**Глава 4. Практическое применение динамических показателей временных рядов для прогнозирования котировок акций**

В данной главе приведен краткий анализ современных методов и моделей прогнозирования стоимости акций, основных характеристик процессов прогнозирования и направлений дальнейшего их развития. По результатам проведенных исследований, представлены данные сравнительного анализа наиболее популярных моделей для прогнозирования временных рядов акций на примере акций компаний разных секторов экономики. Проанализировано влияние некоторых параметров временных рядов котировок акций на качество прогноза (длина исторического периода, скользящее окно и окно прогнозирования) по результатам применения гибридной модели прогнозирования. Проанализированы особенности многомерного процесса прогнозирования, а также особенности применения показателя Херста для прогнозирования временных рядов котировок акций: оценка показателя, основные характеристики, применимость для прогнозирования и практические примеры.

**4.1. Прогнозирование стоимости акций: основные характеристики методов и моделей**

Современные методы и модели прогнозирования на фондовом рынке и, в частности, прогнозирования цен на волатильные активы, такие как акции, криптовалюта и пр., являются достаточно востребованными и интенсивно развиваются. В повышении качества прогноза и эффективности методов и моделей прогнозирования заинтересованы как инвесторы, трейдеры, финансовые аналитики, менеджеры управляющих компаний, управляющие инвестиционных портфелей и различных фондов, так и специалисты программной инженерии, программисты и математики, поскольку развитие моделей прогнозирования акций является достаточно интересной и сложной задачей. Успешный и точный прогноз цен на акции в конечном итоге приводит к максимизации прибыли и снижению рисков. За последние десятилетия значительно возросло число исследований по прогнозированию доходности акций, в которых применяются современные интеллектуальные модели и алгоритмы. Значимость таких работ обоснована необходимостью снижения рисков для растущего числа инвесторов, с различным уровнем подготовки, а также для населения, которое все чаще использует услуги фондового рынка для инвестирования.

Эффективное прогнозирование доходности акций и изучение лежащих в его основе причинных факторов по-прежнему остается заведомо сложным из-за сложной, динамичной и хаотичной природы фондового рынка. Анализы технический и фундаментальный составляют две группы методов прогнозирования, относящихся к наиболее старым традиционным методам. Основу технического анализа составляют различные показатели временных рядов, а фундаментальный анализ использует различные фундаментальные переменные. Фундаментальный анализ для целей прогнозирования применяется намного реже, чем технический. Модели фундаментального анализа наиболее применимы для построения долгосрочных прогнозов и прогнозов долгосрочных инвестиционных стратегий. В фундаментальном анализе используются различные фундаментальные переменные (мультипликаторы), которые формируются на базе годовых отчетов компаний, рыночных показателей, различных финансовых мультипликаторов. Фундаментальные факторы имеют большую ценность при прогнозировании на временной период 3-4 года и более, но практически не применимы при построении прогнозов от нескольких недель и месяцев до двух лет (Edson Gould) [48]. Как фундаментальный анализ, так и технический имеют свои уникальные преимущества при прогнозировании доходности акций.

Имеются данные о хорошей прогностической возможности наиболее простых и популярных инструментов технического анализа, который прогнозирует будущие цены на акции, как правило, на основе исторических ценовых тенденций [48,68,85,153,182]. В настоящее время модели технического анализа получили свое развитие за счет применения современных достижений интеллектуального моделирования и машинного обучения, которые используют авторегрессию и метод опорных векторов, а также нейронные сети, модели случайного леса и пр. Таким образом, для прогнозирования котировок акций в основном применяются различные современные методы прогнозирования временных рядов.

Достичь необходимого качества прогноза и сформировать эффективные методы и модели прогнозирования курса акций является достаточно сложной задачей в связи с рядом особенностей, присущих финансовым временным рядам, таких как нестационарность и нелинейность. Результаты многих исследований и практического применения методов прогнозирования показали, что финансовые рынки в некоторой степени предсказуемы [17,23,78,96, 102,119,137]. Хотя, следует отметить недостаточную точность современных прогнозов котировок акций. Однако, интенсивно развивающиеся и совершенствующиеся современные методы машинного и глубокого обучения, с применяемыми в них интеллектуальными алгоритмами, способны учитывать большее количество факторов, влияющих на цены акций, и могут существенно повышать качество прогнозных данных, хотя и за счет усложнения алгоритмов и программного обеспечения. Имеются данные по анализу, основанному на 109 исследованиях, полученных из 10 баз данных, который показывает, что в ежедневных прогнозах средняя частота ошибок составляет менее 1,5%, по сравнению с ежемесячными данными, которые достигают этого уровня только при оптимальных обстоятельствах [13]. Многие результаты прогнозирования указывают на то, что эффективность методов прогнозирования котировок акций во многом определяется временным историческим периодом прогнозирования: на длинных временных рядах многие методы более результативны и точны, чем на коротких. Отсюда можно сделать вывод о более высокой степени прогнозируемости динамики котировок акций в долгосрочной перспективе. В пользу данного предположения свидетельствуют и эмпирические исследования. Однако, имеются данные, которые указывают, что существует оптимальный временной исторический период для прогнозирования: если данных меньше, то точность прогноза ниже, а увеличение объема данных свыше оптимальной величины способствует усреднению прогнозных показателей, а значит, и снижению точности прогноза.

В настоящее время не существует единой универсальной классификации методов прогнозирования волатильных временных рядов, к которым относятся котировки акций. В литературе в данном направлении представлено достаточное количество различных вариантов и модификаций и имеется ряд обзорных работ [23, 96,119,137]. Например, в работе [119] проведен анализ наиболее применяемых методов прогнозирования курса акций, а также предложен вариант классификации методов прогнозирования. Для классификации предлагается выделять статистические методы, методы распознавания образов, машинное обучение, анализ тональности текста и гибридные методы. Большинство проведенных исследований по прогнозированию акций показывает, что современные подходы к прогнозированию обеспечивают более высокие прогностические результаты, чем традиционные методы.

*Классические методы прогнозирования* временных рядов котировок акций. В качестве *стандартного статистического* инструмента широко используются модель авторегрессионного интегрированного скользящего среднего *ARIMA* и ее сезонное расширение, *SARIMA* [3,7,8,9,18,59-61,72,74,106,123]. В данных моделях требуется качественная предварительная обработка данных. При этом достаточно сложным вопросом является выбор окна временных рядов для прогнозирования котировок акций. Было установлено, что модели *ARIMA* не теряют своей достоверности, если окна данных тщательно выбраны и, при этом, отсутствие нелинейности модели *ARIMA* не является серьезной проблемой [59]. Показано, что пятилетнее скользящее окно фиксированной ширины *ARIMA* в наибольшей степени соответствует ценам активов на фондовом рынке США (исторический период с 03.01.1996г. по 12.05.2017г). Получаемые ошибки прогнозирования меньше для акций с низкой волатильностью. Однако в данной работе не исследовалась проблема влияния различной ширины окна на точность прогноза. Имеются исследования модели *ARIMA*, в которых проводилось прогнозирование акций для компаний разных секторов отдельно и исследовалось влияние исторического периода на точность прогноза [72,106]. Однако, анализировался небольшой исторический период и небольшое количество секторов, поэтому полученные результаты требуют дальнейшего подтверждения.

Модели *SARIMA* способны работать с сезонными тенденциями в данных и позволяют получить относительно хороший результат прогноза, если анализируются временные ряды близкие к линейным, что позволяет немного повысить точность [18,109]. А поскольку котировки акций, различных ценных бумаг, криптовалюты и пр., в основном представлены нелинейными временными рядами, то в данном случае необходимо применять либо методы машинного и глубокого обучения, которые выявляют сложные нелинейные зависимости во временных рядах самостоятельно, либо дополнять ими модели *ARIMA/SARIMA*, то есть формировать различные гибридные модели. Во всех исследованиях для выбора лучшего варианта модели *ARIMA/SARIMA* используется метод *AIC* и/или *BIC*.

Сравнение моделей *ARIMA/SARIMA* с некоторыми классическими моделями прогнозирования показали, что данные модели позволяют получить более качественный прогноз по ценам акций и криптовалюты. Например, *ARIMA*, как традиционный метод, основанный на стохастической модели извлечения информация из данных, превзошла *CART* (*Classification And Regression Tree*), смоделированную по структуре дерева. Однако, сравнение *ARIMA/SARIMA* с моделями *ML* и *DL* показывает, что модели машинного обучения для краткосрочного прогнозирования превосходят данные модели [7]. Но есть некоторые данные, утверждающие, что в отдельных случаях эффективность применения моделей *ARIMA/SARIMA* сопоставима с *ML* и *DL*.

Результаты прогнозирования показывают, что модели *ARIMA* и *SARIMA* имеют большой потенциал для краткосрочного и среднесрочного прогнозирования (особенно на интервалах прогнозирования до года), легко реализуемы и могут успешно конкурировать с существующими методами прогнозирования цен на акции. Это преимущество успешно протестировано на большом количестве различных вариантов гибридных методов, которые объединили линейные и нелинейные модели для прогнозирования и в составе которых имеются *ARIMA/SARIMA*. Результаты таких гибридов показали хорошее качество прогнозирования котировок волатильных ценных бумаг, в частности, акций, а также прогнозирования цен криптовалюты, в частности, биткоина [1,3,7- 9,41,59, 72,74,83,106,115,116,118,146,158,162].

*Модели машинного ML и глубокого DL обучения* для прогнозирования котировок акций. В настоящее время разработано достаточно много традиционных методов машинного обучения для целей прогнозирования на фондовом рынке, прогнозирования доходности различных ценных бумаг, стоимости акций, криптовалюты и пр. Следует отметить очень высокие темпы, которыми развиваются методы машинного обучения, используются новые подходы к задачам прогнозирования, основное преимущество которых состоит в их способности обрабатывать нелинейные показатели. Для прогнозирования временных финансовых рядов наиболее популярны: метод опорных векторов *SVM*, позволяющий найти функцию, аппроксимирующую выборку наилучшим образом, метод случайного леса *RF*, метод К ближайших соседей *KNN* (k Nearest Neighbor) [23,42,47,81,87,96,119,124,137]. Исследователи оценили эффективность этих методов, провели сравнительный анализ процессов прогнозирования с их помощью.

Некоторые исследования показали, что метод случайного леса превосходит *SVM* и искусственные нейронные сети *ANN*. В ряде исследований было показано, что методы машинного обучения могут превосходить по точности прогнозирования, как статистические модели, так и простые нейронные сети, которые лучше работают в гибридных вариантах [73,95]. Учитывая то, что традиционные статистические модели не позволяют получить необходимую точность прогнозирования, а искусственные нейронные сети привели к высокой временной сложности, то на сегодняшний день подход глубокого обучения *DL* оказался лучшим методом решения таких проблем, в которых важны временная сложность и качественные прогнозы. Результаты многих современных исследований по прогнозированию акций показывают, что по эффективности прогнозов модели глубокого обучения значительно превосходят аналоги машинного обучения. В большинстве исследований модели *DL* были лучше, чем модели *ML*. Тем не менее, также было много исследований, в которых результаты были сопоставимы. Предпочтение реализаций *DL* моделям машинного обучения растет. Прогресс в вычислительной мощности, доступность больших данных, превосходная производительность, неявные возможности обучения функциям и удобная среда разработки моделей для *DL* являются одной из основных причин этого роста [7,42,92,96,137].

Наибольшее распространение для прогнозирования котировок акций получили рекуррентные нейронные сети *RNN*, среди которых особое значение имеет нейронная сеть с долгой краткосрочной памятью *LSTM*, как отдельный тип сверточной нейронной сети [7,55,75,93,112,114,120,137]. Согласно статистике публикаций, *LSTM* был предпочтительным выбором большинства исследователей для прогнозирования финансовых временных рядов. Это связано с тем, что сети *LSTM* хорошо обрабатывают последовательные данные, извлекают полезную информацию и удаляют ненужную, лучше способны извлекать нелинейные связи, и позволяют получить относительно более стабильные результаты. Более высокая производительность прогнозирования временных рядов моделью *LSTM* и ее вариантами достигается за счет использования обратной связи. Во многих исследованиях показано, что *LSTM* превосходит многие другие модели в эффективности прогноза, например, позволяет построить более точный прогноз, чем *GRU* (*Gated Recurrent Unit*) и классическая *RNN* [36,82, 105,136,139,150,152]. *LSTM* обеспечивает лучшую точность по сравнению с *SVR* [49], лучше прогнозирует котировки акций, чем модели *ARIMA* [52,164]. Было показано, что модели *LSTM* особенно подходят для данных временных рядов из-за их способности включать прошлую информацию, в то время как обнаружено, что ансамбли нейронных сетей уменьшают изменчивость результатов и улучшают обобщение [75].

Для прогнозирования финансовых временных рядов используются также и сверточные нейронные сети (*CNN*), эффективность прогнозирования которых превосходит *LSTM, SVM* и другие методы машинного обучения [57,122].

В работах [23,96,119] предоставлен всесторонний обзор литературных источников по результатам исследований *DL* в отношении прогнозирования финансовых временных рядов. Разработаны варианты нескольких возможных классификаций данных моделей прогнозирования в соответствии с их предполагаемыми областями применения и выбранными моделями *DL*, такими как сверточные нейронные сети (*CNN*), сети глубокого убеждения *DBN* (*Deep Belief Networks*) и долгосрочной краткосрочной памяти (*LSTM*). Авторы обзоров отметили, что фактически, более половины опубликованных работ по прогнозированию временных рядов попадают в категорию моделей *RNN*, предпочтения отдаются *RNN, GRU* и *LSTM*. При этом особое внимание уделено различным вариантам гибридных методов.

*Гибридные модели прогнозирования.* Наиболее эффективными гибридными моделями прогнозирования курсовых стоимостей акций являются модели с включением *ARIMA/SARIMA*, в частности модели *ARIMA-LSTM* и *SARIMA-LSTM*, в которых линейная модель *ARIMA* и нелинейные модели *LSTM* объединяются с целью повышения точности извлекаемой информации о данных [41,93,108,116]. Модель *ARIMA-LSTM* продемонстрировала значительно лучшую чувствительность предсказания цен акций, чем другие финансовые модели. Интегрированная модель *ARIMA-LSTM* более совершенна и надежнее, чем одиночные *ARIMA* и *LSTM*, работает намного лучше, чем другие модели и чаще используется в прогнозировании временных рядов цен акций [41,115]. Имеются данные о том, что ошибки прогнозирования с помощью гибридной модели *ARIMA-LSTM* снижаются в 1,5-2 раза по сравнению с традиционными методами, но увеличение прогностической эффективности часто не перевешивает увеличение стоимости внедрения [118].

Известны также гибридные модели, сочетающие *ARIMA* с *GRU*, то есть использующие управляемый рекуррентный блок *GRU* [122]. Модель *ARIMA-GRU* выявляет закономерности и анализирует, какие данные подходят для прогнозирования, а затем предсказывает подходящее значение. Предлагаемая модель эффективно работает при отсутствии ограничения по времени, поскольку модель требует обучения на очень большом количестве наборов данных. Есть предположение, что гибрид *ARIMA* и *GRU* может превзойти любой другой алгоритм, представленный на рынке.

Во всех гибридных моделях с использованием *ARIMA*, ее оптимальный вариант выбирается с учетом информационного критерия Акайке (*AIC*), байесовского информационного критерия (*BIC*) и критерия Ханнана-Квинна. Поскольку *ARIMA* не учитывает сезонность, поэтому часто применяется вариант *SARIMA*, а для повышения точности прогноза к временному ряду котировок добавляются и дополнительные факторы, которые влияют на цены акций.

Среди гибридных вариантов объединения классических моделей прогнозирования, можно выделить, к примеру, гибридную модель, используемую для прогнозирования индексов китайского фондового рынка и включающую комбинированное *SVM* и *K*-ближайшего соседа со взвешенными признаками [28].

Имеется практика *включения в гибридные модели современных методов или технологий*. Например, была предложена гибридная модель для краткосрочного прогнозирования *VMD-mRMR-LSTM* [157]. Для выбора связанного признака путем анализа корреляции между каждым компонентом и признаком, а также избыточности между признаками использовался алгоритм *mRMR*, который может устранить избыточность между признаками и показать влияющие факторы модальных компонентов. Экспериментальные результаты доказывают, что признаки, выбранные алгоритмом *mRMR*, имеют высокую точность предсказания и хорошую интерпретируемость. С помощью модели *VMD* производится разложение исходной последовательности так, что разложенная последовательность становится стабильной. Результаты прогнозирования гибридной модели с использованием *VMD* качественнее, чем использование *LSTM* для процесса прогнозирования. По сравнению с моделями одиночного прогнозирования гибридные модели обладают более высокой точностью и более надежны и имеют широкую перспективу применения для краткосрочного прогнозирования.

На примере прогнозирования биткойна были протестированы гибридные модели: *VMD-ARIMA-Catboost* и *VMD-ETS-Catboost*. По результатам прогноза их показатели незначительно, но лучше модели *LSTM*, используемой для аналогичной задачи прогноза. *CatBoost* предоставляет множество режимов для регулирования прогнозирования. *CatBoost* представляет собой контролируемый метод машинного обучения, который использует деревья решений для классификации и регрессии. Модель *ETS* (*Exponential Smoothing*) – модель экспоненциального сглаживания можно рассматривать как обобщение простого экспоненциального сглаживания для временных рядов, содержащих тренды и сезонности. Кроме этого, *ETS* имеют базовую модель пространства состояний. Все модели *ETS* нестационарные, в отличие от *ARIMA*. Поскольку в *ETS* учитываются тенденции и сезонные закономерности в данных, то это повышает их точность по сравнению с простыми моделями линейной регрессии. А это гарантирует, что прогнозы будут более надежными и могут быть использованы для принятия более эффективных решений.

В некоторых гибридных моделях используется более двух составляющих. Например, в исследовании [151] предлагается метод для прогнозирования цены закрытия акции на следующий день, который включает сверточную нейронную сеть (*CNN*), двунаправленную долговременную кратковременную память *BiLSTM* (*Bidirectional long-short term memory bi-lstm*) и механизм внимания *AM* (*attention mechanism*), то есть применяется гибридная модель *CNN-BiLSTM-AM*. В этой модели *CNN* используется для извлечения признаков входных данных, *BiLSTM* использует извлеченные данные о признаках для прогнозирования цены закрытия акции на следующий день, *AM* используется для учета влияния состояний признаков на цену закрытия акций в разное время в прошлом, чтобы повысить точность прогнозирования. Результаты показывают, что производительность этого гибрида наилучшая, *MAE* и *RMSE* наименьшие (21,952 и 31,694), а *R²* — самый большой (его значение равно 0,9804).

В гибридных моделях *ARIMA(SARIMA)-CNN-LSTM* используется модель *ARIMA(SARIMA)* и структура глубокой нейронной сети, которая объединяет слои *CNN* и *LSTM* для обработки линейных и нелинейных функций данных. При этом *ARIMA(SARIMA)* используется для обработки линейных функций, сверточная нейронная сеть (*CNN*) используется для обработки иерархической структуры данных, в то время как сеть долговременной памяти (*LSTM*) используется для обработки долгосрочных зависимостей в данных. Тонкие нелинейные детали в остатке лучше моделируются с помощью модели глубокого обучения [77,155,162].

В некоторых случаях для повышения эффективности прогнозирования использовалась модель *GARCH* (*Generalized AutoRegressive Conditional Heteroskedasticity*), которая комбинировалась с *ARIMA* [4]. Однако, лучший результат прогнозирования с использованием модели *GARCH* получен при применении гибридной модели *BiLSTM-ARIMA-GARCH*.

Чтобы повысить точность и стабильность прогнозирования цен на акции, в исследовании [126] предлагается новая гибридная модель *BiLSTM-MTRAN-TCN*. Эта модель состоит из *BiLSTM* (двунаправленная долговременная кратковременная память) и блока *MTRAN-TCN*, который может полностью использовать преимущества трех моделей: *BiLSTM*, трансформатора *MTRAN* (*Modify the transformer model*) и *TCN* (*Time convolutional network*). Трансформатор хорошо работает при получении полной информации о расстоянии, но его способность получать информацию о последовательности слаба. *BiLSTM* может фиксировать двунаправленную информацию в последовательностях, в то время как *TCN* может фиксировать зависимости последовательностей и улучшать способность модели к обобщению. Показатель *R2* данной гибридной модели составляет в 85,7%, *RMSE* уменьшается на 24,3%. Кроме того, модель *BiLSTM-MTRAN-TCN* имеет относительно стабильную производительность прогнозирования в разные периоды времени и лучше подходит для прогнозирования цен на акции с высокой точностью и способностью к обобщению.

Была предложена и исследована гибридная система, получившая название динамического остаточного прогнозирования *DReF* (*dynamic residual forecasting*), которая использует модифицированный алгоритм динамического выбора *DS* (*Dynamic Selection*) и направлена на снижение неопределенности выбора модели *ML* и предотвращение ухудшения прогноза временных рядов [66]. Предлагаемый метод *DReF* использует гибридную сеть из пяти широко распространенных в литературе моделей машинного обучения: многослойного персептрона, опорной векторной регрессии, радиальной базисной функции, долговременной кратковременной памяти и сверточной нейронной сети.

К гибридным моделям прогнозирования относится и сеть *GAN*, которая использует сверточные нейронные сети (*CNN*) в качестве дискриминатора и *GRU* в качестве генератора, а выходные данные сети *GAN* используются как гиперпараметры для дальнейшей настройки и контроля параметров сети [42]. Модель *GRU* в генераторе *GAN* показала очень хорошую производительность с точки зрения запоминания последовательностей временных рядов и создания новых данных для генератора. Полученные результаты прогнозирования показывают, что модель *GAN* является лучшей в прогнозировании финансовых временных рядов и показала лучшую производительность при малом *MSE*. Сеть *GAN* может помочь улучшить прогнозирование временных рядов на фондовом рынке и повысить эффективность принятия решений инвесторами.

В имеющихся исследованиях на сегодняшний день проверено множество различных вариантов гибридных моделей для прогнозирования стоимости акций, криптовалюты. Но, следует отметить, что практически во всех работах исследовали какой-то отдельный вопрос применения гибридной модели, и практически нет системных исследований в данном направлении, хотя имеются различного уровня обзорные работы. К тому же очень важно проверить работу гибридных моделей не только в стабильные периоды функционирования фондового рынка, но и в кризисные периоды.

*Технические и фундаментальные показатели в прогнозировании акций*. Современные исследования в области прогнозирования котировок волатильных ценных бумаг (акций) и криптовалюты показали, что расширение количества учитываемых при прогнозировании факторов может повысить качество прогнозов. Примером таких методов является применение многофакторных и многомерных моделей прогнозирования [11,12,15,26,29,33,34,54,58,80,94,98,113,125]. В данных моделях часто используются различные технические и фундаментальные показатели. В предпринятых попытках интеграции, как правило, вначале агрегировалась последовательность временных рядов в индикаторы технического анализа (например, скользящие средние, линии тренда и индекс относительной силы), а затем обучались прогностические модели с применением, как индикаторов технического анализа, так и переменных фундаментального анализа. При этом временная зависимость представлена (сжата) специальными индикаторами технического анализа, наиболее популярные из которых приведены в главе 2 и в Приложении 3 (см. Приложение 3).

Использование методов и показателей технического и фундаментального анализа в прогнозировании цен акций имеет ряд особенностей и преимуществ. Большее число имеющихся исследований в данном направлении проведено с применением показателей технического анализа, однако некоторые исследователи при прогнозировании временных рядов применяли фундаментальные переменные. Это связано с тем, что индикаторы технического анализа и показатели фундаментального анализа значительно различаются частотой, масштабом и типом. Технический анализ, как правило, имеет дело с данными временных рядов, а фундаментальный использует данные финансовых отчетов и разные рыночные мультипликаторы. При этом инструменты анализа разные: в техническом анализе, как правило, классические модели – скользящее среднее и авторегрессия, а в фундаментальном – различные модели сортировки акций по характеристикам компаний. Стандартное направление применения технического анализа – прогноз лучших торговых показателей для одного или двух активов, а фундаментального анализа – прогноз длинных-коротких позиций, который открывает длинные позиции по акциям с более высокой ожидаемой доходностью и короткие по акциям с более низкой ожидаемой доходностью. Аналитики отмечают, что методы фундаментального анализа эффективны для прогнозирования на средне и долгосрочных интервалах и не сильно эффективны при прогнозировании на коротких интервалах. Может быть перспективной для прогнозирования непосредственная интеграция технического и фундаментального анализов, однако она является достаточно сложной задачей. Несмотря на это данные исследования необходимо продолжать, поскольку они имеют большой потенциал развития методов прогнозирования для волатильных временных рядов котировок акций, стоимости криптовалюты и пр.

Для прогнозирования могут применяться методы технического и фундаментального анализов как самостоятельно, так и в качестве дополнительных показателей. В некоторых исследованиях технические индикаторы используются в качестве основного механизма формирования данных для анализа и прогнозирования цен и могут помочь трейдерам вовремя обнаружить скрытые тренды и закономерности временного ряда. В настоящее время предложены и апробированы модели прогнозирования на основе методов *ML*, *DL* и дополнительных технических или фундаментальных показателей. Например, предложен ансамбль методов технического анализа и машинного обучения [113]. В одном из проведенных исследований, в качестве входных данных для генерации сигналов в системе поддержки принятия решений используется шесть основных технических индикаторов для модели искусственных нейронных сетей (*ANN*) [33]. Было проведено исследование по прогнозированию стоимости акций, в котором использовалось более 270 функций (факторов), относящихся к показателям технического и количественного анализов [33]. При этом используется большинство существующих технических индикаторов (базовые и продвинутые), а количественный анализ в основном базируется на статистических моделях, которые могут предоставить статистику, скрытую в данных. При этом проверяется их достоверность для краткосрочного прогнозирования среднего значения цены акций. Недостаток данного исследования в одновременно используемом большом количестве технических индикаторов, которые могут информационно перекрывать друг друга. Желательно проводить предварительный отбор индикаторов.

Для прогнозирования цен акций могут использоваться различные фундаментальные показатели, характеризующие компании-эмитенты [33,58,94,98,113,125]. Показано, что точность различных моделей прогнозирования улучшается, когда они скорректированы финансовыми характеристиками [33]. В некоторых исследованиях для прогнозирования были взяты следующие фундаментальные показатели: данные о балансовой стоимости компании, курс одномесячных казначейских векселей, балансовая стоимость одной акции (ликвидационная стоимость обыкновенных акций, деленная на количество обыкновенных акций в обращении на конец финансового года) [58]. Кроме этого, были в прогнозирование включены: доходность и цены отдельных акций, доходность индекса S&P 500, отраслевые категории, количество акций в обращении, код акции, код биржи и объем торгов. Прогнозирование, проведенное с использованием гибридного метода *LSTM-DNN* и фундаментальных индикаторов, однозначно показало свое преимущество по сравнению с моделями, в которые не были включены данные показатели. В гибридной модели *LSTM-DNN* для объединения, как данных временных рядов, так и фундаментальной информации, смешанная структура моделирует их по отдельности, а затем объединяет промежуточные результаты для получения окончательных прогнозов. *LSTM* прогнозирует данные временных рядов, а фундаментальные данные обрабатываются *DNN*, что позволяет создавать более сложные отношения между различными функциями. Этот подход помогает модели *LSTM-DNN* извлекать больше информации и повышать качество прогноза.

В исследовании [11] показано, что наблюдается наибольшая связь динамики индекса фондового рынка с изменениями уровня доходов корпораций, сглаженными на 10 лет. Соответственно – *P/E10* является наиболее эффективным фундаментальным показателем с точки зрения построения прогнозов. Это же подтверждено и в работе [34], результаты исследований которой показали, что индикатор *P/E10* демонстрирует высокую прогностическую силу при построении прогнозов на 10 – 20 лет.

Несмотря на достаточно большое количество работ по прогнозированию котировок акций, цен криптовалюты, в которых применяются технические и фундаментальные индикаторы, однозначного вывода относительно их эффективности до сих пор не существует. Это можно объяснить тем, что в проведенных исследованиях используется, как правило, либо один-два индикатора, либо набор индикаторов, некоторые, из которых являются излишними, которые предоставляют противоречащие друг другу сигналы и просто вводят в заблуждение. При этом нет предварительного отбора, исследования и обоснования относительно значимости данных индикаторов для прогнозирования. Следует также отметить и слабую проработанность теоретической базы при разработке и подборе индикаторов технического и фундаментального анализа. Однако, имеющиеся в настоящее время исследования, в большинстве своем признают прогностическую силу отдельных индикаторов технического и фундаментального анализа.

Результаты исследований, приведенные в данной монографии, подтверждают полезность применения технических индикаторов для целей прогнозирования, в частности, в многомерных моделях.

Кроме различных финансовых показателей на курсовую стоимость акций влияют различные экономические факторы, а также новостные факторы, социальные сети, политическая информация, эмоции и настроения инвесторов. Использование данных факторов при прогнозировании цен на акции может способствовать повышению производительности методов прогнозирования и точности прогноза. В настоящее время имеется ряд исследований, результаты которых это подтверждают [89,90,117]. Например, в исследовании [149] получены результаты, которые показали, что функция настроений улучшает точность прогнозирования алгоритмов машинного обучения на 0–3%, а функция политической ситуации повышает точность прогнозирования примерно на 20%. Кроме этого, было установлено, что влияние настроения наиболее эффективно на 7-й день, тогда как влияние политической ситуации наиболее эффективно на 5-й день. Было обнаружено, что алгоритм последовательной минимальной оптимизации *SMO* (*Sequential minimal optimization*) показывает лучшую производительность, в то время как *ASC* (*Adaptive Stability Certification*) и *Bagging* (*Bootstrap aggregating*) показывают низкую производительность.

Объединение финансовой информации и информации о настроениях в сообщениях социальных сетей повышают эффективность прогнозирования фондового рынка, достигая уровня 74,3% (прогноз на следующий час) [90]. При этом использовалась модель на основе ансамбля многослойного персептрона, долговременной кратковременной памяти и модели сверточных нейронных сетей (для оценки настроений). Таким образом, на сегодняшний день подход глубокого обучения, базирующийся на моделях обучения на основе *RL* в сочетании с интеллектуальным анализом текста для учета психологии инвестора, оказался лучшим методом решения таких проблем, в которых важны временная сложность и качественные прогнозы [111].

Достаточно часто для прогнозирования цен на акции с учетом влияния показателей настроения к различным новостям компании, настроениям рынка используются гибридные модели прогнозирования с применением нейронной сети с глубоким обучением *LSTM* или *SVM* [64]. Например, для прогнозирования тренда акций с учетом сочетания новостного контента с историческими ценами на акции в исследовании [145] анализируется модель глубокого обучения. При этом используется многоуровневая долговременная кратковременная память (*LSTM*) для последовательного изучения информации о ценах на акции, а гибридные сети внимания *HAN* (*Hybrid Attention Networks*) используются для того, чтобы определить относительную важность слов из новостных сообщений. Экспериментальные результаты показывают, что может быть достигнут наилучший результат показателя *макро-F1* в 79,0%. Метрика *F1* достаточно часто используется, поскольку она определяется как гармоническое, а не арифметическое среднее значение точности и полноты. Результат 79% по метрике *макро-F1* является достаточно хорошим (более 90% результат считается отличным). По сравнению с отдельными моделями можно получить улучшение производительности до 40%. Ряд исследований указывают на то, что гибридные модели на основе *NLP* (*Natural Language Processing*), семантики и интеллектуального анализа текста в сочетании с данные временных рядов могут стать более распространенными в ближайшем будущем [96]. *NLP* — это сочетание машинного обучения и математической лингвистики, которое изучает методы анализа и синтеза естественного языка и предоставляет компьютерам возможность интерпретировать, манипулировать и понимать человеческий язык.

**4.2. Влияние параметров временных рядов котировок акций на качество прогнозирования, многомерное прогнозирование**

Методы прогнозирования котировок акций строятся на разных техниках и используют разного рода показатели. Учитывая нестационарность, присущую финансовым временным рядам, следует отметить, что далеко не все разработанные модели прогнозирования временных рядов эффективны для предсказания курса акций. Имеющиеся в настоящее время исследования отмечают недостаточную точность прогнозов, используемых в прогнозировании динамики фондовых рынков [23,85,96,119]. Точность прогнозов котировок акций зависит от множества факторов: волатильности; характера периода прогнозирования (стабильный или кризисный); горизонта прогнозирования; субъекта прогнозирования и т.п. При этом качество прогноза зависит от:

* характера периода прогнозирования (гораздо ниже при смене фазы деловых циклов),
* показателя, который прогнозируется: гораздо ниже точность для финансовых показателей, а также ряда нефинансовых показателей, например, уровня запасов, выше для ВВП, инфляции, уровня безработицы и пр.;
* исторического периода для формирования модели прогнозирования.

При оценке точности прогноза учитывается и специфика прогнозирования акций, которая зависит от сложности функционирования фондовых (финансовых) рынков, влияния огромного количества факторов на их динамику, репутационных рисков для субъектов прогнозирования при существенных отклонениях прогноза от действительности. С учетом имеющейся специфики аналитики выделяют следующие особенности прогнозирования динамики биржевых индексов акций:

* преобладание краткосрочных прогнозов (на несколько дней, месяцев, до года);
* отсутствие точных значений в средне‑ и долгосрочных прогнозах;
* доступность во многом только текущих прогнозов и их закрытость вследствие коммерческой ценности.

На рынках акций, как правило, преобладают прогнозы на более короткие сроки (менее года, поскольку год можно считать среднесрочной перспективой), которые базируются на данных технического анализа. Для данных методов, как правило, точность прогнозирования снижается при возрастании горизонта прогнозирования и для более волатильных макроэкономических показателей [20,140]. Методы фундаментального анализа эффективны на средне и долгосрочной перспективе и недостаточно точны при прогнозировании на коротких интервалах [85].

Даже с учетом значительного числа работ в области исследования методов прогнозирования котировок акций, следует отметить, что большая их часть фокусируется на сравнении нескольких методов на примере малого числа компаний, а чаще всего, на примере биржевых индексов. Кроме того, имеются разрозненные данные по прогностической способности отдельных моделей прогнозирования котировок акций. Поэтому актуальным и востребованным является исследование прогностической способности наиболее эффективных моделей прогнозирования котировок акций. В настоящее время практически нет работ, исследующих влияние различных характеристик временных рядов котировок акций на прогностическую способность методов прогнозирования, формирующих оптимальные параметры для моделей прогнозирования и пр.

Многочисленные исследования и сопоставления различных методик прогнозирования временных рядов котировок акций показали, что наиболее перспективным направлением развития является разработка и внедрение гибридных моделей [2,23,96,119,137]. В основе гибридных методов прогнозирования, кроме объединения различных моделей, может использоваться и объединение различных техник, использующих более углубленный математический аппарат. Согласно одной из классификаций гибридные методы прогнозирования можно разделить на четыре группы в зависимости от техник, используемых при объединении различных моделей и их параметров, а также техник дополнительной обработки временных рядов и результатов прогнозирования [161]:

* модели, использующие предварительную обработку временных рядов;
* модели, использующие оптимизацию параметров;
* модели, основанные на сочетании различных компонентов;
* модели на основе постобработки результатов.

В настоящее время существует большое количество сервисов, предоставляющих услуги по автоматическому формированию прогноза для выбранной акции на необходимый промежуток времени, что повышает и подтверждает популярность гибридных моделей прогнозирования. Учитывая превосходство гибридных методов для прогнозирования временных рядов котировок акций, большое количество наших исследований было направлено на практическое изучение особенностей реализации гибридных моделей и анализ качества прогнозирования в зависимости от параметров используемых временных рядов. Первая часть исследований касалась сопоставления качества прогноза котировок акций с использованием различных моделей прогнозирования. На основе различных обзорных исследований и множества современных работ, были выбраны наиболее популярные модели прогнозирования: статистическая модель *SARIMA*, сеть прямого распространения *MLP*, рекуррентная нейронная сеть *LSTM* и две гибридные модели: *SARIMA/MLP, SARIMA/LSTM* [1,2,5,7,71,75,93,108,155,162]. В данных исследованиях было отдано предпочтение модели *SARIMA* поскольку при предварительном тестировании модели были получены лучшие результаты, чем по модели *ARIMA.* Более подробно анализ параметров модели и выбор варианта архитектуры приведены в работах [131,132]. Исследование прогнозирования котировок акций в стабильный временной период, результаты которого приведены в статьях [131,132] было повторно проведено на более широком массиве данных и с учетом кризисного периода. Результаты отличались незначительно в стабильный период функционирования фондовой биржи и имели большие различия в кризисный период. Различия результатов прогнозирования в кризисный период носили не системный характер, а объяснялись лишь воздействием кризисных факторов на конкретные компании, что отражалось и на временных рядах котировок их акций.

Проведем подробное рассмотрение полученных результатов прогнозирования относительно используемых временных рядов котировок акций по выбранным компаниям. Исследовался процесс прогнозирования котировок акций компаний, относящихся к двум группам экономических секторов IT и автомобильной. Данные секторы были выбраны, как наиболее различающиеся по стабильности временных рядов котировок акций среди крупных компаний, входящих в индекс S&P500. Данные по котировкам акций взяты с Нью-Йоркской фондовой биржи за период 2016-2019 г.г. (период стабильного функционирования биржи). На основе предварительного анализа временных рядов исходных данных для обеих отраслей был выбран один исторический период в 3,5 года с 01.07.2016 года по 31.05.2019 год включительно (обучение моделей). Прогнозирование проводилось для временного отрезка с 01.06.2019 года по 30.06.2019 год, то есть на месяц вперед. В данном исследовании были выбраны компромиссные решения относительно качества прогноза по параметрам модели и по историческому периоду для прогнозирования (3,5 года). Проведем сравнительный анализ полученных данных. На рисунке 4.1 представлены усредненные данные по метрикам прогнозирования котировок акций всем компаниям (всего было проанализировано 57 компаний).

 а)

 b)

*Рис. 4.1. Усредненные данные по метрикам прогнозирования котировок акций компаний IT сектора и автомобильного: а) показатель MAPE, % и b) показатель MDA, %*

По метрике *MAPE* гибридные модели показали наилучший результат прогнозирования (погрешность наименьшая) как для акций компаний IT сектора, так и для компаний автомобильного сектора, но самый лучший результат имеем по модели *SARIMA/LSTM*. Что касается метрики *MDA*, то лучший результат для акций компаний IT сектора по данным модели *SARIMA/MLP*, а для акций компаний автомобильного сектора – у модели *MLP*. Что касается автомобильной отрасли, то следует отметить, что *SARIMA/LSTM* лучше справляется с прогнозированием курсов акций по цене, а *SARIMA/MLP* с распознаванием направления движения курса акций. В IT секторе меньшую процентную ошибку также допускает гибридная модель *SARIMA/LSTM*, однако, определяет направление роста или падения цены лучше простая сеть прямого распространения *MLP*. Такой же рейтинг результатов по точности прогнозирования акций получен и для прогноза в кризисный период, только значение погрешности по метрикам больше и больше разброс данных по различным методикам.

Что касается различий между значениями метрик прогноза по разным моделям прогнозирования котировок акций относительно компаний одного сектора, то разность между самым большим значением абсолютной ошибки и самым низким ее значением (*MAPE*) для компаний IT сектора составляет в среднем 2п.п., а для компаний автомобильного сектора 3п.п. В кризисный период эти показатели практически не менялись относительно отдельно взятых компаний одного сектора. Разность между самым большим значением правильного определения направления динамики прогноза и самым низким его значением (*MDA*) для компаний IT сектора составляет в среднем 8п.п., а для компаний автомобильного сектора 11п.п. Разница между значением метрик по разным моделям прогнозирования в наших исследованиях совсем невелика, однако гибридные модели практически во всех, анализируемых нами, случаях показывали лучший результат по прогнозированию. Поэтому можно утверждать об однозначном преимуществе гибридных моделей. Учитывая этот факт и цели прогнозирования, можно подбирать более оптимальные варианты модели.

*Таблица 4.1. Процент различий по средним показателям метрик MAPE и MDA для прогноза котировок акций компаний IT сектора и автомобильного сектора*

|  |  |
| --- | --- |
| Модель | Различие средних значений метрик по моделям прогнозирования между разными секторами, % |
| *MAPE* | *MDA* |
| *SARIMA* | 14,87 | 7,50 |
| *MLP* | 12,18 | 11,46 |
| *LSTM* | 19,06 | 9,48 |
| *SARIMA/MLP* | 1,51 | 16,34 |
| *SARIMA/LSTM* | 15,68 | 0,19 |

Метрики *MAPE* и *MDA* показали, что различие в точности прогноза котировок акций между компаниями разных секторов по выбранным моделям прогнозирования может достигать по метрике *MAPE* 20% (в кризисный период – до 30%) и немногим меньше значения по метрике *MDA*. Показатели по метрике *MDA* имеют больший разброс значений. Меньший уровень различий по гибридным моделям: лучший результат по метрике *MAPE* у модели *SARIMA/MLP*, а по метрике MDA – у модели *SARIMA/LSTM*.

Следует отметить, что и модель *MLP* показывает очень близкие результаты по прогнозированию акций компаний разных секторов. Поэтому, если стоит задача сравнения показателей прогнозирования для акций различных секторов, то лучше применять модели с *MLP*. Это объясняется особенностью данной модели прогнозирования. В общем виде *MLP* (многослойный персептрон) — это классический тип нейронной сети с прямой связью, состоящей из полностью связанных нейронов с нелинейным типом функции активации, организованных как минимум в три слоя и отличающихся способностью различать данные, которые не являются линейно разделимыми. *MLP* подходят для задач прогнозирования с помощью классификации, когда входным данным назначается класс или метка, а также для задач регрессионного прогнозирования в машинной области. *MLP* применяются и для прогнозирования временных рядов. Проблема с использованием *MLP* для прогнозирования временных рядов заключается в подготовке данных: наблюдения запаздывания должны быть сведены к векторам признаков.

Что касается модели *LSTM,* то в целом по всем показателям данная модель позволяет получить лучшие результаты прогнозирования котировок акции, чем *MLP*. В отличие от сетей прямого распространения, рекуррентные сети имеют обратные связи между нейронами в скрытых слоях, что позволяет лучше справляться с обработкой последовательных данных. Однако, наилучшую прогностическую способность показала гибридная модель *SARIMA/LSTM* поскольку основная часть прогноза формируется более устойчивой и надежной техникой, которая определяет общее направление движения котировки и показывает более стабильные данные при автоматическом подборе параметров, а машинное обучение используется для того, чтобы снизить усредненность прогноза линейной моделью, а также добавить динамическую изменчивость и некоторый компонент зашумленности, которые свойственны временным рядам котировок акций. А за счет наличия обратных связей между нейронами и способности формировать внутреннюю память, рекуррентные сети лучше справляются с обработкой последовательностей данных, и поэтому модель *SARIMA/LSTM* превосходит по результатам прогнозирования модель *SARIMA/MLP*.

Таким образом, результатами наших исследований, наряду с многочисленными обзорными работами, по результатам сравнительного анализа различных методик прогнозирования для временных рядов котировок акций подтвержден факт превосходства гибридных моделей. Однако, при этом надо учитывать ряд особенностей при построении гибридных моделей и относительную сложность практической их реализации. Качество прогноза в значительной степени зависит от настроек модели, а также от предварительной обработки исходных данных временных рядов и учете некоторых их динамических показателей.

Второй этап работ по прогнозированию касался исследования некоторых показателей временного ряда котировок акций, которые могут влиять на точность прогнозирования [131,132,133]. Исследовались следующие параметры временного ряда, которые могут влиять на качество прогноза:

* длина исторического интервала котировок акций;
* величина скользящего окна;
* величина окна прогнозирования.

Еще один вопрос, который был исследован, это – зависят ли все эти показатели от отраслевого сектора компании, акции которой прогнозируются.

Для проведения этого этапа работ исследования была выбрана гибридная модель прогнозирования *SARIMA/LSTM*,которая в большинстве современных исследований показывает наилучшие результаты для прогноза котировок акций. Согласно исследуемым показателям, в работе были использованы следующие значения параметров:

* скользящее окно: 10 (2 недели), 20 (около месяца), 40 (около 2 месяцев), 60 (около 3 месяцев) и 80 (около 4 месяцев);
* исторический период: от 2 до 5 лет с шагом в полгода;
* период прогнозирования (окно прогнозирования): 1 день, 1 неделя, 2 недели и 1 месяц.

Более подробно анализ параметров модели и выбор варианта архитектуры приведены в работах [131,132]. Акции компаний выбирались из тех, что входили в индекс S&P500 на Нью-Йоркской фондовой бирже. Исследовались акции компаний следующих секторов: энергетический, финансовый, технологический и IT сектор. Первичное исследование было проведено для временного интервала стабильной работы фондовой биржи с 01/12/2014 по 01/01/2020, где декабрь 2019 года используется для формирования прогнозов. Вторичное исследование включало кризисный период: с 01/12/2015 по 01/01/2021, где декабрь 2020 года используется для формирования прогнозов.

*Подготовка модели прогнозирования более подробно приведена в работе* [132]*.* Фильтр скользящего среднего и соответствующее окно сглаживания выбирались автоматически для каждого временного ряда путем сравнения коэффициента эксцесса получившегося ряда с критическим значением. Выборка для обучения *SARIMA* была предварительно логарифмирована. Параметры для моделей были подобраны автоматически путем минимизации *RMSE*. Для нейронной сети *LSTM* была выбрана архитектура с одним скрытым слоем и 1024 нейронами. С целью уменьшения проблемы переобучения был добавлен слой *DropOut*. Обучение модели проходило на 100 эпохах, данные были предварительно нормированы. Формирование прогноза части *LSTM* было построено пошагово путем дообучения модели на каждом своем последующем прогнозировании. Перед формированием конечного прогноза выводы моделей были приведены к исходному масштабу. По данному алгоритму подбиралась модель для прогнозирования котировок акций, как в стабильный, так и кризисный периоды функционирования фондовой биржи. Полученные прогнозные показатели были усреднены в рамках отрасли и проанализированы результаты относительно средних отраслевых данных.

Проанализируем и сопоставим полученные прогнозные данные.

1. Вначале рассмотрим, как влияет величина исторического периода, на котором обучается модель, на качество прогноза котировок акций с учетом различных окон прогнозирования. На рис. 4.2 приведены показатели метрик MAPE и *MDA* в зависимости от соотношения исторического периода и окна прогнозирования. Имеется четкая прямая зависимость качества прогнозирования от исторического периода для акций компаний всех трех секторов: чем больше длина исторического периода, тем ниже значение метрики *MAPE*. Анализ графиков рис. 4.3. а) – d) показывает, что большее влияние на данную зависимость оказывает размер окна прогнозирования, чем сам исторический период. Из данных графиков а) и b), где показано изменение метрики *MAPE* в зависимости исторического периода для одинаковых окон прогнозирования (30 дней и 1 день), можно сделать вывод, что для пятилетнего исторического периода показатель метрики однозначно лучше, но не так значимо, как влияние окна прогнозирования (см. рис. 4.3.c) и d)). Что касается различий качества прогнозирования для акций компаний рассматриваемых отраслей, то прогноз для акций компаний финансового сектора имеет значительно лучшие результаты по всем метрикам. Установлено, что это связано с показателем стабильности временного ряда акций. Для компаний финансового сектора данный показатель существенно выше, чем для энергетического и технологического. Самый низкий показатель стабильности имеют временные ряды котировок акций технологического сектора. В кризисный период эта разница в показателе стабильности еще значимее и поэтому различия в точности прогноза еще больше. Следует также отметить, что разброс показателей качества прогнозирования по акциям компаний различных секторов значительно зависит от окна прогнозирования, чем больше окно, тем больше разброс показателей (см. рис. 4.3. c) и d)). Разброс оценок прогнозирования по метрике *MAPE* в зависимости от соотношения исторического периода и окна прогнозирования достаточно большой и составляет 80 и более процентов (см. рис. 4.4).

Что касается показателей метрики *MDA*, то имеем неоднозначную ситуацию. Для акций компаний финансового и энергетического секторов имеется слабая прямая зависимость качества прогнозирования от исторического периода, а для акций компаний технологического сектора такой зависимости нет. Чем ниже показатель стабильности курсовой стоимости акций, тем сложнее прогнозировать курсовой рост или снижение стоимости акций. Однако, четко прослеживается следующая ситуация: чем больше окно прогнозирования, тем меньше различия между прогнозными данными по метрике *MDA* между акциями компаний различных отраслей (ситуация противоположная метрике *MAPE*). Разброс оценок прогнозирования по метрике *MDA* в зависимости от соотношения исторического периода и окна прогнозирования достаточно большой и составляет от 40 до 60 процентов (см. рис. 4.4).

|  |  |
| --- | --- |
| а) |  |

 b)

*Рис. 4.2. Показатели метрик MAPE и MDA в зависимости от соотношения* *исторического периода и окна прогнозирования*

|  |  |
| --- | --- |
|  |  |
| a) | e) |
| b) | f) |
| c) | g) |
| d) | h) |

*Рис. 4.3. Показатели метрик MAPE и MDA в зависимости от соотношения исторического периода и окна прогнозирования*

*Рис. 4.4. Уровень разброса прогнозных оценок для акций рассматриваемых отраслей в зависимости от выбранного исторического периода и окна прогнозирования*

1. Проанализируем отдельно по каждой отрасли влияние исторического периода на качество прогнозирования акций компаний отрасли в соответствие с принятым окном прогнозирования (рассматриваются усредненные данные рис. 4.5). Для компаний всех отраслей наименьшее значение средней абсолютной ошибки прогнозирования котировок акций находится в диапазоне исторического периода 4-5 лет. Показатель *MAPE* для компаний всех отраслей постепенно, но с разной величиной увеличивается в зависимости от увеличения окна прогнозирования. Наименьший разброс данных *MAPE* по историческим периодам относительно окна прогнозирования у компаний технологического сектора. Поскольку наименьший разброс данных для окна прогнозирования 7 дней, то если необходимо провести какой-то сравнительный анализ (например, сравнение различных методик прогнозирования), это лучше проводить для данных с окном в 7 дней. Что касается метрики, оценивающей направления динамики прогноза *MDA*, то для акций компаний энергетического и финансового секторов лучшим является прогноз в историческом диапазоне 4-5 лет для окна прогнозирования 1 день. А для компаний технологического сектора – лучший исторический период тоже 4-5 лет, но для окна прогнозирования 30 дней. Если по метрике *MAPE* для каждого исторического периода лучшие показатели для окна прогнозирования 1 день, а худшие – 30 дней, то по метрике *MDA* такое разделение не просматривается и все показатели достаточно часто меняются между собой. Наименьший разброс данных по метрике *MDA* имеем для окна прогнозирования 30 и 14 дней.

|  |  |
| --- | --- |
| По метрике MAPE | По метрике MDA |
| 1. энергетический сектор
 | e) энергетический сектор |
| b) финансовый сектор | f) финансовый сектор |
| c) технологический сектор | g) технологический сектор |
| d) | h) |

*Рис. 4.5. Показатели влияния исторического периода на качество прогнозирования акций в соответствие с принятым окном прогнозирования по каждой отрасли отдельно*.

1. Рассмотрим метрики *MAPE* и *MDA* для прогнозирования котировок акций компаний разных отраслей в зависимости от соотношения скользящего окна и окна прогнозирования. В исследовании использовался следующий размер скользящего окна: 10 (2 недели), 20 (около месяца), 40 (около 2 месяцев), 60 (около 3 месяцев) и 80 (около 4 месяцев). Оценим, какая величина скользящего окна наиболее приемлема в зависимости от окна прогнозирования по выбранным метрикам (см. рис. 4.6). Ситуация с показателями *MAPE* и *MDA* для окна скольжения практически такая же, как и в случае с историческим временным периодом. Метрика *MAPE* имеет наилучшее значение для окна скольжения 10 и растет с ростом окна скольжения для котировок акций всех отраслей. При этом данные *MAPE* у акций компаний финансового сектора самые наилучшие, а у акций компаний технологического сектора наихудшие, но ненамного хуже, чем для акций компаний энергетического сектора. Очень четко у всех котировок акций прослеживается следующая зависимость: наилучшие показатели по *MAPE* для окна прогнозирования 1 день и наихудшие – для окна прогнозирования 30 дней. Разброс данных по метрике *MAPE* для акций компаний разных отраслей меньше для окна прогнозирования 1 день, а больше для окна прогнозирования 30 дней, а по метрике *MDA* – все результаты имеют обратную тенденцию. Однако и по метрике *MDA* лучшие показатели прогнозирования для окна скольжения 10. Наихудшие результаты по метрике *MDA* у акций компаний технологического сектора, а результаты по акциям компаний финансового и компаний энергетического секторов чередуются.

|  |  |  |
| --- | --- | --- |
| a) | b) |  c) |
| d) | e) |  f) |

*Рис. 4.6. Показатели метрик MAPE и MDA в зависимости от соотношения скользящего окна и окна прогнозирования*

*Рис. 4.7. Уровень разброса прогнозных оценок для акций рассматриваемых отраслей в зависимости от выбранного окна скольжения и окна прогнозирования*

Что касается уровня разброса значений метрик, то, как и для исторического периода, уровень разброса прогнозных оценок по метрике *MAPE* для акций компаний разных отраслей достаточно большой (более 80%) в зависимости от окна скольжения и окна прогнозирования, но примерно одинаковый для акций компаний разной отраслевой принадлежности. По метрике *MDA* уровень разброса ниже, чем для *MAPE*. При этом значительно ниже для акций компаний энергетического и финансового секторов (20%-35%) и выше для акций компаний технологического сектора (60%-75%). Это объясняется показателем нестабильности временных рядов котировок акций компаний технологического сектора.

1. Проанализируем отдельно по каждой отрасли влияние величины скользящего окна на качество прогнозирования акций в соответствие с принятым окном прогнозирования (рассматриваются усредненные данные рис. 4.8). Для акций компаний всех отраслей наименьшее значение средней абсолютной ошибки прогнозирования котировок акций определяется размером скользящего окна 10. Такое же значение величины скользящего окна является наилучшим и для метрики *MDA*. Показатель *MAPE* для акций компаний всех отраслей постепенно, но с разной величиной увеличивается в зависимости от увеличения окна прогнозирования и роста величины скользящего окна. Наименьший разброс значений *MAPE* по данным величины скользящего окна относительно окна прогнозирования у акций компаний энергетического сектора. Поскольку наименьший разброс данных для акций компаний всех отраслей для окна прогнозирования 30 дней, то для сравнительного анализа (например, сравнение различных методик прогнозирования) лучше проводить прогнозирование с окном в 30 дней. Наименьший разброс данных по метрике *MDA* для окна прогнозирования 30 дней.

|  |  |
| --- | --- |
| 1. энергетический сектор
 | e) энергетический сектор |
| 1. финансовый сектор
 | f) финансовый сектор |
| 1. технологический сектор
 | g) технологический сектор |
| 1. разброс оценок по метрике MAPE
 | h) разброс оценок по метрике MDA |

*Рис. 4.8. Показатели влияния величины окна скольжения на качество прогнозирования акций в соответствие с принятым окном прогнозирования по компаниям каждой отрасли отдельно*

Таким образом, для гибридной модели прогнозирования *SARIMA*/*LSTM*, кроме подбора параметров модели и выбора варианта архитектуры, необходимо учитывать и подбирать параметры временных рядов котировок акций с учетом поставленных целей исследования. При этом необходимо учитывать и динамические характеристики прогнозируемых временных рядов, например показатель стабильности. Полученные в данном исследовании данные требуют дальнейших уточнений, но их значимость для процессов прогнозирования неоспорима.

В нашем исследовании для гибридной модели прогнозирования *SARIMA/LSTM* с установленными параметрами и архитектурой [58] получены оптимальные данные по величине исторического периода 4-5 лет и величине скользящего окна 10. А также показано, для какого окна прогнозирования получены оптимальные значения. В целом, по результатам эксперимента, можно отметить и следующий факт, что как недостаток данных для обучения, так и их избыток негативно сказывается на качестве прогноза.

*Таблица 4.2. Усредненные показатели метрик относительно исторического периода (a) и скользящего окна (b)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Исторический период, лет | *MDA*, % | Рейтинг значений *MDA* | *MAPE*, % | Рейтинг значений *MAPE* |
| 2 | 42,93 | 7 | 3,41 | 7 |
| 2,5 | 45,97 | 4 | 3,02 | 5 |
| 3 | 45,65 | 5 | 3,03 | 6 |
| 3,5 | 45,43 | 6 | 2,93 | 4 |
| 4 | 50,98 | 2 | 2,66 | 1 |
| 4,5 | 52,62 | 1 | 2,71 | 2 |
| 5 | 49,59 | 3 | 2,72 | 3 |

a) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Скользящее окно | *MAPE*, % | Рейтинг значений *MAPE* | *MDA*, % | Рейтинг значений *MDA* |
| 10 | 2,46 | 1 | 49,61 | 1 |
| 20 | 2,80 | 2 | 47,20 | 4 |
| 40 | 2,97 | 3 | 48,39 | 2 |
| 60 | 3,10 | 4 | 47,30 | 3 |
| 80 | 3,29 | 5 | 45,53 | 5 |

b) |

По усредненным данным (см. табл. 4.2) однозначно можно утверждать, что по обоим метрикам одновременно оптимальный исторический период для прогнозирования котировок акций компаний энергетического, финансового и технологического секторов по модели *SARIMA*/*LSTM* для стабильного периода функционирования фондовой биржи составляет 4-5 лет, а наихудший период 2 года. Что касается скользящего окна, то тоже имеем наилучшую его величину 10, а наихудшая величина 80. Было установлено, что подобранные оптимальные показатели исторического периода, движущего окна позволяют снизить среднюю по акциям рассмотренных компаний ошибку прогнозирования по метрике *MAPE* на 10%-30%. Наши исследования подтвердили существующую в настоящее время проблему в недостаточности работ, в которых сравниваются между собой различные типы гибридных моделей с анализом их оптимальных параметров.

*Использование одномерных и многомерных данных в прогнозировании с помощью гибридных моделей.* В исследованиях по использованию гибридных моделей прогнозирования для котировок акций нами была проанализирована гибридная модель, сочетающая в себе две различных архитектуры нейросетей *CNN/LSTM.* В данном гибриде использовалась сверточная нейронная сеть *CNN* (*ConvNet*), которая представляет собой подтип нейронных сетей. *CNN* хорошо улавливают локальные закономерности для моделирования краткосрочных зависимостей. В общем случае метод *CNN* в гибриде *CNN-LSTM* можно использовать тремя способами: использовать изображения (*GAF, MTF, RP* и т. д.); матричный способ, то есть провести преобразование временных рядов в матрицу *X* и *Y*; провести размещение *CNN* перед *LSTM*. В модели *CNN-LSTM* локальное восприятие и разделение веса *CNN* используется для уменьшения количества параметров в данных до того, как данные будут обработаны *LSTM*. То есть, исходные данные проходят слои *CNN* и после этого передаются в модуль *LSTM*, после чего формируется численный прогноз при помощи полносвязного слоя из одного нейрона. Метод апробирован на примере акций российской фондовой биржи за период ее стабильного функционирования 2009-2019г.г. и за период 2011-2021г.г., включающий кризисные ситуации. При этом были рассмотрены два варианта прогнозирования:

* по временному ряду цены закрытия акций, как наиболее распространенного показателя, и
* по временному ряду *OHLCV* (*open, high, low, close and volume*), который построен на усредненном значении нескольких показателей (цена закрытия акций компании, цена открытия, максимальная и минимальная цена за день, а также объем проданных за день акций) за рассматриваемый период.

*OHLCV* или сокращенная форма *OHLC* (*open, high, low, and close prices*) — это агрегированная форма рыночных данных, обозначающая цену открытия торгов, максимум и минимум по цене за день, цену закрытия и объем торгов по данному активу за определенный период. Показатель *OHLCV* включает следующие данные (составляющие усредняются в один показатель за каждый единичный временной интервал и образуют временной ряд):

* open – цена первой сделки;
* high – самая высокая торгуемая цена;
* low – самая низкая торгуемая цена;
* close – окончательная цена сделки;
* объем – общий объем актива, проданный по всем сделкам.

Иногда в качестве данных используется средневзвешенная цена по объему *VWAP* (*volume-weighted average price*) — это торговый эталон, который дает среднюю цену, по которой ценная бумага торговалась в течение дня, по данным, как объема, так и стоимости.

При этом в обоих случаях прогнозирования предсказывалась только цена закрытия. Полученные данные подтвердили, что погрешность прогнозирования при использовании гибридной модели *CNN/LSTM* по сравнению с негибридными моделями, входящими в нее *CNN* и *LSTM,* ниже.При этом модель *CNN* показала немного лучшие результаты, чем модель *LSTM*,но не для всех акций проанализированных компаний. Как было уже сказано выше, если результаты прогноза по одной модели незначительно превышают результаты, полученные по другой модели 10% – 15% и, при этом, данный результат не стабилен по всем объектам прогнозирования, то нельзя говорить однозначно, что данная модель лучше. Причиной тому являются следующие положения:

* модели машинного обучения требуют кропотливой настройки (оптимизации) своих параметров для процессов прогнозирования и до сих пор нет стандартных рекомендаций в отношении этого процесса (вряд ли они могут появиться в настоящее время) и поэтому могут применяться в разных исследованиях по-разному настроенные модели;
* вторая причина, косвенно связанная с первой, состоит в большом разнообразии временных рядов, которыми характеризуются объекты прогнозирования, а также в эффективности и особенностях их предварительной обработки.

Таким образом, по результатам нашего исследования однозначно можно сделать вывод, что гибридные модели (в данном исследовании *CNN/LSTM* по одному или нескольким показателям) лучше, поскольку их результаты были лучшими стабильно по всем объектам прогнозирования, даже, невзирая на то, что одиночные модели показали не значительное улучшение результатов. Полученные данные подтверждают, что использование нескольких показателей одновременно снижает погрешность прогнозирования для стабильного временного периода до 50% (см. рис. 4.9), а для кризисного периода до 60% и более.

*Рис. 4.9. Метрики моделей прогнозирования CNN, LSTM и CNN/LSTM для цен закрытия акций компании Газпром на российской фондовой бирже за период ее стабильного функционирования 2009-2019г.г.*

Для гибридной модели *CNN/LSTM* дополнительные признаки позволяют повысить качество прогнозирования.

*Гибридные методы прогнозирования, состоящие из объединения трех моделей.* Если объединение двух моделей в одну гибридную модель прогнозирования способно повысить качество прогноза, то, можно предположить, что объединение трех моделей будет еще более успешным. Однако, следует помнить, что при объединении моделей их точность не аддитивна, а величина изменения точности может не соответствовать увеличению сложности реализации модели. В общем случае гибридные модели позволяют объединить преимущества составляющих их более простых моделей и перекрыть некоторые их недостатки. Также, в процессе объединения моделей могут появляться принципиально новые типы моделей прогнозирования. В качестве преимуществ гибридных моделей отмечается также меньшая потребность в экспертных знаниях и участии специалиста в обучении, более простой процесс дообучения, обнаружение большего количества и более сложных зависимостей в данных и адаптация к изменению характеристик временного ряда. В настоящее время уже имеется практика разработки и применения гибридов для прогнозирования с объединением трех простых моделей [23,96,119,137]. Было проанализировано качество прогнозирования с помощью гибридной модели *VMD-CNN-LSTM*. Полученные результаты показали, что данная модель способна снизить ошибку прогноза *RMSE* в среднем на 13%, *MAPE* в среднем на 15% по сравнению с результатами *LSTM*. Сравнение тройного гибрида с *CNN*, *VMD-LSTM* и *CNN-LSTM* показало, что данная модель способна снизить среднюю ошибку по *RMSE* на 12%, 13% и 3% соответственно, а также среднюю ошибку по метрике *MAPE* на 12%, 15% и 6% соответственно. При разработке таких моделей прогнозирования основной проблемой является подбор и согласование параметров простых моделей, включенных в гибридную. Кроме этого, необходимо разрабатывать и использовать такие сложные модели в том случае, если это обосновано, поскольку, как показали обзорные данные по гибридам с тремя моделями, улучшение точности прогнозирования с помощью них по сравнению с гибридами с двумя моделями, как правило, незначительное.

*Многомерный многошаговый метод прогнозирования*. Был исследован многомерный многошаговый метод прогнозирования на примере применения модели машинного обучения *LSTM* для прогнозирования курса биткоина на 7 дней вперед. Модель *LSTM* отличается от всех остальных в отношении многомерности, поскольку она может принимать на вход несколько временных рядов и сразу же выдавать на выходе многошаговое предсказание. Благодаря своей способности распознавать сложные временные корреляции и закономерности сети *LSTM* продемонстрировали эффективность в задачах многоэтапного прогнозирования временных рядов. Модель *LSTM*, например, можно обучить, используя ряд исторических ценовых входных данных для прогнозирования цены Биткойна. Скользящие окна или методы последовательности за последовательностью (*seq2seq*) могут использоваться для адаптации модели *LSTM* к созданию многошаговых прогнозов цены биткойна. Для прогнозирования использовался метод прямого прогноза. Формируется 7 моделей для многошаговых прогнозов на 7 дней. Каждая модель эффективно предсказывает свой собственный период. После этого происходит объединение всех результатов и получается прогноз на 7 дней вперед. Этот подход требует больших вычислительных мощностей. Сеть двунаправленной долговременной кратковременной памяти (*LSTM*) для данного исследования была разработана для исследования в бакалаврской работе Саидом Дандамаевым с использованием библиотеки Keras.

Данные взяты за период 2014-2023г.г. (16.06.2023), при этом за период 01.11.2022-31.05.2023 – тестовая выборка. Целевая цена биткоина в данном исследовании – это цена закрытия, которую и прогнозируем.

Для прогнозирования использовались три модели: в одной модели – одномерной использовались только показатели *OHLCV*, в двух других моделях (многомерных) использовались два временных ряда данных: в одном *OHLCV* и временной ряд кумулятивного показателя стабильности цен биткойна, а в другом – *OHLCV* и временной ряд кумулятивного показателя просадок цен биткойна. Дополнительным временным рядом в модели многомерного прогнозирования является динамический ряд кумулятивных показателей стабильности исходного временного ряда котировок биткойна. Динамический ряд кумулятивных индикаторов стабильности характеризует тенденцию относительно стабильности уровней временного ряда котировок биткойна (более полное описание приведено в главе 2 монографии). Стабильность определяется как значение уровня временного ряда, равное или превышающее предыдущие уровни. Этот кумулятивный подход к оценке стабильности цены биткойна позволяет нам учитывать рост/падение цены и величину этого роста/падения. Совокупные показатели стабильности временных рядов цен на биткойн можно рассматривать как информативный индикатор качества поведения цены биткойна на фондовом рынке. Этот индикатор сочетает в себе как рисковую, так и ценовую составляющие. Косвенно кумулятивные показатели стабильности также учитывают и просадки. Графики временных рядов для анализируемых цен биткоина показаны на рис. 4.10.



*Рис. 4.10. Временные ряды кумулятивного показателя стабильности и кумулятивной просадки цен биткойна*

Что касается метрики по индексу стабильности, то следует подчеркнуть, что временной ряд Биткойн имеет высокий показатель индекса стабильности, 87,3% (рассчитывается как отношение суммы показателей стабильности временного ряда к сумме показателей стабильности полностью стабильного ряда за рассматриваемый временной период). Можно видеть, что в периоды бычьего рынка цены биткойнов достигают новых верхних значений.

*Рис. 4.11. Метрика MAE по прогнозированию цены закрытия биткоина по модели LSTM на июнь 2023г.*

*Рис. 4.12. Метрика SMAPE по прогнозированию цены закрытия биткоина по модели LSTM на июнь 2023г.*

*Рис. 4.13. Метрика MDA по прогнозированию цены закрытия биткоина по модели LSTM на июнь 2023г.*

Выходной переменной, как результатом прогнозирования, является цена закрытия торгового дня по биткоину. Результаты показывают, что при небольшом горизонте прогнозирования модель, обученная на торговых данных, показывает наилучшие результаты. В то же время модель, обученная на данных с функцией стабильности, показала несколько лучшие результаты по процентной ошибке. Следует отметить, что временной ряд показателей стабильности достаточно сильно коррелирует с временным рядом показателей *OHLCV*, но он отражает показатели отклонений от стабильности. Это позволяет уточнить и скорректировать прогнозные показатели. Что касается временного ряда кумулятивных просадок, то этот показатель ухудшил результаты модели, из чего можно сделать вывод о его незначительности влияния на прогностическую способность модели. По-видимому, он будет хорошо работать в случае нестабильного временного ряда цен биткоина.

Подобранная архитектура многомерной многоэтапной модели позволила повысить точность прогноза по сравнению с одноэтапными моделями на 10-15%. Применение многомерных многоэтапных моделей может позволить получить более точный прогноз. В нашем случае многомерная многоэтапная модель с дополнительным временным рядом кумулятивных показателей устойчивости позволила получить лучшие результаты на 5 и более шагов прогнозирования. Это показывает, что необходимо дальнейшее исследование многомерных многоэтапных моделей прогнозирования с дополнительными данными, например техническими индикаторами, дополнительными финансовыми показателями, например, временными рядами цен на золото, нефть и прочее.

**4.3. Анализ применения показателя Херста для прогнозирования**

В настоящее время возобновился интерес по исследованию фрактальности на фондовых рынках, применению коэффициента Херста в процессах прогнозирования, поскольку имеющиеся практические результаты по данным проблемам недостаточны для их использования. Особенности использования коэффициента Херста в прогнозировании акций и иных ценных бумаг исследовались в ряде научно-экспериментальных работ. В работе [147] данные фондового рынка Шэньчжэня изучались с использованием метода *DFA*(*detrended fluctuation analysis*) и, анализируя изменение показателя Херста, было установлено, что повышается эффективность фондового рынка.

В исследовании [159] использовался метода *DFA* в анализе доходности (дневной) фондового индекса Шанхайской биржи. Результаты позволили во временных рядах установить и проанализировать источники показателя мультифрактальности. Для сравнения финансовых рисков рынка использовались обобщенные показатели Херста. В результате было обнаружено, что для данных показателей характерна сильная изменчивость при резком росте или падении индекса.

Для прогнозирования значений уровней временных рядов в исследовании Митра [88] была оценена полезность коэффициента Херста. Полученные в исследовании результаты показали, что более высокую торговую прибыль можно ожидать в том случае, если временные ряды котировок показывают более высокое значение коэффициента Херста. Однако, полученные результаты нуждаются в более точном обосновании.

В настоящее время можно отметить совсем небольшое количество исследований, в которых были попытки оценить непосредственное влияние показателя Херста на прогнозирование финансовых временных рядов. Результаты таких исследований не очень очевидны и требуют дальнейших уточнений. Авторами монографии был проведен ряд исследований по оценке показателя Херста для акций IT компаний и компаний, входящих в индекс S&P500, за период 2015-2021г.г. А также были проведены исследования для оценки возможности применения данного показателя в анализе и прогнозировании котировок акций.

Первым вопросом, который необходимо было решить – это вопрос о том, как показатель Херста во временном ряду котировок акций соотносится с другими показателями, характеризующими временной ряд и как этот показатель изменяется во времени (реагирует ли он на кризисные ситуации). Исследования по данным вопросам авторы монографии проводят с 2019 года [130,134,176]. В данном разделе монографии систематизируем полученные за пять лет результаты по исследованию показателя Херста.

Известно положение о том, что коэффициент Херста характеризует наличие тренда во временном ряду. Однако, для изучения тренда во временных рядах используются коэффициенты ранговых корреляций Спирмена, Кендалла, а также можно применять коэффициент конкордации рангов. Например, устойчивость тенденции изменений уровней временного ряда определяется по коэффициенту Спирмена, а коэффициент Кендалла оценивает наличие тенденции (тренда) во временном ряду. Если коэффициент Херста указывает на наличие тренда, то его показатели должны коррелировать с показателями Спирмена и Кендалла. Проверим, совпадают ли показатели Спирмена и Кендалла с коэффициентом Херста.

Сравнивая логику оценки показателей ранговой корреляции, можно сделать вывод, что, если данная корреляция будет высокой, то имеется устойчивый тренд во временном ряду и, на основе этого, можно, используя принципы наивного прогноза, сделать предположение (прогноз) о возрастании или падении уровня временного ряда на будущий шаг. Хотя ранговая корреляция и по Спирмену и по Кендаллу предоставляет данные по одинаковой характеристике временного ряда. В проводимом нами исследовании оценим показатели ранговой корреляции и коэффициент Херста и проанализируем полученные результаты. Исследование проводилось на большом массиве данных: анализировались временные ряды котировок акций более семидесяти компаний, входящих в индекс S&P500 и относящихся к разным секторам экономики. Отдельно оценивались показатели для четырех лет 2015-2018 г.г. и 2020г как кризисный период. Поскольку показатели оценивались за год или полугодие (по торговым дням), то для расчета коэффициента Херста применялась формула Неймана по методу *R/S* анализа [130,134,176].

В качестве примера приведены показатели Херста на рис 4.14 (2015-2017 г.г., расчет по месяцам, таблица с данными помещена в Приложении 1) и рисунке 4.15 (2020г., расчет по данным закрытия торговых дней).

*Рис.4.14. Показатели Херста, рассчитанные на 2015-2017г.г., ед.*

*Рис.4.15. Показатели Херста, рассчитанные на 2020г. (расчет по данным закрытия торговых дней на первое и второе полугодия), ед.*

Первый вопрос, с которого начнем анализ коэффициента Херста, – это сравнение его значений для временных рядов котировок акций в стабильный и кризисный периоды. Полученные данные показывают:

* большинство из проанализированных компаний (временных рядов котировок акций компаний) примерно 70% имеют значение Херста более 0,5, но менее 0,7 в стабильный период, что указывает на персистентность (наименее вероятна смена предыдущего направления движения котировок акций);
* большинство из проанализированных компаний (временных рядов котировок акций компаний) около 80% имеют значение Херста более 0,2, но менее 0,35 в кризисный период, что указывает на антиперсистентность (смена предыдущего направления движения котировок акций более вероятна);
* не замечено значимое отличие данных по коэффициенту Херста относительно отраслей;
* разброс значений коэффициента Херста внутри рассматриваемого временного диапазона в кризисный период немного возрастает, примерно на 10%-20%.

Была проанализирована корреляция значений коэффициента Херста между анализируемыми временными периодами: ни в стабильном периоде, ни в кризисном не выявлена коррелированность значений. Сравнить динамику изменений значений коэффициентов Херста для некоторых компаний за период 2015-2017 г.г. можно по данным рис. 4.14 - 4.16.

Рис. 4.16. Значения коэффициентов Херста по годам 2015г., 2016г., 2017г., ед. для некоторых компаний

Сравнение значений Херста по годам показало, что зависимости между данными показателями нет. Это можно увидеть и на рис. 4.16.

Необходимо учитывать следующее: в исследованиях использовался коэффициент корреляции, который позволяет установить коррелированность тенденций изменения показателя, а не саму причинно-следственную взаимосвязь. То есть устанавливается коррелированность изменений данных, но это не всегда означает взаимосвязь данных. Если коррелированность данных высокая и имеет устойчивую тенденцию, тогда можно анализировать и установить взаимосвязь данных.

Были сравнены значения коэффициента Херста, рассчитанные по методу *R/S* анализа, и по методу *DFA*. Результат показал, что данные по *DFA* немного завышены по сравнению с данными метода *R/S*. Это объясняется методом расчета показателя Херста.

Вторым этапом анализа показателя Херста было сопоставление его значения с коэффициентом Спирмена, критерием Кендалла и коэффициентом конкордации (см. рис. 4.17, 4.18).

*Рис. 4.17. Значения коэффициентов Спирмена, Кендалла и Херста для интернет-компаний за 2017* *год.*

Рис. 4.18. Значения коэффициента Спирмена, Кендалла, конкордации и Херста некоторых компаний из S&P500 за 2020г.

Корреляция между значениями коэффициентов Спирмена, Кендела и конкордации выше 90%, поскольку данные коэффициенты оценивают одинаковый показатель ранговой корреляции и используют сопоставимый методический подход к оценке, который характеризует наличие тренда. Что касается коэффициента Херста, то в наших оценках его значение не коррелирует с перечисленными выше показателями ранговой корреляции. Таким образом, можно сдклать вывод, что коэффициент Херста не показывает наличие тренда во временном ряду данных. Что противоречит заключениям ряда аналитиков о способности по коэффициенту Херста установить наличие или отсутствие тренда [30,97,169,171].

Однако, для прогноза направления изменения значений уровней временного ряда на 1-2 прогнозные точки в будущем (на очень короткий период) можно применять коэффициент Херста. По данным наших исследований можно сделать вывод, что точность определения направления движения акции на 1 прогнозный шаг составляет более 80%, на второй шаг составляет более 70%, а на третий примерно 20%. В кризисный период эти показатели были немногим ниже. Стоит отметить, что данная прогнозная способность коэффициента Херста базируется не на оценке трендовости временного ряда, как у коэффициентов корреляции Спирмена и Кендела. Это указывает на то, что прогнозная способность коэффициента Херста базируется на свойствах фрактальности. А, как отмечают некоторые аналитики, фондовый рынок не всегда фрактальный.

Поскольку ранее в ниших исследованиях анализировалось большое количество различных динамических показателей временного ряда котировок акций, то был проведен сравнительный анализ данных показателей с коэффициентом Херста. При этом выполнение многомерного факторного анализа по всем динамическим показателям и коэффициенту Херста дает возможность определить содержательную характеристику коэффициента Херста.

В исследовании Э.Петерсона [174] было сделано заключение о том, что преимущество при анализе риска временных рядов котировок акций надо отдавать не стандартному отклонению, а фрактальному риску по коэффициенту Херста. По результатам, проведенного нами исследования данные, полученные по коэффициенту Херста, среднеквадратическому отклонению и доходности акций, оцененной, как среднее значение стоимости акций за анализируемый период, можно сравнить наглядно по рис. 4.19 и 4.20.

Рис. 4.19. *Значения коэффициента Херста, среднеквадратического отклонения и доходности акций некоторых компаний за период 2020 г.*

*Рис. 4.20. Значения коэффициента Херста, среднеквадратического отклонения и доходности акций некоторых компаний за период 2017 г.*

Корреляция среднеквадратического отклонения и коэффициента Херста наблюдалась только в данных за 2017 год со значением 30% и только для котировок акций интернет-компаний. Но данный факт требует дальнейшего исследования для установления причин и устойчивости данной корреляции. В дальнейших исследованиях такая связь не подтверждалась. Во всех остальных случаях корреляция была очень незначительной и неустойчивой. В данном исследовании установлено: стандартное отклонение и показатель Херста оценивают относительно разные характеристики риска, поэтому эти показатели лучше использовать совместно, как дополняющие друг друга. Таким образом, заключение, сделанное Э.Петерсом [174] о том, что при анализе риска акций преимущественно надо использовать не стандартное отклонение, а фрактальное измерение, к которому и относится показатель Херста, не подтвердилось. Вероятно, константу Херста можно использовать как меру фрактального риска при прогнозировании по временным рядам. Стоит отметить, что небольшая корреляция (30%-40%) имеется между коэффициентом Херста и показателем колеблемости временного ряда, но в кризисный период этот показатель не подтвердился.

Многомерный факторный анализ по динамическим показателям и коэффициенту Херста был проведен отдельно по годам с 2015г. по 2018г. и за 2020 г. Получены следующие результаты:

* по всем результатам анализа не было установлено ни одного факта более или менее устойчивой корреляции коэффициента Херста с динамическими показателями ряда котировок акций, то есть показатель Херста выделялся в отдельную группу, не объединяясь ни с одним из динамических показателей ряда;
* в кризисном году отмечена была невысокая коррелированность коэффициента Херста с показателем коэффициента устойчивости (25%-35%);
* и в стабильном периоде, и в кризисном наблюдалось одинаковое распределение анализируемых показателей по комплексным факторам.

Все это указывает на то, что показатель Херста не коррелирован с динамическими характеристиками ряда, а отражает фрактальную характеристику, т.е. вероятность прогнозного показателя, который можно использовать как показатель фрактального риска.

В кризисный период отмечена была невысокая коррелированность с показателем коэффициента устойчивости (25%-35%). Это подтверждает факторный анализ пяти динамических показателей, которые были отобраны (предварительно удалены показатели, которые имеют высокую корреляцию с отобранными). Факторный анализ с выделением 3 и 4 факторов помещает показатель Херста вместе с коэффициентом устойчивости в один комплексный фактор.

В таблицах 4.3 и 4.4 приведены примеры многофакторного анализа данных за 2020г. Приведены варианты выделения трех и двух факторов. Причем, только в случае выделения трех или двух факторов можно увидеть небольшую корреляцию показателя Херста с коэффициентом устойчивости. Что касается среднего квадратического отклонения, то ни в одном из проведенных нами факторных анализов данный показатель не объединялся с показателем Херста. В данных таблицах приведены результаты заключительного этапа факторного анализа, в котором участвуют пять основных показателей временных рядов котировок акций.

Таблица 4.3. Объясненная совокупная дисперсия

|  |  |
| --- | --- |
| Компонент | Начальные собственные значения |
| Всего | % дисперсии | Суммарный % |
| 1 | 2,102 | 42,049 | 42,049 |
| 2 | 1,310 | 26,202 | 68,251 |
| 3 | 0,759 | 15,187 | 83,438 |
| 4 | 0,423 | 8,455 | 91,893 |
| 5 | 0,405 | 8,107 | 100,000 |

По таблице 4.3. выбиралось либо три, либо два фактора для формирования матрицы компонент, которые проиллюстрированы в табл. 4.5.

*Таблица 4.4. Результаты факторного анализа: матрица компонент повернутая*

|  |  |  |
| --- | --- | --- |
| Показатели  | Три комплексных фактора | Два комплексных фактора |
| F1 | F2 | F3 | F1 | F2 |
| Среднее квадратическое отклонение | -0,157 | -0,052 | 0,959 | -0,657 | -0,137 |
| Коэффициент устойчивости | 0,375 | 0,673 | -0,414 | 0,556 | 0,673 |
| Индекс кумулятивной просадки | 0,875 | 0,232 | -0,048 | 0,763 | 0,151 |
| Показатель Херста | -0,113 | 0,917 | 0,060 | -0,106 | 0,912 |
| Индекс сходства структур | 0,810 | -0,230 | -0,272 | 0,822 | -0,276 |

По данным таблиц 4.3 и 4.4 можно выделить комплексные факторы (см. табл. 4.5):

*Таблица 4.5. Комплексные факторы и их характеристики*

|  |  |  |  |
| --- | --- | --- | --- |
| Факторы | % дисперсии | Собственные значения | Показатели |
| F1 фактор структуры | 42% | 2,102 | Индекс кумулятивной просадки (0,88);Индекс сходства структур (0,81) |
| F2 фактор устойчивости | 26,2% | 1,310 | Показатель Херста (0,92),Коэффициент устойчивости (0,67) |
| F3 фактор риска | 15,2% | 0,759 | Среднее квадратическое отклонение (0,96) |

Факторные нагрузки примерно одинаковые.

Поскольку коэффициент Херста является одним из основных значений фрактального анализа и может повысить эффективность прогнозных данных при анализе временных рядов, то в нашем исследовании ставилась цель изучения возможности применения коэффициента Херста для повышения информативности современных методов прогнозирования курсов акций, в частности методов машинного обучения. Наиболее востребовано решение данного вопроса различными автоматизированными системами на фондовых рынках, автоматизированными системами для консультаций инвесторов, робоэдвайзерами и пр.

Таким образом, следующим этапом исследований показателя Херста был этап непосредственного прогнозирования и выяснения полезности данного показателя для предсказания направленности изменений курсов акций. Данный анализ был проведен с применением классического статистического метода прогнозирования и с применением современных методов машинного обучения для прогнозирования.

Используем метод прогнозирования Хольта-Уинтерса. В качестве исходных данных, были выбраны для прогнозирования средние ежемесячные показатели котировок акций 30 ведущих информационных компаний за период 2015-2017 г.г. В общем случае имеем временной ряд котировок акций, в структуре данных которого есть сложившийся тренд и сезонность. Для процесса прогнозирования был выбран метод Хольта-Уинтерса. Модель строилась по данным котировок акций помесячно за три года 2015-2017 г.г., а прогноз проводился на 2018 г. [150].

Сезонный метод Холта-Уинтерса включает в себя уравнение прогноза и три уравнения сглаживания — одно экспоненциальное сглаживание для $l\_{t} $, другое для тренда $b\_{t}$ и третье для сезонной составляющей $s\_{t}$, с соответствующими параметрами сглаживания α, β и γ. Для обозначения повторяемости сезонности используем *m*, т. е. числа сезонов в году. Например, для квартальных данных *m*=4 и для ежемесячных данных *m*=12.

Существует два варианта этого метода, которые различаются характером сезонной составляющей. Аддитивный метод предпочтителен, когда сезонные колебания примерно постоянны в ряду, тогда как мультипликативный метод предпочтителен, когда сезонные колебания изменяются пропорционально уровню ряда. При аддитивном методе сезонная составляющая выражается в абсолютном выражении в масштабе наблюдаемого ряда, а в уравнении уровня ряд сезонно корректируется путем вычитания сезонной составляющей. В течение каждого года сезонная составляющая в сумме будет равна примерно нулю. При использовании мультипликативного метода сезонная составляющая выражается в относительных величинах (процентах), а ряд сезонно корректируется путем деления на сезонную составляющую.

Таким образом, применялась трехпараметрическая с мультипликативной сезонностью модель Хольта-Уинтерса с расчетами по формуле: [150]

$$\hat{y}\_{t+h}=\left(l\_{t}+hb\_{t}\right)s\_{t-m+h},$$

где $\hat{y}\_{t+h}$ *–* прогнозное значение уровня ряда на $t+h$ периодов;

$l\_{t} $ *–* экспоненциальное сглаживание уровней ряда;

$h$ *–* порядковый номер периода, на который делается прогноз;

$b\_{t}$ *–* сглаживание тренда;

$s\_{t-m+h}$ *–* сглаживание для сезонности;

*m* *–* обозначение повторяемости сезонности, т. е. числа сезонов в году (для квартальных данных *m*=4 и для ежемесячных данных *m*=12).

*Таблица 4.6. Формулы для расчета компонент сглаживания в моделях Хольта-Уинтерса с мультипликативной сезонностью*

|  |  |  |
| --- | --- | --- |
| Показатель сглаживания | Формула расчета | Значение коэффициента |
| Экспоненциальное сглаживание уровней ряда | $$l\_{t}=α\*\frac{y\_{t}}{s\_{t-m}}+\left(1-α\right)\*\left(l\_{t-1}+b\_{t-1}\right),$$$l\_{t}$ *–* экспоненциальное сглаживание уровней ряда;$α$ *–* коэффициент сглаживания уровней ряда;$s\_{t-m}$ *–* сглаживание для сезонности$ t-m$ периода;$y\_{t}$ *–* текущее значение ряда;$l\_{t-1}$ *–* сглаженная величина уровней ряда за предыдущий период;$b\_{t-1}$ *–* сглаженный тренд за предыдущий период. | $0\leq α\leq 1$  |
| Тренд | $$b\_{t}=β\*\left(l\_{t}-l\_{t-1}\right)+\left(1-β\right)b\_{t-1},$$$b\_{t}$ *–* сглаженный тренд за текущий период;$β$ *–* коэффициент сглаживания тренда. | $0\leq β\leq 1$  |
| Сезонность | $$s\_{t}=γ\*\frac{y\_{t}}{l\_{t-1}+b\_{t-1}}+\left(1-γ\right)s\_{t-m}, $$$s\_{t}$ *–* сглаживание для сезонности$ $текущего периода;$γ$ *–* коэффициент сглаживания сезонности. | $0\leq $ $γ\leq 1$ |

Рис. 4.21. Пример подобранных коэффициентов α, β и γ для моделей Хольта-Уинтерса некоторых IT-компаний на 2018 год

Используя автоматизированные системы, можно подобрать поправочные коэффициенты α, β, γ и минимизировать возможные ошибки. На рис. 4.21 приведены значения подобранных коэффициентов сглаживания для модели прогнозирования по некоторым компаниям.

Модели Хольта-Уинтерса позволяют получить относительно хорошую точность прогноза (см. метрики прогноза для некоторых интернет-компаний на рис. 4.22).

*Рис. 4.22. Метрики прогноза для некоторых интернет-компаний по модели Хольта-Уинтерса*

Анализ метрик показал, что отсутствуют значимые показатели погрешностей и модель позволяет получить хороший результат прогнозирования. Таким образом, еще раз было подтверждено, что использование методов *HWES* для прогнозов котировок акций в средне и долгосрочных временных периодах может быть рекомендовано для практического применения.

Точность прогноза методом *HWES* приемлема для проведения исследований относительно показателя Херста. Был рассчитан коэффициент Херста для временных рядов котировок акций. В ходе прогнозирования, кроме оценки прогнозных значений на несколько будущих уровней временного ряда котировок акций, проверялись и вероятности смены направления движения уровней котировок акций на прогнозных уровнях в зависимости от значений коэффициента Херста. Было установлено, что для первой прогнозной точки зависимость выполняется на, более чем, 80%, для второй прогнозной точки – вероятность более 70%, далее данная зависимость резко снижается (эти показатели снижаются в кризисных периодах). Таким образом, установлено на многочисленных (более 300) примерах прогнозирования котировок акций в разные временные периоды (стабильные и кризисные), что использовать коэффициент Херста при оценке прогноза на период 6 месяцев (6 точек) и более не представляется возможным. Выявлено было наличие среднего уровня коррелированности показателей Херста с показателем точности прогноза. В нашем исследовании подтвердилось предположение Митра о том, что методы прогнозирования дают лучшие результаты в периоды высокого показателя Херста [88,130,134,176].

Данное исследование еще раз подтвердило, стохастичность характера взаимосвязи показателя Херста и прогнозных показателй временного ряда котировок акций. Однако, многочисленные исследования показали, что длина прогонозирования по показателю Херста (предсказания направления движения уровней временного ряда) достаточно мала. Но, по-видимому, временной диапазон прогнозирования с помощью коэффициента Херста может различаться в зависимости от характеристик временного ряда и стабильности временного периода для прогнозирования.

Еще одно полезное направления использования показателя Херста – фрактальная мера риска при прогнозировании котировок акций [130,134,174,176]. Причем данный показатель рекомендуется использовать как дополнение к стандартным показателям оценки риска котировок акций. Это позволит повысить эффективность прогнозирования и принятия последующих решений.

В проведенных исследованиях кроме использования для прогноза методов *HWES* и изучения оценки вероятностного направления изменения динамики временного ряда по показателю Херста, был проведен анализ и сделано обоснование эффективности использования данного показателя при прогнозировании современными методами, среди которых применялись модели: *ARIMA,* *ESM*, *LSTM*. Данная часть исследований проведена на примере котировок акций ведущих IT компаний и телекоммуникационных компаний фондовой биржи Нью-Йорка (прогнозный период 2020г.). Проанализированы котировки акций 61 компании. Данные по курсам акций взяты с фондовых бирж по показателям закрытия торговых дней. Для моделей прогнозирования использовалось программное обеспечение, разработка которого выполнена в среде Google Colab на языке программирования Python версии 3.7.13 с использованием библиотек Pandas, NumPy, Matplotlib, Fathon, Keras[[1]](#footnote-1). Экспериментальное вычисление показателя Херста проводилось с помощью методов *R/S* и *DFA*.

В рамках данного исследования по результатам подбора гиперпараметров и прогнозирования, были получены результаты, еще раз подтвердившие, что модель *LSTM* позволяет получить более качественные результаты прогнозирования по сравнению с *ARIMA,* *ESM* и *HWES*. Также полученный результат подтверждает предположения Митра [88] о том, что методы прогнозирования дают лучшие результаты в периоды высокого показателя Херста (см. показатели рис. 4.23).

*Таблица 4.23. Пример результатов показателей MAPE и значений коэффициента Херста для некоторых IT компаний.*

Поскольку временные ряды котировок акций, характеризуются своей нестационарностью, то в данном исследовании кроме метода *R/S* использовался метод оценки показателя Херста по модели *DFA*. Значения показателя Херста близки к 0,5, но, в основном, были немного ниже этого показателя. Данные по *DFA* немного завышены по сравнению с данными метода *R/S*, поскольку количество точек было немного недостаточным для точной оценки *DFA*. Подтвержден тот факт, что показатель Херста, для оценки вероятности смены предыдущего направления временного ряда котировок акций, целесообразно использовать только на коротких прогнозных интервалах (1-3 прогнозные точки). На средних и длинных интервалах результат очень неточный. Анализ полученных значений показал, что большинство интернет-компаний и телекоммуникационных (примерно 60%) имеют значение Херста более 0,5 в стабильные периоды, но данный показатель снижается в кризисный период.

По результатам данного исследования получено:

* величина временного интервала прогнозирования, для которого миожно использовать показатель Херста достаточно мала 1-3 прогнозные точки;
* показатель Херста можно использовать как фрактальную меру риска по временным рядам котировок акций;
* имеется корреляция между значениями показателя Херста и прогнозной точностью, что показывает полезность оценки значения коэффициента Херста при предварительном анализе;
* использование фрактального анализа позволяет повысить эффективность прогнозирования для принятия последующих решений.

Таким образом, применение коэффициента Херста в отношении оценки прогнозного состояния временного ряда по курсам акций и оценки устойчивости его тренда способно улучшить получаемые результаты, но только на краткосрочном временном периоде [130]. Показатели Херста могут использоваться как дополнительные (риск данных по прогнозу) и позволяют повысить надежность прогнозных данных.

1. Программное обеспечение для данного исследования было разработано при выполнении ВКР бакалавром Будановой Н. [↑](#footnote-ref-1)