

38. ПРОГНОЗИРОВАНИЕ ЦЕНЫ БАЗОВОГО АКТИВА НА ОСНОВЕ S&P 500

Сацута Д.В., студент гр.172303, Голда О.А., доцент каф.ПОИТ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Голда О. А.– канд. экон. наук

Аннотация. В статье рассматривается прогнозирование цены базового актива на основе индекса S&P 500 с использованием эконометрических моделей. Изучаются модели семейства GARCH для анализа волатильности и прогнозирования доходности.

Ключевые слова. Прогнозирование цены, S&P 500, эконометрические модели, GARCH, волатильность, ARMA, асимметричные модели, логарифмическая доходность, ARCH-эффект, EGARCH.

Прогнозирование цен базовых активов, таких как индекс S&P 500, остается ключевой задачей в финансовой аналитике. Современные исследования сочетают традиционные эконометрические подходы с методами машинного обучения, что позволяет повысить точность прогнозов и адаптироваться к изменяющимся рыночным условиям [1].

Модель ARIMA успешно применяется для анализа временных рядов S&P 500. Исследование Dun [2] демонстрирует, что стационарность ряда достигается после дифференцирования первого порядка, а наилучшая спецификация модели (ARIMA(1,1,1)) обеспечивает точность прогноза на 30 дней с ошибкой менее 2%. Автор подчеркивает, что сезонные компоненты индекса минимальны, что упрощает моделирование.

Федеральная резервная система США (2024) разработала макроэконометрическую модель, объясняющую 25-45% дисперсии прибыли S&P 500 через ВВП, инфляцию зарплат и курс доллара. Расхождения между прогнозами аналитиков и моделью позволяют предсказывать ошибки с точностью 75% [3].

Сравнение LSTM и градиентного бустинга (GB) показало преимущество последнего: RMSE GB на 15% ниже, а портфели на его основе приносят на 3.2% больше годовой доходности. Включение технических индикаторов (RSI, MACD) улучшает точность обоих алгоритмов, тогда как PCA увеличивает ошибку прогноза [4].

Прогнозирование волатильности S&P 500 с использованием асимметричных GARCH-моделей демонстрирует высокую эффективность в условиях рыночных шоков. Исследование Nagendra Marisetty (2024) на данных за 2004–2023 гг. выявило, что TGARCH оптимально фиксирует леввередж-эффект для FTSE 100 и Hang Seng, в то время как APARCH лучше адаптируется к S&P 500 благодаря чувствительности к прошлым шокам. Автор подчеркивает, что асимметричные модели (EGARCH, GJR-GARCH) на 18–25% точнее предсказывают экстремальную волатильность в периоды кризисов, таких как 2008 г. и пандемия COVID-19, за счет учета разной реакции рынка на позитивные и негативные новости [5].

В качестве объекта исследования в данной работе используется индекс, рассчитываемых компанией Standart & Poor's. Данные для анализа представляют собой совокупность всех ежедневных биржевых характеристик акций этих компаний начиная с 2010 г по сегодняшний день.

Таблица 1 – Идентификация переменных

Идентификатор	Описание	Классификация по характеру
Data	Дата предоставления информации по акции	Качественная
Symbol	Символ акций на бирже	Количественная
Adj Close	Скорректированная цена закрытия	Количественная
Close	Цена закрытия	Количественная
High	Пик цены акции	Количественная
Low	Минимум цены акции	Количественная
Open	Цена открытия	Количественная

Проанализировав какие данные имеются в наборе за рассматриваемый период, можно сделать вывод о том, что данные необходимо предварительно подвергнуть обработке, поскольку цена базового актива будет формироваться из ежедневной скорректированной цены закрытия по всем акциям входящим в индекс и будет представлять собой временной ряд. Для обработки и анализа данных будет использован Python, так как он располагает специализированным набором библиотек и инструментов.

Осуществим загрузку данных с исходного файла «sp500_stocks.csv» с помощью библиотеки «pandas» (рисунок 1).

```
[ ] import pandas as pd

df = pd.read_csv('/content/drive/MyDrive/S&P500/sp500_stocks.csv', delimiter=',', header=0)

[ ] df
```

	Date	Symbol	Adj Close	Close	High	Low	Open	Volume
0	2010-01-04	MMM	40.835052	69.414719	69.774246	69.122070	69.473244	3640265.0
1	2010-01-05	MMM	40.579269	68.979935	69.590302	68.311035	69.230766	3405012.0
2	2010-01-06	MMM	41.154747	69.958191	70.735786	69.824417	70.133781	6301126.0
3	2010-01-07	MMM	41.184269	70.008362	70.033447	68.662209	69.665550	5346240.0
4	2010-01-08	MMM	41.474464	70.501671	70.501671	69.648827	69.974915	4073337.0
...
1813813	2024-04-25	ZTS	153.360001	153.360001	153.589996	150.039993	150.970001	4567200.0
1813814	2024-04-26	ZTS	158.419998	158.419998	158.740005	152.630005	152.630005	5236500.0
1813815	2024-04-29	ZTS	159.699997	159.699997	165.039993	158.419998	158.419998	4849500.0
1813816	2024-04-30	ZTS	159.240005	159.240005	162.130005	158.919998	159.639999	3647900.0
1813817	2024-05-01	ZTS	158.500000	158.500000	160.779999	158.039993	158.169998	3771661.0

1813818 rows x 8 columns

Рисунок 1 – Загрузка исходных данных

Проведем усечение выборки, для получения информации исключительно за последние 10 лет. Выборка была проверена на наличие пропусков, которые были устранены. Наибольший интерес для прогнозирования представляет ряд цены закрытия, поэтому целесообразно для моделирования использовать именно указанный ряд, который и будет являться объектом анализа.

Для проведения исследования была получена логарифмическая дневная доходность отраслевых индексов, которая рассчитывается по следующей формуле [6]:

$$r_t = \ln[P_t / P_{(t-1)}] , \quad (1)$$

График, отражающий дневную логарифмическую доходность в разрезе рассматриваемого десятилетия приведен на рисунке 2.

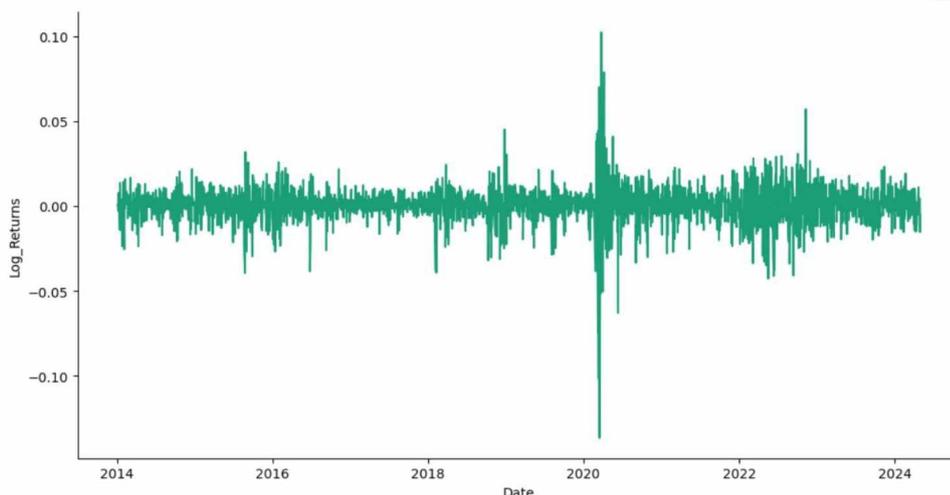


Рисунок 2.6 – Динамика доходности S&P 500 за период с 2010 года по 2024 год

На графике видно, что в течении рассматриваемого периода наблюдались колебания с примерной амплитудой около 1.5% в положительную и отрицательную сторону.

После получения описательных статистик ряда особое внимание стоит уделить показателям асимметрии и эксцесса. Показатель асимметрии составил -0,92 что говорит о наличии асимметрии, наблюдается склонность к более частым крупным падениям цены по сравнению с ростами, что отражает скошенность распределения влево. Показатель эксцесса, как и свойственно финансовым временным рядам, сильно превышает нормальное значение, равное 3, а именно имеет значение 17.9358. Можно утверждать, что распределение ряда доходности из выборки не подчиняется нормальному закону распределения. В этой связи, в дальнейшем анализе будет уделено особое внимание выбору подходящего распределения. Проверим ряд на стационарность при помощи расширенного теста Дики-Фуллера (рисунок 3) [7].



Рисунок 3 – Результат ADF теста

Исходя из результатов теста, можно сделать вывод о том, что по всем параметрам нулевая гипотеза отвергается на любом уровне значимости, что говорит о стационарности рядов доходности фондового индекса S&P 500.

Проведем разложение временного ряда на детерминированные составляющие (рисунок 4).

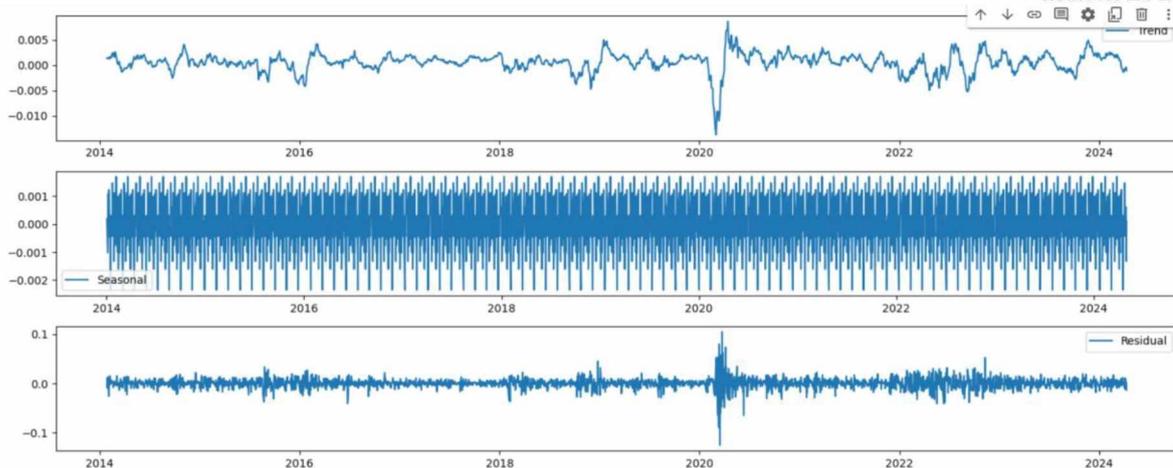


Рисунок 4 – Визуализация разложения

Анализируя первый график можно сделать вывод, что в данном временном ряду отсутствует тренд. Изучив второй график, можно заметить, что при разложении ряда четко выделяется сезонность. Сезонность обычно предполагает наличие повторяющихся паттернов в данных за определенный период времени. В контексте стационарных временных рядов, сезонность может присутствовать, но только если она постоянна во времени, то есть амплитуда и частота сезонных колебаний не меняются что мы и наблюдаем на графике. Третий график говорит о стационарности остатков ряда.

Переходя к моделированию, первым этапом будет являться подбор подходящей спецификации, который заключается в анализе автокорреляционной функции (ACF) и частной автокорреляционной функции (PACF), с помощью которых определяют наиболее значимые лаги (рис.5). Корреллограмма строится для индекса.

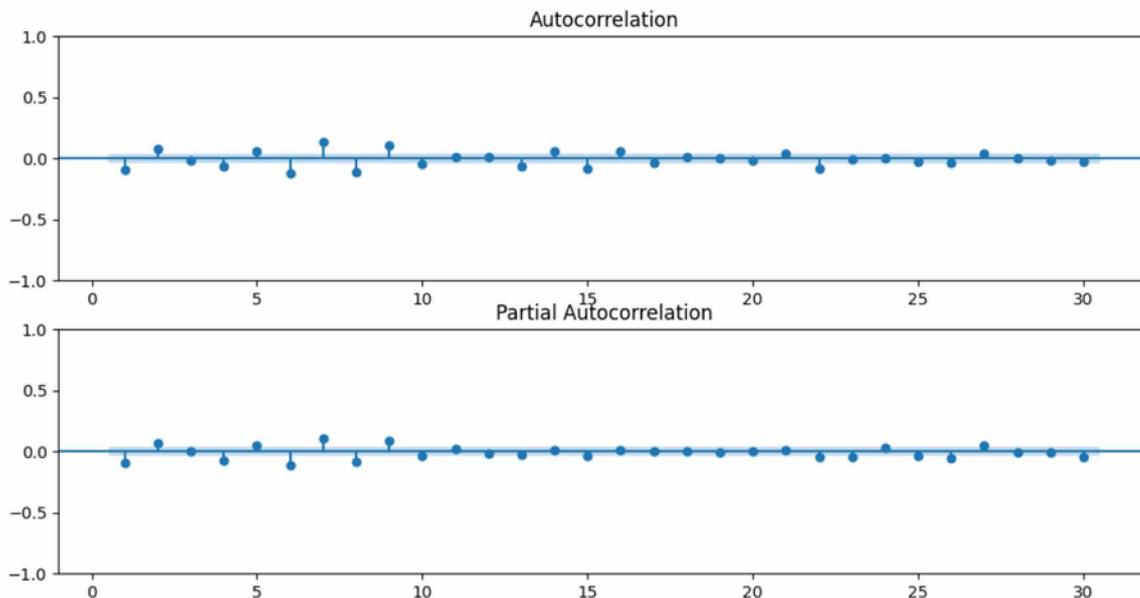


Рисунок 5 – Графики ACF и PACF

Для того чтобы выбрать лучшую модель ARMA, было принято, что наилучшей признавалась модель с наименьшими относительными значениями информационных критериев Акаике (AIC), Шварца (BIC) и Ханна Квина (HQ) и наивысшим показателем функции правдоподобия (Log likelihood).

При дальнейшем построении моделей было выявлено, что главные претенденты на лучшую модель AR(5), MA(5), AR(6), MA(6) оказались не самыми лучшими вариантами, поскольку включали статистически незначимые параметры, а основные показатели, по которым принимается решение о выборе лучшей модели, оставляли желать лучшего. Однако была замечена одна закономерность при построении этих моделей, которая указывала, что необходимо строить комбинированную модель

ARMA(1,1), что и было сделано. Данная модель показала наилучшие результаты по выбранным критериям среди всех (рисунок 6).

SARIMAX Results						
Dep. Variable:	Log>Returns	No. Observations:	2599			
Model:	ARIMA(1, 0, 1)	Log Likelihood	7982.006			
Date:	Sat, 04 May 2024	AIC	-15956.011			
Time:	13:02:44	BIC	-15932.560			
Sample:	0	HQIC	-15947.514			
	- 2599					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
const	0.0005	0.000	2.437	0.015	0.000	0.001
ar.L1	-0.5740	0.047	-12.095	0.000	-0.667	-0.481
ma.L1	0.4871	0.051	9.523	0.000	0.387	0.587
sigma2	0.0001	1.43e-06	87.750	0.000	0.000	0.000
Ljung-Box (L1) (Q):		0.03	Jarque-Bera (JB):	30531.88		
Prob(Q):		0.86	Prob(JB):	0.00		
Heteroskedasticity (H):		1.72	Skew:	-0.91		
Prob(H) (two-sided):		0.00	Kurtosis:	19.69		

Рисунок 6 – Сводное представление модели ARMA(1,1)

Анализируя результаты модели ARMA(1,1), она же модель SARIMAX(ARIMA(1,0,1)) для выбранного временного ряда, можно сделать следующие выводы:

- 1 Для константы ρ -значение составляет 0.015, что указывает на её статистическую значимость.
- 2 Коэффициенты AR(1) и MA(1) являются статистически значимыми с ρ -значениями 0.000 для каждого, что указывает на то, что они важны для модели.
- 3 Значение σ^2 очень маленькое (0.0001), что указывает на низкую остаточную дисперсию модели.
- 4 Тест Ljung-Box с ρ -значением 0.86 не отвергает нулевую гипотезу о том, что остатки не автокоррелированы.
- 5 Тест Jarque-Bera с ρ -значением 0.00 отвергает нулевую гипотезу о нормальности распределения остатков, что указывает на то, что остатки не распределены нормально.
- 6 Показатель Heteroskedasticity (H) с ρ -значением 0.00 указывает на наличие гетероскедастичности, то есть непостоянства дисперсии остатков модели.

Эти результаты показывают, что модель хорошо описывает изучаемый процесс, но распределение остатков не принадлежит к нормальному, а так же некоторые модельные предпосылки нарушены.

Прежде чем приступить к построению ARCH/GARCH-модели, необходимо проверить наличие ARCH-эффектов в построенной модели ARMA. Согласно предпосылке ARCH/GARCH-модели, остатки модели ARMA не должны коррелировать между собой, однако должна наблюдаться значимая корреляция квадратов остатков, свидетельствующая о наличии условной гетероскедастичности [8].

Проверка на наличие ARCH-эффектов осуществляется с использованием ARCH-LM теста остатков. Результаты представлены на рисунке 7.

Heteroskedasticity Test: ARCH
F-statistic 158.58049
Prob.F(1.0232630213455701e-259, 2599.0) 0
Obs*R-squared 986.0329
Prob. ChiSquare(1) 0

Рисунок 7 – Результаты теста ARCH-LM

По результатам теста можно увидеть, что Prob.F., вероятность, связанная с F-статистикой, стремится к нулю. Это говорит о том, что мы можем отвергнуть нулевую гипотезу о наличии гомоскедастичности и сделать вывод, о присутствии в данных ARCH-эффектов. Наличие ARCH-эффекта, говорит о необходимости, применение ARCH/GARCH-моделей.

Для исследуемого индекса осуществим построение трех моделей (GARCH, TAR, EGARCH) с максимальным количеством запаздываний равным 2.

Как было сказано выше, анализ распределения остатков модели ARMA показал серьезную островершинность распределений. Для борьбы с островершинностью и «тяжелыми хвостами» произведено тестирование моделей с различными распределениями остатков: с нормальным распределением, распределением Стьюдента и обобщенным распределением ошибок (GED). Хвосты t-распределения также являются тяжелыми, однако, в отличие от варианта GED ($\nu < 2$), пик имеет округлую, а не острую вершину.

На рисунках показана последняя модель в семействе EGARCH, используя обобщенное распределение ошибки, она же GED (рисунки 8-9).

```
# Модель EGARCH(1,1) G
egarch_model = arch_model(summed_data['Log>Returns'], vol='EGARCH', p=1, q=1, dist='ged')
egarch_result = egarch_model.fit(dispatch='off')
print(egarch_result.summary())
```

Рисунок 8 – Построение модели EGARCH(1,1) с обобщенным распределением ошибки

Constant Mean - EGARCH Model Results					
Dep. Variable:	Log>Returns	R-squared:	0.000		
Mean Model:	Constant Mean	Adj. R-squared:	0.000		
Vol Model:	EGARCH	Log-Likelihood:	8674.55		
Distribution:	Generalized Error Distribution	AIC:	-17339.1		
Method:	Maximum Likelihood	BIC:	-17309.8		
Date:	Thu, May 09 2024	No. Observations:	2599		
Time:	19:49:07	Df Residuals:	2598		
		Df Model:	1		
Mean Model					
	coef	std err	t	P> t	95.0% Conf. Int.
mu	1.0056e-03	1.327e-04	7.576	3.564e-14	[7.455e-04, 1.266e-03]
Volatility Model					
	coef	std err	t	P> t	95.0% Conf. Int.
omega	-0.3788	8.977e-02	-4.219	2.450e-05	[-0.555, -0.203]
alpha[1]	0.3320	3.656e-02	9.081	1.075e-19	[0.260, 0.404]
beta[1]	0.9585	9.659e-03	99.232	0.000	[0.940, 0.977]
Distribution					
	coef	std err	t	P> t	95.0% Conf. Int.
nu	1.3563	5.619e-02	24.137	1.034e-128	[1.246, 1.466]

Рисунок 9 – Сводное представление модели EGARCH(1,1) с обобщенным распределением ошибки

Для того чтобы выбрать лучшую модель, все показатели полученных моделей, на которые необходимо опираться при принятии решения, были занесены в сводную таблицу 2.

Таблица 2 – Результирующие показатели моделей

	Нормальное распределение	Распределение Стьюдента	Обобщённое распределение ошибок
GARCH(1,1)			
Log-Likelihood	7271.34	4348.35	4638.80
AIC	-14532.7	-8686.70	-9267.61
BIC	-14503.4	-8657.38	-9238.30
TGARCH(1,1)			
Log-Likelihood	-14619.6	-281404	-289478
AIC	29249.2	562820	578969
BIC	29278.5	562855	579004
EGARCH(1,1)			
Log-Likelihood	8617.23	-15161.2	8674.55
AIC	-17226.5	30332.4	-17339.1
BIC	-17203.0	30361.7	-17309.8

По сводным данным четко видно, что модель EGARCH с обобщенным распределением ошибки показала наилучшие результаты среди всех. В дальнейшем прогнозировании будет использована именно эта модель в связке с моделью ARMA. Также результаты говорят о том, что негативная доходность оказывает большее влияние на волатильность индекса, чем позитивная.

Общий анализ показывает, что модель хорошо подходит для данных с высокой степенью точности, учитывая высокие значения логарифмического правдоподобия и статистически значимые оценки параметров модели. Однако, R-квадрат равен 0, что говорит о том, что модель не объясняет вариабельность логарифмической доходности, что типично для моделей волатильности.

Точечный прогноз индивидуального значения показателя с использованием модели EGARCH (Exponential Generalized Autoregressive Conditional Heteroskedasticity) включает в себя оценку будущей волатильности на основе исторических данных о доходности (рисунок 10) [9].

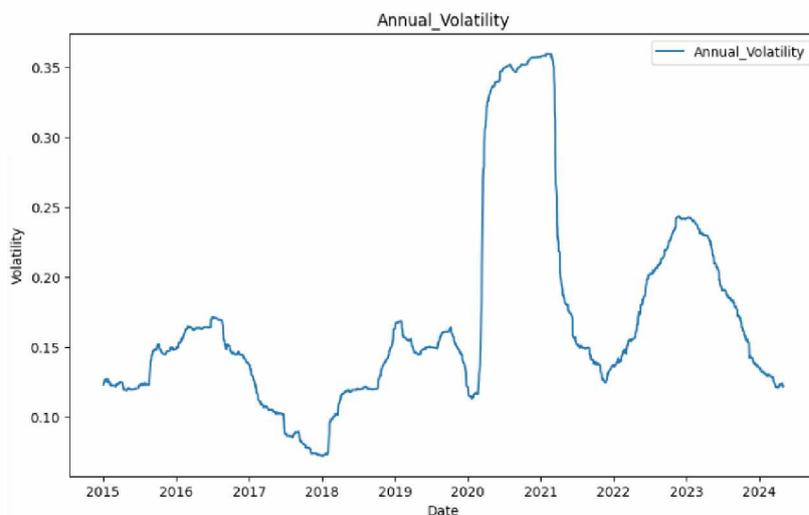


Рисунок 10 – Годовая оцененная волатильность индекса S&P 500

На графике видны резкие скачки в период пандемии 2020-2022 год, что говорит о повышении уровня риска в данный период. Стоит отметить, что анализ исторических графиков годовых волатильностей актива является очень полезным, так как помогает инвестору оценить степень колебания ценной бумаги в спокойные и кризисные периоды. Также прослеживается ярко выраженная кластеризация волатильности, которая, как упоминалось выше, учитывается моделью,

Модель EGARCH учитывает асимметрию и “рычаги” в данных, что позволяет более точно прогнозировать волатильность, особенно в финансовых рынках, где эти характеристики являются обычным явлением.

Проведем точечный прогноз волатильности индивидуального значения показателя (рисунок 11).

```
▶ forecast = egarch_result.forecast(horizon=1)

#Прогнозируемая волатильность
predicted_volatility = forecast.variance.iloc[-1]**0.5

print(f'Прогнозируемая волатильность: {predicted_volatility}')
```

➔ Прогнозируемая волатильность: h.1 0.008997
Name: 2599, dtype: float64

Рисунок 11 – Прогноз волатильности на 2024.05.02

По полученному значению, можно сказать, что ожидаемое стандартное отклонение доходности актива в следующем периоде составляет примерно 0.8997%. Однако, интерес представляет не только волатильность, которая описывает риск актива, но и его прогнозируемая доходность и цена.

Для прогнозирования цены базового актива применяется связка моделей ARMA(1,1)-EGARCH(1,1). Проведем прогнозирование дневной логарифмической доходности с помощью ARMA модели (рисунок 12).

```
[78] #Прогнозируемая дневная логарифмическая доходность
arma_forecast = results_arma_1_1.forecast(horizon=1).values[0]
print(f'Прогнозируемая доходность: {arma_forecast}')
```

Прогнозируемая доходность: 0.00029329202962370866

Рисунок 12 – Прогноз цены дневной логарифмической доходности на 2024.05.02

Используя полученные данные из комбинированной модели ARMA(1,1)-EGARCH(1,1), проведем необходимые преобразования для получения цены базового актива, состоящего из акций компании входящих в состав фондового индекса S&P 500 (рисунок 13).

```
▶ # Текущая цена актива
current_price = summed_data['Adj Close'][2599]
print(f'Текущая цена актива: {current_price}')
```

```
# Расчет будущей цену актива
predicted_price = current_price * np.exp(arma_forecast + predicted_volatility)
print(f'Будущая цена актива: {predicted_price[0]}')
```

Текущая цена актива: 104783.06980323792
Будущая цена актива: 105761.10044492078

Рисунок 13 – Прогноз цены базового актива на 2024.05.02

Согласно прогнозу модели, если приобрести базовый актив 2024.05.01 по цене 104783.06\$, то уже на следующий день его стоимость будет составлять 105761.10\$. Тем самым цена выбранного базового актива увеличится уже за одни сутки на 978.03\$ или другими словами вырастет на 0.93%.

Список использованных источников:

1. Асатуров К.Г., Теплова Т.В. Эффекты перетекания волатильности и заражения на фондовых рынках: определение глобальных и локальных лидеров // Вестник Московского университета. – 2014. - №6 – С. 3-34
2. Vadym Dun. Analysis and Forecasting the Price of The S&P 500 Index Using the Arima Model. Financial Markets, Institutions and Risks, Volume 7, Issue 4, 2023 [Электронный ресурс]. – Режим доступа:

3. Sharpe, Steven A., and Antonio Gil de Rubio Cruz (2024). "Predicting Analysts' S&P 500 Earnings Forecast Errors and Stock Market Returns using Macroeconomic Data and Nowcasts," *Finance and Economics Discussion Series 2024-049*. Washington: Board of Governors of the Federal Reserve System [Электронный ресурс]. – Режим доступа: <https://www.federalreserve.gov/econres/feds/files/2024049pap.pdf>

4. LSTM versus Gradient Boosting for Asset Pricing: A Case Study of the S&P 500. *The Journal of Financial Data Science*, Vol. 6, No. 4, pp. 49- 71, Fall 2024. [Электронный ресурс]. – Режим доступа: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5064862

5. Marisetty, Nagendra. 2024. "Applications of GARCH Models in Forecasting Financial Market Volatility: Insights from Leading Global Stock Indexes". *Asian Journal of Economics, Business and Accounting* 24 (9):63-84.

6. Трифонов А.Ю., Кривицкий О.Л., Бельснер О.А. Моделирование динамических корреляций: общее приложение к исследованию финансовых рынков // *Экономико-математическое моделирование*. – 2012. – 39(294). – С. 58-62.

7. Хацкевич, Г. А. Эконометрика : учебник для студентов учреждений высшего образования по специальностям «Управление информационными ресурсами», «Информационные системы и технологии (в экономике)» / Г. А. Хацкевич, Т. В. Русилко. – Минск : РИВШ, 2021. – 450 с.: ил., табл.

8. Emsbo-Mattingly, Hofshire D., Litvak A., Lund-Wilde J., *The Business Cycle Approach to Equity Sector Investing: Market Research*, 2014

9. Rob J Hyndman, George Athanasopoulos *Forecasting: principles and practice*. – Melbourne, 2013. 292 с.

UDC 336.761.533

FORECASTING THE PRICE OF THE UNDERLYING ASSET BASED ON S&P 500

Satsuta D. V., student of gr.172303, Golda O., Associate Professor, Department of ITS.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Golda O. A.– PhD in Economics

Annotation. The article deals with forecasting the price of the underlying asset on the basis of the S&P 500 index using econometric models. The models of GARCH family for volatility analysis and yield forecasting are studied.

Keywords. Price forecasting, S&P 500, econometric models, GARCH, volatility, ARMA, asymmetric models, logarithmic returns, ARCH effect, EGARCH.