

МИНОБРНАУКИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической экономики
наименование кафедры

**Прогноз реализованной волатильности на российском рынке с
использованием HAR-RV моделей**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента _____ 4 _____ курса _____ 451 _____ группы

направления _____ 38.03.05 - Бизнес - информатика _____
код и наименование направления

_____ механико-математического факультета
наименование факультета, института, колледжа

_____ Цуканова Валентина Юрьевича _____
фамилия, имя, отчество

Научный руководитель

_____ доцент, к.ф-м.н., доцент _____
должность, уч. степень, уч. звание

_____ дата, подпись

_____ Новиков В. В. _____
инициалы, фамилия

Заведующий кафедрой

_____ д.ф-м.н., профессор _____
должность, уч. степень, уч. звание

_____ дата, подпись

_____ Дудов С. И. _____
инициалы, фамилия

Саратов 2020

Введение. Целью настоящей работы является прогнозирование реализованной волатильности на российском рынке с использованием моделей HAR-RV.

Для достижения цели работы были поставлены следующие задачи:

- изучение теоретических основ концепции реализованной волатильности в ее историческом развитии;
- исследование специфики класса HAR-RV моделей;
- сравнение существующих моделей прогнозирования реализованной волатильности;
- углубленное изучение объектно-ориентированного языка программирования Python в качестве инструмента практической реализации исследуемых моделей;
- построение и оценка моделей прогнозирования реализованной волатильности на российском рынке с использованием моделей класса HAR-RV на языке программирования Python;
- анализ перспектив применения семейства HAR-RV моделей для прогнозирования реализованной волатильности на российском рынке.

Актуальность работы обусловлена постоянной необходимостью изучения и совершенствования методов и технологий управления финансовыми рисками в условиях нестабильной финансово-экономической обстановки. Прогнозирование волатильности имеет ключевое значение при решении разнообразных задач управления рыночными и кредитными рисками, включая управление портфелями финансовых активов, хеджирование риска, оценку банками размера резервного капитала для покрытия риска активных операций. Если можно сделать относительно точный прогноз, то минимизируется риск принятых управленческих решений.

Модели прогнозирования реализованной волатильности являются одним из наиболее оптимальных инструментов в сфере управления рисками благодаря схожести доходностей, стандартизированных по реализованной волатильности, с нормальным распределением и простоте расчета. Модели семейства HAR-RV уже более пятнадцати лет успешно применяются для прогнозирования реализованной волатильности в мировой практике, однако практически отсутствуют исследования эффективности моделей в условиях рос-

сийского рынка финансовых активов.

Источниками статистических данных для проведения исследования являются котировки российских активов, которые обращаются на московской межбанковской валютной бирже (далее ММВБ). Вследствие необходимости наличия котировок с высокой частотой для корректной оценки моделей реализованной волатильности в анализе используются «голубые фишки» – активы наиболее крупных, ликвидных и надежных компаний со стабильными показателями доходности, торги по которым проводятся каждую секунду.

Работа имеет следующую структуру:

1. первый раздел работы содержит описание теоретических основ концепции реализованной волатильности в ее историческом развитии, модифицированных методов расчета реализованной волатильности;
2. второй раздел содержит исследование специфики семейства HAR-RV моделей, литературный обзор работ, посвященных сравнению моделей прогнозирования реализованной волатильности;
3. в третьем разделе представлены результаты построения и оценки моделей для прогноза реализованной волатильности на российском рынке с использованием моделей HAR-RV на языке программирования Python, а также аналитическое исследование перспектив использования данного семейства моделей для российского рынка финансовых активов.

Основное содержание работы. Волатильность актива – ключевой статистический финансовый показатель, который характеризует амплитуду колебаний доходности актива относительно ожидаемого значения, и тем больше, чем больше неопределенность относительно будущих изменений, а следовательно, и доходности актива. Данный показатель является основным для правильной оценки финансового риска. Выражается волатильность в абсолютном или в относительном от начальной стоимости значении.

Эмпирические исследования свойств финансовых рядов показали, что волатильность обладает такими свойствами, как кластеризация («группирование» периодов с высокой и низкой волатильностью), цикличность, а также так называемый эффект «рычага» – уменьшение волатильности при росте цены, и наоборот.

Одной из ключевых особенностей современных валютного и фондово-

го рынков является существенная неоднородность волатильности: периоды относительно низкой волатильности чередуются с периодами высокой волатильности.

Волатильность является широко используемым показателем, который имеет важное практическое значение. Она применяется для решения таких задач, как:

- определение вероятного запаса движения;
- определение риска на сделку;
- расчет многих индикаторов технического анализа;
- выявление активности на рынке и наличия интереса к определенному торговому инструменту;
- исследование рыночной конъюнктуры, на основании которой зачастую вырабатывается торговая тактика.

Реализованная волатильность. Основная идея концепции реализованной волатильности заключается в том, что зачастую в распоряжении исследователя имеются не только дневные статистические данные, но и внутридневная (иногда вплоть до ежесекундной) информация по многим финансовым инструментам. Использование высокочастотных временных рядов позволяет значительно улучшить качество оценки волатильности как на дневном интервале, так и на более длительных промежутках времени.

Пусть логарифмическая дневная доходность актива задается процессом:

$$r_t = \sigma_t z_t,$$

где $r_t = \ln p_t - \ln p_{t-1}$, p_t — последовательность цен; z_t — последовательность независимых одинаково распределенных гауссовских случайных величин $z_t \sim NID(0, 1)$. В этом случае $r_t \sim N(0, \sigma_t^2)$.

Далее рассмотрим n цен $p_{t,i}$, $i = 1 \dots n$ внутри торгового дня, интервал между наблюдениями равен Δ , где $n = \frac{1}{\Delta}$ — количество доступных внутри дня наблюдений. Дополнительно предполагаем, что $r_{t,i} = \sigma_{t,i} z_{t,i}$, где $r_{t,i} = \ln p_{t,i} - \ln p_{t,i-1}$, $z_{t,i} \sim NID(0, n^{-1})$, тогда для дневных доходностей и дисперсий выполнено:

$$r_t = \sum_{i=1}^n r_{t,i},$$

$$\sigma_t^2 = \frac{1}{n} \sum_{t=1}^n \sigma_{t,i}^2.$$

Формула расчета реализованной волатильности:

$$RV_t = \sum_{n=1}^{N_t} r_{n,t}^2, \quad (1)$$

где RV_t – оценка волатильности по модели реализованной волатильности в день t , $r_{n,t}^2$ – квадрат логарифмической доходности в момент n дня t , N_t – число наблюдений доходности внутри дня t .

Семейство HAR-RV моделей. Первая HAR-RV (The Heterogeneous Autoregressive model of the Realized Volatility) модель была предложена Ф. Корси в 2004 году. Эта модель основана на концепции поведения агентов на финансовых рынках, согласно которой они отличаются своим восприятием волатильности в зависимости от их инвестиционных горизонтов и делятся на краткосрочных, среднесрочных и долгосрочных. Гипотеза о существовании таких гетерогенных структур на финансовых рынках (гипотеза о гетерогенности рынков) была сформулирована в 1993 году.

HAR-RV модель представляет волатильность как сумму дневного, недельного и месячного компонентов. Обычно HAR-RV модель рассчитывается с параметрами (1, 5, 21), что соответствует лагам дневной, недельной и месячной RV.

$$RV_{t+1} = \beta_0 + \beta_d RV_t^d + \beta_w RV_t^w + \beta_m RV_t^m + \varepsilon_{t+1}, \quad (2)$$

где ε_{t+1} – случайный шок, RV_t^d – реализованная волатильность в день t , RV_t^w – лагированная недельная волатильность за предыдущую неделю, определяемая как $RV_t^w = \frac{1}{4} \sum_{i=1}^4 RV_{t-i}$, и RV_t^m – месячная волатильность, определяемая

похожим образом как $RV_t^m = \frac{1}{17} \sum_{i=5}^{21} RV_{t-i}$.

Модели этого семейства способны воспроизводить наблюдаемые свойства волатильности, такие как: «длинная память» и устойчивость волатильности, «каскад волатильности» (иерархичная структура волатильности, при которой долгосрочная волатильность оказывает сильное влияние на кратко-

срочную волатильность, однако краткосрочная волатильность почти не влияет на долгосрочную), «толстые хвосты», а также имеют экономическую интерпретацию. Особенностью HAR-RV модели является рассмотрение волатильности как ненаблюдаемой величины, для которой существует наблюдаемая оценка – реализованная волатильность.

Формальная постановка задачи работы. Пусть имеется стохастический процесс r_t с дискретным временем, $t = 1, \dots, T$, и предысторией $\Omega_T = (r_{T-1}, r_{T-2}, \dots, r_1, \dots)$, где r_t – логарифмическая доходность некоторого заранее определенного финансового инструмента.

Требуется оценить значение RV_t – реализованную волатильность – для периодов времени $t = T + 1, \dots, T + m, m \geq 1$ на основе моделей семейства HAR-RV.

Этапы построения моделей прогнозирования реализованной волатильности на российском рынке. Для решения поставленной задачи были выделены следующие этапы:

1. сбор и предварительная обработка статистических данных о котировках активов;
2. выбор соответствующих моделей и интервалов времени, метрик оценки моделей, алгоритмов построения на языке программирования Python (версия 3.5);
3. построение моделей прогнозирования и оценка их эффективности;
4. сравнение и анализ полученных результатов.

Сбор и анализ данных. Построение прогнозных моделей выполнялось на базе исторических данных котировок десяти наиболее ликвидных российских активов, обращающихся на ММВБ, эмитенты которых являются представителями различных отраслей хозяйствования (в скобках указывается листинг на ММВБ):

1. акции ПАО «НК «Роснефть» (ROSN) – российская нефтегазовая компания, контрольный пакет акций которой принадлежит государственному АО «Роснефтегаз»;
2. акции ПАО «Татнефть» (TATN) – российская нефтяная компания, занимает пятое место в России по добыче нефти;
3. акции ПАО «ГМК Норильский никель» (GNWK) – российская горно-

- металлургическая компания;
4. акции ПАО «Нефтяная компания «Лукойл» (ЛКОН) – российская нефтяная компания;
 5. акции ПАО «Сургутнефтегаз» (SNGS) – одна из крупнейших российских нефтяных и газодобывающих компаний;
 6. акции ПАО «Новатэк» (NVTK) – российская газовая компания;
 7. акции ПАО «Северсталь» (CHMF) – российская вертикально интегрированная сталелитейная и горнодобывающая компания, владеющая Череповецким металлургическим комбинатом, вторым по величине сталелитейным комбинатом России;
 8. акции ООО «Яндекс» (YNDX) – российская транснациональная компания в отрасли информационных технологий, которая владеет одноименной системой поиска в Сети, интернет-порталами и службами в нескольких странах;
 9. акции ПАО «Сбербанк» (SBER) – российский государственный финансовый конгломерат, крупнейший транснациональный и универсальный банк Российской Федерации;
 10. акции группы НЛМК (NLMK) – международная сталелитейная компания с активами в России, США и странах Европы, основным активом Группы является Новолипецкий металлургический комбинат.

В качестве периода был взят временной промежуток со 2 февраля 2017 года по 4 апреля 2019 года, что в целом составляет 548 торговых дней.

Для сравнительного анализа прогнозной силы моделей использовались данные котировок на закрытии торговых сессий (цена CLOSE), а также внутридневные данные – цены закрытия внутридневных пятиминутных интервалов (для отдельных экспериментов использовались также интервалы в пятнадцать или тридцать минут). При этом данные первых 398 дней использовались исключительно для построения и оценки параметров моделей, а «out-of-sample» – тестовым периодом служили следующие 150 дней, на которых происходило сравнение прогнозной силы моделей.

Для анализа ключевых статистических характеристик эмпирических данных в работе были использованы следующие показатели:

— среднее значение

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i;$$

— стандартное отклонение

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2};$$

— коэффициент смещения

$$s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}};$$

— куртозис

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}.$$

Первичный статистический анализ и обработка данных были произведены с помощью высокоуровневого языка программирования общего назначения Python в интегрированной среде разработки Spyder. Выбор данного инструмента обусловлен тем, что Python является одним из лидеров среди свободно распространяемых систем статистического анализа, однако все же уступает языку R, который иногда называют стандартом статистических программ.

Оценки эффективности моделей. Несмотря на то, что реализованная волатильность является всего лишь оценкой волатильности, а следовательно подвержена некоторой ошибке измерения, данная оценка обеспечивает достаточную надежность и ее можно считать фактической волатильностью. Такое допущение позволяет использовать весь инструментарий стандартного анализа временных рядов.

Для оценки качества моделей используется следующий алгоритм:

1. оценка реализованной волатильности по соответствующим внутрисуточным

- ным доходностям;
2. оценка параметров модели HAR-RV по обучающей выборке (398 торговых дней);
 3. построение прогноза реализованной волатильности на каждый день тестовой выборки в соответствии с полученной моделью;
 4. оценка качества модели с помощью R^2 , F -статистики, средней ошибки аппроксимации (доверительная вероятность равна 0.95).

Реализация моделей прогнозирования реализованной волатильности на языке программирования Python. *Первая группа* экспериментов заключается в расчете реализованной волатильности по историческим данным, построении стандартных HAR-RV моделей в соответствии с формулой 2 для всех выбранных ценных бумаг и сравнении полученных результатов. Предполагается использование гауссовского белого шума в качестве ошибки ε_t . Для оценивания параметров модели используется классический метод наименьших квадратов.

По первой группе экспериментов были сделаны следующие выводы: наиболее близкий к историческим данным прогноз реализованной волатильности модель HAR-RV-5MIN обеспечила для акций SBER, LKOH и SNGS; для большинства финансовых инструментов построенные модели обеспечивали низкое качество прогноза реализованной волатильности, следовательно классическая модель HAR-RV-5MIN не может быть названа эффективной для данного набора акций российского рынка.

В качестве основных конкурирующих моделей прогнозирования *для второй группы* экспериментов были выбраны стандартные модели HAR-RV для разных интервалов времени и тех финансовых активов, для которых в предыдущем эксперименте были получены наиболее качественные модели (SBER, LKOH и SNGS), так как часто эффективность модели зависит от длины интервала и на некоторых интервалах возможны сильные искажения:

1. HAR-RV-5MIN;
2. HAR-RV-15MIN;
3. HAR-RV-30MIN.

Эксперименты второй группы продемонстрировали явное преимущество моделей HAR-RV-5MIN для всех выбранных финансовых инструмен-

тов, это позволяет заключить, что снижение частоты котировок существенно уменьшает точность модели.

Далее было принято решение провести сравнение моделей HAR-RV, HAR-RV-SQRT и HAR-RV-LN, которые были построены по формулам 2, 3 и 4 соответственно. Данное решение обусловлено тем, что качество прогноза реализованной волатильности может повышаться при использовании $\sqrt{RV_t^{(n)}}$ и $\ln(RV_t^{(n)})$ в формуле стандартной HAR-RV модели.

$$\sqrt{\sigma_{t+ld}^{(d)}} = c + \beta^{(d)}\sqrt{RV_t^{(d)}} + \beta^{(w)}\sqrt{RV_t^{(w)}} + \beta^{(m)}\sqrt{RV_t^{(m)}} + \varepsilon_t. \quad (3)$$

$$\ln(\sigma_{t+ld}^{(d)}) = c + \beta^{(d)}\ln(RV_t^{(d)}) + \beta^{(w)}\ln(RV_t^{(w)}) + \beta^{(m)}\ln(RV_t^{(m)}) + \varepsilon_t. \quad (4)$$

По результатам проведенных экспериментов было выявлено, что логарифмическая модификация и модификация с использованием квадратного корня стандартной модели HAR-RV позволяют получать более точные оценки реализованной волатильности для выбранных финансовых инструментов; при сравнении отдельно взятых HAR-RV-LN и HAR-RV-SQRT модификаций преимущественное использование зависит от конкретных данных, однако в настоящем исследовании более высокую эффективность продемонстрировала логарифмическая модификация модели.

Основной вывод по оценке эффективности использования семейства HAR-RV моделей для прогнозирования реализованной волатильности на российском рынке по результатам эмпирических исследований: несмотря на незначительное увеличение точности прогнозирования благодаря выбору определенного временного интервала, а также использованию HAR-RV-LN и HAR-RV-SQRT модификаций модели, HAR-RV модель не может обеспечить достаточную эффективность при определении реализованной волатильности на российском рынке, предсказательная сила построенных моделей все еще остается на низком уровне.

Заключение. В работе были построен прогноз реализованной волатильности на российском фондовом рынке с использованием моделей семейства HAR-RV на примере финансовых инструментов, торговля которыми осуществляется на ММВБ, а также оценено качество моделирования реализованной волатильности.

Были решены следующие задачи:

- изучение теоретических основ концепции реализованной волатильности в ее историческом развитии;
- исследование специфики класса HAR-RV моделей;
- сравнение существующих моделей прогнозирования реализованной волатильности;
- углубленное изучение объектно-ориентированного языка программирования Python в качестве инструмента практической реализации исследуемых моделей;
- построение и оценка моделей прогнозирования реализованной волатильности на российском рынке с использованием моделей класса HAR-RV на языке программирования Python;
- анализ перспектив применения семейства HAR-RV моделей для прогнозирования реализованной волатильности на российском рынке.

Проведенные эксперименты позволили выявить зависимость предсказательной силы моделей семейства HAR-RV от временного интервала (длина наиболее оптимального интервала составляет пять минут), а также от выбранных модификаций расчета реализованной волатильности (модификации с добавлением натурального логарифма и квадратного корня имеют большую силу по сравнению со стандартным расчетом).

В качестве ключевого вывода стоит отметить, что построенные модели прогнозирования реализованной волатильности российского фондового рынка не могут обеспечивать высокий уровень точности в связи с его зависимостью от цены на нефть. Фондовый рынок РФ начиная с 2014 г. находился под давлением как санкций, так и резкого падения цен на нефть, что привело к его повышенной волатильности. Вышеуказанная зависимость наиболее ярко выражается в кризисные периоды, а текущий процесс адаптации российского фондового рынка к значимым экономическим изменениям усложняет

процедуру расчета реализованной волатильности и построение оптимальных стратегий минимизации риска. Также политизированность российской биржи составляет более 80%, что существенно влияет на точность прогнозирования ценовой динамики.

Также стоит отметить, что HAR-RV модель не может применяться для определения реализованной волатильности на российском рынке, так как данный показатель определяется после закрытия позиции, что объясняет низкую эффективность рассмотренных и протестированных в работе моделей. Рассчитать реализованную волатильность можно после того, как траектория движения цены базового актива уже реализовалась, следовательно в качестве необходимого условия выступает устранение чувствительности позиции к изменению цены базового актива, что модели семейства HAR-RV не учитывают.