

Финансовые рынки

Прогноз волатильности криптовалют с использованием Google Trends и GDELT

Максим Алексеевич Тетерин

ORCID: 0000-0003-2750-9264

Аспирант департамента прикладной экономики факультета экономических наук,
Национальный исследовательский университет «Высшая школа экономики»
(РФ, 109028, Москва, Покровский бул., 11)
E-mail: mateterin@hse.ru

Аннотация

Криптовалюты стали важной частью современного финансового рынка. Одними из крупнейших являются Bitcoin, Ether, Litecoin и XRP, их капитализация составляет львиную долю рынка цифровых активов. При этом рынок криптовалют отличается от рынка традиционных финансовых активов более высоким уровнем волатильности цен, что делает методологию построения более точных прогнозов дневной волатильности для этих активов особенно значимой. Оценка дневной волатильности финансового актива является важным аспектом при разработке инвесторами торговых стратегий, определении направления политики риск-менеджмента, а также для учета в регуляторной политике рынка цифровых активов со стороны финансовых регуляторов. Настоящее исследование фокусируется на интеграции в HAR-RV-модель дополнительной нефинансовой информации, извлеченной из сети Интернет, для улучшения качества прогнозов реализованной волатильности четырех крупнейших криптовалют. В настоящем исследовании используются переменные, построенные на основе источников GDELT и Google Trends в качестве экзогенных для учета общественного интереса и сентимента (эмоциональных, поведенческих факторов). Переменные включаются в стандартную HAR-log-RV-модель, все вариации моделей для каждого актива оцениваются в скользящем окне на периоде с 1 января 2018 года по 23 июня 2024-го. Было получено более 2000 вневыборочных прогнозов на один день вперед, на основе которых рассчитываются функции потерь. Набор вневыборочных ошибок моделей для каждого актива тестируется с помощью процедуры MCS для определения набора статистически лучших прогнозных моделей. Предложенные в исследовании источники данных и построенные на их основе переменные позволяют значительно улучшить прогнозы реализованной волатильности четырех крупнейших криптовалют.

Ключевые слова: реализованная волатильность, реализованная ковариация, прогнозирование волатильности, HAR-RV-модель, сентимент, блокчейн

JEL: C32, C58, G15, G17

Автор благодарен Российскому научному фонду за финансовую поддержку исследования (project 24-28-00280, <https://rscf.ru/project/24-28-00280/>).

Автор также выражает признательность профессору А. А. Пересецкому за ценные комментарии.

Financial Markets

Cryptocurrency Volatility Forecasting Using Google Trends and GDELT

Maxim A. Teterin

ORCID: 0000-0003-2750-9264

Postgraduate Student, Applied Economics Department,
National Research University Higher School of Economics,^a e-mail: mateterin@hse.ru

^a 11, Pokrovskiy bul., Moscow, 109028, Russian Federation

Abstract

Bitcoin, Ether, Litecoin and XRP are among the largest cryptocurrencies and together have a market capitalization that constitutes a substantial portion of the digital asset market. At the same time, the cryptocurrency market differs from that of traditional financial assets because it has greater price volatility, which makes constructing more accurate forecasts of daily volatility for these assets especially useful. A more accurate estimate of the daily volatility of a financial asset is an important aspect of trading strategies and arriving at a risk management stance. Volatility forecasts are also a factor in the regulatory policy that financial authorities pursue for the digital asset market. The study collects supplementary non-financial information from the internet and feeds it into an HAR-RV model to improve the accuracy of realized volatility forecasts for the four largest cryptocurrencies. Variables are then derived from GDELT and Google Trends data sources as exogenous indicators of public interest and sentiment. These variables are incorporated into a standard HAR-log-RV model, and various versions of the model for each asset are applied to a moving window from 1 January 2018 to 23 June 2024 to yield more than 2,000 out-of-sample one-day-ahead forecasts on which loss functions are calculated. The set of out-of-sample model errors for each asset is tested using the MCS procedure to select the set of statistically superior forecast models. The data sources proposed in the study and the variables constructed from them significantly improve realized volatility forecasts for the four largest cryptocurrencies.

Keywords: realized volatility, realized covariance, volatility forecasting, HAR-RV model, sentiment, blockchain

JEL: C32, C58 G15, G17

Acknowledgements

The author is grateful for the financial support of the Russian Science Foundation (project 24-28-00280, <https://rscf.ru/project/24-28-00280/>) and would like to thank Professor A. Peresetsky for his valuable comments.

Введение

Криптовалюты, такие как Bitcoin, Ether, Litecoin и XRP, стали важной частью современного финансового рынка. Bitcoin (далее — BTC) является первой и крупнейшей по рыночной капитализации криптовалютой и во многом определяет динамику всего крипторынка. Ethereum¹ (далее — ETH), вторая по рыночной капитализации криптовалюта, предоставляет платформу для смарт-контрактов и децентрализованных приложений, благодаря чему обладает широким набором функций, выходящих за рамки простой расчетной единицы. Litecoin (далее — LTC) — один из пионеров в сфере альтернативных криптовалютных активов (альткоинов), который по своей технической инфраструктуре схож с BTC, но имеет более высокую скорость совершения транзакций. XRP² отличается целевой направленностью на обеспечение быстрых трансграничных транзакций и сотрудничеством с финансовыми институтами. Несмотря на различия в назначении и технической реализации, все названные криптоактивы занимают значительные доли рынка и отличаются от традиционных финансовых активов высокой волатильностью цен, которая может создавать как инвестиционные возможности, так и серьезные риски для инвесторов. Эта особенность обуславливает актуальность исследования подходов к прогнозированию волатильности криптовалют, способных учитывать природу и факторы таких ценовых колебаний.

Учет поведенческих факторов играет важную роль в современной практике управления инвестициями и научных исследованиях. Особенно актуальна задача учета поведенческих факторов на рынке криптовалют, характеризующемся высокой волатильностью и сильной зависимостью ценовых колебаний от новостного фона и интереса со стороны инвесторов. В связи с этим актуальным направлением исследований является включение внешней нефинансовой информации в модели прогнозирования волатильности криптоактивов.

Целью настоящей работы является улучшение качества прогнозов реализованной волатильности четырех крупнейших криптовалют: Bitcoin, Ether, Litecoin и XRP — посредством расширения базовой модели HAR-RV рядом дополнительных переменных. В качестве таких переменных используются индикаторы, характеризующие интерес пользователей (данные Google Trends), а также

¹ В академической литературе и в среде неквалифицированных инвесторов может упоминаться под названием Ethereum, что на самом деле является блокчейном (технологической платформой), на которой работает актив Ether.

² Ранее этот актив имел название Ripple и может упоминаться под старым названием.

новостная активность и средний сентимент по каждой криптовалюте (данные базы GDELT). Помимо этого, учитываются возможные перетоки волатильности между криптовалютами путем интеграции реализованных попарных ковариаций активов. Оценивание моделей проводится на скользящем временном окне для контроля за возможными структурными изменениями в данных.

Новизна исследования обусловлена совместным учетом факторов, ранее не использовавшихся в задачах прогнозирования волатильности криптовалют. Полученные результаты демонстрируют, что интеграция упомянутых внешних переменных позволяет существенно повысить точность краткосрочных прогнозов волатильности.

Результаты исследования могут быть полезны инвесторам и трейдерам для более точного оценивания волатильности актива и эффективного выстраивания стратегии по риск-менеджменту при работе с криптоактивами, основанной не только на прошлой динамике цен, но и на индикаторах новостного фона. Для регуляторов финансовых рынков представленная методология обеспечивает инструмент для более детального анализа природы резких ценовых колебаний и влияния новостных факторов на рынки цифровых активов, что может способствовать разработке новых мер по мониторингу и снижению системных рисков в этом секторе. Наконец, исследование представляет интерес и для академического сообщества, поскольку демонстрирует возможность интеграции альтернативных (нефинансовых) данных в модели реализованной волатильности и тем самым закладывает основу для дальнейших исследований, направленных на повышение точности прогнозов и понимания механизмов формирования волатильности на рынках криптовалют.

Работа имеет следующую структуру. В первом разделе приводится обзор литературы по тематике прогнозирования цен и волатильности криптовалют с использованием нефинансовых данных, а также данных с рынка другого актива. Во втором содержится описание использованных в работе данных, а также методы их предварительной обработки. В третьем разделе описана методология построения прогнозных моделей и алгоритма прогнозирования. В четвертом приведены результаты построенных моделей по каждому активу и анализ полученных результатов. В заключении подводятся итоги исследования.

1. Обзор литературы

В работе [Liang et al., 2022] исследуется возможность использования внешних данных для улучшения прогнозов реализованной

волатильности BTC. В качестве таких данных авторы используют временные ряды финансовых индексов, индексов неопределенности, а также данные Google Trends. Показано, что, несмотря на то что один из финансовых показателей (GVZ, индекс волатильности цен на золото) является одним из лучших предикторов для волатильности BTC, модель с использованием данных Google Trends в периоды высокой волатильности цен может давать более точные прогнозы, чем модель с индексом GVZ.

Исследование [Arratia, López-Barrantes, 2021] анализирует способность данных Google Trends улучшать прогнозы доходности BTC, ETH, XRP и LTC. Авторы выявляют наличие двусторонней причинности по Грейнджеру в обе стороны (от Google Trends к доходности и от доходности — к Google Trends) на нескольких горизонтах (лагах) и переходят к задаче прогнозирования. В качестве прогнозных моделей авторы используют стандартные статистические модели (ARIMA) и методы машинного обучения (нейросетевую авторегрессию, NNAR). Исследователи показывают, что включение данных Google Trends как экзогенных переменных в модели для каждой криптовалюты значительно снижает ошибку прогноза их доходностей.

В работе [Aslanidis et al., 2022] исследуется взаимосвязь между рынком криптовалют (на примере Bitcoin) и динамикой запросов сервиса Google Trends. Авторы конструируют индекс на основе данных поисковых запросов, связанных с криптовалютами (например, “cryptocurrency”, “blockchain”, “bitcoin” и др.), и исследуют взаимосвязь этого показателя с доходностью BTC с помощью энтропийных методов, которые позволяют учесть потенциальные нелинейные взаимосвязи двух случайных процессов. Выявлено наличие двунаправленной взаимосвязи между поисковыми запросами и доходностью BTC. Более того, показано, что эта связь остается устойчивой на протяжении нескольких дней: влияние поисковых запросов на доходность сохраняется дольше, чем обратное влияние.

В статье [Teterin, Peresetsky, 2024] исследуется возможность использования интернет-данных для улучшения однодневных прогнозов реализованной волатильности BTC. В качестве источника интернет-данных авторы используют сервис Google Trends по набору различных ключевых слов и фраз в качестве экзогенных переменных для HAR-log-RV-моделей. Результаты работы моделей показывают, что включение данных этого сервиса в модели реализованной волатильности помогает улучшить прогнозы по нескольким мерам точности прогнозов. Более того, модель с переменными, сконструированными на основе значений Google Trends, оказывается единственной, которая вошла в набор лучших моделей по методологии MCS.

Работа [Anamika, Subramaniam, 2022] исследует влияние эмоциональной окраски (сентимента) новостей на волатильность (среднеквадратичное и среднее абсолютное отклонение доходностей) ряда криптовалют. Авторы извлекают средний сентимент новостных заголовков по каждой криптовалюте из базы данных GDELT. Обнаружена отрицательная взаимосвязь разброса доходности и сентимента: при позитивном тоне новостных заголовков наблюдается уменьшение волатильности и более высокая степень согласованности ожиданий между инвесторами. Таким образом, показано, что эмоциональная окраска новостных заголовков является важным индикатором, который может быть использован для прогнозирования показателей рынка криптовалют.

В [Brauneis, Sahiner, 2024] также исследуется способность сентимента новостных заголовков улучшать качество вневыборочных прогнозов для набора из восьми самых популярных криптовалют. Авторы извлекают сентимент из агрегатора новостей *headlinhunter.ai*. В качестве прогнозных моделей используется набор моделей машинного обучения (нейронные сети и градиентные бустинги над деревьями), а также HAR-RV-модели. Каждая модель оснащается матрицей экзогенных предикторов, включающих количество новостных статей по криптовалюте и средний сентимент этих новостей. Не обнаружено заметного улучшения качества прогнозов волатильности от включения данных переменных в HAR-RV-модель для всех криптовалют, кроме BTC. Однако некоторые модели машинного обучения, оснащенные данными новостей, превзошли свои аналоги, не включающие данные переменные, что подтверждает важность их учета при анализе волатильности криптовалют.

Ряд моделей для прогнозирования волатильности BTC, ETH, LTC и XRP: GARCH(1,1), случайный лес (Random Forest), модификации нейронной сети LSTM (Long Short-Term Memory) — рассматриваются в [Wang et al., 2023]. В модели включаются экзогенные переменные, в том числе сентимент и данные Google Trends. Авторы показали, что увеличение частоты поисковых запросов, связанных с криптоактивами, часто предшествует росту волатильности анализируемых активов. Кроме того, выявлена зависимость между изменением тональности финансовых новостей и скачками волатильности. Также проведен дополнительный анализ важности переменных для LSTM-моделей с помощью подхода SHAP (SHapley Additive exPlanations), который показал, что одной из самых важных переменных является прошлая (лагированная) волатильность, но при этом данные Google Trends и тональность являются важным фактором, который положительно влияет на волатильность исследуемых активов.

Стоит отметить, что для изучения влияния новостей на финансовые показатели криптовалют исследователи могут использовать разные наборы баз данных. Например, в [Al Guindy, 2021; Alipour, Charandabi, 2023; Kyriazis et al., 2023; Ullah et al., 2022] широко используются данные социальных сетей, например сети X (ранее Twitter).

Помимо внешних источников для прогнозных моделей также может быть использовано наличие потенциальной сонаправленности в движении цен криптовалют — перетока волатильности (*volatility spillover*).

Например, в работе [Chen, 2024] показано, что точность прогнозов реализованной волатильности ETH можно улучшить при учете ряда внешних факторов, включая волатильность BTC. В своей работе исследователь применяет четыре варианта HAR-RV-модели для прогнозирования волатильности ETH. Первая модель использует только лаговые дневные, недельные и месячные значения волатильности ETH (стандартная HAR-RV-модель). Вторая модель включает фиктивные переменные для структурных сдвигов. В третью модель добавляются экзогенные переменные, такие как реализованная волатильность BTC. Последняя модель включает все введенные переменные. Результаты показывают, что волатильность BTC помогает уточнить прогнозы волатильности ETH на различных горизонтах прогнозирования, что подтверждает важность волатильности BTC как экзогенной переменной для ETH.

В [Harb et al., 2024] исследуется динамика перетока волатильности (*volatility spillover*) между четырьмя самыми торгуемыми криптовалютами (BTC, ETH, LTC, XRP) и анализируется их взаимосвязь с рынками акций и облигаций США. Авторы применяют двухшаговую процедуру оценивания: моделируется волатильность криптовалют, а затем используются одновременные уравнения для оценивания взаимосвязи этих активов. Результаты показывают, что существует сложная структура перетока волатильности между криптовалютами, что подтверждает важность использования данных одной криптовалюты для анализа волатильности другой.

В [Teterin, Peresetsky, 2025] показано, что использование данных одной криптовалюты может улучшать прогнозы другой. Авторы применяют лагированные данные реализованной волатильности BTC как экзогенную переменную в HAR-RV-модели для ETH, и наоборот. Кроме того, используются реализованные ковариации двух криптовалют в качестве дополнительных предикторов для HAR-RV-моделей каждой криптовалюты. Показано, что добавление реализованной ковариации BTC и ETH для модели BTC улучшает однодневные вневыборочные прогнозы. Для ETH лучше оказывается модель, которая помимо реализованных

ковариаций включает лагированные значения реализованной волатильности BTC.

В настоящей работе данные Google Trends и базы GDELT используются как источники для генерации экзогенных переменных для HAR-RV-моделей прогноза волатильности четырех криптовалют на день вперед. Кроме того, для учета наличия взаимосвязи криптовалют друг с другом и потенциального перетока волатильности использованы попарные реализованные ковариации доходностей активов в качестве независимых переменных модели. В следующем разделе содержится более подробное описание этих источников, особенности предобработки данных и генерации экзогенных переменных.

2. Данные

Реализованные волатильности и ковариации

Моделирование волатильности финансового актива — один из важных аспектов для риск-менеджмента и разработки торговых стратегий на финансовых рынках. В академической литературе и практической области одним из самых распространенных подходов является моделирование волатильности как ненаблюдаемого (латентного) процесса. Для этой цели традиционно используют модели из семейства GARCH, введенные в [Engle, 1982; Engle, Bollerslev, 1986], а также их расширения T-GARCH, E-GARCH из [Nelson, 1996; Rabemananjara, Zakoian, 1993]. Такой подход к моделированию волатильности используется также и для рынка криптовалют [Маневич и др., 2022; Dyhrberg, 2016; Gyamerah, 2019].

Альтернативный подход к моделированию волатильности (Heterogeneous AutoRegressive Realized Volatility, HAR-RV), получивший в последнее время широкое распространение в научной литературе, предложен в [Corsi, 2009]. В отличие от параметрического подхода в моделях GARCH альтернативный подход основан на непараметрическом оценивании волатильности (реализованной волатильности, Realized Volatility, RV) с помощью высокочастотных (в основном пятиминутных) торговых данных финансового актива. Эта модель, несмотря на довольно простую структуру, позволяет успешно моделировать и прогнозировать волатильность финансового актива. Она способна эффективно моделировать эффекты долгой памяти (long memory), которые часто встречаются во временных рядах доходностей высокорисковых финансовых активов, в том числе криптовалют. Эта модель стала широко использоваться не только для традиционного финансового рынка, но и для рынка криптовалют [Аганин

и др., 2023; Bergsli et al., 2022; Bouri et al., 2021; Dudek et al., 2024; Ma et al., 2014; Manevich, Ignatov, 2023; Mastro, 2014]. В настоящей работе используется подход, предложенный в [Andersen, Bollerslev, 1998; Andersen et al., 2006; Liu et al., 2015]: рассчитывать дневные реализованные волатильности на основе пятиминутных торговых интервалов для всех четырех криптовалют:

$$RV_t = \sqrt{\sum_{i=1}^{288} r_{t,i}^2},$$

(1)

где t — индекс времени (здесь день), $r_{t,i} = \ln(P_{t,i}) - \ln(P_{t,i-1})$ — логарифмическая доходность актива в пятиминутном интервале i , всего каждый день наблюдается 288 пятиминутных интервалов при отсутствии пропусков. Высокочастотные данные торгов BTC, ETH, LTC и XRP получены с помощью CryptoDataDownload.com — ресурса, который предоставляет данные торгов с различных криптобирж. В настоящей работе использованы данные криптовалютной биржи Bitstamp.

Нужно сказать, что хотя отсутствие торговых данных в каком-либо интервале является довольно редким событием, в программном коде всё же выполнена процедура, позволяющая в таком случае корректно оценивать реализованную волатильность [Teterin, Peresetsky, 2024].

Описательные статистики и динамика реализованной волатильности, рассчитанная по вышеописанной процедуре, представлены в табл. 1 и на рис. 1 соответственно.

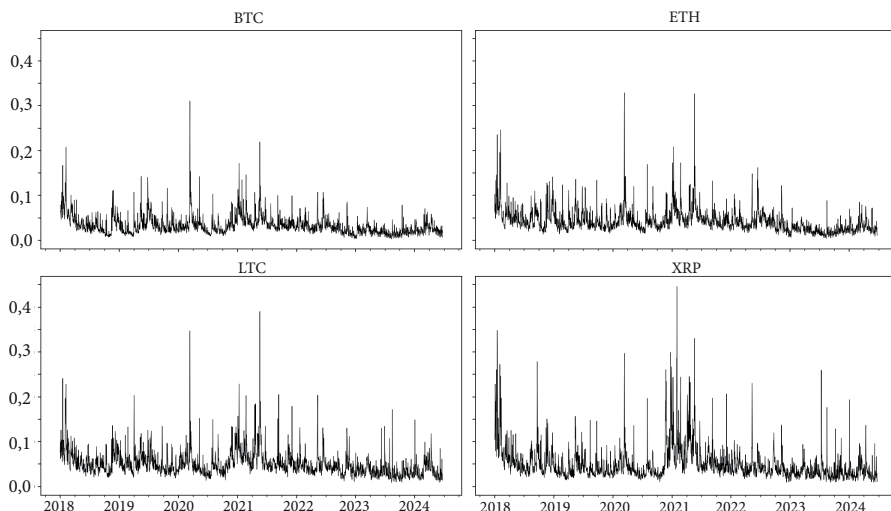
Т а б л и ц а 1

Описательные статистики реализованной волатильности
и реализованной доходности криптовалют

T a b l e 1

Statistical Profile for the Realized Volatility and Realized Return of Cryptocurrencies					
Показатель	Описательная статистика	BTC	ETH	LTC	XRP
RV — реализованная волатильность	T	2366	2366	2366	2366
	mean	0,03340	0,04270	0,05030	0,04990
	min	0,00270	0,00410	0,00860	0,00940
	max	0,31040	0,32890	0,39010	0,44600
	std	0,02200	0,02660	0,02910	0,03780
RR — реализованная доходность	T	2366	2366	2366	2366
	mean	0,00140	0,00180	0,00070	0,00100
	min	−0,3846	−0,4338	−0,3574	−0,4271
	max	0,19900	0,27160	0,33450	0,73280
	std	0,03670	0,04700	0,05020	0,05760

Источник: составлено автором.



Источник: составлено автором на основе данных CryptoDataDownload.com.

Рис. 1. Реализованная волатильность BTC, ETH, LTC, XRP
с 1 января 2018 до 23 июня 2024 года

Fig. 1. Realized Volatility of BTC, ETH, LTC, XRP From 1 January 2018 to 23 June 2024

Из описательных статистик, приведенных в табл. 1, можно сделать предварительные выводы о различных паттернах в волатильности рассматриваемых криптовалют.

Так, LTC и XRP характеризуются более высокими значениями средней волатильности по сравнению с BTC и ETH, что может быть связано с тем, что последние являются одними из самых популярных монет на крипторынке с большим объемом торгов, что сильно ограничивает возможности манипуляции рынком. Несмотря на эти различия в среднем уровне волатильности, максимальные значения для всех криптовалют оказываются крайне высокими, что говорит о наличии резких скачков в ценовой динамике активов.

По динамике реализованных волатильностей криптовалют на рис. 1 можно выделить несколько периодов, когда волатильность была особенно высокой. Так, в начале 2018 года был зафиксирован резкий скачок волатильности всех четырех валют, что является следствием пузыря на рынке криптовалют. Другой ярко выраженный период высокой волатильности наблюдается в первой половине 2020 года, и он связан с началом пандемии COVID-19. Этот шок оказал значительное влияние на все финансовые рынки, в том числе и на рынок криптовалют. Резких изменений в волатильности всех активов в начале 2022 года не наблюдалось.

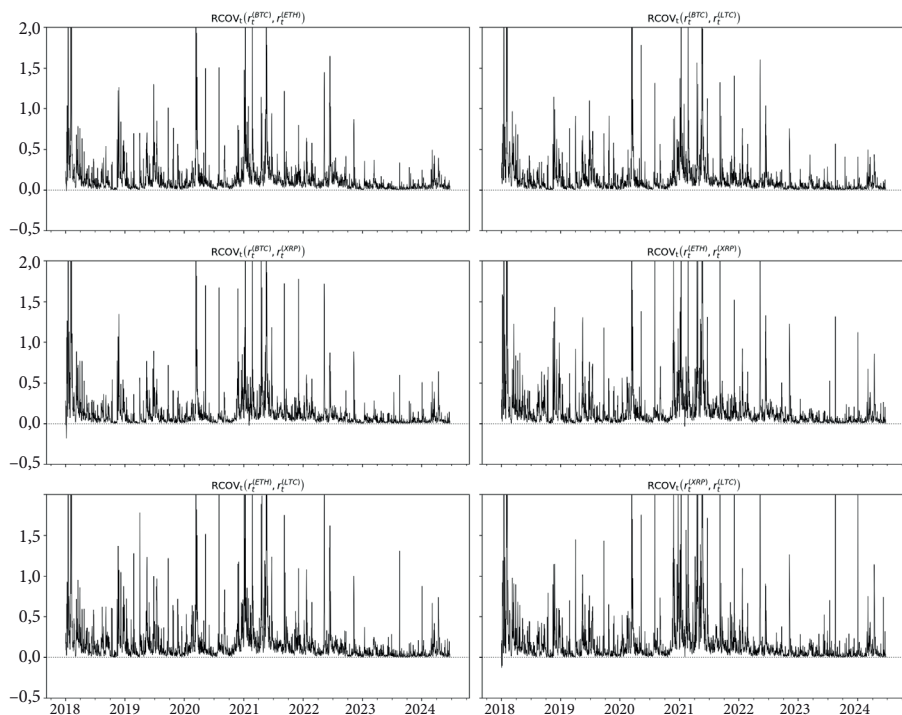
Сравнивая поведение реализованной волатильности четырех активов, можно увидеть некоторые различия. Так, в данных по LTC и XRP наблюдаются резкие краткосрочные скачки, в то время как BTC и ETH характеризуются более стабильным поведением в этот же период. Это особенно заметно в период 2021–2022 годов, когда BTC и ETH имели относительно малые значения волатильности по сравнению с двумя другими криптовалютами. Этот факт может объясняться более высокой рыночной капитализацией BTC и ETH и более широким признанием, что обычно способствует повышению стабильности цен активов.

Поскольку в рамках настоящего исследования предполагается улучшение прогнозных моделей с помощью данных других криптоактивов, при анализе необходимо ввести определение реализованной ковариации, которая рассчитывается по процедуре из [Barndorff-Nielsen, Shephard, 2004], схожей с расчетом реализованной волатильности одного актива по формуле :

$$RCov_t(r_t^k, r_t^j) = \sum_{i=1}^M r_{t,i}^k r_{t,i}^j, \quad k \neq j, M = 288, \quad (2)$$

где t — индекс времени (здесь день), k и j — индексы двух активов, пара $(r_{t,i}^k, r_{t,i}^j)$ — логарифмические доходности двух активов с индексами k и j в пятиминутном интервале i , M — число пятиминутных интервалов (288). Динамика реализованных ковариаций представлена на рис. 2.

По динамике попарных реализованных ковариаций, представленной на рис. 2, для рассматриваемых криптовалют можно сделать важные выводы. Практически на всем рассматриваемом интервале наблюдается положительная ковариация для всех пар, только несколько наблюдений в начале 2018 года принимают отрицательные значения. В периоды шоков в начале 2018 и 2020 годов ковариация между активами резко возрастает. Этот факт может указывать на то, что в периоды шоков криптовалюты становятся более взаимозависимыми и имеют схожую динамику цен. Особенно ярко это проявляется в парах BTC — ETH и BTC — LTC, где наблюдается резкий рост реализованных ковариаций в указанные периоды времени. Примечательно, что есть и периоды относительно низких ковариаций в 2019 году и начале 2020-го (до начала пандемии COVID-19), когда ковариация между активами колеблется около нуля. Однако даже в эти относительно стабильные периоды криптовалюты продолжают реагировать схожим образом на внешние шоки, что подтверждает их тесную взаимосвязь и предоставляет возможность использовать эту информацию для прогнозных моделей.



Примечание. Значения умножены на 100, вертикальная ось ограничена значением 2 для лучшей визуализации.

Источник: составлено автором на основе данных CryptoDataDownload.com.

Рис. 2. Попарные реализованные ковариации BTC, ETH, LTC, XRP с 1 января 2018 до 23 июня 2024 года

Fig. 2. Pairwise Realized Covariances of BTC, ETH, LTC, XRP From 1 January 2018 to 23 June 2024

Данные Google Trends

Одним из источников данных для генерации экзогенных переменных для прогнозных моделей является сервис Google Trends. Он позволяет исследовать частоту запросов по определенным тематикам, ключевым словам или словосочетаниям и широко применяется не только в коммерческих целях, но и в научных исследованиях.

Платформа предоставляет доступ к случайной выборке поисковых запросов пользователей Google Search в режиме реального времени. Поскольку Google не раскрывает данные сырых поисковых запросов, внутри платформы Google Trends данные агрегированы путем создания специального индекса. Этот индекс показывает соотношение количества запросов по определенной тематике

к общему числу запросов за выбранный период и в выбранной географической локации. Значения индекса лежат в интервале от 0 до 100 и интерпретируются как относительная популярность той или иной поисковой выдачи. Деление на общее количество запросов важно, так как оно учитывает возможные колебания в числе пользователей, которые делают поисковые запросы.

Сервис предоставляет данные с 2004 года с различными уровнями детализации — от годового до почасового. В настоящем исследовании используются ежедневные данные. Работа с ежедневными данными сервиса за длинный промежуток времени сопряжена с определенными проблемами. Одна из них заключается в ограниченном количестве дневных данных, которые можно извлечь за один раз, — 270 дней. Если количество дней превышает эту величину, данные будут автоматически приведены к недельной гранулярности. Это ограничение достаточно просто преодолеть, загружая данные итеративно по частям. Более серьезной проблемой является непосредственно методика построения индекса. Так как количество запросов делится на общее количество всех поисков, данные из двух выгрузок за различные периоды будут несопоставимы из-за нормирования разными знаменателями.

Для преодоления этой проблемы в работе использована процедура, описанная в [Teterin, Peresetsky, 2024]. Алгоритм, предложенный авторами, заключается в итеративной загрузке данных из сервиса с пересекающимися интервалами. Данные из соседних выгрузок за один и тот же период используются для построения шкалирующего коэффициента, который отражает соотношение знаменателей исходных данных. Далее данные за более ранний период домножаются на шкалирующий коэффициент, что делает их сопоставимыми. Отметим, что в этом случае данные могут выходить за верхнюю границу в 100 пунктов, но тогда можно интерпретировать это значение. Например, если за сегодняшний день показатель Google Trends по тематике равен 200, а за вчера был равен 50, то количество запросов по тематике возросло в четыре раза.

В табл. 2 и на рис. 3 приведены описательные статистики и динамика запросов, шкалированных по описанной выше методологии, по ключевым словам bitcoin, ethereum, litecoin, ripple.

Сопоставление динамики запросов Google Trends (рис. 3) с динамикой реализованной волатильности соответствующей криптовалюты (рис. 1) показывает, что, как и динамика реализованной волатильности, динамика поисковых запросов соответствующих активов возрастает в периоды шоков в начале 2018 года и в первом квартале 2020-го. Это свидетельствует о наличии потенциальной связи между ростом интереса к криптовалютам в определен-

ный период и увеличением волатильности цен. Стоит отметить, что резкие скачки поисковых запросов криптовалют могут как предшествовать, так и сопровождать периоды резких ценовых изменений. Сравнение динамики данных Google Trends и реализованной волатильности показывает, что изменение активности в поиске тесно связано с периодами повышенной волатильности на рынке криптовалют. Анализ и использование поисковых тенденций может служить полезным инструментом для прогнозирования динамики реализованной волатильности.

Т а б л и ц а 2

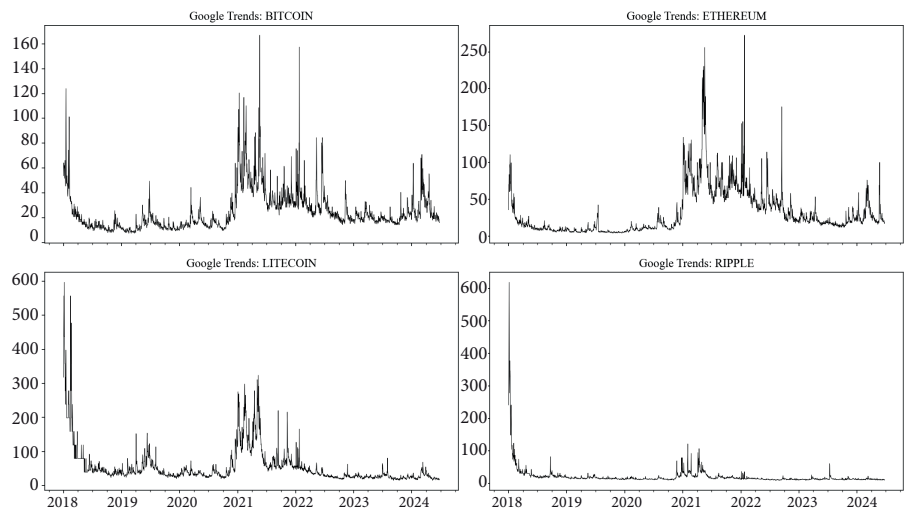
Описательные статистики масштабированных данных Google Trends
по каждому ключевому слову

Table 2

Statistical Profile of the Scaled Google Trends Data for Each Keyword

Описательная статистика	Bitcoin	Ethereum	Ripple	Litecoin
T	2366	2366	2366	2366
mean	24,83	30,968	20,80	56,23
std	16,08	30,808	29,47	56,93
min	7,42	3,77	9,45	15,85
max	166,98	272,61	618,83	596,41

Источник: составлено автором на основе данных Google Trends.



Источник: составлено автором на основе данных Google Trends.

Рис. 3. Значения шкалированного индекса Google Trends по ключевым словам
для BTC, ETH, LTC, XRP с 1 января 2018 до 23 июня 2024 года

Fig. 3. Values of the Scaled Google Trends Index for Keywords for BTC, ETH, LTC, XRP
from 1 January 2018 to 23 June 2024

Данные GDELT

В базе данных GDELT (Global Database of Events, Language, and Tone) хранятся и анализируются новостные сообщения со всего мира. Источник данных охватывает большой спектр новостных ресурсов: онлайн-издания, телевидение, печатную прессу, — предоставляя структурированную в виде таблицы информацию о событиях, их участниках, географической локации и эмоциональной окраске (сентименте, или тоне) каждой новости. База данных GDELT пополняется в режиме реального времени и позволяет пользователям анализировать динамику общественных настроений и их направленность.

Данные в GDELT собираются из тысяч источников новостей на разных языках. Каждый новостной материал проходит обработку, в результате которой выделяются ее основные атрибуты: дата публикации, идентификатор источника, тематика статьи, географическая принадлежность новостного события, а также его тональность (сентимент), что позволяет проводить количественный анализ медиа по интересующей тематике.

Записи в GDELT представлены в формате таблицы, где каждая строка соответствует конкретной статье (документу). Структура таблицы включает несколько полей (столбцов), которые представляют особую важность в настоящем исследовании:

- время публикации — временная метка публикации новостного материала;
- уникальный идентификатор документа;
- URL-адрес статьи;
- тональность (сентимент) — числовой показатель, характеризующий эмоциональную окраску каждого документа от резко негативного (–10) до крайне положительного (10).

База данных поддерживается, в частности, Google, что позволило автору настоящего исследования использовать информацию этого ресурса с помощью инструментов, предоставляемых корпорацией. Так, на основе указанных полей был сформирован запрос на языке SQL в системе Google BigQuery (система управления базами данных с собственным диалектом языка SQL) по следующей процедуре:

- 1) с помощью регулярных выражений отобраны все документы из базы данных, где упоминаются *cryptocurrency, blockchain, bitcoin, btc, ethereum, litecoin, ltc, ripple, xrp*;
- 2) рассчитано общее количество уникальных документов (новостей) с упоминанием криптовалют и средний тон по всем документам за анализируемый период;

3) рассчитано общее количество уникальных документов с упоминанием каждой криптовалюты в отдельности, а также средний тон документов по ним за анализируемый период.

Описательные статистики собранных данных представлены в табл. 3. Динамика количества документов (статей) по каждой криптовалюте представлена на рис. 4, а динамика средней тональности этих статей по каждой криптовалюте — на рис. 5.

Т а б л и ц а 3

Описательные статистики данных GDELT по каждому ключевому слову

Table 3

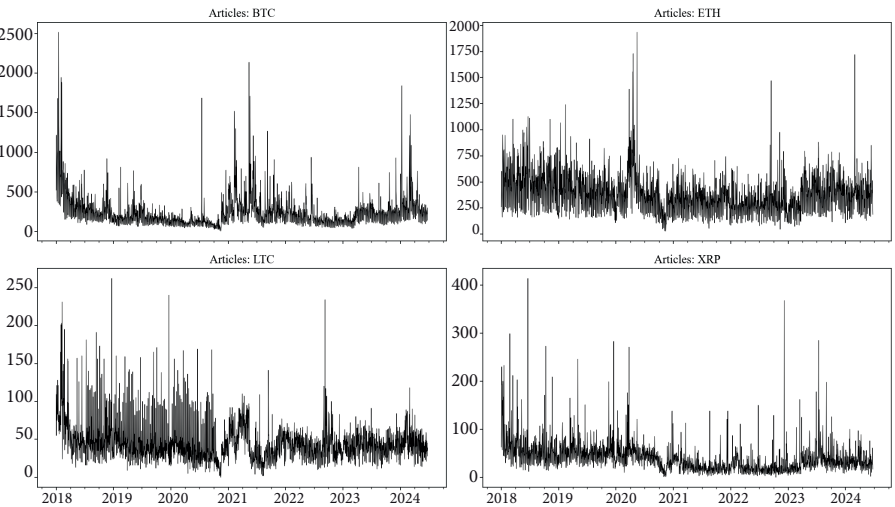
Statistical Profile of the GDELT Data for Each Keyword					
Переменные	Описательная статистика	BTC	ETH	LTC	XRP
A (Articles) — количество новостных статей	T	2366	2366	2366	2366
	mean	245,4	382,6	45,9	40,4
	std	208,6	186,6	26,7	31,6
	min	5	29	0	0
	max	2516	1934	262	414
S (Sentiment) — среднедневной сентимент (тон) статей	T	2366	2366	2366	2366
	mean	−0,3944	0,1921	−0,5063	−1,2591
	std	0,7424	0,8336	1,0154	1,1601
	min	−4,3688	−4,8094	−6,1242	−7,0593
	max	3,2731	6,8093	3,7098	5,8644

Источник: составлено автором на основе данных GDELT.

Визуальный анализ рис. 4 демонстрирует паттерны, схожие с теми, что получены и для данных Google Trends. Резкий рост публикаций сопровождается резкими скачками в поисковых запросах.

Надо заметить, что средняя тональность этих статей (рис. 5) колеблется вблизи нейтрального уровня, однако в периоды высокой информационной активности демонстрирует значительные колебания. Тональность новостей адаптируется к текущей рыночной ситуации, усиливаясь в периоды нестабильности.

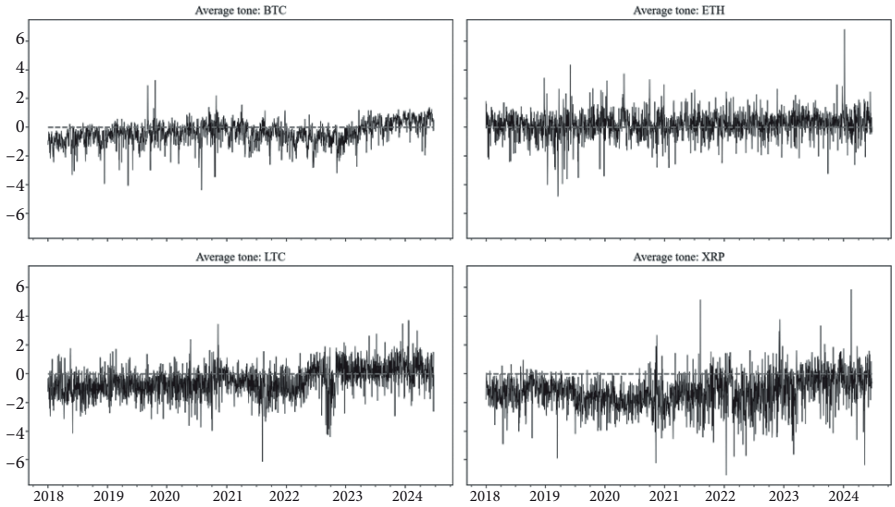
Сопоставляя данные GDELT с динамикой реализованной волатильности на рис. 1, можно заметить, что резкие пики волатильности для всех анализируемых криптовалют наблюдаются примерно в те же временные периоды, когда происходит увеличение количества новостных публикаций и усиление колебаний тональности. Этот факт указывает на возможную связь между объемом и эмоциональной окраской новостного потока и волатильностью криптовалютного рынка.



Источник: составлено автором на основе данных GDELT.

Рис. 4. Динамика количества новостных статей по тематике BTC, ETH, LTC, XRP с 1 января 2018 до 23 июня 2024 года

Fig. 4. Dynamics of the Number of News Articles on BTC, ETH, LTC, and XRP From 1 January 2018 to 23 June 2024



Примечание: пунктирная линия — нулевое значение.

Источник: составлено автором на основе данных GDELT.

Рис. 5. Динамика среднего сентимента новостных статей по тематике BTC, ETH, LTC, XRP с 1 января 2018 до 23 июня 2024 года

Fig. 5. Dynamics of the Average News Sentiment on BTC, ETH, LTC, XRP From 1 January 2018 to 23 June 2024

Рост числа новостных публикаций и изменение их тональности зачастую предшествуют резким колебаниям волатильности, что позволяет предположить, что внимание медиа к криптовалютам, выражающееся в количестве публикаций, может выступать в качестве индикатора будущих изменений цен криптовалют. Интенсивное освещение криптовалютного рынка (и отдельных криптовалют, в частности) в СМИ может отражать изменяющиеся ожидания инвесторов и влиять на волатильность цен. Таким образом, мониторинг новостного фона, фиксируемого GDELT, может выступать ценным предиктором для значений волатильности криптовалют.

3. Методология

Основной модельный инструментарий настоящего исследования — HAR-RV-модель, предложенная в [Corsi, 2009]. Уравнение модели обычно записывается в следующем виде:

$$RV_{t+1} = \beta_0 + \beta_d RV_t^d + \beta_w RV_t^w + \beta_m RV_t^m + \varepsilon_{t+1}, \quad (3)$$

где RV_t — реализованная волатильность исследуемого актива, рассчитанная с помощью формул (1) и (2), $RV_t^d = RV_{t-1}$ — лагированное значение реализованной волатильности, $RV_t^w = \frac{1}{w} \sum_{j=1}^w RV_{t-j}$ — среднее значение реализованной волатильности за неделю (среднее значение за w дней); $RV_t^m = \frac{1}{m} \sum_{j=1}^m RV_{t-j}$ — среднее значение реализованной волатильности за месяц (среднее значение за m дней).

В качестве w и m обычно выбираются значения в 5 и 22 дня, что соответствует количеству торговых дней за неделю и месяц. Хотя криптовалюты торгуются ежедневно, выбор большего количества дней для расчета указанных средних величин не оказывает, как показано в [Маневич и др., 2022], значительного влияния на качество прогнозов, поэтому в настоящей работе использованы значения величин w и m , равные 5 и 22 дням соответственно.

Также нужно отметить, что в работе в уравнении (3) вместо RV_t используются натуральные логарифмы реализованной волатильности $\ln(RV_t)$. При расчете точности прогнозов с помощью функций потерь (ошибок) прогнозные значения приведены к исходному масштабу обратным преобразованием. Логарифмирование значения реализованной волатильности обусловлено несколькими причинами. Во-первых, гарантируется неотрицательность прогнозов волатильности. Во-вторых, в работах [Аганин и др., 2023; Aalborg et al., 2019; Bergsli et al., 2022; Chiriac, Voev, 2011; Clements, Preve, 2021; Sapkota, 2022] показано, что использование модели в логарифмах дает более точные прогнозы. Это может объясняться тем, что реализованная волатильность RV_t имеет лог-нормальное распределение.

На исследуемом в настоящей работе интервале вероятно наличие структурных изменений на рынке криптовалют, связанных с периодами экономической неопределенности, кризисами, изменениями в правилах торговли. В связи с этим трудно предполагать постоянство параметров модели на всем интервале, поэтому используется алгоритм прогнозирования в скользящих окнах, широко применяемый для такого рода анализа [Аганин и др., 2023; Dudek et al., 2024; Manevich, Ignatov, 2023]. В работе [Teterin, Peresetsky, 2024] оценен оптимальный (минимизирующий ошибки вневыборочного прогноза) размер скользящего окна прогнозирования, равный 300 дням. В настоящем исследовании также используются размер окна в 300 дней.

Полный алгоритм прогнозирования, используемый в исследовании, предполагает, что для каждой модели и каждой криптовалюты необходимо:

- 1) оценить модель на первых 300 наблюдениях;
- 2) сделать прогноз вневыборочного наблюдения с индексом $t+1$;
- 3) рассчитать ошибку прогноза для наблюдения $t+1$;
- 4) сдвинуть выборку (окно) на 1 шаг вперед: добавить наблюдение с индексом $t+1$ и выбросить первое наблюдение с индексом t_0 ;
- 5) повторять шаги 2–4 до достижения последнего наблюдения T ;
- 6) рассчитать средние значения функций потерь;
- 7) определить набор лучших моделей по каждой функции потерь.

В работе используются стандартные функции потерь (ошибок): RMSE, MAE, MAPE:

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (h_i - \hat{h}_i)^2}, MAE = \frac{1}{T} \sum_{i=1}^T |h_i - \hat{h}_i|, MAPE = \frac{1}{T} \sum_{i=1}^T \left| \frac{h_i - \hat{h}_i}{h_i} \right|,$$

где h_i — истинное значение анализируемой величины, $\hat{h}_i$ — прогнозируемое значение.

Для удобства записи регрессионных уравнений для четырех криптовалют в работе используются следующие обозначения.

Логарифм реализованной волатильности для зависимого (с точки зрения записи уравнений) актива j обозначим как $y_t = \ln(RV_t^j)$, $j \in (1, 4)$.

Логарифмированная реализованная волатильность для остальных активов, которые будут выступать экзогенными для актива j , записывается в виде $x_{k,t} = \ln(RV_t^k)$, $k \neq j$.

Например, для уравнения BTC его реализованная волатильность будет записана как y_t , а волатильность ETH, LTC и XRP будет записана в уравнении как $x_{k,t}$, $k \neq j$.

В уравнениях использованы попарные реализованные ковариации доходностей активов друг с другом. Поскольку практически на всем анализируемом промежутке времени попарные ковариации являются неотрицательными, в уравнение включается абсолютное значение ковариации зависимого актива j и экзогенного актива k : $C_{k,t} = RCov_t(r_t^k, r_t^j)$.

Для удобства записи NAR-структуры регрессионных уравнений вводится также множество $\{d, w, m\}$, которое отвечает за усредненные дневные (первый лаг наблюдения), недельные и месячные значения анализируемой величины.

Все введенные обозначения, а также обозначения переменных, извлеченных из внешних источников (Google Trends, GDELT), вынесены в табл. 4.

Переменная ΔS_t — это изменение значения сентимента в момент времени t , рассчитанное по формуле:

$$\Delta S_t = \text{sgn}(S_t) \times \ln(1 + |S_t|) - \text{sgn}(S_{t-1}) \times \ln(1 + |S_{t-1}|).$$

Это уравнение является аналогом логарифмического прироста абсолютного значения сентимента S_t , но, поскольку S_t может менять свой знак (рис. 5), в расчет введен оператор знака $\text{sgn}(\cdot)$ для учета этого изменения. Прибавление единицы гарантирует отсутствие стремления логарифма к бесконечности.

Т а б л и ц а 4

Описание используемых в моделях переменных

Table 4

Variables Employed in the Models		
Переменные	Описание переменной	Источник
$x_{k,t}$	Значение реализованной волатильности «экзогенного» актива k	Рыночные данные
$C_{k,t}$	Абсолютное значение реализованной ковариации целевой переменной с активом k	Рыночные данные
$GT_{i,t}$	Значение индекса Google Trends для соответствующего актива в период i , где $i \in \{d, w, m\}$	Google Trends
A_t	Количество новостных статей с упоминанием соответствующего актива	GDELT
A_t^{crypto}	Количество новостных статей с упоминанием всех активов и общей тематики криптовалют	GDELT

О к о н ч а н и е т а б л и ц ы 4

Переменные	Описание переменной	Источник
S_t	Среднедневное значение сентимента (тона) всех новостных статей по тематике	GDELT
S_t^{crypto}	Среднедневное значение сентимента (тона) всех новостных статей по анализируемым активам и общей тематике криптовалют	GDELT
ΔS_t	Изменение значения сентимента (тона) с учетом знака	GDELT
ΔS_t^{crypto}	Изменение значения сентимента (тона) с учетом знака по анализируемым активам и общей тематике криптовалют	GDELT

Источник: составлено автором.

Рассматриваются следующие регрессионные уравнения для прогнозирования реализованной волатильности каждой криптовалюты. Первая вводимая модель (E1) — это обычная HAR-log-RV(1, 5, 22)-модель (далее — HAR-RV), переписанная в соответствии с введенными обозначениями. В настоящем исследовании модель (E1) является базовой. Значения функций потерь каждой следующей модели будут сравниваться со значениями функций потерь базовой.

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \varepsilon_{t+1}.$$

(E1)

Модель (E2) расширяет базовую модель (E1), включая лагированные значения реализованной волатильности других криптовалют. Так, для BTC это ETH, LTC и XRP; для ETH — BTC, LTC и XRP. Включение этих переменных необходимо для учета потенциального эффекта перетока волатильности (volatility spillover), когда волатильность одного актива может влиять на волатильность другого. Учитывая высокую взаимосвязанность рынков криптовалют, предполагается, что такие связи будут полезны при прогнозировании реализованной волатильности:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \theta_i x_{i, t} + \varepsilon_{t+1}.$$

(E2)

В модели (E3) предпринята попытка учесть эффект перетока волатильности с использованием логарифмов реализованных ковариаций $C_{k, i}$:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \alpha_i \ln(C_{i, t}) + \varepsilon_{t+1}.$$

(E3)

Учитывая наличие резких скачков в попарных реализованных ковариациях (рис. 2), в модели (E4) вводится логарифмированное изменение реализованных ковариаций. Предполагается, что ковариация может отражать долгосрочные структурные зависимости, которые уже учитываются в (E3), резкие скачки ковариации могут нести критически важную информацию для прогнозирования волатильности.

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \gamma_i \ln \left(\frac{C_{i, t}}{C_{i, t-1}} \right) + \varepsilon_{t+1}. \quad (E4)$$

Модель (E5) является естественным расширением моделей (E3) и (E4), в ней учитывается долгосрочное влияние попарных ковариаций, а также резкие изменения ее динамики, которые потенциально могут уловить скачки будущей волатильности:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \alpha_i \ln(C_{i, t}) + \sum_{i=1}^3 \gamma_i \ln \left(\frac{C_{i, t}}{C_{i, t-1}} \right) + \varepsilon_{t+1}. \quad (E5)$$

В моделях (E6)–(E10) введены дополнительные экзогенные переменные, которые отражают интерес инвесторов. В этом наборе используются данные Google Trends для соответствующей криптовалюты: запрос “bitcoin” для уравнений BTC, запрос “ethereum” — для уравнений ETH и т. д. Первой моделью в данном наборе является модель (E6) с логарифмированным значением индекса Google Trends за предыдущий день ($GT_{i, t}$, где $i = d$). Предполагается, что переменная учитывает интерес инвесторов и может оказывать влияние на будущую волатильность:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_d, t) + \varepsilon_{t+1}. \quad (E6)$$

Модель (E7) использует HAR-структуру для величины индекса Google Trends и подразумевает наличие гетерогенности (краткосрочного и долгосрочного влияния) влияния интереса инвесторов на реализованную волатильность. В данном случае используются такие же значения $w = 5$ и $m = 22$, как и для реализованной волатильности.

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_i, t) + \varepsilon_{t+1}. \quad (E7)$$

В модель (E8) вводится логарифмический прирост запросов Google Trends по соответствующей криптовалюте аналогично тому, как это было сделано для реализованных ковариаций в моделях (E4) и (E5). Эта переменная призвана уловить пики в поисковых запросах, которые могут отражать изменения в настроениях инвесторов и служить опережающими индикаторами скачков волатильности:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \varepsilon_{t+1}. \quad (E8)$$

Модель (E9) представляет собой сокращенную модель (E8) — в ней используется только прошлое логарифмированное значение Google Trends и его логарифмическое изменение:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \varepsilon_{t+1}. \quad (E9)$$

Уравнение (E10) расширяет модели с использованием Google Trends, вводя лог-приросты реализованных ковариаций наряду с динамикой поисковых запросов. Логика здесь такова, что в то время, как внимание инвесторов сигнализирует о предстоящей волатильности, резкие изменения в совместном движении активов (ковариации) обеспечивают дополнительную прогнозную силу:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \sum_{i=1}^3 \gamma_i \ln\left(\frac{C_{i, t}}{C_{i, t-1}}\right) + \varepsilon_{t+1}. \quad (E10)$$

Следующий набор спецификаций (E11)–(E15) представляет собой HAR-RV-модели, оснащенные экзогенными переменными, сгенерированными на основе информации из базы данных GDELT. Так, спецификация (E11) включает логарифмированное

количество статей, упоминающих криптовалюту. Таким образом, эта спецификация учитывает влияние интенсивности освещения актива в СМИ на его ценовые колебания:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \delta_1 \ln(A_t) + \varepsilon_{t+1}. \quad (E11)$$

Аналогично предыдущим спецификациям в модель (E12) добавлен лог-прирост количества статей, подразумевая, что резкие изменения в тональности освещения в СМИ актива могут представлять дополнительную ценность при прогнозе в периоды высокой волатильности:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \varepsilon_{t+1}. \quad (E12)$$

В уравнении (E13) в качестве экзогенной переменной вводится средний сентимент (тональность) статей, вышедших в день t по тематике анализируемой криптовалюты. Поскольку торговля, основанная на настроениях, может усиливать волатильность, эта спецификация обеспечивает учет поведенческих аспектов инвесторов в прогнозной модели:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \tau_1 S_t + \varepsilon_{t+1}. \quad (E13)$$

С помощью спецификации (E14) в HAR-модель в качестве экзогенной переменной вводится изменение сентимента (с учетом знака). Предполагается, что резкое изменение абсолютного значения и тем более смена знака средней тональности сигнализируют об изменении настроений инвесторов, приводя к резким скачкам цен на рынке криптовалют. Таким образом, учет этого фактора может быть полезен при прогнозе реализованной волатильности:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E14)$$

Модель (E15) является естественным расширением моделей (E11)–(E15) и использует все переменные, сгенерированные с помощью данных, извлеченных из базы данных GDELT:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \delta_1 \ln(A_t) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E15)$$

Модели (E16)–(E31) представляют собой различные комбинации ранее введенных переменных, интегрирующих в базовую HAR-RV-модель экзогенные предикторы, такие как данные индекса Google Trends соответствующей криптовалюте $GT_{i, t}$, сентимент S_t и ΔS_t , количество статей A_t , реализованные попарные ковариации $C_{i, t}$, а также алгебраические преобразования этих переменных. Эти модели основываются на комбинациях переменных в уравнениях (E2)–(E15), которые введены для полноты проводимого анализа. Две последние модели (E30)–(E31) являются самыми длинными спецификациями, включающими практически все ранее введенные переменные. Отдельно нужно отметить, что последняя модель (E31) отличается от (E30) тем, что в ней используются данные GDELТ по общей тематике криптовалют (то есть эти переменные для различных криптовалют одинаковые). Логика этой спецификации заключается в том, что на динамику волатильности определенной криптовалюты могут влиять новости не только по соответствующей криптовалюте, но и по другим криптовалютам, а также по общей тематике крипторынка.

Ниже приведен список указанных моделей:

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_1 \ln(A_t) + \varepsilon_{t+1}. \quad (E16)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \varepsilon_{t+1} + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right). \quad (E17)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \tau_1 S_t + \varepsilon_{t+1}. \quad (E18)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E19)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_1 \ln(A_t) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E20)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_1 \ln(A_t) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E21)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \gamma_i \ln\left(\frac{C_{i, t}}{C_{i, t-1}}\right) + \delta_1 \ln(A_t) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E22)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \delta_3 \ln(A_t) + \tau_1 S_t + \varepsilon_{t+1}. \quad (E23)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \varepsilon_{t+1}. \quad (E24)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho_d \ln(GT_{d, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \delta_3 \ln(A_t) + \tau_1 S_t + \varepsilon_{t+1}. \quad (E25)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \rho_d \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \delta_3 \ln(A_t) + \tau_1 S_t + \varepsilon_{t+1}. \quad (E26)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \delta_3 \ln(A_t) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E27)$$

$$y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \delta_3 \ln(A_t) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E28)$$

$$\begin{aligned}
y_{t+1} = & \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \\
& + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \\
& + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \delta_3 \ln(A_t) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E29)
\end{aligned}$$

$$\begin{aligned}
y_{t+1} = & \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \gamma_i \ln\left(\frac{C_{i, t}}{C_{i, t-1}}\right) + \\
& + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_2 \ln\left(\frac{A_t}{A_{t-1}}\right) + \\
& + \delta_3 \ln(A_t) + \tau_1 S_t + \tau_2 \Delta S_t + \varepsilon_{t+1}. \quad (E30)
\end{aligned}$$

$$\begin{aligned}
& \left| y_{t+1} = \beta_0 + \sum_{i \in \{d, w, m\}} \beta_i y_{i, t} + \sum_{i=1}^3 \gamma_i \ln\left(\frac{C_{i, t}}{C_{i, t-1}}\right) + \right. \\
& + \sum_{i \in \{d, w, m\}} \rho_i \ln(GT_{i, t}) + \rho \ln\left(\frac{GT_{d, t}}{GT_{d, t-1}}\right) + \delta_2 \ln\left(\frac{A_t^{crypto}}{A_{t-1}^{crypto}}\right) + \\
& \left. + \delta_3 \ln(A_t^{crypto}) + \tau_1 S_t^{crypto} + \tau_3 \Delta S_t^{crypto} + \varepsilon_{t+1}. \quad (E31) \right.
\end{aligned}$$

4. Результаты

С помощью моделей (E1)–(E31) по описанному в начале предыдущего раздела алгоритма для каждой криптовалюты сделано около 2000 вневыборочных прогнозов (общее число построенных конкурирующих прогнозов для каждой криптовалюты по всем моделям — около 65 000). На основе прогнозов рассчитаны функции потерь (ошибок), результаты для каждой криптовалюты представлены в табл. 5. Для выявления статистически значимо лучших моделей для каждой криптовалюты используется процедура MCS (Model Confidence Set), предложенная в [Hansen et al., 2011]. Она заключается в итеративном исключении худших (по функции ошибок) моделей из исходного набора моделей, до тех пор пока не останется один (лучший) набор моделей для заданного

Метрики качества (функции потерь) оцененных моделей (E1)–(E31) для BTC, ETH, LTC и XRP

Т а б л и ц а 5

Т а б л и ц а 5

Performance Metrics (Loss Functions) of Estimated Models (E1)–(E31) for BTC, ETH, LTC, and XRP

Модели	BTC			ETH			LTC			XRP		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
E1	0.01310	0.00766	24.66	0.01569	0.00903	21.91	0.01893	0.01091	21.68	0.02236	0.01223	24.46
E2	0.01314	0.00774	24.98	0.01574	0.00911	22.20	0.01900	0.01098	21.87	0.02247	0.01235	24.76
E3	0.01314	0.00772	24.89	0.01577	0.00909	22.06	0.01897	0.01094	21.71	0.02262	0.01230	24.58
E4	0.01183	0.00719	23.42	0.01449	0.00864	21.19	0.01806	0.01065	21.26	0.02177	0.01198	23.90
E5	0.01186	0.00722	23.66	0.01431*	0.00859	21.14	0.01781	0.01059	21.19	0.02214*	0.01199	23.89
E6	0.01310	0.00773	24.84	0.01574	0.00908	22.13	0.01899	0.01090	21.60	0.02247	0.01212	24.07
E7	0.01315	0.00771	24.81	0.01562	0.00899	21.98	0.01904	0.01082	21.49	0.02252	0.01201	24.00
E8	0.01200	0.00732	23.80	0.01501	0.00874	21.39	0.01823	0.01062	21.31	0.02280	0.01199	23.76
E9	0.01212	0.00731	23.78	0.01494	0.00872	21.40	0.01815	0.01059	21.22	0.02208	0.01184	23.64
E10	0.01156	0.00711	23.31	0.01432	0.00849	20.94	0.01800	0.01050	21.05	0.021349*	0.01168	23.34
E11	0.01204	0.00707	22.46	0.01546	0.00885	21.10	0.01881	0.01087	21.52	0.02218	0.01216	24.09
E12	0.01287	0.00747	23.99	0.01561	0.00895	21.60	0.01893	0.01089	21.62	0.02212	0.01218	24.30
E13	0.01310	0.00766	24.68	0.01571	0.00904	21.94	0.01890	0.01093	21.79	0.02245	0.01225	24.43
E14	0.01312	0.00767	24.68	0.01570	0.00904	21.94	0.01894	0.01092	21.73	0.02239	0.01225	24.53
E15	0.01206	0.00707	22.17	0.01550	0.00888	21.16	0.01878	0.01089	21.59	0.02210	0.01219	24.12
E16	0.01128	0.00674	21.55	0.01474	0.00858	20.77	0.01803	0.01054	21.07	0.02181*	0.01179	23.35
E17	0.01204	0.00728	23.65	0.01488	0.00868	21.23	0.01817	0.01057	21.14	0.02178*	0.01178	23.47
E18	0.01211	0.00731	23.84	0.01495	0.00874	21.44	0.01812	0.01061	21.31	0.02214	0.01186	23.64
E19	0.01213	0.00732	23.83	0.01494	0.00874	21.44	0.01816	0.01058	21.23	0.02211	0.01187	23.71
E20	0.01110	0.00671	21.25	0.01477	0.00864	20.89	0.01802	0.01054	21.06	0.02173*	0.01181	23.40
E21	0.01108	0.00675	21.45	0.01484	0.00866	20.88	0.01807	0.01057	21.16	0.02197	0.01192	23.49
E22	0.01102	0.00673	21.40	0.01437	0.00850	20.59	0.01787	0.01061	21.12	0.02163*	0.01196	23.71
E23	0.01178	0.00686	21.35	0.01545	0.00886	21.03	0.01877	0.01088	21.54	0.02221	0.01213	23.93
E24	0.01107	0.00666	20.98	0.01464	0.00858	20.53	0.01791	0.01052	21.05	0.02237	0.01179	23.34
E25	0.01103	0.00663	20.88*	0.01471	0.00861	20.74	0.01796	0.01052	21.03	0.02193	0.01180	23.31
E26	0.01105	0.00666	20.89	0.01480	0.00863	20.72	0.01801	0.01056	21.14	0.02219	0.01192	23.42
E27	0.01183	0.00689	21.32	0.01547	0.00888	21.07	0.01876	0.01088	21.57	0.02221	0.01216	24.03
E28	0.01111	0.00667	20.89*	0.01467	0.00860	20.59	0.01790	0.01053	21.07	0.02237	0.01182	23.43
E29	0.01106	0.00666	20.81*	0.01482	0.00866	20.77	0.01798	0.01056	21.16	0.02223	0.01195	23.51
E30	0.01054*	0.00653*	20.67*	0.01413*	0.00841	20.43	0.01750*	0.01037	20.81	0.02173*	0.01171	23.15
E31	0.01120	0.00675	21.23	0.01401*	0.00817*	19.46*	0.01744*	0.01016*	20.03*	0.021353*	0.01147*	22.34*

Примечания: 1. Курсивом выделена лучшая модель по каждой функции потерь для каждого актива. 2. Знак «*» указывает, что модель вошла в набор лучших моделей согласно процедуре MCS с уровнем доверия 0,9. 3. Период вневыборочных прогнозов: 28.10.2018–22.06.2024. 4. Функция потерь MAPE указана в процентах, RMSE и MAE — в исходных значениях реализованных волатильностей.

Источник: составлено автором.

уровня значимости. Эта процедура применена для набора моделей (E1)–(E31) для каждой криптовалюты и для каждой функции ошибок: квадратичной ошибки (их усреднение дает RMSE), абсолютной ошибки (их усреднение дает MAE) и процентной ошибки (их усреднение дает MAPE).

Анализ точности прогнозов моделей для BTC, ETH, LTC и XRP демонстрирует значительное улучшение качества прогнозов по сравнению с базовой моделью HAR-RV (E1).

Для BTC наилучшие результаты демонстрирует модель (E30), достигая минимальных значений RMSE, MAE, MAPE в 0,01054, 0,00653, 20,67% соответственно. Существенное снижение ошибок прогноза по сравнению с базовой спецификацией свидетельствует о том, что интеграция данных Google Trends, реализованных ковариаций, новостей и их тональностей (GDELT) существенно повышает точность модели. Нужно отметить, что эта модель является одной из самых длинных среди всех. Более того, модель (E31), где вместо новостей по BTC, как в (E30), используются все новости по криптовалютной тематике, проигрывает по качеству даже более коротким моделям по всем функциям ошибок. Модели, которые исключают наборы этих предикторов из (E30), такие как (E24)–(E27), не входят в множество лучших моделей согласно процедуре MCS для квадратичных (RMSE) и абсолютных ошибок (MAE).

Интересно, что для процентной ошибки (MAPE) в набор лучших моделей также входят (E25), (E28), (E29). Модель (E25) не использует, в отличие от (E30), HAR-структуру для переменных Google Trends, попарные реализованные ковариации с доходностями других криптовалют, а также изменения сентимента. В (E28) используется лог-прирост индекса Google Trends и отсутствуют реализованные ковариации, в (E29) также отсутствуют реализованные ковариации. Предполагается, что эти переменные действительно могут помогать в прогнозировании резких пиков волатильности, но их вклад в улучшение робастной к высоким ошибкам MAPE не столь велик, чтобы только (E30) осталась в наборе лучших моделей на 10-процентном уровне значимости.

Для ETH лучшей моделью является модель (E31), которая дает следующий набор ошибок: RMSE — 0,01401, MAE — 0,00817 и MAPE — 19,46%. Интересно, что для этой криптовалюты (E31) является единственной моделью, входящей в набор лучших моделей для абсолютной (MAE) и процентной (MAPE) ошибок. Включение практически всех сгенерированных переменных позволяет модели превосходить альтернативные спецификации, в том числе модели (E28) и (E29), которые не учитывают часть экзогенных переменных и, соответственно, демонстрируют менее точные

прогнозы. Для MAE и MAPE в набор лучших моделей не входит (E30), где из базы данных GDELТ используются только новости с упоминанием ЕТН, что говорит о том, что учет общей тематики криптовалют, в том числе новостные публикации и их тональность по другим криптовалютам, может влиять на качество прогнозов. Это явление не обнаружено в моделях для ВТС, поскольку (E31) не входит в набор лучших моделей (MCS) для этого актива по всем функциям ошибок. Это может говорить о том, что ВТС является ведущей и самой торгуемой криптовалютой, а изменения в новостных заголовках по тематике ВТС могут влиять также и на другие криптовалюты, и на весь рынок в целом.

Интересные результаты моделей для ЕТН состоят в том, что, в отличие от MAE и MAPE, для RMSE находится несколько моделей, входящих в набор лучших. Одной из таких моделей является компактная по количеству параметров модель (E5), которая не содержит переменных, связанных с интересом инвесторов (данные Google Trends и GDELТ). В качестве экзогенных содержится только информация о попарных ковариациях и их изменениях.

Для LTC модель (E31) снова демонстрирует наилучшие результаты, достигая RMSE 0,01744, MAE 0,01016 и MAPE 20,03%. Высокая точность прогнозов модели для этого актива вновь подчеркивает важность учета как рыночных индикаторов (реализованные ковариации), так и переменных, связанных с динамикой общественного восприятия криптовалюты (Google Trends и GDELТ). Альтернативные спецификации, такие как (E24)–(E29), которые исключают часть данных о настроениях, показывают худшие результаты и не входят в MCS. Для LTC (E31) является единственной спецификацией, которая включена в набор лучших моделей на 10-процентном уровне значимости для абсолютной и процентной ошибок. Для RMSE в MCS также включена и (E30), где в качестве независимых переменных используются данные GDELТ только по тематике LTC.

Интерпретация качества прогнозов по RMSE для XRP существенно отличается от результатов для других криптовалют. Так, лучшей моделью оказывается короткая модель (E10) с RMSE 0,021349, которая превосходит по качеству модель (E31) с большим количеством предикторов. При этом (E31) также входит в набор лучших моделей. Более того, в набор лучших моделей по RMSE входят также модели (E5), (E16), (E20) и (E22). Модель (E5) не использует в качестве предикторов данные GDELТ и Google Trends, квадратичные ошибки прогнозов этой модели MCS. Общей чертой спецификаций (E16), (E20) и (E22) является наличие в качестве экзогенных переменных логарифмированного значения количества публикаций по тематике XRP – $\ln(A_t)$. Лучшая спецификация (E10) в качестве экзогенных переменных

включает только предикторы, построенные на основе данных Google Trends, а также лог-изменение реализованных ковариаций доходностей с другими криптовалютами. Надо заметить, что RMSE всех моделей для этого актива примерно в два раза превосходят более робастные MAE. Более того, для MAE и MAPE в лидерах, как и для других криптовалют, остается только модель (E31).

Таким образом, анализ точности прогнозов по криптовалютам показывает, что модели, использующие внешние источники данных, такие как Google Trends и GDELT, всегда превосходят стандартную HAR-RV-спецификацию (поскольку ни для одной криптовалюты по всем функциям ошибок модель (E1) не вошла в MCS). В лидерах практически всегда присутствует модель с применением всех использованных в настоящем исследовании источников данных, что указывает на важность учета новостного фона и интереса инвесторов при прогнозировании волатильности. Результаты исследования подтверждают, что интеграция альтернативных нефинансовых данных в базовую HAR-RV-модель приводит к более точным предсказаниям и может быть полезна для построения торговых стратегий и управления рисками на криптовалютном рынке.

Заключение

Рынок криптовалют выделяется среди рынков традиционных финансовых активов повышенной неопределенностью, следствием которой является высокая волатильность цен активов. Использование источников нефинансовых данных, в том числе информации о совместных ценовых движениях, поисковых запросах, освещении цифровых активов в СМИ и их эмоциональной окраске, может быть полезным при построении прогнозных моделей волатильности и оптимизации на их основе рисковой и регуляторной стратегии игроков рынка.

Настоящее исследование использует HAR-RV-модели для прогнозирования реализованной волатильности четырех крупнейших криптовалют — BTC, ETH, LTC, XRP. Стандартная модель оснащается дополнительными экзогенными переменными, построенными как на основе данных торгов (реализованные ковариации), так и на источниках интернет-данных (Google Trends, GDELT). В дополнение к обычной HAR-RV-модели было построено 30 моделей с использованием комбинаций переменных, сгенерированных на основе указанных источников данных. С применением подхода прогнозирования в скользящих окнах для каждой модели и каждой криптовалюты было сделано более 2000 вневыборочных прогнозов, для оценки качества модели использованы

три стандартные функции ошибок. Для нахождения множества статистически лучших прогнозов применялся подход MCS.

Результаты работы моделей показали, что включение нефинансовых данных наряду с информацией о совместной динамике активов позволяет добиться значительного улучшения качества прогноза по сравнению с базовой HAR-RV-моделью, которая использует только данные о прошлой динамике актива. Показано, что в прогнозах почти для всех криптовалют лидируют модели, использующие большую часть переменных, описанных в настоящем исследовании. При этом для части криптовалют можно добиться сопоставимого с длинными моделями качества с помощью достаточно компактной по параметрам спецификации.

Литература

1. Аганин А. Д., Маневич В. А., Пересецкий А. А., Погорелова П. В. Сравнение моделей прогноза волатильности криптовалют и фондового рынка // Экономический журнал ВШЭ. 2023. № 27(1). С. 49–77. DOI: 10.17323/1813-8691-2023-27-1-49-77.
2. Маневич В. А., Пересецкий А. А., Погорелова П. В. Волатильность фондового рынка и волатильность криптовалют // Прикладная эконометрика. 2022. № 65. С. 65–76. DOI: 10.22394/1993-7601-2022-65-65-76.
3. Aalborg H. A., Molnár P., de Vries J. E. What Can Explain the Price, Volatility and Trading Volume of Bitcoin? // Finance Research Letters. 2019. No 29. P. 255–265. DOI: 10.1016/j.frl.2018.08.010.
4. Al Guindy M. Cryptocurrency Price Volatility and Investor Attention // International Review of Economics & Finance. 2021. No 76. P. 556–570. DOI: 10.1016/j.iref.2021.06.007.
5. Alipour P., Charandabi S. E. Analyzing the Interaction Between Tweet Sentiments and Price Volatility of Cryptocurrencies // European Journal of Business and Management Research. 2023. No 8(2). P. 211–215. DOI: 10.24018/ejbmr.2023.8.2.1865.
6. Anamika A., Subramaniam S. Do News Headlines Matter in the Cryptocurrency Market? // Applied Economics. 2022. No 54. P. 6322–6338. DOI: 10.1080/00036846.2022.2061904.
7. Andersen T. G., Bollerslev T. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts // International Economic Review. 1998. No 39(4). P. 885–905. DOI: 10.2307/2527343.
8. Andersen T. G., Bollerslev T., Christoffersen P. F., Diebold F. X. Volatility and Correlation Forecasting // Handbook of Economic Forecasting. 2006. No 1. P. 777–878. DOI: 10.1016/S1574-0706(05)01015-3.
9. Arratia A., López-Barrantes A. X. Do Google Trends Forecast Bitcoins? Stylized Facts and Statistical Evidence // Journal of Banking and Financial Technology. 2021. No 5(1). P. 45–57. DOI: 10.1007/s42786-021-00027-4.
10. Aslanidis N., Bariviera A. F., López Ó. G. The Link Between Cryptocurrencies and Google Trends Attention // Finance Research Letters. 2022. No 47. Article 102654. DOI: 10.1016/j.frl.2021.102654.
11. Barndorff-Nielsen O. E., Shephard N. Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics // Econometrica. 2004. No 72(3). P. 885–925. DOI: 10.1111/j.1468-0262.2004.00515.x.
12. Bergsli L. Ø., Lind A. F., Molnár P., Polasik M. Forecasting Volatility of Bitcoin // Research in International Business and Finance. 2022. No 59. Article 101540. DOI: 10.1016/j.ribaf.2021.101540.
13. Bouri E., Gkillas K., Gupta R., Pierdzioch C. Forecasting Realized Volatility of Bitcoin: The Role of the Trade War // Computational Economics. 2021. No 57. P. 29–53. DOI: 10.1007/s10614-020-10022-4.
14. Brauneis A., Sahiner M. Crypto Volatility Forecasting: Mounting a HAR, Sentiment, and Machine Learning Horse race // Asia-Pacific Financial Markets. 2024. 13 December. P. 1–33. DOI: 10.1017/s10690-024-09510-6.

15. *Chen R.* Forecasting Ethereum's Volatility: An Expansive Approach Using HAR Models and Structural Breaks // *Cogent Economics & Finance*. 2024. No 12(1). Article 2300925. DOI: 10.1080/23322039.2023.2300925.
16. *Chiriac R., Voev V.* Modelling and Forecasting Multivariate Realized Volatility // *Journal of Applied Econometrics*. 2011. No 26(6). P. 922–947. DOI: 10.1002/jae.1152.
17. *Clements A., Preve D. P.* A Practical Guide to Harnessing the HAR Volatility Model // *Journal of Banking & Finance*. 2021. No 133. Article 106285. DOI: 10.1016/j.jbankfin.2021.106285.
18. *Corsi F.* A Simple Approximate Long-Memory Model of Realized Volatility // *Journal of Financial Econometrics*. 2009. No 7(2). P. 174–196. DOI: 10.1093/jjfinc/nbp001.
19. *Dudek G., Fiszeder P., Kubus P., Orzeszko W.* Forecasting Cryptocurrencies Volatility Using Statistical and Machine Learning Methods: A Comparative Study // *Applied Soft Computing*. 2024. No 151(2). Article 111132. DOI: 10.1016/j.asoc.2023.111132.
20. *Dyhrberg A. H.* Bitcoin, Gold and the Dollar — A GARCH Volatility Analysis // *Finance Research Letters*. 2016. No 16. P. 85–92. DOI: 10.1016/j.frl.2015.10.008.
21. *Engle R. F.* Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation // *Econometrica*. 1982. № 54(4). P. 987–1007. DOI: 10.2307/1912773.
22. *Engle R. F., Bollerslev T.* Modelling the Persistence of Conditional Variances // *Econometric Reviews*. 1986. No 5(1). P. 1–50. DOI: 10.1080/07474938608800095.
23. *Gyamerah S. A.* Modelling the Volatility of Bitcoin Returns Using GARCH Models // *Quantitative Finance and Economics*. 2019. No 3(4). P. 739–753. DOI: 10.3934/QFE.2019.4.739.
24. *Hansen P. R., Lunde A., Nason J. M.* The Model Confidence Set // *Econometrica*. 2011. No 79(2). P. 453–497. DOI: 10.3982/ECTA5771.
25. *Harb E., Bassil C., Kassamany T., Baz R.* Volatility Interdependence Between Cryptocurrencies, Equity, and Bond Markets // *Computational Economics*. 2024. No 63(3). P. 951–981. DOI: 10.1007/s10614-022-10318-7.
26. *Kyriazis N., Papadamou S., Tzeremes P., Corbet S.* The Differential Influence of Social Media Sentiment on Cryptocurrency Returns and Volatility During COVID-19 // *The Quarterly Review of Economics and Finance*. 2023. No 89. P. 307–317. DOI: 10.1016/j.qref.2022.09.004.
27. *Liang C., Zhang Y., Li X., Ma F.* Which Predictor Is More Predictive for Bitcoin Volatility? And Why? // *International Journal of Finance & Economics*. 2022. No 27(2). P. 1947–1961. DOI: 10.1002/ijfe.2252.
28. *Liu L. Y., Patton A. J., Sheppard K.* Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes // *Journal of Econometrics*. 2015. No 187(1). P. 293–311. DOI: 10.1016/j.jeconom.2015.02.008.
29. *Ma F., Wei Y., Huang D., Chen Y.* Which Is the Better Forecasting Model? A Comparison Between HAR-RV and Multifractality Volatility // *Physica A: Statistical Mechanics and Its Applications*. 2014. No 405. P. 171–180. DOI: 10.1016/j.physa.2014.03.007.
30. *Manevich V., Ignatov D.* Machine Learning, Neural Networks and Econometric Models for Prediction the Realized Volatility of Bitcoin and E-Mini S&P500. SSRN. Working Paper 4334006. 2023. DOI: 10.2139/ssrn.4334006.
31. *Mastro D.* Forecasting Realized Volatility: ARCH-Type Models vs. the HAR-RV Model. SSRN Working Paper 2519107. 2014. DOI: 10.2139/ssrn.2519107.
32. *Nelson D. B.* Asymptotic Filtering Theory for Multivariate ARCH Models // *Journal of Econometrics*. 1996. No 71(1–2). P. 1–47. DOI: 10.1016/0304-4076(94)01679-8.
33. *Rabemananjara R., Zakoian J. M.* Threshold ARCH Models and Asymmetries in Volatility // *Journal of Applied Econometrics*. 1993. No 8(1). P. 31–49. DOI: 10.1002/jae.3950080104.
34. *Sapkota N.* News-Based Sentiment and Bitcoin Volatility // *International Review of Financial Analysis*. 2022. No 82. Article 102183. DOI: 10.1016/j.irfa.2022.102183.
35. *Teterin M., Peresetsky A.* Google Trends and Bitcoin Volatility Forecast // *Journal of the New Economic Association*. 2024. No 4(65). P. 118–135. DOI: 10.31737/22212264_2024_4_118-135.
36. *Teterin M., Peresetsky A.* Can Ethereum Predict Bitcoin's Volatility? // *Applied Econometrics*. 2025. No 77. P. 74–90. DOI: 10.22394/1993-7601-2025-77-74-90.

37. Ullah S., Attah-Boakye R., Adams K., Zaeferian G. Assessing the Influence of Celebrity and Government Endorsements on Bitcoin's Price Volatility // *Journal of Business Research*. 2022. No 145. P. 228–239. DOI: 10.1016/j.jbusres.2022.01.055.
38. Wang Y., Andreeva G., Martin-Barragan B. Machine Learning Approaches to Forecasting Cryptocurrency Volatility: Considering Internal and External Determinants // *International Review of Financial Analysis*. 2023. No 90. Article 102914. DOI: 10.1016/j.irfa.2023.102914.

References

1. Aganin A.D., Manevich V.A., Peresetsky A.A., Pogorelova P.V. Sravnenie modeley prognoza volatil'nosti kriptovalyut i fondovogo rynka [Comparison of Cryptocurrency and Stock Market Volatility Forecast Models]. *Ekonomicheskii zhurnal VShE [HSE Economic Journal]*, 2023, no. 27(1), pp. 49–77. DOI: 10.17323/1813-8691-2023-27-1-49-77. (In Russ.)
2. Manevich V.A., Peresetsky A.A., Pogorelova P.V. Volatil'nost' fondovogo rynka i volatil'nost' kriptovalyut [Stock Market and Cryptocurrency Market Volatility]. *Prikladnaya ekonometrika [Applied Econometrics]*, 2022, no. 65, pp. 65–76. DOI: 10.22394/1993-7601-2022-65-65-76. (In Russ.)
3. Aalborg H.A., Molnár P., de Vries J.E. What Can Explain the Price, Volatility and Trading Volume of Bitcoin? *Finance Research Letters*, 2019, no. 29, pp. 255–265. DOI: 10.1016/j.frl.2018.08.010.
4. Al Guindy M. Cryptocurrency Price Volatility and Investor Attention. *International Review of Economics & Finance*, 2021 no. 76, pp. 556–570. DOI: 10.1016/j.iref.2021.06.007.
5. Alipour P., Charandabi S.E. Analyzing the Interaction Between Tweet Sentiments and Price Volatility of Cryptocurrencies. *European Journal of Business and Management Research*, 2023, no. 8(2), pp. 211–215. DOI: 10.24018/ejbmr.2023.8.2.1865.
6. Anamika A., Subramaniam S. Do News Headlines Matter in the Cryptocurrency Market? *Applied Economics*, 2022, no. 54, pp. 6322–6338. DOI: 10.1080/00036846.2022.2061904.
7. Andersen T.G., Bollerslev T. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 1998, no. 39(4), pp. 885–905. DOI: 10.2307/2527343.
8. Andersen T.G., Bollerslev T., Christoffersen P.F., Diebold F.X. Volatility and Correlation Forecasting. *Handbook of Economic Forecasting*, 2006, no. 1, pp. 777–878. DOI: 10.1016/S1574-0706(05)01015-3.
9. Arratia A., López-Barrantes A.X. Do Google Trends Forecast Bitcoins? Stylized Facts and Statistical Evidence. *Journal of Banking and Financial Technology*, 2021, no. 5(1), pp. 45–57. DOI: 10.1007/s42786-021-00027-4.
10. Aslanidis N., Bariviera A.F., López Ó.G. The Link Between Cryptocurrencies and Google Trends Attention. *Finance Research Letters*, 2022, no. 47, article 102654. DOI: 10.1016/j.frl.2021.102654.
11. Barndorff-Nielsen O.E., Shephard N. Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics. *Econometrica*, 2004, no. 72(3), pp. 885–925. DOI: 10.1111/j.1468-0262.2004.00515.x.
12. Bergsli L.Ø., Lind A.F., Molnár P., Polasik M. Forecasting Volatility of Bitcoin. *Research in International Business and Finance*, 2022, no. 59, article 101540. DOI: 10.1016/j.ribaf.2021.101540.
13. Bouri E., Gkillas K., Gupta R., Pierdzioch C. Forecasting Realized Volatility of Bitcoin: The Role of the Trade War. *Computational Economics*, 2021, no. 57, pp. 29–53. DOI: 10.1007/s10614-020-10022-4.
14. Brauneis A., Sahiner M. Crypto Volatility Forecasting: Mounting a HAR, Sentiment, and Machine Learning Horserace. *Asia-Pacific Financial Markets*, 13 December 2024, pp. 1–33. DOI: 10.1016/j.econlet.2018.02.001.
15. Chen R. Forecasting Ethereum's Volatility: An Expansive Approach Using HAR Models and Structural Breaks. *Cogent Economics & Finance*, 2024, no. 12(1), article 2300925. DOI: 10.1080/23322039.2023.2300925.

16. Chiriac R., Voev V. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics*, 2011 no. 26(6), pp. 922-947. DOI: 10.1002/jae.1152.
17. Clements A., Preve D. P. A Practical Guide to Harnessing the HAR Volatility Model. *Journal of Banking & Finance*, 2021, no. 133, article 106285. DOI: 10.1016/j.jbankfin.2021.106285.
18. Corsi F. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 2009, no. 7(2), pp. 174-196. DOI: 10.1093/jffinec/nbp001.
19. Dudek G., Fiszeder P., Kubus P., Orzeszko W. Forecasting Cryptocurrencies Volatility Using Statistical and Machine Learning Methods: A Comparative Study. *Applied Soft Computing*, 2024, no. 151(2), article 111132. DOI: 10.1016/j.asoc.2023.111132.
20. Dyhrberg A. H. Bitcoin, Gold and the Dollar - A GARCH Volatility Analysis. *Finance Research Letters*, 2016, no. 16, pp. 85-92. DOI: 10.1016/j.frl.2015.10.008.
21. Engle R. F. Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 1982, no. 54(4), pp. 987-1007. DOI: 10.2307/1912773.
22. Engle R. F., Bollerslev T. Modelling the Persistence of Conditional Variances. *Econometric Reviews*, 1986, no. 5(1), pp. 1-50. DOI: 10.1080/07474938608800095.
23. Gyamerah S. A. Modelling the Volatility of Bitcoin Returns Using GARCH Models. *Quantitative Finance and Economics*, 2019, no. 3(4), pp. 739-753. DOI: 10.3934/QFE.2019.4.739.
24. Hansen P. R., Lunde A., Nason J. M. The Model Confidence Set. *Econometrica*, 2011, no. 79(2), article 453497. DOI: 10.3982/ECTA5771.
25. Harb E., Bassil C., Kassamany T., Baz R. Volatility Interdependence Between Cryptocurrencies, Equity, and Bond Markets. *Computational Economics*, 2024, no. 63(3), pp. 951-981. DOI: 10.1007/s10614-022-10318-7.
26. Kyriazis N., Papadamou S., Tzeremes P., Corbet S. The Differential Influence of Social Media Sentiment on Cryptocurrency Returns and Volatility During COVID-19. *The Quarterly Review of Economics and Finance*, 2023, no. 89, pp. 307-317. DOI: 10.1016/j.qref.2022.09.004.
27. Liang C., Zhang Y., Li X., Ma F. Which Predictor Is More Predictive for Bitcoin Volatility? And Why? *International Journal of Finance & Economics*, 2022, no. 27(2), pp. 1947-1961. DOI: 10.1002/ijfe.2252.
28. Liu L. Y., Patton A. J., Sheppard K. Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes. *Journal of Econometrics*, 2015, no. 187(1), pp. 293-311. DOI: 10.1016/j.jeconom.2015.02.008.
29. Ma F., Wei Y., Huang D., Chen Y. Which Is the Better Forecasting Model? A Comparison Between HAR-RV and Multifractality Volatility. *Physica A: Statistical Mechanics and Its Applications*, 2014, no. 405, pp. 171-180. DOI: 10.1016/j.physa.2014.03.007.
30. Manevich V., Ignatov D. Machine Learning, Neural Networks And Econometric Models for Prediction the Realized Volatility of Bitcoin and E-Mini S&P500. SSRN, Working Paper 4334006, 2023. DOI: 10.2139/ssrn.4334006.
31. Mastro D. Forecasting Realized Volatility: ARCH-Type Models vs. the HAR-RV Model. SSRN, Working Paper 2519107, 2014. DOI: 10.2139/ssrn.2519107.
32. Nelson D. B. Asymptotic Filtering Theory for Multivariate ARCH Models. *Journal of Econometrics*, 1996, no. 71(1-2), pp. 1-47. DOI: 10.1016/0304-4076(94)01679-8.
33. Rabemananjara R., Zakoian J. M. Threshold ARCH Models and Asymmetries in Volatility. *Journal of Applied Econometrics*, 1993, no. 8(1), pp. 31-49. DOI: 10.1002/jae.3950080104.
34. Sapkota N. News-Based Sentiment and Bitcoin Volatility. *International Review of Financial Analysis*, 2022, no. 82, article 102183. DOI: 10.1016/j.irfa.2022.102183.
35. Teterin M., Peresetsky A. Google Trends and Bitcoin Volatility Forecast. *Journal of the New Economic Association*, 2024, no. 4(65), pp. 118-135. DOI: 10.31737/22212264_2024_4_118-135.
36. Teterin M., Peresetsky A. Can Ethereum Predict Bitcoin's Volatility? *Applied Econometrics*, 2025, no. 77, pp. 74-90. DOI: 10.22394/1993-7601-2025-77-74-90.
37. Ullah S., Attah-Boakyie R., Adams K., Zaefarian G. Assessing the Influence of Celebrity and Government Endorsements on Bitcoin's Price Volatility. *Journal of Business Research*, 2022, no. 145, pp. 228-239. DOI: 10.1016/j.jbusres.2022.01.055.
38. Wang Y., Andreeva G., Martin-Barragan B. Machine Learning Approaches to Forecasting Cryptocurrency Volatility: Considering Internal and External Determinants. *International Review of Financial Analysis*, 2023, no. 90, article 102914. DOI: 10.1016/j.irfa.2023.102914.