

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**СОЦИАЛЬНЫЕ МЕДИА КАК ИСТОЧНИК
НЕТРАДИЦИОННЫХ ЭКОНОМИЧЕСКИХ ИНДИКАТОРОВ**
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 412 группы
направления 01.03.02 – Прикладная математика и информатика

механико-математического факультета
Фефилова Михаила Ивановича

Научный руководитель

доцент, к. ф.-м. н., доцент

Л. В. Борисова

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2026

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Основное содержание работы	6
ЗАКЛЮЧЕНИЕ	17

ВВЕДЕНИЕ

В последние годы финансовые рынки всё в большей степени зависят не только от фундаментальных экономических показателей, но и от информационного фона. Новости, публичные заявления политиков, сообщения государственных органов и публикации в социальных сетях способны существенно влиять на ожидания участников рынка. Особенно заметно это проявляется на валютных рынках, где изменение ожиданий относительно внешнеполитической ситуации, санкционных ограничений, цен на сырьё и глобального спроса на риск может быстро отражаться в динамике обменного курса.

Курс рубля к доллару США является важным индикатором состояния российской экономики. Его динамика связана с ценами на сырьевые товары, состоянием мировых финансовых рынков, денежно-кредитной политикой, санкционными рисками и политической неопределённостью. При этом краткосрочные изменения курса не всегда могут быть объяснены только классическими макроэкономическими факторами. Поэтому актуальной является задача учёта дополнительной информации, поступающей на рынок в виде новостей и публичных сообщений.

Одним из источников такой информации являются социальные сети. Сообщения политиков и других публичных лиц могут восприниматься участниками рынка как сигналы о возможном изменении внешнеполитической позиции, санкционного режима или общего уровня международной напряжённости. В частности, сообщения Дональда Трампа, связанные с Россией, Украиной, санкциями, военными конфликтами и переговорами, представляют интерес как источник политически значимой информации. Подобные тексты могут содержать как негативную, так и позитивную тональность, что делает возможным их количественный анализ с помощью методов обработки естественного языка.

Для включения текстовых данных в эконометрическую модель необходимо преобразовать сообщения в числовые признаки. В данной работе для этого используется анализ тональности текста. Такой подход позволяет оценить, преобладает ли в сообщении положительная или отрицательная окраска, а затем построить временные ряды sentiment-показателей. При этом результат зависит от выбранного словаря, способа обработки текста и метода

агрегации сообщений по дням. Поэтому в работе используются несколько лексиконов и несколько схем временного сглаживания, что позволяет сравнить различные варианты построения текстовых признаков.

Особенность рассматриваемой задачи состоит в том, что влияние политического сообщения на валютный курс не обязано быть постоянным во времени. В спокойные периоды курс рубля может определяться преимущественно рыночными факторами, такими как цена нефти или общее состояние мировых рынков. В периоды повышенной политической неопределённости роль новостного фона может возрасти. В связи с этим в работе применяются не только регрессионные модели с отбором признаков, но и модели условной волатильности, а также модель марковского переключения режимов.

Цель бакалаврской работы состоит в построении и исследовании математико-статистической модели, позволяющей оценить связь между тональностью политических сообщений Дональда Трампа, связанных с Россией и Украиной, и краткосрочной динамикой курса рубля к доллару США.

Задачи бакалаврской работы:

1. рассмотреть основные подходы к анализу влияния публичной информации на валютный курс;
2. сформировать корпус сообщений, связанных с Россией, Украиной, санкциями и международными конфликтами;
3. выполнить предварительную обработку текстов и построить sentiment-показатели на основе нескольких лексиконов;
4. объединить текстовые признаки с финансовыми временными рядами;
5. провести отбор объясняющих переменных с помощью метода Elastic Net;
6. оценить модель динамики курса USDRUB с учётом условной гетероскедастичности;
7. построить модель марковского переключения режимов;
8. провести локальный анализ отдельных сообщений и сопоставить его с результатами эконометрического моделирования.

Объектом исследования является динамика курса рубля к доллару США. Предметом исследования являются методы построения и применения текстовых sentiment-признаков при моделировании краткосрочных изменений

валютного курса.

Данная бакалаврская работа состоит из четырёх разделов. В первом разделе рассматриваются теоретические основы анализа текстовой информации и её возможного влияния на валютный курс. Во втором разделе описывается формирование корпуса сообщений и построение sentiment-переменных. В третьем разделе проводится эконометрическое моделирование динамики курса USDRUB. В четвёртом разделе выполняется локальный анализ отдельных событий и интерпретация полученных результатов.

1 Основное содержание работы

В первом разделе работы рассматриваются основные понятия, необходимые для построения модели. Валютный курс рассматривается как цена одной денежной единицы, выраженная в другой денежной единице. В дальнейшем под курсом USDRUB понимается стоимость одного доллара США, выраженная в рублях. Рост данного показателя соответствует ослаблению рубля по отношению к доллару США, а снижение — укреплению рубля.

При анализе финансовых временных рядов обычно рассматриваются не сами значения цен или курсов, а их относительные изменения. Пусть P_t — значение курса USDRUB в момент времени t . Тогда логарифмическая доходность определяется формулой

$$r_t = \ln P_t - \ln P_{t-1} = \ln \frac{P_t}{P_{t-1}}. \quad (1.1)$$

Если $r_t > 0$, то курс USDRUB вырос, то есть рубль ослаб относительно доллара. Если $r_t < 0$, то курс USDRUB снизился, то есть рубль укрепился.

Для использования текстов в математической модели каждому сообщению сопоставляется числовой признак тональности. Рассмотрим текст T , состоящий из набора токенов

$$T = \{w_1, w_2, \dots, w_m\}.$$

Пусть для некоторого лексикона l задано разбиение слов на положительные, отрицательные и нейтральные. Обозначим через $N_{T,l}^+$ число положительных слов в тексте T , через $N_{T,l}^-$ — число отрицательных слов, а через $N_{T,l}^0$ — число нейтральных слов, распознанных данным лексиконом. Тогда показатель тональности текста относительно лексикона l можно определить следующим образом:

$$S_{T,l} = \frac{N_{T,l}^+ - N_{T,l}^-}{N_{T,l}^+ + N_{T,l}^0 + N_{T,l}^-}. \quad (1.2)$$

Значение $S_{T,l}$ лежит в промежутке от -1 до 1 . Положительное значение соответствует преобладанию положительной тональности, отрицательное — преобладанию отрицательной.

Если в один день было опубликовано несколько сообщений, то их оценки тональности агрегируются. Пусть в день t опубликованы сообщения $T_{1,t}, \dots, T_{n_t,t}$. Тогда дневной sentiment-показатель можно определить как средневзвешенное значение:

$$S_{t,l} = \sum_{i=1}^{n_t} \omega_{i,t} S_{T_{i,t},l}, \quad \sum_{i=1}^{n_t} \omega_{i,t} = 1. \quad (1.3)$$

В работе используются веса, пропорциональные длине текста. Такой выбор объясняется тем, что более длинное сообщение содержит больше слов и, следовательно, больше информации для оценки тональности.

Кроме дневной агрегации, учитывается возможное сохранение эффекта сообщения во времени. Для этого вводится сглаженный sentiment-показатель

$$\tilde{S}_{t,l} = \sum_{k=0}^K a_k S_{t-k,l}, \quad \sum_{k=0}^K a_k = 1, \quad a_k \geq 0. \quad (1.4)$$

При экспоненциальном затухании веса имеют вид

$$a_k(\lambda) = \frac{e^{-\lambda k}}{\sum_{j=0}^K e^{-\lambda j}}, \quad \lambda > 0. \quad (1.5)$$

Чем больше параметр λ , тем быстрее уменьшается вес более старых сообщений.

После построения sentiment-переменных возникает задача выбора признаков, которые будут включены в модель. Если используются несколько лексиконов и несколько схем сглаживания, то число объясняющих переменных быстро возрастает. При этом многие из них могут быть коррелированы между собой, поскольку все они строятся на основе одного и того же набора сообщений. Для решения этой проблемы применяется метод Elastic Net. Пусть $y = (y_1, \dots, y_T)^\top$ — вектор значений зависимой переменной, а X — матрица объясняющих переменных. Тогда оценка Elastic Net определяется как решение задачи минимизации

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{T} \sum_{t=1}^T (y_t - X_t^\top \beta)^2 + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right) \right\}. \quad (1.6)$$

Здесь $\lambda > 0$ — параметр регуляризации, а $\alpha \in [0, 1]$ определяет соотношение между L_1 - и L_2 -штрафами.

Для моделирования финансового ряда используется модель ARMAX–EGARCH. Уравнение среднего имеет вид

$$r_t = \mu + \sum_{i=1}^p \varphi_i r_{t-i} + X_t^\top \beta + \varepsilon_t, \quad (1.7)$$

где r_t — логарифмическая доходность курса, X_t — вектор объясняющих переменных, ε_t — ошибка модели. Предполагается, что

$$\varepsilon_t = \sigma_t z_t, \quad \mathbb{E}z_t = 0, \quad \mathbb{D}z_t = 1. \quad (1.8)$$

Для модели EGARCH(1,1) условная дисперсия задаётся уравнением

$$\ln \sigma_t^2 = \omega + \alpha (|z_{t-1}| - \mathbb{E}|z_{t-1}|) + \gamma z_{t-1} + \delta \ln \sigma_{t-1}^2. \quad (1.9)$$

Коэффициент δ описывает устойчивость волатильности, коэффициент α отвечает за влияние величины прошлой инновации, а коэффициент γ позволяет учитывать асимметрию реакции волатильности на положительные и отрицательные шоки.

Также в работе используется модель марковского переключения режимов. Пусть s_t — скрытое состояние системы в момент времени t , принимающее значения из множества $\{1, 2, \dots, M\}$. Предполагается, что последовательность $\{s_t\}$ является марковской цепью:

$$\mathbb{P}(s_t = j \mid s_{t-1} = i, s_{t-2}, \dots) = \mathbb{P}(s_t = j \mid s_{t-1} = i) = p_{ij}. \quad (1.10)$$

В простейшем случае с двумя режимами модель можно записать в виде

$$r_t = \mu_{s_t} + X_t^\top \beta_{s_t} + \varepsilon_{t,s_t}, \quad \varepsilon_{t,s_t} \sim N(0, \sigma_{s_t}^2). \quad (1.11)$$

Это позволяет описывать ситуации, когда в одном режиме ряд имеет низкую волатильность, а в другом — высокую.

Во втором разделе работы была построена процедура перехода от исход-

ных текстовых сообщений к числовым временным рядам. Исходный массив сообщений был очищен от технических элементов, после чего был проведён первичный отбор по ключевым словам. Для повышения качества отбора дополнительно использовалась семантическая фильтрация на основе косинусной близости векторных представлений текстов.

В качестве ключевых слов использовались слова, связанные с Россией, Украиной, санкциями, военными конфликтами и международными переговорами: *russia, ukraine, putin, zelensky, nato, sanction, missile, kremlin, ruble, war, ceasefire* и другие близкие по смыслу слова. Данный список не является окончательной разметкой релевантности, а используется только для первичного отбора кандидатов.

Для уточнения релевантности используется семантическая близость текста к заранее заданным описаниям интересующей тематики. Пусть x — очищенный текст сообщения, а q_1, \dots, q_m — набор эталонных текстов. Обозначим через $e(x)$ векторное представление текста x . Тогда релевантность сообщения определяется величиной

$$\rho(x) = \max_{1 \leq j \leq m} \frac{\langle e(x), e(q_j) \rangle}{\|e(x)\| \cdot \|e(q_j)\|}. \quad (1.12)$$

В итоговый корпус включаются сообщения, удовлетворяющие условию

$$\rho(x) \geq \rho_0, \quad (1.13)$$

где $\rho_0 = 0.30$ — выбранный порог релевантности.

Общая схема отбора сообщений представлена в таблице 1.

Таблица 1 – Этапы формирования корпуса релевантных сообщений

Этап	Описание	Количество
1	Исходные сообщения после очистки и удаления дубликатов	26858
2	Сообщения, содержащие хотя бы одно ключевое слово	968
3	Сообщения, прошедшие семантический порог релевантности	464

Таким образом, итоговый корпус содержит 464 сообщения. Это существенно меньше исходного числа наблюдений, однако такой отбор позволяет сосредоточиться именно на сообщениях, потенциально связанных с рассматриваемой политико-экономической тематикой.

Для каждого сообщения были построены показатели тональности на основе нескольких лексиконов. В работе использовались General Inquirer (GI), Loughran–McDonald (LM), NRC, SentiWordNet (SWN), Jockers–Rinker (JR). Использование нескольких лексиконов связано с тем, что политические сообщения имеют неоднозначную природу: одни словари лучше отражают общеязыковую эмоциональную окраску, другие — финансовую или деловую лексику.

После построения sentiment-переменных они были объединены с финансовыми временными рядами. В качестве зависимой переменной рассматривалась логарифмическая доходность курса USDRUB. В качестве дополнительных рыночных факторов использовались цена нефти Brent, индекс доллара DXY, индекс волатильности VIX, индекс S&P 500 и цена золота. Итоговая выборка содержит 1034 дневных наблюдения. Общее число потенциальных объясняющих переменных равно 42.

В третьем разделе работы проводится эконометрическое моделирование динамики курса USDRUB. На первом этапе применяется Elastic Net. В результате перебора параметров регуляризации было получено значение

$$\lambda = 0.003562, \quad \alpha = 0.7.$$

Методом Elastic Net было выбрано 7 переменных. Результаты представлены в таблице 2.

Таблица 2 – Переменные, выбранные методом Elastic Net

№	Переменная	Коэффициент
1	LM_LIN	0.012500
2	SP500	-0.006046
3	LM_EXP01	-0.004865
4	VIX	0.004586
5	Gold	0.004459
6	LM_EQ14	-0.002591
7	GI_LIN	0.000346

Из таблицы видно, что среди выбранных переменных присутствуют как рыночные показатели, так и текстовые признаки. Среди рыночных переменных были выбраны SP500, VIX и Gold. Среди sentiment-переменных были выбраны три показателя, построенные на основе лексикона Loughran–McDonald: LM_LIN, LM_EXP01, LM_EQ14. Кроме того, была выбрана переменная GI_LIN, построенная на основе словаря General Inquirer.

После отбора признаков рассматривается модель для логарифмической доходности курса USDRUB. Для сравнения были оценены две модели: без авторегрессионного лага и с одним лагом зависимой переменной. Значения информационных критериев представлены в таблице 3.

Таблица 3 – Сравнение EGARCH-моделей

Модель	Наблюдения	AIC	BIC
ARMAX(0,0)–EGARCH(1,1)	1034	3763.75	3832.93
ARMAX(1,0)–EGARCH(1,1)	1033	3734.06	3808.16

Модель с одним авторегрессионным лагом имеет меньшее значение AIC и BIC. Поэтому в дальнейшем рассматривается спецификация ARMAX(1,0)–EGARCH(1,1). Результаты оценки уравнения среднего приведены в таблице 4.

Таблица 4 – Оценки коэффициентов уравнения среднего

Переменная	Коэффициент	t-статистика	p-value
Const	-0.0048	-0.017	0.986
USDRUB [1]	-0.1497	-4.486	< 0.001
LM_LIN	0.0943	0.521	0.602
SP500	-0.1672	-0.267	0.789
LM_EXP01	-0.1781	-2.771	0.006
VIX	-0.1374	-0.266	0.790
Gold	0.0316	0.163	0.870
LM_EQ14	0.0120	0.039	0.969
GI_LIN	-0.0018	-0.003	0.998

Из таблицы 4 видно, что лаг зависимой переменной статистически значим. Его коэффициент отрицателен, что указывает на наличие краткосрочного возвратного эффекта. Среди sentiment-переменных статистически значимой оказалась переменная LM_EXP01. Её коэффициент отрицателен. Поскольку рост курса USDRUB соответствует ослаблению рубля, отрицательный коэффициент означает, что увеличение данного sentiment-показателя связано со снижением доходности USDRUB, то есть с укреплением рубля. Соответственно, ухудшение тональности по этому показателю связано с ростом курса USDRUB и ослаблением рубля.

Результаты оценки EGARCH-части модели приведены в таблице 5.

Таблица 5 – Оценки параметров EGARCH(1,1)

Параметр	Коэффициент	t-статистика	p-value
ω	2.0800	3.055	0.002
α	1.4929	3.592	< 0.001
γ	-0.4043	-2.103	0.035
δ	0.6527	4.684	< 0.001

Все параметры уравнения условной волатильности статистически значимы на уровне 5%. Это подтверждает необходимость использования модели с условной гетероскедастичностью. Значимый параметр δ показывает, что

волатильность обладает инерционностью: высокий уровень волатильности в предыдущий момент времени влияет на текущий уровень условной дисперсии. Значимый параметр γ означает наличие асимметрии в реакции волатильности на положительные и отрицательные инновации.

Дополнительно была рассмотрена модель марковского переключения режимов. В работе рассматривается модель с двумя режимами. Основные характеристики оценённой модели представлены в таблице 6.

Таблица 6 – Основные характеристики модели марковского переключения режимов

Показатель	Значение
Число наблюдений	1034
Log-Likelihood	-1965.123
AIC	3978.247
BIC	4096.835
σ_0^2	1.9415
σ_1^2	914.6239
p_{00}	0.9852
p_{10}	0.6153

Полученные оценки показывают, что два режима существенно отличаются по дисперсии. В режиме 0 дисперсия равна приблизительно 1.94, тогда как в режиме 1 она равна приблизительно 914.62. Следовательно, режим 0 можно интерпретировать как обычный режим динамики курса, а режим 1 — как режим повышенной нестабильности. Вероятность $p_{00} = 0.9852$ близка к единице, следовательно, обычный режим является устойчивым. Значение $p_{10} = 0.6153$ показывает, что из режима повышенной нестабильности процесс с достаточно большой вероятностью возвращается в обычный режим.

В четвёртом разделе работы выполняется локальный анализ событий. Пусть сообщение опубликовано в момент времени t_0 . Если сообщение опубликовано в выходной день, то ему сопоставляется ближайший следующий торговый день. Для горизонта h относительное изменение курса определяется формулой

$$R_h(t_0) = \frac{P_{t_0+h}}{P_{t_0}} - 1. \quad (1.14)$$

Так как рост курса USDRUB означает ослабление рубля, то $R_h(t_0) > 0$ соответствует ослаблению рубля, а $R_h(t_0) < 0$ — укреплению рубля. В качестве основного горизонта используется $h = 3$.

Сообщение считается согласующимся с последующей реакцией курса, если знак тональности соответствует ожидаемому направлению движения USDRUB:

$$S(t_0) < 0, \quad R_3(t_0) > 0, \quad (1.15)$$

или

$$S(t_0) > 0, \quad R_3(t_0) < 0. \quad (1.16)$$

Первое условие означает, что негативный информационный фон сопровождается ослаблением рубля. Второе условие означает, что позитивный информационный фон сопровождается укреплением рубля.

Для локального анализа были рассмотрены отдельные сообщения из итогового корпуса. Каждому сообщению соответствует окно наблюдения вокруг даты публикации. В таблице 7 приведены примеры сообщений, для которых доступна трёхдневная реакция курса.

Таблица 7 – Примеры локального анализа сообщений

№	Дата	LM_LIN	R_3 USDRUB	Интерпретация
1	06.12.2022	-0.0119	+4.71%	Негативная тональность сопровождалась ослаблением рубля.
2	12.03.2024	-0.0024	+0.94%	Негативная тональность сопровождалась ростом USDRUB.
3	23.08.2024	-0.0137	-0.00%	Направление реакции курса не соответствует негативному сигналу.
4	18.09.2024	+0.0061	+0.28%	Позитивная тональность не сопровождалась укреплением рубля.

№	Дата	LM_LIN	R_3 USDRUB	Интерпретация
5	02.10.2024	-0.0116	-1.10%	Негативная тональность сопровождалась снижением USDRUB.
6	17.03.2025	-0.0025	-0.67%	Направление реакции курса не соответствует негативному сигналу.
7	28.07.2025	-0.0433	+3.77%	Негативная тональность сопровождалась ослаблением рубля.
8	25.11.2025	+0.0031	-0.70%	Позитивная тональность сопровождалась укреплением рубля.
9	20.04.2026	+0.0092	-0.55%	Позитивная тональность сопровождалась укреплением рубля.

Из таблицы 7 видно, что в части случаев знак sentiment-показателя соответствует направлению последующего изменения курса. Например, 6 декабря 2022 года значение LM_LIN было отрицательным и составило -0.0119 , а курс USDRUB за три торговых дня вырос на 4.71%. Так как рост USDRUB означает ослабление рубля, данный случай согласуется с гипотезой о негативном влиянии отрицательной политической тональности.

Аналогичная ситуация наблюдается 28 июля 2025 года. Значение LM_LIN равно -0.0433 , что является выраженным отрицательным сигналом. За три торговых дня курс USDRUB вырос на 3.77%. Такой пример также можно отнести к случаям, где негативный информационный фон сопровождается ослаблением рубля.

Положительный пример наблюдается 25 ноября 2025 года. Значение LM_LIN= 0.0031 , а трёхдневная реакция USDRUB равна -0.70% . Поскольку снижение курса USDRUB соответствует укреплению рубля, данный случай согласуется с предположением о том, что более позитивная тональность может сопровождаться снижением давления на рубль.

Вместе с тем существуют случаи, в которых знак реакции курса не соответствует знаку sentiment-показателя. Это показывает, что связь между тональностью сообщения и динамикой валютного курса не является механической. Валютный курс зависит от множества факторов, и даже в коротком окне на него могут влиять цены на нефть, глобальный спрос на риск, новости центральных банков, санкционные заявления и другие события. Поэтому локальный анализ показывает не причинность, а согласованность направления движения с текстовым сигналом.

Результаты локального анализа согласуются с выводами эконометрического моделирования в следующем смысле. В модели ARMAX–EGARCH значимой оказалась sentiment-переменная LM_EXP01, а в локальных примерах часто используется близкий по смыслу показатель LM_LIN. Это указывает на то, что признаки, построенные на основе лексикона Loughran–McDonald, являются информативными для рассматриваемой выборки. В то же время локальный анализ показывает, что нельзя утверждать, что любое негативное сообщение обязательно приводит к ослаблению рубля, а любое позитивное — к его укреплению.

Таким образом, основные результаты работы состоят в следующем. Во-первых, был сформирован корпус политически релевантных сообщений. Во-вторых, были построены sentiment-переменные по нескольким лексиконам и схемам сглаживания. В-третьих, метод Elastic Net выделил набор признаков, включающий как рыночные показатели, так и текстовые переменные. В-четвёртых, EGARCH-модель подтвердила наличие условной гетероскедастичности в динамике USDRUB. В-пятых, модель марковского переключения режимов выявила режимы, существенно различающиеся по дисперсии. Наконец, локальный анализ событий показал, что знак sentiment-показателя в части случаев согласуется с последующей краткосрочной динамикой курса.

ЗАКЛЮЧЕНИЕ

В данной бакалаврской работе была рассмотрена задача анализа связи между тональностью политических сообщений Дональда Трампа, относящихся к России, Украине, санкциям и международным конфликтам, и краткосрочной динамикой курса рубля к доллару США. Основная сложность данной задачи состоит в том, что исходная информация имеет текстовую форму и не может быть непосредственно включена в эконометрическую модель. Поэтому в работе была построена последовательная схема перехода от неструктурированных сообщений к числовым временным рядам, пригодным для статистического анализа.

В ходе работы были получены следующие основные результаты:

1. сформирован корпус релевантных политических сообщений, связанных с Россией, Украиной, санкциями, военными конфликтами и международными переговорами;
2. построены sentiment-показатели на основе нескольких лексиконов, что позволило сравнить различные варианты оценки тональности текста;
3. введены схемы временного сглаживания sentiment-переменных, учитывающие возможное сохранение информационного эффекта в течение нескольких дней;
4. текстовые признаки были объединены с финансовыми временными рядами, включающими курс USDRUB, цену Brent, индекс DXY, VIX, S&P 500 и цену золота;
5. методом Elastic Net среди 42 потенциальных объясняющих переменных были выбраны 7 признаков, включая как рыночные показатели, так и sentiment-переменные;
6. построена модель $ARMAX(1,0)$ – $EGARCH(1,1)$, в которой значимой оказалась sentiment-переменная LM_EXP01, а параметры условной волатильности оказались статистически значимыми;
7. построена модель марковского переключения режимов, выявившая два режима динамики курса, существенно различающиеся по уровню дисперсии;
8. проведён локальный анализ отдельных сообщений, показавший, что в части случаев знак sentiment-показателя согласуется с направлением

трёхдневной реакции курса USDRUB.

Полученные результаты позволяют сделать общий вывод: sentiment-показатели, построенные на основе политических сообщений, содержат информацию, связанную с краткосрочной динамикой курса USDRUB. При этом данная информация не заменяет рыночные факторы, а дополняет их. Наиболее корректно рассматривать текстовые признаки как часть общего информационного фона, который может усиливать или ослаблять ожидания участников рынка в отдельные периоды времени.

В то же время предложенный подход имеет ряд ограничений. Во-первых, словарный анализ тональности не всегда учитывает контекст, сарказм, цитаты и политическую риторику. Во-вторых, не каждое сообщение, содержащее слова, связанные с Россией или Украиной, действительно является экономически значимым для валютного рынка. В-третьих, курс рубля зависит от множества факторов, и даже на коротком горизонте его изменение может быть связано с нефтяными ценами, глобальной волатильностью, действиями регуляторов или другими новостями. Поэтому результаты работы следует интерпретировать как выявление статистической связи, а не как строгое доказательство причинного влияния отдельных сообщений.

Дальнейшее развитие исследования может идти в нескольких направлениях. Во-первых, можно расширить набор текстовых источников, включив сообщения других политиков, официальные заявления государственных органов и новостные публикации. Во-вторых, можно использовать более сложные методы анализа текста, учитывающие контекст и семантическую структуру сообщения. В-третьих, можно построить более детальную модель событийного анализа, включающую контроль за новостями нефтяного рынка, санкционными объявлениями и макроэкономическими публикациями.

Таким образом, в работе была построена и реализована схема анализа, объединяющая методы обработки текстов и эконометрики финансовых временных рядов. Проведённые расчёты показали, что политические сообщения могут быть формализованы в виде числовых признаков и использованы при моделировании валютного курса. Несмотря на ограничения, такой подход представляет интерес для задач анализа информационных рисков и исследования реакции финансовых рынков на политические события.