды сентиментальной окраски с использованием пяти упоминавшихся ранее словарей. Таким образом были сформированы $l \cdot q = 25$ сентиментальных переменных $S_t^{l,q}$, где q – индекс тематической группы новостей. ADF-тест (с константой, но без тренда) показал, что гипотеза о наличии единичного корня для каждой из них отклоняется на 5-процентном уровне, что позволяет говорить об их стационарности и включать в итоговую эконометрическую модель без

каких-либо дополнительных преобразований. По этой же причине мы не включали временные ряды тональностей новостей в анализ коинтеграции, так как в отличие от экономических показателей в них отсутствует долгосрочная динамика (см. рис. 1 и рис. 2).

Обсудим кратко возможную проблему эндогенности данных в нашем исследовании. С одной стороны, тональность новостей о России в международных СМИ может оказывать воздействие на поведение экономических агентов, принимающих решение о покупке или продаже российского рубля на валютном рынке, т.е. влиять на динамику курса. С другой стороны, ситуация на рынке также может сказываться на эмоциональной окраске сообщений в СМИ. Однако важно отметить, что новостной поток в данном случае будет запаздывать относительно событий на рынке, а его смысловое содержание будет отражать фактически произошедшие события. Проще говоря, сообщения о состоянии валютного рынка в новостных источниках публикуются либо после завершения торгового дня, либо только на следующий день. Кроме того, в рассматриваемых нами тематических группах новостей содержатся сообщения о значительном количестве различных событий, которые по своему смыслу являются экзогенными относительно валютного рынка. С учетом сказанного, мы считаем оправданным предположение об отсутствии влияния динамики валютного курса на тональность новостей «день в день», т.е. отсутствия эндогенности в данных.

3.3. Эконометрическая модель рыночного курса рубля

Построение модели рыночного курса российского рубля выполнялось нами с учетом полученных выше результатов предварительного анализа исследуемых временных рядов и включало два этапа.

На первом этапе из всего множества регрессоров (два фундаментальных и 25 сентиментальных) мы отбирали только те, которые потенциально значимы для объяснения логарифмированной доходности курса рубля. Данная необходимость была продиктована двумя причинами. Во-первых, это существенное количество объясняющих сентиментальных переменных, которые необходимо включить в модель, и как следствие, значительное число степеней свободы, теряемое при такой спецификации модели. Во-вторых, многие из полученных временных рядов эмоциональной тональности сообщений ввиду общности словарей и схемы затухания достаточно сильно коррелированы между собой, что приводит к проблеме мультиколлинеарности и, как следствие, неустойчивости оценок параметров модели к малым изменениям выборки.

Для отбора объясняющих переменных мы использовали регуляризацию посредством метода эластичной сети [Zou, Hastie, 2005]. Для классической линейной модели, заданной уравнением

$$(4) \quad y_t = \alpha_0 + \sum_k \alpha_k x_t^k + \varepsilon_t,$$

отбор регрессоров x_t^k методом эластичной сети сводится к оценке параметров данной модели через минимизацию следующего функционала (переменные предварительно должны быть стандартизированы):

$$(5) \quad \min_{\alpha_0, \alpha_k} \left\{ \frac{1}{T} \sum_{t=1}^T \left(y_t - \alpha_0 - \sum_k \alpha_k x_t^k \right)^2 + \right. \\ \left. + \gamma_1 \left(\gamma_2 \left[\alpha_0 + \sum_k \alpha_k \right] + (1 - \gamma_2) \left[\alpha_0^2 + \sum_k \alpha_k^2 \right] \right) \right\}.$$

Первое слагаемое здесь – классическая функция потерь для линейной регрессии, в то время как второе добавляет регуляризацию, вводя штраф за количество включенных в модель объясняющих переменных. Параметр γ_1 определяет общий коэффициент штрафа: для $\gamma_1 = 0$ мы получаем обычный МНК, а при $\gamma_1 \rightarrow \infty$ все коэффициенты модели стремятся к нулю; промежуточное значение позволяет отобрать только потенциально значимые переменные. В свою очередь, γ_2 определяет баланс между двумя «встроенными» в метод эластичных сетей техниками регуляризации: гребневая или Ridge [Hoerl, Kennard, 1970] регрессия (квадрат L_2 -нормы вектора коэффициентов модели) при $\gamma_2 = 0$ и LASSO [Tibshirani, 1996] регрессия (L_1 -норма вектора коэффициентов модели) при $\gamma_2 = 1$. Для выбора конкретных значений параметров γ_1 и γ_2 мы использовали поиск на сетке кандидатов, минимизируя информационный критерий Акаике (AIC), адаптированный для регрессий с большим количеством объясняющих переменных [Tibshirani, Taylor, 2012]. Параметр γ_2 варьировался от 0,1 до 0,9 с шагом 0,1, и для каждого γ_2 генерировалось 100 значений γ_1 в соответствии со стратегией, предложенной Фридманом и др. [Friedman et al., 2010]. Таким образом, выполнялся перебор 900 различных комбинаций параметров.

На втором этапе мы оценивали модель ARMAX(0,0) – GARCH(1,1), учитывающую ранее выявленное отсутствие автокорреляции, но наличие ARCH-эффекта во временном ряду доходностей рыночного курса рубля, а также результаты отбора объясняющих переменных с помощью метода эластичной сети:

$$(6) \quad y_t = \alpha_p p_t \cdot \mathbb{I}_p + \alpha_r r_t \cdot \mathbb{I}_r + \sum_l \sum_q \beta_{l,\tau,q} S_t^{l,\tau,q} \cdot \mathbb{I}_{l,\tau,q} + \sigma_t^2 \varepsilon_t, \\ \varepsilon_t \sim t(\mu = 0, v, \xi), \\ \sigma_t^2 = c + \phi_1 \varepsilon_{t-1}^2 + \psi_1 \sigma_{t-1}^2,$$

где y_t – логарифмированная доходность рыночного курса рубля; p_t – логарифмированная доходность нефтяной корзины ОПЕК; $S_t^{l,\tau,q}$ – агрегированная тональность новостей из группы тематических категорий q , рассчитанная по лексикону l для периода затухания влияния τ ; \mathbb{I}_x – индикаторная функция, принимающее значение единица, если соответствующая объясняющая переменная x была отобрана методом эластичной сети на предыдущем шаге, и ноль – в обратном случае; $\alpha_p, \alpha_r, \beta_{l,\tau,q}$ – коэффициенты ARMA-ком-

поненты модели; σ_t^2 – условная дисперсия инноваций; ε_t – стандартизированные инновации, имеющие скошенное распределение Стьюдента t с центром $\mu = 0$; ν – ширина распределения t ; ξ – асимметрия распределения t ; c, ϕ_1, ψ_1 – константа и коэффициенты GARCH-компоненты модели. Оценка параметров модели выполнялась посредством метода максимального правдоподобия. При этом выбор именно скошенного распределения Стьюдента для инноваций обоснован рядом численных экспериментов, которые мы провели, оценивая модель с различными распространенными формами распределений, после чего анализировались графики «квантиль-квантиль» остатков, а также сравнивались значения функции правдоподобия.

4. Результаты

Рассмотрим результаты эконометрического анализа влияния сентиментальной окраски освещения Российской Федерации в зарубежных средствах массовой информации на рыночный курс рубля к доллару США. Рисунок 2 показывает полученную динамику тональности новостей о России, усредненную по всем используемым в работе словарям (GI, JR, LM, NRC, SWN) в разрезе пяти тематических групп: Business, Market, World, Policy, Common (для случая отсутствия затухания влияния новостей).

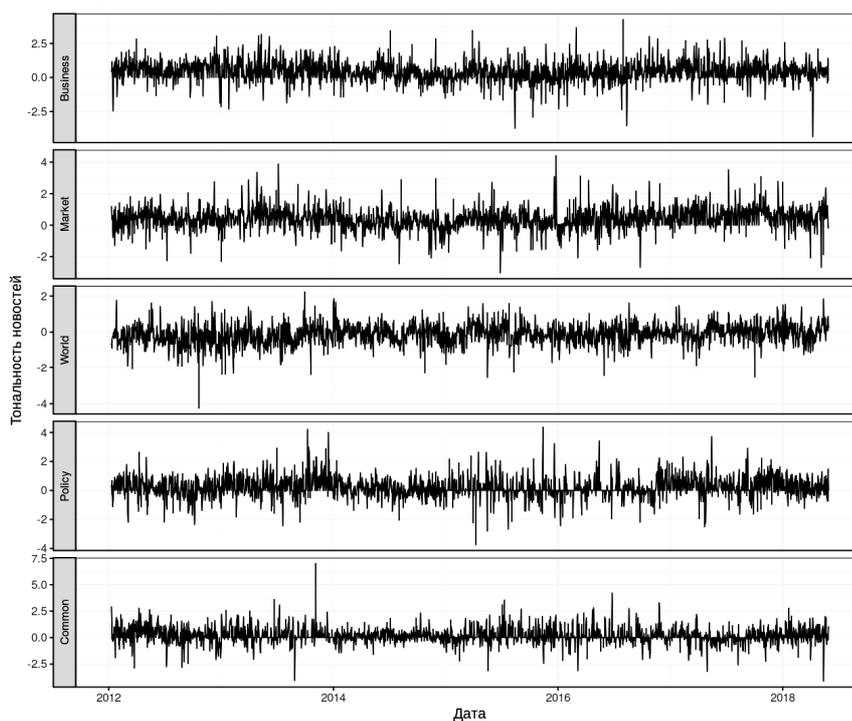


Рис. 2. Сентиментальная окраска тематических групп новостей о Российской Федерации в международных СМИ (для случая отсутствия затухания), усредненная по лексиконам GI, JR, LM, NRC, SWN (для наглядности значения умножены на 100)

Таблица 5 содержит описательную статистику сентиментальных переменных. Можно отметить следующие интересные факты. В среднем в изучаемый период времени новости в группах Business и Market имели позитивную эмоциональную окраску, в то время как Policy и Common демонстрировали нейтральную тональность (особенно четко это прослеживается при анализе медианных значений). Группа World единственная продемонстрировала отрицательную тональность в среднем, что говорит о преобладании негативных настроений в сообщениях СМИ, рассматривающих международные события вокруг Российской Федерации. При этом стандартное отклонение эмоциональной окраски для всех групп примерно одинаково (кроме World, для которой оно несколько меньше) и во всех случаях значительно превышает среднее значение. Как следствие, можно утверждать, что разброс настроений в международных СМИ являлся достаточно существенным в период с 2012 по 2018 гг. Наконец, заметим, что практически все группы показывают ту или иную асимметричность распределения тональности, за исключением Business, для которой этот показатель практически равен нулю, а распределение в целом близко к нормальному (куртозис также относительно несущественно отличается от трех). Из этого можно сделать вывод, что эмоциональная окраска группы Business формировалась под влиянием множества достаточно различных новостей (факторов), а ее распределение близко к нормальному.

Таблица 5.
Описательная статистика временных рядов сентиментальной окраски новостей о Российской Федерации в международных СМИ (без пролонгированного влияния), усредненных по лексиконам GI, JR, LM, NRC, SWN

Статистика	Тематические группы новостей				
	Business	Market	World	Policy	Common
Среднее	0,41	0,40	-0,12	0,20	0,18
Стандартное отклонение	0,73	0,67	0,56	0,69	0,74
Медиана	0,37	0,39	-0,09	0,00	0,00
Среднее абсолютное отклонение	0,55	0,57	0,50	0,45	0,40
Минимум	-4,35	-3,03	-4,25	-3,74	-4,12
Максимум	4,25	4,41	2,23	4,37	7,00
Разброс	8,61	7,45	6,48	8,11	11,12
Ассиметричность	-0,01	0,23	-0,36	0,39	0,54
Куртозис	3,43	3,68	2,61	3,92	6,72

Примечание. Для наглядности перед расчетом описательной статистики все временные ряды сентиментальных переменных были умножены на 100.

Таблица 6 показывает комбинированные результаты двухэтапного моделирования рыночного курса российского рубля (логарифмированной доходности) для рассмотренных нами периодов затухания влияния эмоциональной тональности новостей τ : отбор эко-

номических (p_t, r_t) и сентиментальных ($S_t^{l,\tau,q}$) объясняющих переменных с помощью метода эластичной сети (5) на первом шаге и оценка модели (6) посредством метода максимального правдоподобия – на втором. Анализ полученных результатов моделирования позволяет сделать следующие выводы.

Во-первых, обе рассматриваемые нами экономические переменные – котировки нефтяной корзины ОПЕК и ставка межбанковского кредитования RUONIA – были отобраны методом эластичной сети в качестве потенциально значимых регрессоров для объяснения динамики рыночного курса российского рубля, в независимости от используемого периода затухания влияния сентиментальных переменных τ . В то же время состав отобранных факторов, отражающих тональность новостного фона, существенно варьируется в зависимости от τ и составляет от 4 ($\tau = 3, 7, 28$) до 7 (модель без затухания) из 25. Наиболее часто среди потенциально значимых сентиментальных переменных встречаются две: тематическая группа Policy, оцененная по словарю GI, и группа Business, оцененная по словарю LM. Отметим, что на этапе отбора факторов для всех τ мы нашли в качестве оптимального (в смысле минимизации AIC) значения $\gamma_2 = 0,9$, т.е. в моделях с регуляризацией методом эластичной сети наблюдается преобладание L_1 -нормы вектора коэффициентов (т.е. редукция в сторону LASSO-регрессии).

Во-вторых, из рассмотренных нами экономических факторов значимое влияние на динамику рыночного курса рубля оказывает только поведение цены нефтяной корзины ОПЕК. При этом соответствующий коэффициент является статистически значимым на 1-процентном уровне и составляет $-0,24$ для всех построенных моделей. Таким образом, относительное увеличение цены на нефть на 1% ведет к укреплению российского рубля на 0,24%. Зависимость российской валюты от котировок нефти была выявлена ранее во многих работах, среди которых можно упомянуть [Навоев, 2016; Миркин и др., 2016] и др. Стоит отметить, что коэффициент модели не зависит от выбранного периода затухания и для всех моделей идентичен. Это позволяет утверждать, что влияние цены на нефть и сентиментальных переменных происходит независимо друг от друга (это также следует и из простого расчета корреляционной матрицы регрессоров).

Что касается ставки межбанковского кредитования RUONIA, то в соответствии с нашими результатами она не может быть признана статистически значимым детерминантом рыночного курса рубля, хотя и была отобрана как потенциальная объясняющая переменная на первом этапе моделирования. С учетом этого можно предположить, что в течении 2012–2018 гг. данный косвенный инструмент денежно-кредитной политики Банка России либо применялся несистемно (т.е. крайне эпизодически), либо не давал какого-либо ощутимого эффекта.

В-третьих, набор статистически значимых сентиментальных переменных отличается для моделей с различными периодами затухания влияния тональности на поведение экономических агентов. Так, для периода в одну неделю только одна из них оказалась значимой, в то время как модель без затухания демонстрирует уже три значимых сентиментальных переменных. Так как отсутствуют объективные причины выбора единственного значения периода, то дальнейший анализ влияния тональности новостей о России в международных СМИ на курс рубля должен проводиться комплексно по всем построенным моделям (т.е. для различных значений τ), а не на основании только единственной из них.

Таблица 6.

Результаты двухэтапного моделирования рыночного курса российского рубля (логарифмированной доходности) в период с 10.01.2012 г. по 30.05.2018 г.

Параметр	Период затухания влияния sentimentальных переменных τ				
	нет	3 дня	1 неделя	2 недели	1 месяц
Этап 1: Отбор переменных методом эластичной сети					
<i>Параметры регуляризации</i>					
γ_1	0,0159	0,0253	0,0159	0,0175	0,0255
γ_2	0,9	0,9	0,9	0,9	0,9
Этап 2: Оценка параметров модели ARMAX(0,0) – GARCH(1,1)					
<i>Коэффициенты при экономических переменных</i>					
Цена на нефть (лог-доходность), α_p	-0,239*** (0,017)	-0,240*** (0,017)	-0,239*** (0,017)	-0,239*** (0,017)	-0,239*** (0,017)
Ставка RUONIA (приращение), α_r	0,029 (0,063)	0,037 (0,063)	0,035 (0,064)	0,034 (0,064)	0,033 (0,065)
<i>Коэффициенты при sentimentальных переменных, $\beta_{l,\tau,q}$</i>					
GI, Business		-0,685 (0,700)	-1,021 (0,969)		
GI, Policy		1,101 (0,715)	1,110 (0,980)	1,885 (1,309)	1,895 (1,465)
GI, Common	0,389 (0,494)				
LM, Business		-2,181*** (0,853)	-2,390*** (0,880)	-3,028** (1,529)	-4,989*** (1,921)
LM, Market	-1,337*** (0,431)				
LM, World		0,155 (0,661)			
LM, Policy	-0,804 (0,532)				
LM, Common	-0,796 (0,522)		-0,972 (0,845)	-1,570 (1,148)	
NRC, Business				-2,644** (1,16)	-3,603*** (1,468)

Окончание табл. 6.

Параметр	Период затухания влияния сен­ти­мен­таль­ных пере­мен­ных τ				
	нет	3 дня	1 неделя	2 недели	1 месяц
NRC, World	-1,066** (0,479)				
NRC, Common				0,202 (1,247)	1,408 (1,461)
JR, Business	-1,170** (0,606)				
JR, Policy	0,864 (0,675)				
<i>Коэффициенты GARCH-компоненты</i>					
c	0,007*** (0,002)	0,007*** (0,002)	0,007*** (0,002)	0,007*** (0,002)	0,007*** (0,002)
ϕ_1	0,045*** (0,005)	0,044*** (0,004)	0,044*** (0,004)	0,045*** (0,004)	0,044*** (0,004)
ψ_1	0,954*** (0,004)	0,955*** (0,003)	0,955*** (0,003)	0,954*** (0,003)	0,955*** (0,003)
<i>Параметры распределения ошибок</i>					
Асимметрия, ν	1,017*** (0,012)	1,023*** (0,014)	1,027*** (0,015)	1,025*** (0,016)	1,027*** (0,016)
Ширина, ξ	2,536*** (0,074)	2,521*** (0,07)	2,522*** (0,07)	2,528*** (0,074)	2,523*** (0,074)
<i>Параметры модели</i>					
LL	-1961,3	-1961,3	-1959,2	-1953,1	-1941,8
AIC	1,6934	1,6922	1,6933	1,6940	1,6937
LB-статистика	2,910	2,653	2,730	2,533	2,532

Примечания. Уровни статистической значимости: *** – 1%, ** – 5%, * – 10%. В скобках приведены стандартные ошибки коэффициентов модели. Для сен­ти­мен­таль­ных пере­мен­ных показаны только те, которые были предварительно отобраны методом эластичной сети (из общего числа в 25 временных рядов) и далее включены в модель ARMA(0,0) – GARCH(1,1). LL – значение логарифмированной функции правдоподобия, AIC – значение информационного критерия Акаике, LB-статистика – статистика Q-теста Льюнга – Бокса на отсутствие автокорреляции в остатках модели до 7 лага.

В-четвертых, среди переменных, отражающих тональность новостей, наиболее часто статистически значимое влияние на курс российского рубля обнаруживается со стороны тематической группы Business (тональность которой оценена по словарям LM и NRC). При

этом направление движения сентиментальной окраски новостей данной группы (включающей тематики об отдельных индустриях, банках, финансах, крупных сделках, корпоративном секторе, макроэкономике, инвестициях, освещение сделок слияния и поглощения компаний и т.д.) и относительного изменения курса российской валюты оказывается противоположным, т.е. рост позитивной тональности новостного фона приводит к укреплению рубля, и наоборот. Это хорошо согласуется с общей логикой, так как позитивный настрой в указанной тематической группе потенциально ведет к общему улучшению ожиданий экономических агентов относительно роста экономики России и, как следствие, к возрастанию склонности к инвестированию в российскую валюту.

Отметим, что для ряда словарей сентиментальные переменные группы Business оказались систематически значимы сразу в нескольких моделях: в четырех для лексикона LM и двух для лексикона NRC. При этом модуль оценок соответствующих коэффициентов хотя и растет с возрастанием периода влияния τ , однако с точки зрения 95-процентного доверительного интервала остается эквивалентным среди всех моделей (в рамках одного словаря). Также консистентными являются и знаки коэффициентов. Учитывая это, можно утверждать, что данные оценки устойчивы и отражают систематическое влияние со стороны соответствующих сентиментальных переменных на рыночный курс российской валюты, а полученный выше вывод является обоснованным.

Тональность новостей тематической группы Policy также потенциально могла бы рассматриваться как один из детерминантов рыночного курса российского рубля, так как на первом этапе моделирования метод эластичной сети отбирал ее достаточно часто (особенно для оценки тональности по словарю GI – в четырех моделях из пяти). Однако при построении итоговых моделей все коэффициенты соответствующих переменных оказались незначимы даже на 10-процентном уровне. С учетом этого можно предположить, что участники валютного рынка в большей степени склонны принимать решения под влиянием освещения в СМИ событий из сферы бизнеса и экономики, но не политических событий вокруг России.

Наконец, *в-пятых*, среди словарей, которые использовались для оценки тональности тематических групп новостей и затем были отобраны методом эластичной сети в качестве потенциальных объясняющих факторов, наиболее часто встречаются LM (10 переменных) и GI (7 переменных). Учитывая, что лексикон LM составлялся авторами как специализированный финансовый словарь [Loughran, McDonald, 2011], то он позволяет получить «более качественные» оценки тональности экономических новостей, так как меньшее количество специфичных терминов относит к нейтральным (неизвестным) по сравнению со словарями общего назначения. Вероятно, это и объясняет найденное наибольшее количество отобранных переменных, полученных на его основе, особенно в тематической группе Business. При этом для словаря общего назначения GI достаточно часто отбирались переменные группы Policy, тексты новостей из которой содержат менее специфичную терминологию, чем для группы Business. Отдельно отметим, что сентиментальные переменные на базе лексикона SWN ни для одной модели не были отобраны методом эластичной сети, ввиду чего мы не можем рекомендовать использовать этот словарь при исследовании влияния новостного потока на экономические процессы, как минимум, в части валютных рынков.

5. Заключение

В данной работе выполнено исследование влияния эмоциональной тональности освещения России в международных средствах массовой информации (СМИ) на рыночный курс рубля относительно доллара США в период 2012–2018 гг., наряду с общепризнанным фундаментальным объясняющим фактором – ценой на нефть. Также рассматривается влияние денежно-кредитной политики Банка России через включение в эконометрическую модель такого косвенного инструмента ДКП, как ставка межбанковского кредитования RUONIA. Для целей оценки тональности новостного потока использован метод «мешка слов», пять различных словарей (как общей направленности, так и финансовые словари), а также принята во внимание инерционность влияния информации на экономических агентов – участников валютного рынка. По результатам данного исследования можно сделать ряд ключевых выводов.

1. Конъюнктура нефтяного рынка по-прежнему остается одним из наиболее значимых факторов при формировании рыночного курса рубля к доллару США. При этом относительное увеличение цены на нефть на 1% ведет к укреплению российского рубля на 0,24%. В то же время попытки Банка России косвенно влиять на обменный курс через ставку межбанковского кредитования RUONIA в период с 2012 по 2018 гг. были в основном малоэффективны.

2. Эмоциональная окраска новостного фона, наряду с фундаментальными экономическими факторами, оказывает влияние на поведение участников валютного рынка, что в свою очередь отражается в обменном курсе российского рубля. Однако не все тематические группы новостей одинаково проявляют себя. Так, новости, освещающие события из сферы бизнеса, систематически воздействуют на обменный курс рубля (рост позитивной тональности приводит к укреплению рубля), в то время как сообщения международных СМИ, касающиеся политических событий вокруг России, не оказывают значимого влияния. Вероятно, это связано с тем, что каким бы положительным не был политический информационный фон, он отражает чаще всего лишь намерения стран и их лидеров по отношению к России, в то время как в реальной практике негативная динамика тональности новостного потока в области бизнеса и экономики имеет более существенное значение для участников валютного рынка.

3. При исследовании влияния эмоциональной тональности информационного потока на экономические процессы с использованием метода «мешка слов» необходимо учитывать тематическую направленность текстов и выбирать подходящие для них словари. Так, специализированный финансовый словарь [Loughran, McDonald, 2011] наиболее подходит при оценке тональности экономических новостей, в то время как словарь общей направленности Harvard General Inquirer (Harvard IV-4, объединенный с Laswell) лучше проявляет себя для текстов, касающихся политических событий.

Полученные выше результаты потенциально могут быть полезны при прогнозировании динамики рыночного курса российского рубля инвесторами, Банком России, коммерческими банками, инвестиционными фондами и другими профессиональными участниками рынка. Однако вопрос влияния учета сентиментальных переменных на точность прогноза валютного курса требует отдельного дополнительного исследования, которое может быть выполнено в будущем.

* *
*

СПИСОК ЛИТЕРАТУРЫ

- Крепцев Д., Селезнев С.* Прогнозирование российской экономики с использованием DSGE-моделей с малым количеством уравнений // Деньги и кредит. 2018. Т. 77. № 2. С. 51–67.
- Миркин Я.М., Жукова Т.В., Бахтараева К.Б., Комова А.В.* Курс доллара и мировые цены на сырье (нефть, газ, металлы): средний долгосрочный прогноз // Банковские услуги. 2016. № 1. С. 14–23.
- Навоев А.* О причинно-следственной связи цены на нефть и курса валют // Общество и экономика. 2016. № 5. С. 86–94.
- Федорова Е.А., Линкова М.А.* Прогнозирование курса валюты с использованием нейросетевых технологий // Аудит и финансовый анализ. 2013. № 3. С. 152–155.
- Ahmad I.S., Setiawan S., Masun N.H.* Forecasting of Monthly Inflow and Outflow Currency Using Time Series Regression and ARIMAX: The Idul Fitri Effect // AIP Conference Proceedings. 2015. Vol. 1691.
- Ardia D., Bluteau K., Boudt K.* Questioning the News about Economic Growth: Sparse Forecasting Using Thousands of News-based Sentiment Values // International Journal of Forecasting. 2019. (Forthcoming) (<https://ssrn.com/abstract=2976084>)
- Baccianella S., Esuli A., Sebastiani F.* SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining // Proceedings of the International Conference on Language Resources and Evaluation. LREC, 2010. P. 2200–2204.
- Beckmann J., Czudaj R.* Exchange Rate Expectations and Economic Policy Uncertainty // European Journal of Political Economy. 2017. 47. P. 148–162.
- Bollerslev T.* Generalized Autoregressive Conditional Heteroskedasticity // Journal of Econometrics. 1986. 31. P. 307–327.
- Chatrath A., Miao H., Ramchander S., Villupuram S.* Currency Jumps, Cojumps and the Role of Macro News // Journal of International Money and Finance. 2014. 40. P. 42–62.
- Clements K.W., Lan Y.* A New Approach to Forecasting Exchange Rates // Journal of International Money and Finance. 2010. 29(7). P. 1424–1437.
- Das S.R., Kuhoo Mishra D., Rout M.* An Optimized Feature Reduction Based Currency Forecasting Model Exploring the Online Sequential Extreme Learning Machine and Krill Herd Strategies // Physica A: Statistical Mechanics and its Applications. 2019. 513. P. 339–370.
- Ederington L., Guan W., Zongfei L.* The Impact of the U.S. Employment Report on Exchange Rates // Journal of International Money and Finance. 2019. 90. P. 257–267.
- Friedman J., Hastie T., Tibshirani R.* Regularization Paths for Generalized Linear Models via Coordinate Descent // Journal of Statistical Software. 2010. 33. P. 1–22.
- Hoerl A.E., Kennard R.W.* Ridge Regression: Biased Estimation for Nonorthogonal Problems // Technometrics. 1970. 12. P. 55–67.
- Hu M., Liu B.* Mining and Summarizing Customer Reviews // Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2004. P. 168–177.
- Jockers M.L.* Syuzhet: Extract Sentiment and Plot Arcs from Text. 2017. (<https://github.com/mjockers/syuzhet>)
- Johansen S.* Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford: Oxford University Press, 1995.
- Kawakami K.* Conditional Forecast Selection from Many Forecasts: An Application to the Yen / Dollar Exchange Rate // Journal of the Japanese and International Economies. 2013. 28. P. 1–18.
- Khadjeh Nassirtoussi A., Aghabozorgi S., Ying Wah T., Ngo D.C.L.* Text Mining of News-headlines for FOREX Market Prediction: A Multi-layer Dimension Reduction Algorithm with Semantics and Sentiment // Expert Systems with Applications. 2015. 42(1). P. 306–324.

- Loughran T., McDonald B.* Textual Analysis in Accounting and Finance: A Survey // *Journal of Accounting Research*. 2016. 54(4). P. 1187–1230.
- Lüdering J., Tillmann P.* Monetary Policy on Twitter and Asset Prices: Evidence from Computational Text Analysis // *North American Journal of Economics and Finance*. 2018. (April). P. 1–18.
- Mikova E., Teplova T., Munir Q.* Puzzling Premiums on FX Markets: Carry Trade, Momentum, and Value Alone and Strategy Diversification // *Emerging Markets Finance and Trade*. 2019. DOI: 10.1080/1540496X.2018.156289
- Mohammad S.M., Turney P.D.* Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon // *Proceeding of Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. 2010. P. 26–34.
- Morales-Arias L., Moura G.V.* Adaptive Forecasting of Exchange Rates with Panel Data // *International Journal of Forecasting*. 2013. 29(3). P. 493–509.
- Osterwald-Lenum M.* A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics // *Oxford Bulletin of Economics and Statistics*. 1992. 55(3). P. 461–472.
- Rinker T.W.* Lexicon: Lexicon Data. 2018. (<http://github.com/trinker/lexicon>)
- Sarno L., Sojli E.* Exchange Rate Forecasting, Order Flow and Macroeconomic Information. 2010. P. 72–88.
- Shen D., Urquhart A., Wang P.* Does Twitter Predict Bitcoin? // *Economics Letters*. 2019. 174. P. 118–122.
- Stone P., Dunphy D.C., Smith M.S., Ogilvie D.M.* The General Inquirer: A Computer Approach to Content Analysis. The MIT Press, 1966.
- Sun S., Wang S., Wei Y.* A New Multiscale Decomposition Ensemble Approach for Forecasting Exchange Rates // *Economic Modelling*. 2018. (<https://doi.org/10.1016/j.econmod.2018.12.013>)
- Tibshirani R.* Regression Shrinkage and Selection via the LASSO // *Journal of the Royal Statistical Society: Series B*. 1996. 58. P. 267–288.
- Tibshirani R.J., Taylor J.* Degrees of Freedom in LASSO Problems // *Annals of Statistics*. 2012. 40. P. 1198–1232.
- Yin W., Li J.* Macroeconomic Fundamentals and the Exchange Rate Dynamics: A Non-arbitrage Macrofinance // *Journal of International Money and Finance*. 2014. 41. P. 46–64.
- Zou H., Hastie T.* Regularization and Variable Selection via the Elastic Net // *Journal of the Royal Statistical Society: Series B*. 2005. 67. P. 301–320.

On the Impact of News Tonality in International Media on the Russian Ruble Exchange Rate: Textual Analysis

Dmitriy Afanasyev¹, Elena Fedorova²,
Oleg Rogov³

¹ JSC Greenatom,
10 bld. 1, 1st Nagatinskiy pas., Moscow, 115230, Russian Federation.
E-mail: dmafanasyev@gmail.com

² The Financial University under the Government of the Russian Federation,
49, Leningradskiy av., Moscow, 125993, Russian Federation.
E-mail: ecolena@mail.ru

³ The Moscow Institute of Physics and Technology, Research Centre,
9, Institutskiy Pereulok, Dolgoprudny, 141701, Russian Federation.
E-mail: olegrg@yandex.ru

This paper examines the impact of the Russia-related international news tonality on the market ruble exchange rate (in the US dollar). The dataset features 162 259 news texts from the Thomson Reuters agency, dated from January 10, 2012, to May 30, 2018. The texts are combined into 5 thematic groups: business, market, world, policy and common. To calculate the text tonality, the “bag of words” method and 5 different lexicons were used, as well as the inertia effect of the news impact on the market participants behavior was taken into account. The econometric methodology consisted of two steps: the selection of potentially significant explanatory variables based on the elastic net regression and the estimation of the ARMAX-GARCH model. As a result of modeling, it was confirmed that the dynamics of oil prices significantly affect the market ruble exchange rate, while the Bank of Russia monetary policy tool – the interbank lending rate RUONIA – was not effective in the period studied. The tonality of the news, along with the fundamental economic factors, has a systematic impact on the ruble exchange rate, which nevertheless depends on their subject matter: the most important are business news, while political messages in the media do not statistically influence the market ruble exchange rate. The obtained results are potentially applicable for the ruble exchange rate forecasting by investors, the Bank of Russia, commercial banks, investment funds and other professional market participants.

Key words: textual analysis; ruble exchange rate; news impact; news tonality.

JEL Classification: F31, G41, O24.

* *
*

References

- Kreptsev D., Seleznev S. (2018) Prognozirovanie rossijskoj ekonomiki s ispol'zovaniem DSGE-modelej s malym kolichestvom uravnenij [Forecasting for the Russian Economy Using Small-Scale DSGE Models]. *Russian Journal of Money and Finance*, 77, 2, pp. 51–67
- Mirkin Y.M., Zhukova T.V., Bakhtareva K.B., Komova A.V. (2016) Kurs dollara i mirovye ceny na syr'e (neft', gaz, metally): srednij dolgosrochnyj prognoz [The Dollar Exchange Rate and Commodity Prices (Oils, Gas, Metals): MediumTerm and LongTerm]. *Banking Services*, 1, pp. 14–23.
- Navoev A. (2016) O prichinno-sledstvennoj svyazi ceny na neft' i kursa valyut [Causation of Oil Prices and the Exchange Rate]. *Society and Economics*, 5, pp. 86–94.
- Fedorova E.A., Linkova M.A. (2013) Prognozirovanie kursa valyuty s ispol'zovaniem nejros-etevyh tekhnologij [Forecasting the Case of Price for Currency with the Use of Neuronetwork Technologies]. *Audit and Financial Analysis*, 3, pp. 152–155.
- Ahmad I.S., Setiawan S., Masun N.H. (2015) Forecasting of Monthly Inflow and Outflow Currency Using Time Series Regression and ARIMAX: The Idul Fitri Effect. *AIP Conference Proceedings*, 1691.
- Ardia D., Bluteau K., Boudt K. (2019) Questioning the News about Economic Growth: Sparse Forecasting Using Thousands of News-based Sentiment Values. *International Journal of Forecasting*. (Forthcoming). Available at: <https://ssrn.com/abstract=2976084>
- Baccianella S., Esuli A., Sebastiani F. (2010) SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of the International Conference on Language Resources and Evaluation*. LREC, pp. 2200–2204.
- Beckmann J., Czudaj R. (2017) Exchange Rate Expectations and Economic Policy Uncertainty. *European Journal of Political Economy*, 47, pp. 148–162.
- Bollerslev T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, pp. 307–327.
- Chatrath A., Miao H., Ramchander S., Villupuram S. (2014) Currency Jumps, Cojumps and the Role of Macro News. *Journal of International Money and Finance*, 40, pp. 42–62.
- Clements K.W., Lan Y. (2010) A New Approach to Forecasting Exchange Rates. *Journal of International Money and Finance*, 29, 7, pp. 1424–1437.
- Das S.R., Kuhoo Mishra D., Rout M. (2019) An Optimized Feature Reduction Based Currency Forecasting Model Exploring the Online Sequential Extreme Learning Machine and Krill Herd Strategies. *Physica A: Statistical Mechanics and its Applications*, 513, pp. 339–370.
- Ederington L., Guan W., Zongfei L. (2019) The Impact of the U.S. Employment Report on Exchange Rates. *Journal of International Money and Finance*, 90, pp. 257–267.
- Friedman J., Hastie T., Tibshirani R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33, pp. 1–22.
- Hoerl A.E., Kennard R.W. (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, pp. 55–67.
- Hu M., Liu B. (2004) Mining and Summarizing Customer Reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177.
- Jockers M.L. (2017) *Syuzhet: Extract Sentiment and Plot Arcs from Text*. Available at: <https://github.com/mjockers/syuzhet>
- Johansen S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Kawakami K. (2013) Conditional Forecast Selection from Many Forecasts: An Application to the Yen / Dollar Exchange Rate. *Journal of the Japanese and International Economies*, 28, pp. 1–18.

- Khadjeh Nassirtoussi A., Aghabozorgi S., Ying Wah T., Ngo D.C.L. (2015) Text Mining of News-headlines for FOREX Market Prediction: A Multi-layer Dimension Reduction Algorithm with Semantics and Sentiment. *Expert Systems with Applications*, 42, 1, pp. 306–324.
- Loughran T., McDonald B. (2016) Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54, 4, pp. 1187–1230.
- Lüdering J., Tillmann P. (2018) Monetary Policy on Twitter and Asset Prices: Evidence from Computational Text Analysis. *North American Journal of Economics and Finance*, April, pp. 1–18.
- Mikova E., Teplova T., Munir Q. (2019) Puzzling Premiums on FX Markets: Carry Trade, Momentum, and Value Alone and Strategy Diversification. *Emerging Markets Finance and Trade*. DOI: 10.1080/1540496X.2018.156289
- Mohammad S.M., Turney P.D. (2010) Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. *Proceeding of Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 26–34.
- Morales-Arias L., Moura G.V. (2013) Adaptive Forecasting of Exchange Rates with Panel Data. *International Journal of Forecasting*, 29, 3, pp. 493–509.
- Osterwald-Lenum M. (1992) A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statistics*, 55, 3, pp. 461–472.
- Rinker T.W. (2018) *Lexicon: Lexicon Data*. Available at: <http://github.com/trinker/lexicon>
- Sarno L., Sojli E. (2010) *Exchange Rate Forecasting, Order Flow and Macroeconomic Information*, 80, pp. 72–88.
- Shen D., Urquhart A., Wang P. (2019) Does Twitter Predict Bitcoin? *Economics Letters*, 174, pp. 118–122.
- Stone P., Dunphy D.C., Smith M.S., Ogilvie D.M. (1966) *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.
- Sun S., Wang S., Wei Y. (2018) A New Multiscale Decomposition Ensemble Approach for Forecasting Exchange Rates. *Economic Modelling*. Available at: <https://doi.org/10.1016/j.econmod.2018.12.013>
- Tibshirani R. (1996) Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society: Series B*, 58, pp. 267–288.
- Tibshirani R.J., Taylor J. (2012) Degrees of Freedom in LASSO Problems. *Annals of Statistics*, 40, pp. 1198–1232.
- Yin W., Li J. (2014) Macroeconomic Fundamentals and the Exchange Rate Dynamics: A Non-arbitrage Macrofinance. *Journal of International Money and Finance*, 41, pp. 46–64.
- Zou H., Hastie T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B*, 67, pp. 301–320.