Моделирование и прогнозирование HFT volatility акций российских компаний и его методы подсчета

Bakhyshov Vakhid *[1], Dugin Maxim †[1], Samoilov Roman ‡[1], and Chasov Nikita§[1]

[1]Moscow State University

December 3, 2024

Abstract

В данной работе рассматривается задача прогнозирования волатильности акций российских компаний с использованием высокочастотных (HFT) данных. Основное внимание уделено моделированию волатильности: от методов расчета HFT-волатильности до её прогнозирования с помощью традиционных моделей (GARCH, FIGARCH, HAR) и других моделей. Анализируется связь прогнозируемой волатильности с объемами торгов, что позволяет построить комплексный подход для улучшения прогнозных характеристик. Основной целью работы является разработка комплекса моделей, включающего преобразование цен в волатильность, прогноз волатильности и оценку связи между волатильностью и объемом. В качестве методов прогнозирования используются ARCH, GARCH, HAR, а также модель машинного обучения: LSTM, XGBoost и Random Forest.

Эмпирическая часть основывается на данных Московской биржи, что подчеркивает практическую значимость исследования.

Keywords: ARCH, GARCH (EGARH, FIGARCH), HAR, Volume Series; LSTM, XGBoost, Random Forest, (Realized) Volatility Estimation, High-Frequency

Introduction

Высокочастотная торговля (HFT) занимает ключевую роль в современной финансовой эконометрике, предоставляя возможность анализа ценовых движений с высокой точностью. Волатильность используется как индикатор рыночного риска, а её точное моделирование необходимо для ценообразования деривативов, управления портфелями и оценки рыночной ликвидности. Одной из ключевых метрик в HFT является волатильность, которая описывает степень изменчивости цен на активы.

Однако стандартные методы, такие как implied volatility, часто оказываются неэффективными и неприменимыми в условиях HFT из-за недостаточной гибкости и точности на микроуровне, что требует разработки альтернативных подходов. Поэтому важно исследовать и адаптировать методы расчета и прогнозирования HFT-волатильности.

Настоящая работа направлена на исследование методов расчета HFT-волатильности, ее прогнозирование и оценку связи с объемами торгов. В рамках проекта решены

*e-mail: vakhid.bakhyshov@math.msu.ru
†e-mail: maksim.dugin@math.msu.ru
‡e-mail: roman.samoilov@math.msu.ru
§e-mail: nikita.chasov@math.msu.ru

следующие задачи:

1. Разработка функции преобразования временного ряда цен в HFT-волатильность.

2. Исследование существующих моделей прогнозирования волатильности (ARCH, GARCH, HAR) и их модификаций.

3. Оценка корреляции между прогнозируемой волатильностью и объемами торгов.

Для достижения поставленных целей использовались данные Московской биржи, содержащие временные ряды цен и объемов.

1 Strategy (Econometric approach)

1.1 ARCH and GARCH

Модели ARCH (Autoregressive Conditional Heteroskedasticity - Engle, 1982) и GARCH (Generalized ARCH - Bollerslev, 1986) являются стандартными инструментами для прогнозирования волатильности. Эти модели учитывают автокорреляцию и кластеризацию волатильности, что делает их подходящими для моделирования финансовых временных рядов.

В рамках исследования использовались модификации GARCH, включая EGARCH (Exponential GARCH) и FIGARCH (Fractional Integrated GARCH), которые учитывают асимметричные эффекты и более сложную динамику. Особое внимание уделялось адаптации моделей для работы с HFT-данными, где высокая частота наблюдений требует учета специфики рыночных движений.

1.2 EGARCH and FIGARCH

EGARCH (Exponential GARCH) модель вводит асимметрию, позволяя учитывать, что отрицательные шоки (падение цены) могут оказывать большее влияние на волатильность, чем положительные. Формула включает логарифм волатильности, что устраняет проблему отрицательных значений.

FIGARCH (Fractionally Integrated GARCH) учитывает долгосрочную память, что делает модель подходящей для финансовых рядов с устойчивой автокорреляцией. Это особенно полезно для HFT-данных, где эффекты прошлых наблюдений могут сохраняться дольше.

1.3 HAR

Модель HAR (Heterogeneous Autoregressive) предлагает альтернативный подход к прогнозированию волатильности, учитывая различные временные горизонты. Она позволяет интегрировать дневные, недельные и месячные компоненты волатильности, что особенно важно для высокочастотных данных. Она проста в реализации, но эффективно захватывает межвременные взаимодействия волатильности. Формула модели имеет вид:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^{w} + \beta_3 RV_{t-1}^{m} + \epsilon_t$$

где RV_t — реализованная волатильность, рассчитанная на основе внутридневных данных. Модель адаптирована для HFT-данных за счет использования различных вариантов вычисления RV, таких как амплитудная и дисперсионная оценки.

В работе была реализована модификация HAR, адаптированная для HFT-волатильности, рассчитанной на основе методов, рассмотренных в предыдущем разделе. Преимущество HAR-модели заключается в ее способности учитывать взаимодействие краткосрочной и долгосрочной волатильности, что может улучшить точность прогнозов.

2 Neural networks and machine learning models

2.1 LSTM (Long Short-Term Memory)

LSTM (Long Short-Term Memory) нейронные сети - это тип рекуррентных нейронных сетей (RNN), которые особенно хорошо подходят для работы с временными рядами, где зависимость между предыдущими и последующими значениями может быть значительной. LSTM используются для предсказания будущих значений, таких как объем торгов, в зависимости от исторических данных. Соотвественно, в нашем проекте мы воспользовались нейронкой LSTM. Казалось бы, а почему мы используем нейронку LSTM, потому хорошо подходят для предсказания объема торгов: 1) Временная зависимость: Объемы торгов в финансовых данных часто являются автокоррелированными, т.е. значения в прошлом влияют на будущие значения. LSTM могут “запомнить” важные исторические данные и использовать их для предсказания; 2) Нелинейность: Рынки часто характеризуются нелинейными зависимостями. LSTM способны учитывать нелинейные паттерны в данных и могут более точно предсказывать объемы, чем линейные модели.

Разберём именно нашу модель:

1. Предобработка данных: Исходный временной ряд нормализуется методом Min-Max для ускорения сходимости модели. Мы разрезаем данные на обучающие и тестовые наборы, а также формируем последовательности данных для LSTM на основе скользящего окна.
2. Архитектура модели: Модель состоит из слоя LSTM и линейного выходного слоя, который преобразует выходные значения из скрытого состояния в прогнозы объема. Параметры модели, такие как количество скрытых узлов, слоев и скорость обучения, настраиваются под каждый тикер на основе лучших параметров.
3. Обучение модели: LSTM обучается минимизации ошибки предсказания (MSE) с использованием оптимизатора Adam. На каждом шаге обновляются веса, чтобы минимизировать разницу между прогнозом и реальным значением объема торгов. Процесс повторяется на нескольких эпохах, и результаты обучения контролируются по значению потерь.
4. Прогнозирование: После обучения модель тестируется на последних доступных данных. Мы подаем последовательности с известными значениями объема, сдвигая окно и добавляя реальные данные, чтобы LSTM могла корректировать прогноз на основе актуальных изменений. Прогнозируемые значения восстанавливаются из нормализованного диапазона.
5. Оценка: После прогноза модель оценивается с использованием корня среднеквадратичной ошибки (RMSE) по последнему отрезку данных, что позволяет количественно оценить точность модели для каждого тикера.

Такой подход позволяет получить надежные прогнозы объема торгов с учетом временных зависимостей в данных.

2.2 XGBoost and Random forest

В рамках задачи предобработки данных и создания прогнозов по объему торгов применяются следующие этапы:

1. Загрузка и предобработка данных: Временной ряд для каждого тикера загружается, включая закрывающую цену и объем торгов. На основе данных о цене закрытия рассчитываются основные технические индикаторы, что помогает выделить тренды и волатильность, а также предсказывать будущие изменения.

2. Создание признаков с техническими индикаторами:** Вычисляются такие индикаторы, как:

   - SMA и EMA (простой и экспоненциальный скользящие средние), чтобы уловить краткосрочные и долгосрочные тенденции.
   - RSI (индекс относительной силы) для измерения перепроданности или перекупленности.
   - MACD (конвергенция и дивергенция скользящих средних) для выявления силы тренда и потенциальных разворотов.
   - Bollinger Bands для оценки волатильности.
   - ROC (скорость изменения), Stochastic Oscillator и Momentum для отслеживания скорости изменений и трендов.
   - DPO (осциллятор детрендированных цен), чтобы устранять долгосрочные тренды и фокусироваться на цикличности.

   После расчета индикаторы объединяются в датафрейм для создания полноценного набора признаков.

3. Выбор признаков: Используется регрессор XGBoost, чтобы определить наиболее значимые признаки для прогнозирования объема торгов. Функция визуализирует важность признаков, что позволяет сфокусироваться на тех, которые имеют наибольшее влияние.

4. Оценка: После прогноза модель оценивается с использованием корня среднеквадратичной ошибки (RMSE) по последнему отрезку данных, что позволяет количественно оценить точность модели для каждого тикера.

   Каждая модель обучается на данных и делает прогноз для тестовой выборки. Результаты предсказаний объединяются в стеке, включающем предсказания каждого из алгоритмов и реальные значения.

Такой подход, объединяющий технические индикаторы и машинное моделирование, позволяет учесть особенности и шум данных, обеспечивая более точные и устойчивые прогнозы.

3 Empirical analysis

3.1 Data

Для анализа были использованы данные о минутных значениях объема торгов и цены закрытия акций по конкретному тикеру.

Для проведения эмпирического анализа использовались данные Московской биржи, которые были преобразованы в 8 отдельных временных рядов (включающие временные ряды цен и объемов торгов), где каждый ряд соответствует 10 минутам в течение торгового дня. Например, файл с именем $GAZP\_volume.csv$ содержит данные об объеме торгов акций «Газпром» за каждые 10 минут торгов (с 10:09 до 10:19) за каждый день в период с 1 июня 2022 года по 26 ноября 2024 года. Такой подход позволил выделить особенности временных рядов для каждых 10 минут отдельно.

Данные были очищены от выбросов, синхронизированы по времени и разбиты на тренировочные и тестовые выборки.

Таблицу с основными характеристиками можно видеть ниже **??**

Чтобы применить методы машинного обучения, изначальный временной ряд, включающий данные об объеме торгов за доступные торговые часы, был сохранен в полном объеме.

3.2 Methods

Преобразование цен в HFT-волатильность осуществлялось с использованием следующих методов:

1. Реализованная волатильность RV_t (Realized Volatility) за день, рассчитывается как сумма квадратов внутридневных изменений цены:

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

где $r_{t,j}$ - логарифмическая доходность для периода j для t

2. Хотя оцениватель максимального правдоподобия, приведённый выше, является состоятельным при предположении заданного высокочастотного ценового процесса, он смещён и несостоятелен при более общих свойствах цен. Barndorff-Nielsen et al. (2008a) предложили оцениватель реализованного ядра, который является состоятельным при более общих условиях, таких как, например, зависимости более высокого порядка или эндогенность в шумовом процессе. Идея состоит в том, чтобы учитывать (возможно, шум-индуцированные) серийные корреляции в доходностях сделка-за-сделкой с помощью ядра. Оцениватель реализованного ядра (RK), предложенный Barndorff-Nielsen et al. (2008a), определяется как

$$K(\Delta) = \gamma_0(\Delta) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left\{\gamma_h(\Delta) + \gamma_{-h}(\Delta)\right\},$$

где $\gamma_h(\Delta)$ обозначает h-ую реализованную автоковариацию, определяемую как

$$\gamma_h(\Delta) = \sum_{j=1}^{n}\left(p_{i\Delta_n} - p_{(i-1)\Delta_n}\right)\left(p_{(i-h)\Delta_n} - p_{(i-h-1)\Delta_n}\right),$$

где $h = -H, \ldots, -1, 0, 1, \ldots, H$, а $k(\cdot)$ — это функция ядра, зависящая от ширины полосы H. Barndorff-Nielsen et al. (2008a) предложили использовать ядро Tukey-Hanning $_2$ с

$$k(x) = \sin^2\left\{\pi/2(1-x)^2\right\}.$$

Оптимальный выбор ширины полосы H определяется как

$$H = c\zeta\sqrt{n},$$

где $c = 5.74$ для ядра Tukey-Hanning $_2$, а ζ задаётся как

$$\zeta^2 = \omega^2 / \sqrt{\int_0^1 \sigma_u^4 du}.$$

Величина $\int_0^1 \sigma_u^4 du$ может быть оценена с использованием реализованной квартильности. Для оценки ω^2 Barndorff-Nielsen et al. (2008b) предложили

$$\hat{\omega}^2 = \frac{1}{q} \sum_{i=1}^{q} \hat{\omega}_{(i)}^2,$$

где

$$\hat{\omega}_{(i)}^2 = \frac{RV_{(i)}^n}{2\tilde{n}_{(i)}}, \quad i = 1, \ldots, q,$$

а $RV_{(i)}^n$, $i = 1, \ldots, q$ — это оцениватели реализованной дисперсии $RV_{(i)}^n = \sum_{j=i}^n r_{j\Delta,n}^2$, выборочно учитывающие каждый $q = N/n$-ый шаг с использованием первых q сделок в день в качестве различных стартовых точек, где N — это число транзакций в $[0,1]$, а $\tilde{n}_{(i)}$ — количество ненулевых доходностей, использованных для вычисления $RV_{(i)}^n$. Для повышения устойчивости оценивателя к серийной зависимости в шумовом процессе Barndorff-Nielsen et al. (2008b) предложили выбрать q так, чтобы каждый q-ый шаг в среднем составлял 2 минуты. Этот оцениватель, вероятно, имеет смещение вверх и, таким образом, даёт довольно консервативный выбор ширины полосы. Для получения дополнительной информации см. Barndorff-Nielsen et al. (2008a,b).

3. Принцип, лежащий в основе предварительно усредняющего оценивателя, предложенного Jacod et al. (2009), заключается, грубо говоря, в устранении рыночного микроструктурного шума путём локального усреднения высокочастотных доходностей перед их возведением в квадрат и суммированием. Оцениватель строится путём выбора последовательности k_n целых чисел, удовлетворяющих

$$k_n \Delta^{1/2} = \theta + o\left(\Delta^{\frac{1}{4}}\right),$$

для некоторого $\theta > 0$ и ненулевой вещественной функции $g : [0,1] \to \mathbb{R}$, которая является непрерывной, кусочно непрерывно дифференцируемой, и $g(0) = g(1) = 0$. Типичный пример функции g, используемой в эмпирической части, задаётся как $g(x) = \min\{x, (1-x)\}$ для $x \in (0,1)$. Предварительно усреднённые доходности задаются как

$$\bar{Z}_i^n := \sum_{j=1}^{k_n} g\left(\frac{j}{k_n}\right) r_{(i+j)\Delta,n}, \quad r_{i\Delta,n} = p_{i\Delta} - p_{(i-1)\Delta}.$$

Таким образом, $\bar{Z}_i^n$ соответствует взвешенному среднему приростов $r_{i\Delta,n}$ в локальном окне $[i\Delta, (i+k_n)\Delta]$, что снижает влияние шума. Размер окна k_n выбирается порядка $\Delta_n^{-1/2} = n^{1/2}$, что обеспечивает оптимальные скорости сходимости, см. Jacod et al. (2009).

Аналогом оценивателя реализованной дисперсии является

$$V(Z,n) = \sum_{i=0}^{[1/\Delta]-k_n} \left|\bar{Z}_i^n\right|^2,$$

что даёт предварительно усредняющий оцениватель

$$C^n := \frac{\sqrt{\Delta}}{\theta\psi_2} V(Z,n) - \frac{\psi_1 \Delta}{2\theta^2\psi_2} RV^n,$$

где $\psi_1 := \int_0^1 \left(g'(s)\right)^2 ds = 1$, а $\psi_2 := \int_0^1 (g(s))^2 ds = 1/12$. Jacod et al. (2009) показали, что

$$\frac{\Delta}{2} RV^n \approx \int_0^1 \omega_u^2 du + \frac{\Delta}{2} IV.$$

Ошибка этого приближения имеет порядок Δ и математическое ожидание 0. Следовательно, статистика C^n фактически оценивает

$$\left(1 - \frac{\psi_1^{k_n}\Delta}{2\theta^2\psi_2^{k_n}}\right) IV.$$

Для конечных выборок необходимо скорректировать истинное количество слагаемых в $V(Z, n)$, что приводит к корректировочному термину $[1/\Delta]/\left([1/\Delta] - k_n + 1\right)$. Это даёт предварительно усредняющий оцениватель для конечной выборки

$$C_a^n = \left(1 - \frac{\psi_1^{k_n}\Delta}{2\theta^2\psi_2^{k_n}}\right)^{-1} \left(\frac{[1/\Delta]\sqrt{\Delta}}{\left([1/\Delta] - k_n + 2\right)\theta\psi_2^{k_n}} V(Z, n) - \frac{\psi_1^{k_n}\Delta_n}{2\theta^2\psi_2^{k_n}} RV^n\right),$$

где $\psi_1^{k_n}$ и $\psi_2^{k_n}$ обозначают конечновыборочные аналоги ψ_1 и ψ_2, определяемые как

$$\psi_1^{k_n} = k_n \sum_{j=1}^{k_n} \left(g\left(\frac{j+1}{k_n}\right) - g\left(\frac{j}{k_n}\right)\right)^2,$$

$$\psi_2^{k_n} = \frac{1}{k_n} \sum_{j=1}^{k_n - 1} g^2\left(\frac{j}{k_n}\right).$$

Jacod et al. (2009) показали, что оцениватель является состоятельным и асимптотически нормально распределённым в смешанном виде. Его можно расширить в различных направлениях, например, с учётом серийно зависимого шума и скачков в базовом ценовом процессе. Hautsch и Podolskij (2010) проанализировали зависимость оценивателя от параметра предварительного усреднения θ и предложили минимизирующий MSE выбор на основе данных.

3.3 Initial models

Мы рассматривали случайный лес, градиентный бустинг и $LSTM$.

Их метрики можно видеть в первой таблице по колонкам $rmse\_lstm$ (Корень среднеквадратической ошибки (RMSE) модели LSTM), $rmse\_fe$ (Корень среднеквадратической ошибки (RMSE) модели PC-ARMA), $rmse\_rf$ (Корень среднеквадратической ошибки (RMSE) модели случайного леса (RF)), $rmse\_gb$ (Корень среднеквадратической ошибки (RMSE) модели градиентного бустинга (GB)).

3.4 Creating the model

Модель сначала преобразует данные в hft волатильность, после чего на основе этого временного ряда при помощи ансамблевского метода на основе эконометрических моделей(GARCH и HAR) и ml (описанных выше) прогнозируется волатильность и на основе этого делает предсказание hft волатильности.

3.4.1 Cross-validation

Для эконометрических моделей(GARCH и HAR) реализована кросс-валидация с расширяющимся окном; первая обучающая выборка - дни от первого до двадцать первого с конца, последняя - все дни, представленные в датасете, кроме последнего (то есть окно 20 раз расширялось на один день). Тестовые данные - это следующий день за тренировочными.

Для моделей машинного обучения использовалась кросс-валидация со скользящим окном **??**: размер окна является гиперпараметром в модели LSTM, в остальных моделях первая обучающая выборка - дни от первого до двести восьмидесятого с конца, последняя - дни от 280-го до 14-го с конца (то есть окно 20 раз двигалось на 14 дней).

3.4.2 Optimization of hyperparametrs

Для модели PC-ARMA был выбран размер временного окна так, чтобы метрика MSE, усреднённая по всем акциям и батчам, была минимальной.

Для моделей машинного обучения выбор гиперпараметров реализован с помощью библиотеки optuna.

3.4.3 Stacking

Описать реализацию стейкинга можно в простых шагов:

1. Расчет ошибок: Для каждой модели (LSTM, PC-ARMA, Random Forest, Gradient Boosting) были рассчитаны абсолютные ошибки прогнозов по отношению к фактическому объему торгов. Ошибки определяют, насколько сильно каждая модель отклоняется от истинных значений.
2. Веса на основе ошибок: Для каждой модели был рассчитан вес, обратно пропорциональный ошибке, что позволяет более точным моделям оказывать большее влияние на итоговый прогноз. Вес для каждой модели определялся как:

$$weight = \frac{1}{error}$$

3. Нормализация весов: Чтобы суммарный вес всех моделей был равен единице, веса нормализовались, делением веса каждой модели на сумму всех весов. Это позволяло создать сбалансированный ансамбль, где каждая модель учитывалась пропорционально своей точности.
4. Объединенный прогноз: После нормализации весов итоговый прогноз рассчитывался как взвешенная сумма прогнозов всех моделей, что создавало метапрогноз. Формула комбинированного прогноза выглядела следующим образом:

$$combined\_forecast = (\text{norm.weight LSTM} \times \text{forecast LSTM}) +$$
$$+ (\text{norm.weight PC-ARMA} \times \text{forecast PC-ARMA}) +$$
$$+ (\text{norm.weight RF} \times \text{forecast RF}) +$$
$$+ (\text{norm.weight GB} \times \text{forecast GB})$$

5. Оценка точности объединенной модели: Для анализа точности метапрогноза были рассчитаны метрики: MSE, RMSE, MAE и MAPE, а также RMSE для каждой из моделей и комбинированного прогноза.

3.5 Results

Результаты применения можно видеть на графиках в анендиксе.

Conclusion

1. Методы моделирования hft волатильности выделяют закономерности и патерны в данных. Графики волатильности, посчитанной методами и график исторической волатильности похожи.

2. Разработанная модель прогнозирования волатильности представляет собой ансамбль, объединяющий методы машинного обучения (Random Forest, Gradient Boosting, Linear Regression) и эконометрическую модель ARIMA. Такой подход позволяет учесть как нелинейные зависимости, так и временные структуры данных.

Проведённый сравнительный анализ показал, что комбинированная модель превосходит классическую модель GARCH по метрике RMSE для всех исследуемых тикеров. Например, для тикера NVTK RMSE GARCH составляет 0.001230, в то время как комбинированная модель достигает значительно меньшего значения RMSE — 0.000714. Это демонстрирует, что использование ансамбля моделей позволяет достичь более высокой точности и устойчивости прогнозов.

Основное преимущество разработанного подхода заключается в его гибкости. Машинное обучение эффективно адаптируется к нелинейностям и аномалиям, а ARIMA позволяет корректно моделировать временные зависимости. Объединение прогнозов с помощью ансамбля позволило минимизировать ошибки отдельных моделей, улучшив качество итогового прогноза.

Разработанная модель может быть полезной для задач управления рисками и стратегий торговли, где точное прогнозирование волатильности играет ключевую роль. В будущем планируется дальнейшая оптимизация ансамбля и исследование влияния дополнительных факторов на точность прогнозов.

References

[YZ08] Per A. Mykland Yacine Aït-sahalia and Lan Zhang. *Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise*. 2008.

[EP24] Greeshma Balabhadra El Mehdi Ainasse and Pawel Polak. “High-Frequency Volatility Estimation with Fast Multiple Change Points Detection”. In: (2024).

Appendix A Github

Весь код опубликован у нас на Github: Forecasting Trading Volume

Помимо этого там находятся все расчёты дополнительных статистик как для моделей, используемых в этой статье, так и для моделей, которые в это статью не были включены из-за их плохих предсказательных способностей.

Appendix B Measures to assess the predictive capabilities of models

Средняя абсолютная ошибка (Mean Absolute Error, MAE): $MAE = \frac{1}{H}\sum_{t=1}^{H}|y_t - \hat{y}_t|$

Среднеквадратическая ошибка (Mean Square Error, MSE): $MSE = \frac{1}{H}\sum_{t=1}^{H}(y_t - \hat{y}_t)^2$

Корень из среднеквадратической ошибки (RMSE): $RMSE = \sqrt{\frac{1}{H}\sum_{t=1}^{H}(y_t - \hat{y}_t)^2}$

Средняя абсолютная процентная ошибка (Mean Absolute Percentage Error, MAPE):

$$MAPE = 100\% \cdot \frac{1}{H}\sum_{t=1}^{H}|\frac{y_t - \hat{y}_t}{y_t}|$$

Средняя абсолютная шкалированная ошибка (Mean Absolute Scaled Error, MASE):

$$MASE = \frac{\frac{1}{T}\sum_{t=1}^{H}(y_t - \hat{y}_t)^2}{\frac{1}{T-1}\sum_{t=2}^{H}(y_t - \hat{y_{t-1}})^2}$$

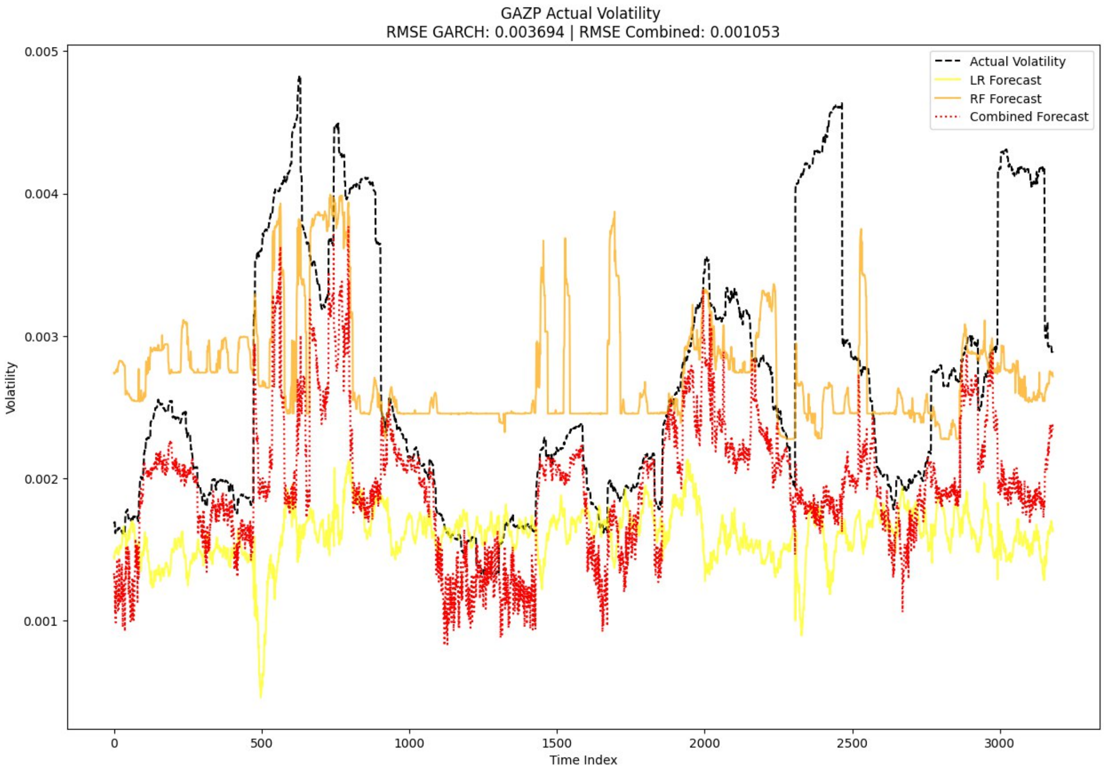
Appendix C Results

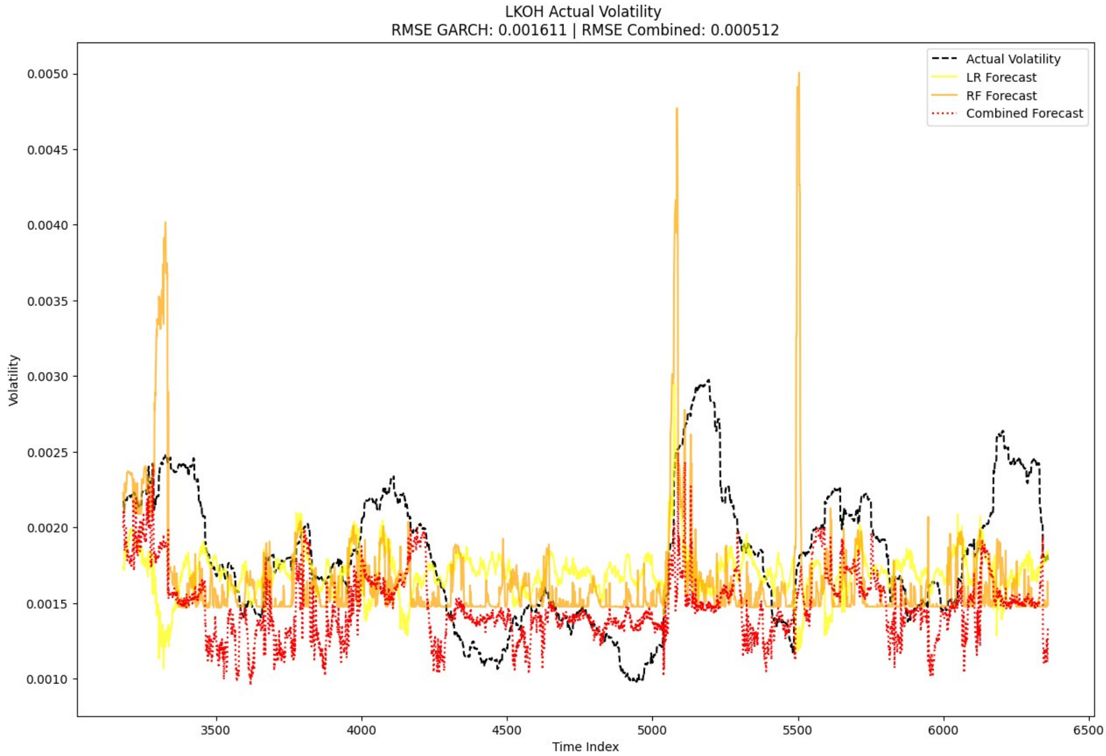
Ticker	MSE	RMSE	MAE	MAPE	RMSE LR	RMSE LSTM
GAZP	1.11×10^{-6}	1.05×10^{-3}	7.29×10^{-4}	23.24%	1.42×10^{-3}	1.42×10^{-3}
LKOH	2.63×10^{-7}	5.12×10^{-4}	4.00×10^{-4}	20.56%	4.89×10^{-4}	4.89×10^{-4}
NVTK	5.10×10^{-7}	7.14×10^{-4}	6.04×10^{-4}	25.75%	7.86×10^{-4}	7.86×10^{-4}
PLZL	1.74×10^{-7}	4.17×10^{-4}	3.08×10^{-4}	13.54%	7.84×10^{-4}	7.84×10^{-4}
ROSN	5.74×10^{-7}	7.58×10^{-4}	6.20×10^{-4}	25.04%	6.29×10^{-4}	6.29×10^{-4}
SBER	4.22×10^{-7}	6.50×10^{-4}	5.24×10^{-4}	25.34%	5.24×10^{-4}	5.24×10^{-4}
SIBN	1.21×10^{-6}	1.10×10^{-3}	6.70×10^{-4}	26.60%	1.00×10^{-3}	1.00×10^{-3}
TATN	1.06×10^{-6}	1.03×10^{-3}	7.70×10^{-4}	27.06%	9.36×10^{-4}	9.36×10^{-4}

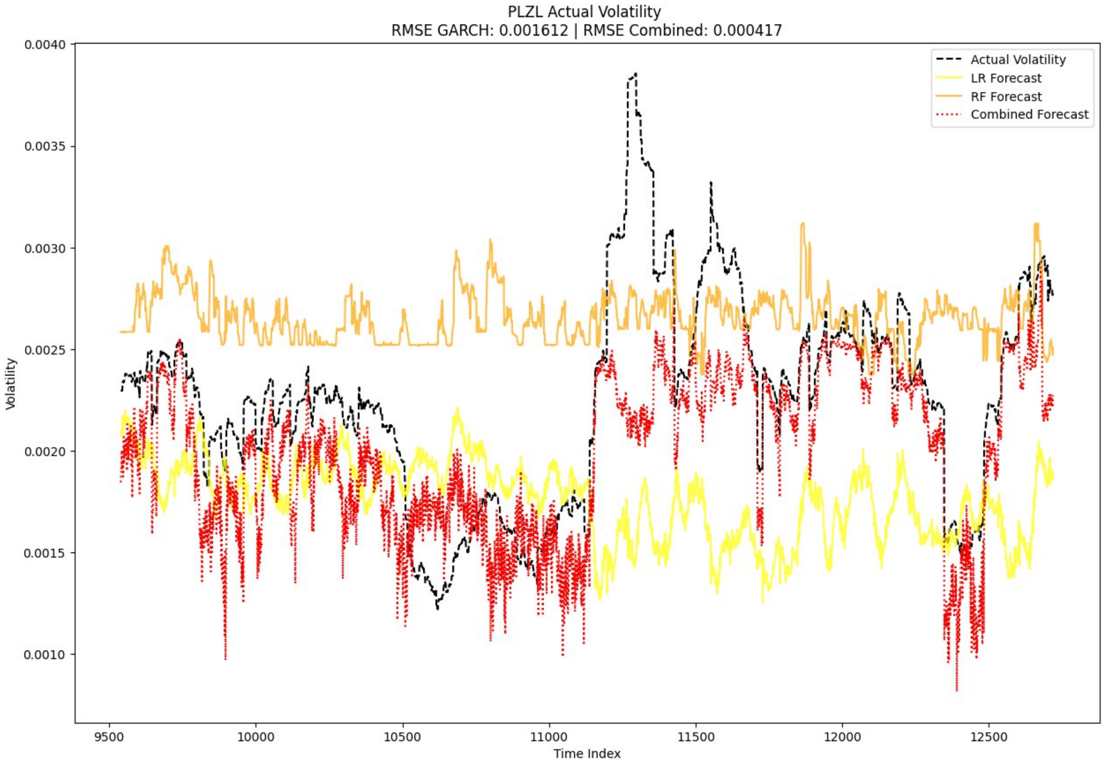
Table C.1: Метрики прогнозов: RMSE, MAE, и MAPE для моделей LSTM и LR

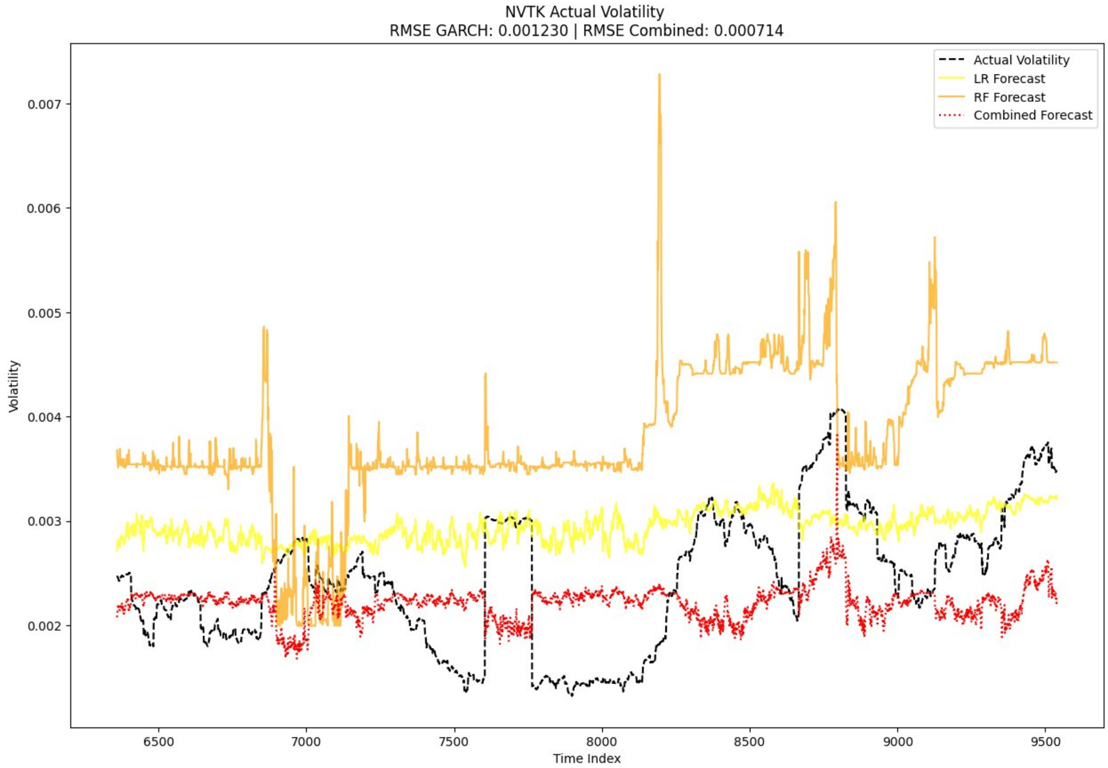
Ticker	RMSE RF	RMSE GB	RMSE GARCH	RMSE Combined
GAZP	8.55×10^{-4}	1.04×10^{-3}	3.69×10^{-3}	1.05×10^{-3}
LKOH	5.30×10^{-4}	4.85×10^{-4}	1.61×10^{-3}	5.12×10^{-4}
NVTK	1.57×10^{-3}	6.67×10^{-4}	1.23×10^{-3}	7.14×10^{-4}
PLZL	6.44×10^{-4}	5.88×10^{-4}	1.61×10^{-3}	4.17×10^{-4}
ROSN	5.44×10^{-4}	5.82×10^{-4}	1.19×10^{-3}	7.58×10^{-4}
SBER	6.80×10^{-4}	4.90×10^{-4}	1.16×10^{-3}	6.50×10^{-4}
SIBN	1.11×10^{-3}	1.06×10^{-3}	2.02×10^{-3}	1.10×10^{-3}
TATN	7.96×10^{-4}	9.24×10^{-4}	1.20×10^{-3}	1.03×10^{-3}

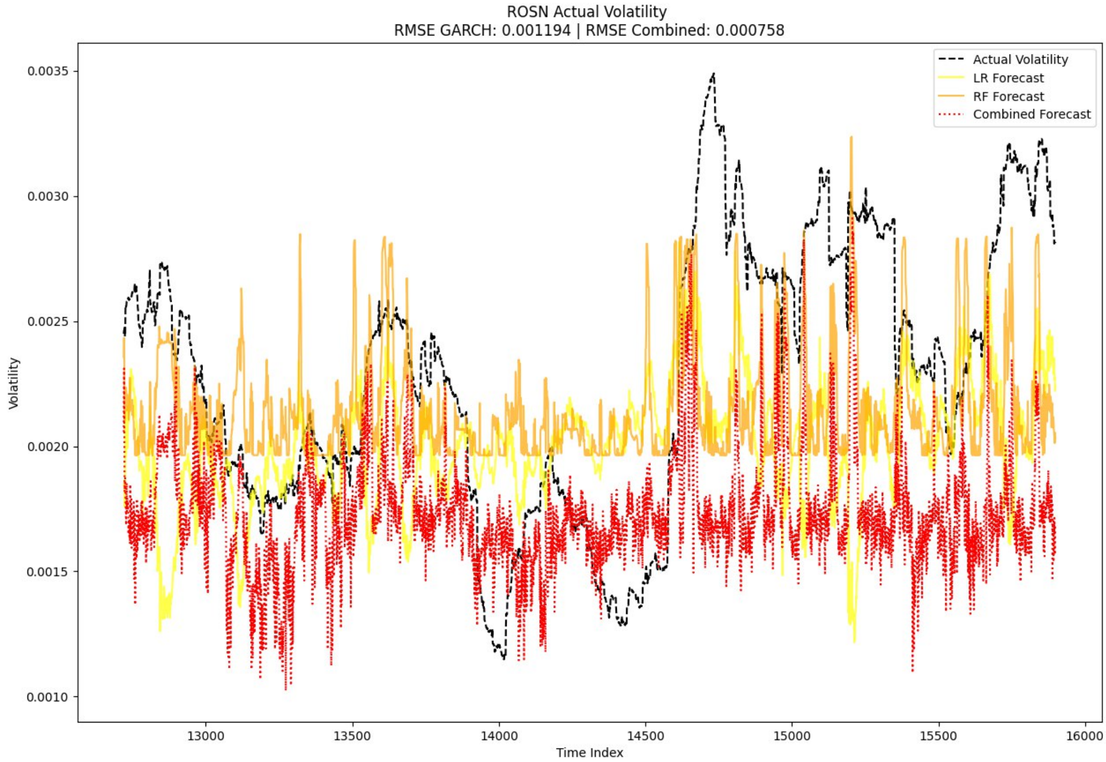
Table C.2: Сравнение RMSE для различных моделей

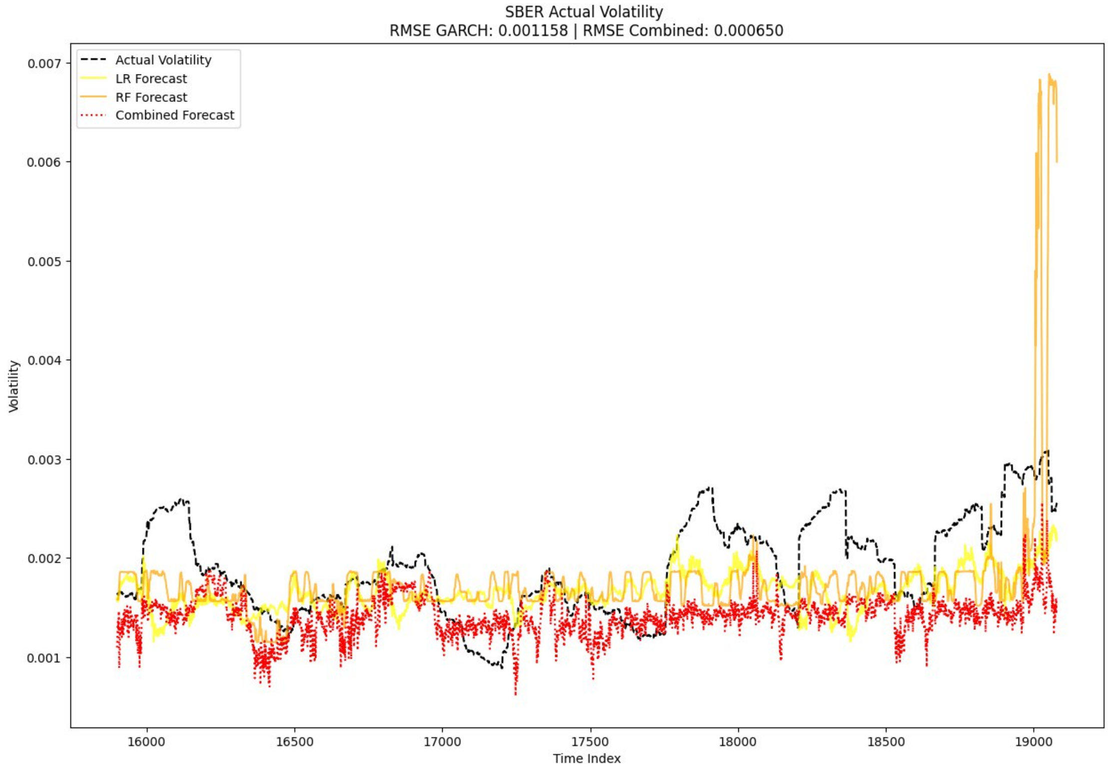


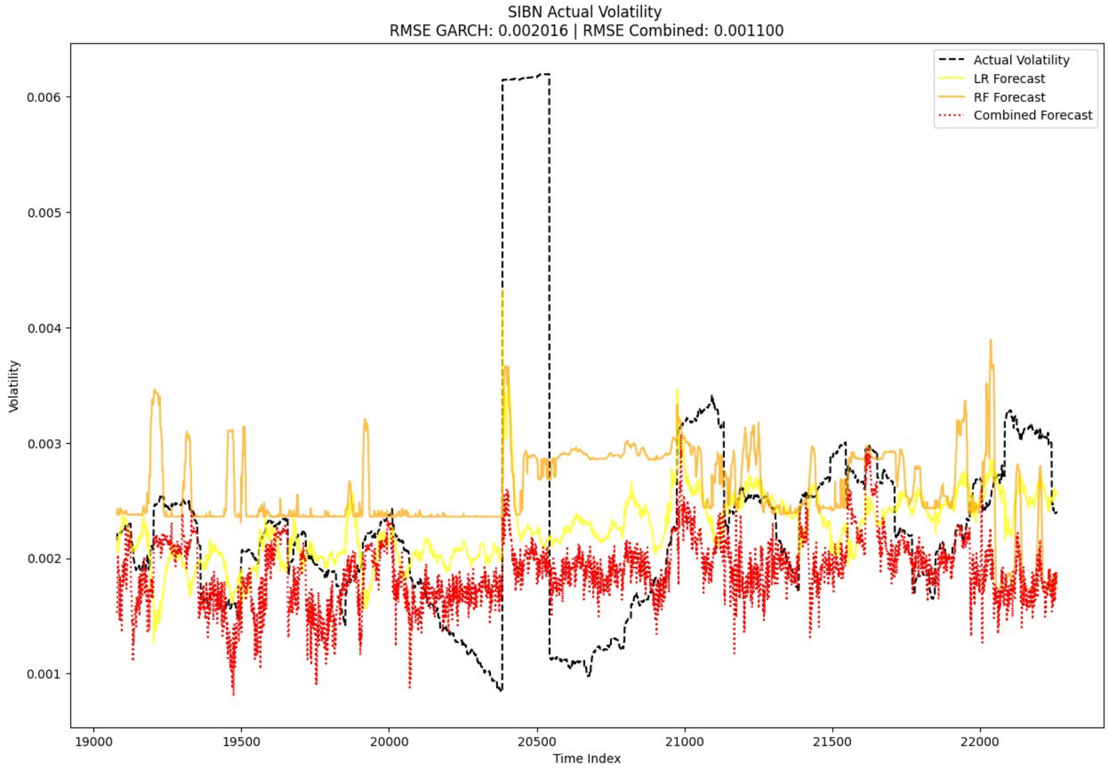

LKOH Actual Volatility
RMSE GARCH: 0.001611 | RMSE Combined: 0.000512
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
Time Index
0.0010
0.0015
0.0020
0.0025
0.0030
0.0035
0.0040
0.0045
0.0050
3500
4000
4500
5000
5500
6000
6500

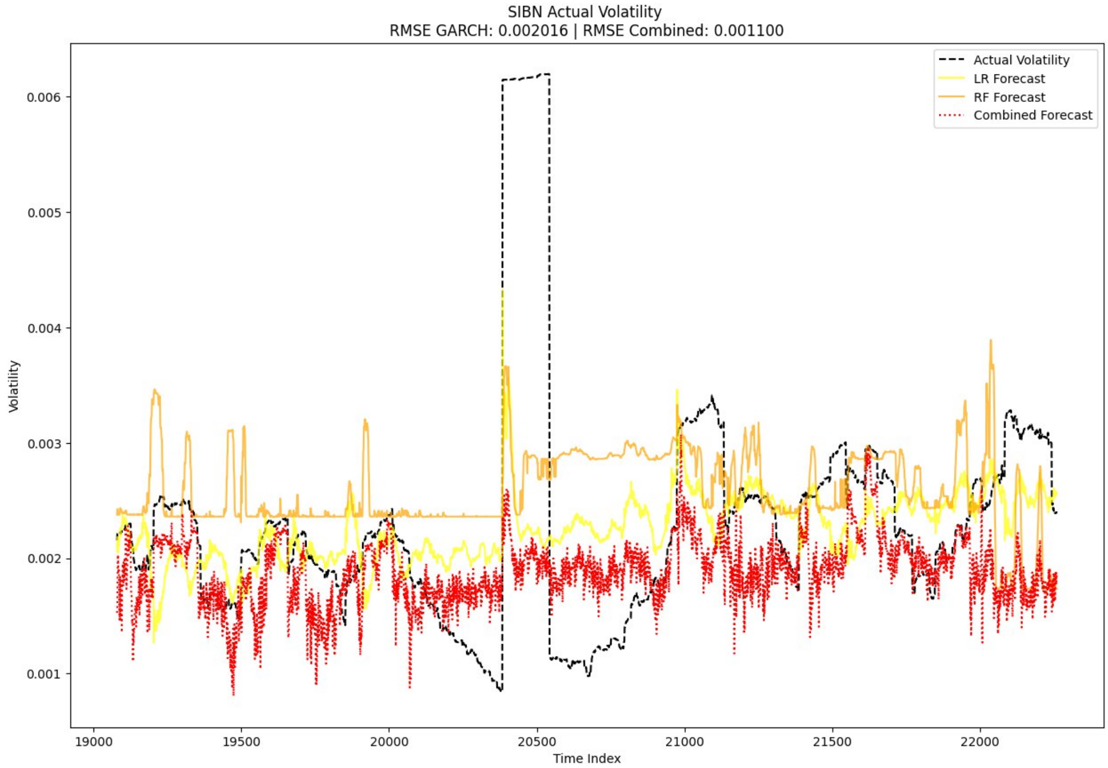

PLZL Actual Volatility
RMSE GARCH: 0.001612 | RMSE Combined: 0.000417
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
Time Index
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
9500
10000
10500
11000
11500
12000
12500


NVTK Actual Volatility
RMSE GARCH: 0.001230 | RMSE Combined: 0.000714
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
0.007
0.006
0.005
0.004
0.003
0.002
Time Index
6500
7000
7500
8000
8500
9000
9500


ROSN Actual Volatility
RMSE GARCH: 0.001194 | RMSE Combined: 0.000758
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
13000
13500
14000
14500
15000
15500
16000
Time Index


SBER Actual Volatility
RMSE GARCH: 0.001158 | RMSE Combined: 0.000650
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
0.007
0.006
0.005
0.004
0.003
0.002
0.001
16000
16500
17000
17500
18000
18500
19000
Time Index


SIBN Actual Volatility
RMSE GARCH: 0.002016 | RMSE Combined: 0.001100
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
Time Index
0.001
0.002
0.003
0.004
0.005
0.006
19000
19500
20000
20500
21000
21500
22000


SIBN Actual Volatility
RMSE GARCH: 0.002016 | RMSE Combined: 0.001100
Actual Volatility
LR Forecast
RF Forecast
Combined Forecast
Volatility
Time Index
0.001
0.002
0.003
0.004
0.005
0.006
19000
19500
20000
20500
21000
21500
22000