

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
Российский экономический университет имени Г.В. Плеханова

Факультет Математической экономики, статистики и информатики

Направление Прикладная математика и информатика

Профиль Прикладная математика и информатика

Кафедра Математических методов в экономике

КУРСОВАЯ РАБОТА

по дисциплине

«Эконометрика и эконометрическое моделирование»

на тему

**«Моделирование и прогнозирование курса акций «The Coca-Cola Company» с
учетом риска повышенной волатильности»**

Выполнена:
студентом 434 группы
очной формы обучения
Эль-Айяссом Дани Валидом

Научные руководители:
д.э.н., профессор Тихомирова Т.М.

Москва, 2018

ОТЗЫВ на курсовую работу

Раздел 1.

Кафедра	«Математические методы в экономике»
Дисциплины	«Эконометрика и эконометрическое моделирование»
Тема	«Моделирование и прогнозирование курса акций «The Coca-Cola Company» с учетом риска повышенной волатильности»

Студент Эль-Айясс Dani Валид, 3 курс, 434 группа

Руководитель курсовой работы д.э.н., проф. Тихомирова Т.М.

Раздел 2. Критерии, при наличии хотя бы одного из которых работа оценивается только на «неудовлетворительно».

1.	Тема и (или) содержание работы не относится к предмету дисциплины
2.	Работа перепечатана из Интернета или других информационных источников
3.	Неструктурированный план курсовой работы
4.	Объем работы менее 15 листов машинописного текста
5.	В работе отсутствуют ссылки (сноски) на нормативные и другие источники
6.	Оформление курсовой работы не соответствует требованиям РЭУ

Раздел 3. Рейтинг работы (при неудовлетворительной оценке не заполняется)

№	Наименование показателя	Баллы
1.	Содержательная составляющая	
1.1	Степень раскрытия темы	
1.2	Полнота охвата научной литературы	
1.3	Использование нормативных актов	
1.4	Индивидуальность подхода к написанию КР	
1.5	Последовательность и логика изложения материала	
Итого по содержательной составляющей (максимум 55 баллов)		
2.	Оформление и информационное сопровождение работы	
2.1	Качество оформления, язык, стиль и грамматический уровень работы	
2.2	Использование иллюстрационного материала (рисунки, таблицы, и т.д.)	
Итого по оформлению и информационному сопровождению работы (максимум 20 баллов)		
3.	Защита (максимум 25 баллов)	
4.	Итого	

Дополнительные замечания

Оценка (прописью)

Научные руководители

« » 2018 г.

Рецензия
на курсовую работу
«Моделирование и прогнозирование курса акций «The Coca-Cola Company» с учетом риска повышенной волатильности»

студента 3-го курса факультета математической экономики, статистики и информатики по направлению «Прикладная математика и информатика» профилю подготовки «Прикладная математика и информатика»

Эль-Айясса Дани Валида

Курсовая работа Эль-Айясса Д.В. выполнена на актуальную тему, так как исследование акций на фондовом рынке играет важную роль в экономике.

Эль-Айяссом Д.В. обработано большое количество научного материала, проведено исследование проблем моделирования и прогнозирования курса акций. Материал в курсовой работе логически структурирован, написан научным стилем изложения. Объем курсовой работы составляет 70 страниц.

В первой главе курсовой работы автором проведен подробный анализ теоретических и практических основ для моделирования и прогнозирования курса акций на основе ежедневных данных за 5 лет (с 01.05.2013 по 30.04.2018). Обоснованы проблемы, возникшие при моделировании. Построен краткосрочный прогноз на 2 дня вперед.

Во второй главе работы построена модель и прогноз для ряда средненедельных данных курса акций. Построен среднесрочный прогноз на 2 недели вперед.

В третьей главе курсовой работы представлено моделирование и прогнозирование курса акций на основе среднемесячных данных. Построен долгосрочный прогноз на 2 месяца.

Автор курсовой работы показал отличную способность формулировать собственную точку зрения по рассматриваемой проблеме. Проделанная работа полностью соответствует всем изначальным целям и задачам. Существенных недостатков в курсовой работе не выявлено.

Аннотация

В данной курсовой работе осуществляется моделирование и прогнозирование курса акций компании «The Coca-Cola Company» к российскому рублю в краткосрочной, среднесрочной и долгосрочной перспективе.

Исходные данные для построения краткосрочного прогноза собраны за 2 года. Для среднесрочного и долгосрочного прогноза, собранные данные за 5 лет анализируются, а также агрегируются по неделям и месяцам соответственно.

В рамках краткосрочного прогноза рассматривается ряд ежедневных данных за последние 2 года. Данный ряд приводится к стационарному виду с помощью взятия разности второго порядка. В конечном итоге, ряд моделируется с помощью модели типа MA(1).

В рамках среднесрочного прогноза строится модель с лаговой зависимой переменной, в качестве зависимого фактора которой используется исходный агрегированный по неделям ряд с лагом на 1, 2 и 3 периода.

В рамках долгосрочного прогноза строится модель с двумя лаговыми зависимыми переменными, в качестве зависимых факторов которой используется временной ряд месячных данных за 5 лет с лагом на 1 и 2 периода.

На основе всех построенных моделей строится прогноз на несколько значений вперед. При этом для каждого из прогнозов строится 95%-й доверительный интервал.

Ключевые слова: ARIMA, MA(1), «белый шум», автокорреляционная функция, стационарность, конечная разность, полупараметрический тест, параметрический тест, эксцесс, асимметрия, прогноз, доверительный интервал, лаг, регрессия, временные ряды.

Abstract

In this work considered the stock of «The Coca-Cola Company» in the short, medium and long terms are modeling and prediction.

The initial data for the short-term forecast were collected for 2 years. For the medium and long-term forecast, the data were collected for 5 years, then analyzed and aggregated by weeks and months.

The short-term forecast considers a number of daily data for the last 2 years. The given series were provided by the stationary view, using the score of difference of second order. In the end, the series simulated while using a model of type MA(1)

In the framework of the medium-term forecast, a model with one dependent variable lag is built, and considered as a dependent factor that use the initial aggregated weekly series with a lag for period 1, 2 and 3.

In the framework of the long-term forecast, a model with two dependent variable lags is built, and considered as the dependent factors using the time series of monthly data for 5 years with a lag of periods 1 and 2.

Based on all the models built, the forecast for several forward values can be built. At the same time, 95% confidence interval is built for each of the forecasts.

Key words: ARIMA, MA (1), "white noise", autocorrelation function, stationary, finite difference, semi parametric test, parametric test, kurtosis, skewness, forecast, confidence interval, lag, regression, time series.

Оглавление

Введение	7
Глава 1 Моделирование и прогнозирование курса акций «The Coca-Cola Company» в краткосрочной перспективе	9
1.1 Анализ данных.....	9
1.2 Приведение исходного ряда к стационарному виду	16
1.3 Построение модели ARIMA(0,2,1)	30
1.4 Тестирование остатков модели ARIMA(0,2,1)	33
1.5 Прогнозирование в краткосрочной перспективе	37
Глава 2 Моделирование и прогнозирование курса акций «The Coca-Cola Company» в среднесрочной перспективе	41
2.1 Анализ исходных данных	41
2.2 Моделирование агрегированных недельных данных.....	45
2.3 Тестирование остатков модели с лаговыми переменными	47
2.4 Прогнозирование в среднесрочной перспективе	51
Глава 3 Моделирование и прогнозирование курса акций «The Coca-Cola Company» в долгосрочной перспективе.....	53
3.1 Анализ исходных данных	53
3.2 Моделирование месячных данных с помощью сплайн функции	56
3.3 Тестирование остатков сплайн функции.....	58
3.4 Прогнозирование в долгосрочной перспективе.....	62
Заключение	64
Список литературы	65
Список иллюстраций.....	67
Список таблиц	69

Введение

Прогнозирование цен на акции широко используется в сфере финансов. Основной целью предсказания будущих цен на акции является получение спекулятивной прибыли, поэтому выявление закономерностей стоимости акций является необходимым условием, обеспечивающим снижение рыночного риска и увеличение доходности.

Тема курсовой работы – «Моделирование и прогнозирование курса акций «The Coca-Cola Company» с учетом риска повышенной волатильности». *Актуальность* выбранной темы заключается в том, что исследование и прогнозирование курса акций играет огромную роль в сфере инвестирования и на фондовом рынке в целом, ведь эффективно функционирующий фондовый рынок – это мощное средство стабилизации экономики.

The Coca-Cola Company — американская пищевая компания, крупнейший мировой производитель и поставщик концентратов, сиропов и безалкогольных напитков. Наиболее известным продуктом компании является напиток Coca-Cola. Входит в список Fortune 500 по итогам 2015 года (62-е место).

Целью данной работы является выявление закономерности в изменении курса акций «The Coca-Cola Company», построение достоверной модели и прогнозирование указанного показателя с учетом риска изменчивости цены.

Задачами курсовой работы являются:

1. Сбор и анализ исходных данных.
2. Построение краткосрочной модели курса акций и прогнозирование котировок на два дня на основе ежедневных данных.
3. Построение модели на основе агрегированных по неделям данных, которая даст более точное описание поведения изучаемого показателя за рассматриваемый период времени и позволит сделать среднесрочный прогноз на несколько недель вперед.
4. Построение модели на основе агрегированных данных по месяцам, которая будет адекватно описывать общую тенденцию в динамике курса

акций и предскажет направление изменения показателя в долгосрочной перспективе.

Предметы исследования: котировки акций банка «The Coca-Cola Company».

При проведении исследования были использованы следующие инструменты: Gretl, MS Excel

Источником исходных данных является интернет-ресурс Yahoo Finance.

Глава 1 Моделирование и прогнозирование курса акций «The Coca-Cola Company» в краткосрочной перспективе

1.1 Анализ данных

Были проанализированы средневзвешенные котировки за период 01.05.2013 - 30.04.2018 г. Общий объём выборки составляет 1259 значений. Для краткосрочного прогнозирования были использованы данные по дням для среднесрочного – агрегированные данные по неделям, долгосрочного – агрегированные данные по месяцам.

Для начала рассмотрим график исходных данных:

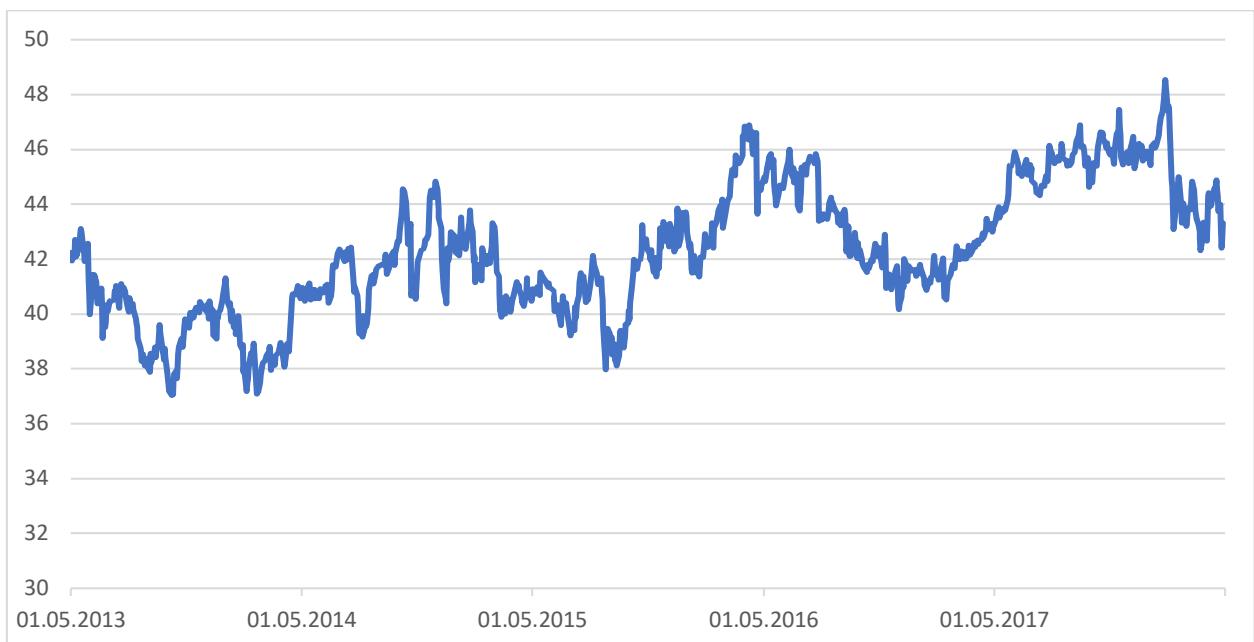


Рисунок 1- Курс акций «The Coca-Cola Company»

Из графика видно, что поведение цены акций имеет циклический характер, но при этом наблюдается общая тенденция к росту. Для построения модели ARIMA используем не весь ряд, а данные за последние два года (503 наблюдения).

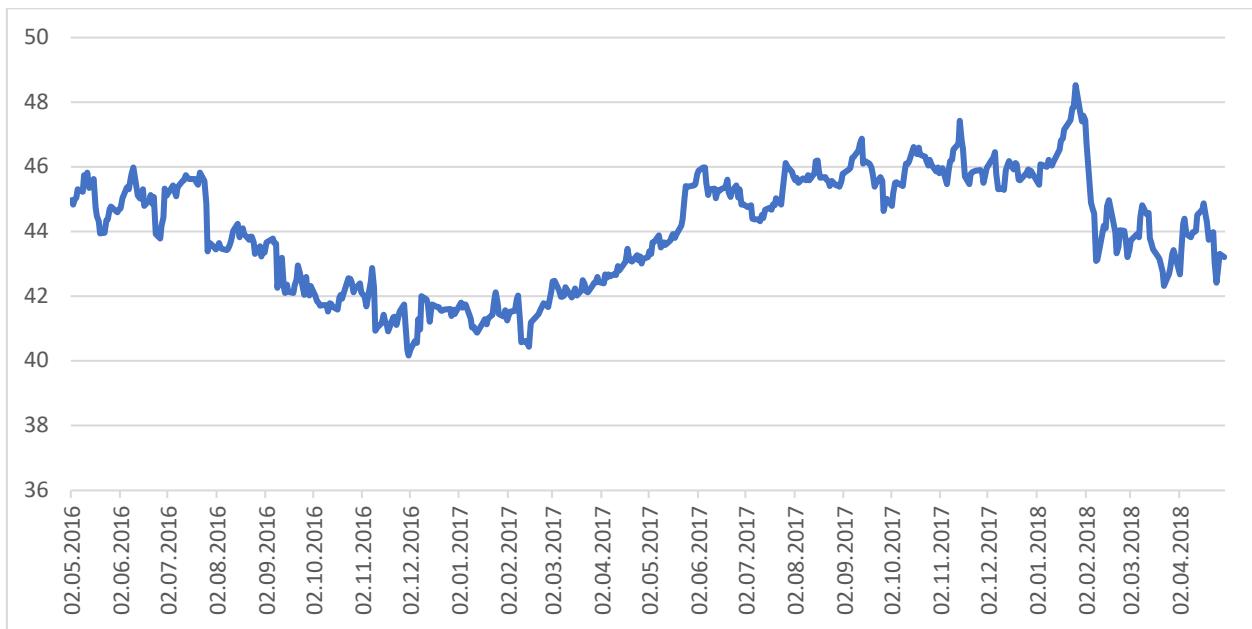


Рисунок 2 – Последние 503 наблюдения ряда «The Coca-Cola Company»

Для того, чтобы более точно определить уровни ряда, необходимо проанализировать его основные характеристики. С помощью статистического пакета получаем следующие результаты:

Описательная статистика, наблюдения 2016-05-02 – 2018-04-30 для переменной 'Close' (использовано 503 наблюдений)

Среднее	44,035
Медиана	44,180
Минимум	40,170
Максимум	48,530
Стандартное отклонение	1,7739
Вариация	0,040283
Асимметрия	-0,15934
Эксцесс	-1,0400
5% перцентиль	41,222
95%-перцентиль	46,506
Interquartile range	3,1600
Пропущенные наблюдения	0

Рисунок 3 – Основные характеристики ряда «The Coca-Cola Company»

При выборке в 503 значений, наш ряд обладает математическим ожиданием 44,035 и стандартным отклонением 1,7739. Остановимся более

подробно на коэффициентах асимметрии и эксцесса. Асимметрия – показатель симметричности кривой распределения относительно среднего значения, эксцесс – мера «размытости» распределения по оси ординат [3, с.26]. Нормальным считают распределение с асимметрией и эксцессом, принадлежащих интервалу [-2;2]. В нашем случае коэффициент эксцесса -1,04 и асимметрии -0,159 не превышают допустимые значения. Для более лучшего понимания структуры нашего ряда рассмотрим его распределение.

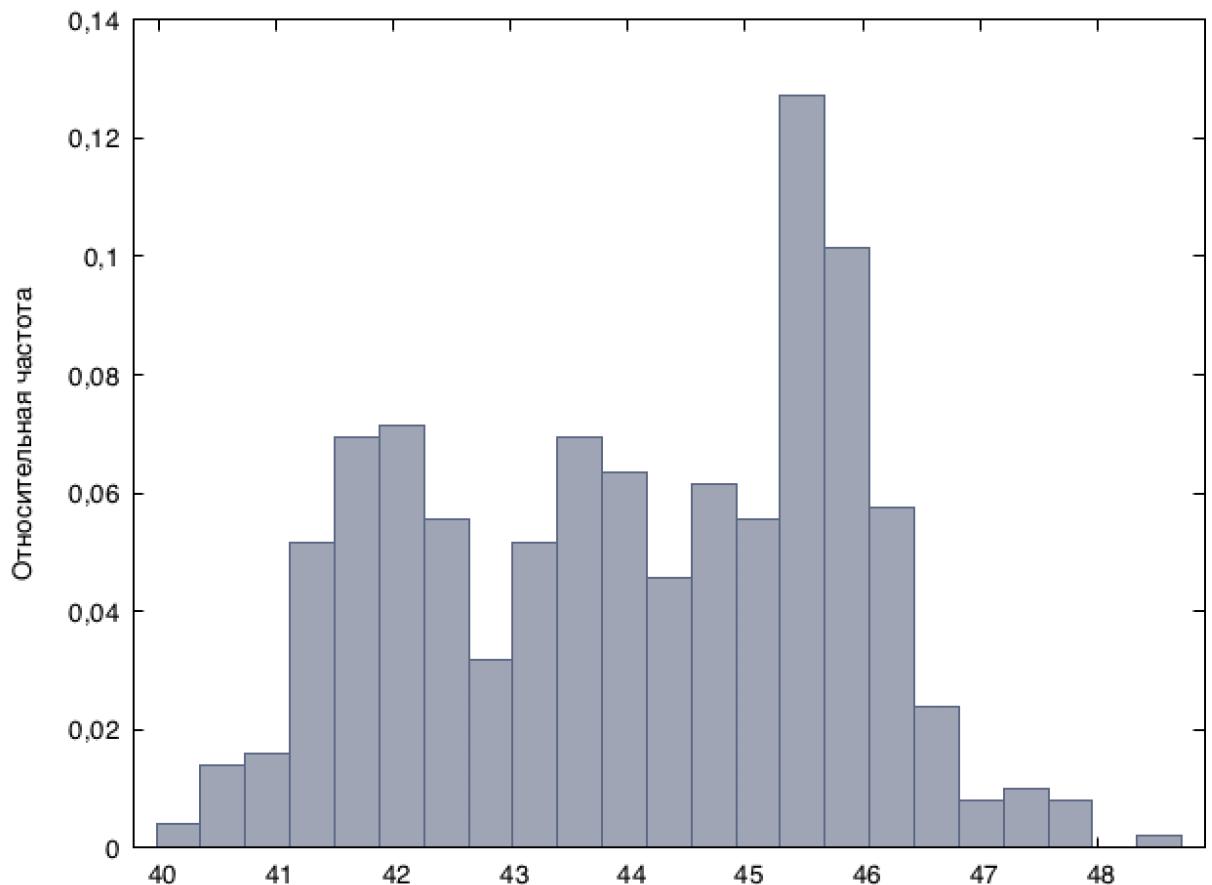


Рисунок 4 – Распределение ряда «The Coca-Cola Company»

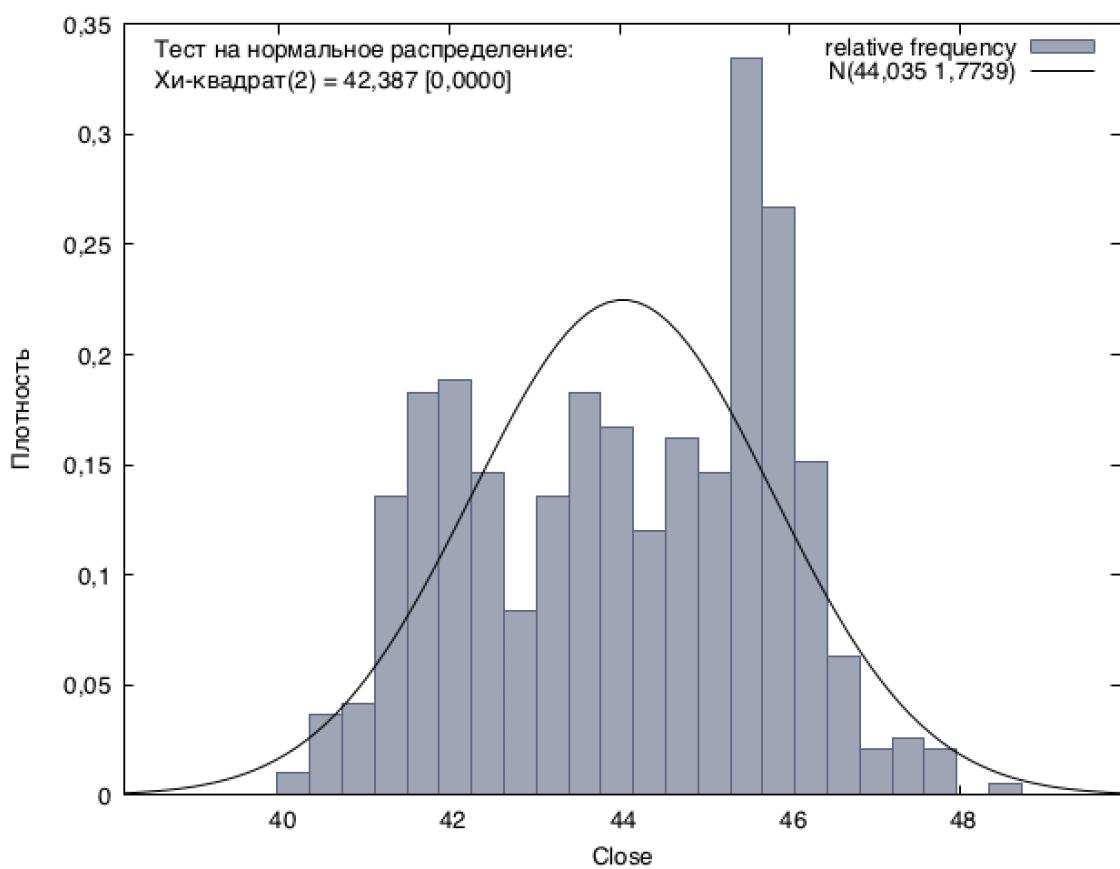


Рисунок 5 – Проверка нормальности ряда «The Coca-Cola Company»

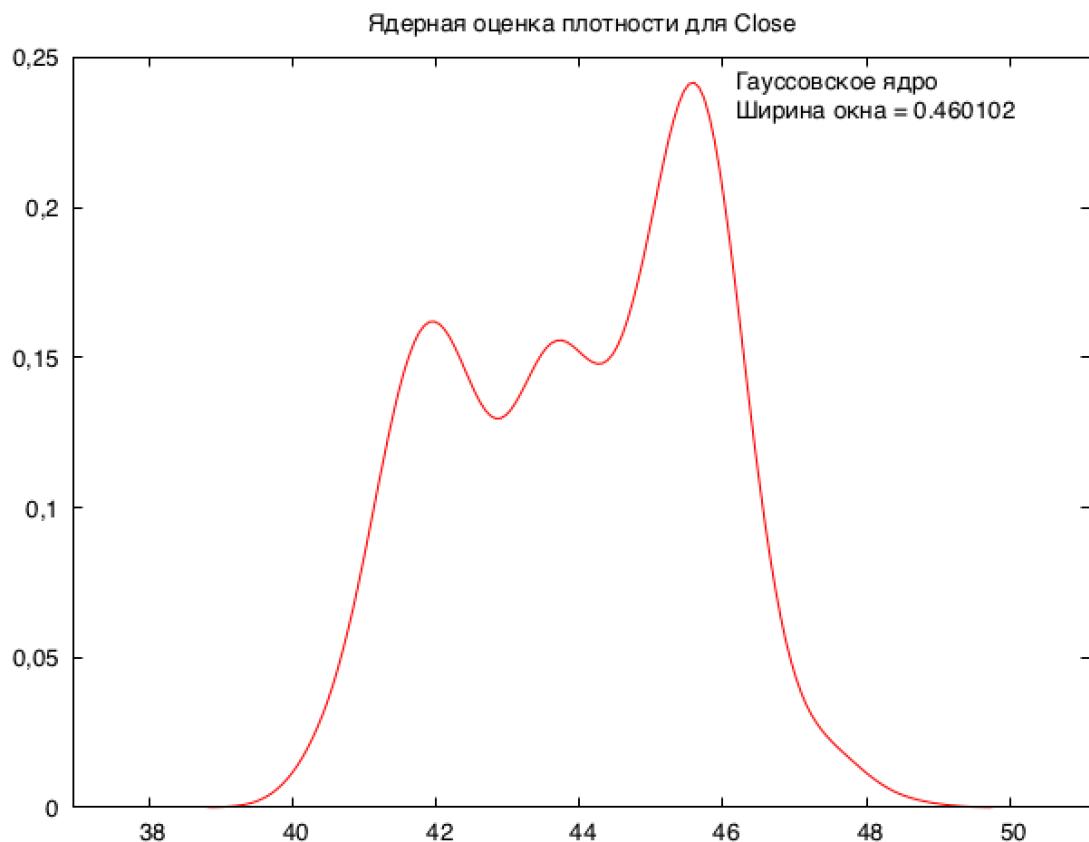


Рисунок 6 – Ядерная оценка плотности ряда «The Coca-Cola Company»

Отрицательное значение эксцесса указывает на менее выраженную остроконечность и более короткие "хвосты". При этом график плотности также показывает наличие бимодальности.

Автокорреляционная функция используется для определения уровней ряда. С её помощью можно установить наличие сезонности, цикличности, тенденции и её тип.

Автокорреляционная функция формируется из коэффициентов автокорреляции первого, второго и других порядков. Коэффициент автокорреляции i -ого порядка считается по формуле:

$$p_i = \frac{cov(y_t; y_{t-i})}{\sigma_{y_t} \sigma_{y_{t-i}}},$$

где i - размер лага,

σ_{y_t} – стандартное отклонение ряда y_t ,

$\sigma_{y_{t-i}}$ – стандартное отклонение ряда y_{t-i}

Проанализируем коррелограмму выборочной автокорреляционной функции (ВАФ) и частной автокорреляционной функции (ЧАФ) нашего ряда:

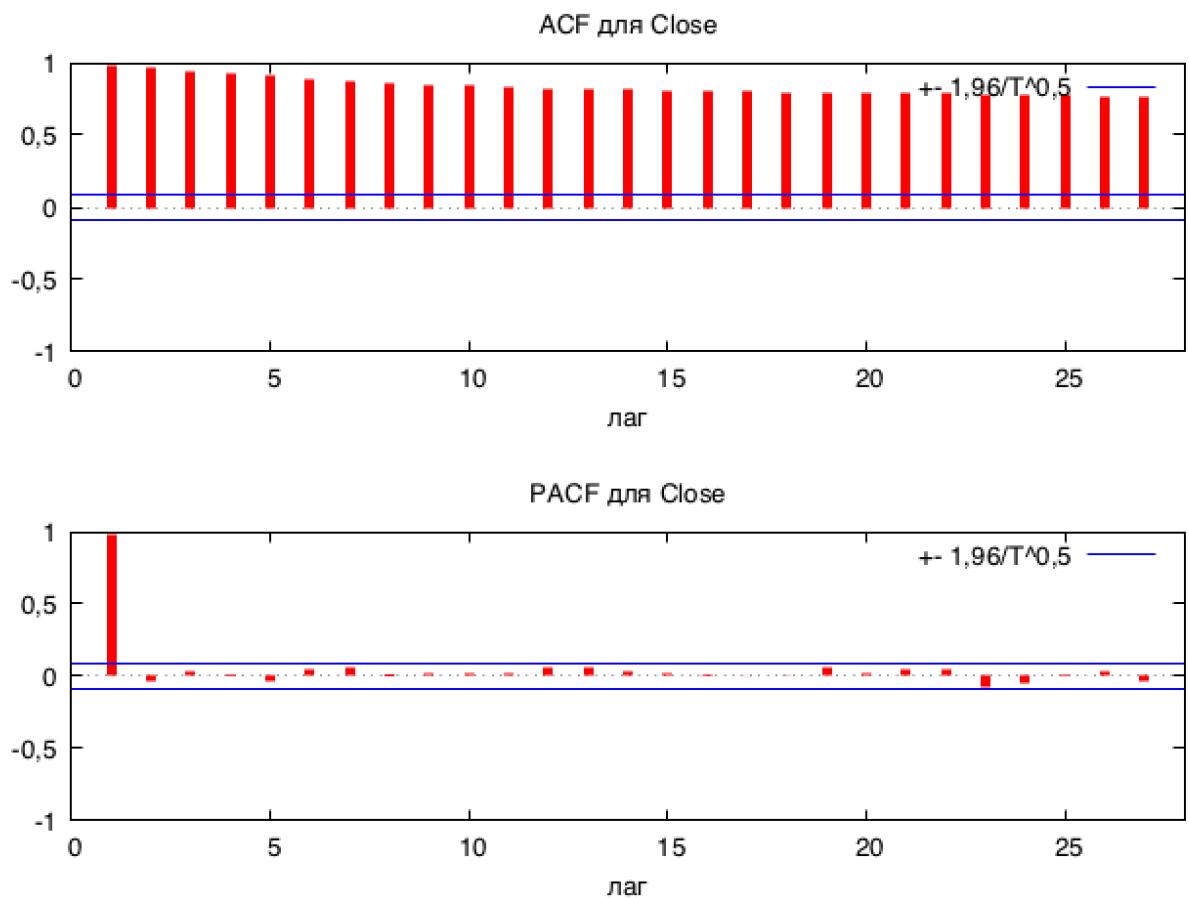


Рисунок 7 – ВАФ и ЧАФ ряда «The Coca-Cola Company»

График ВАФ убывает, что свидетельствует о наличии у ряда тренда. То есть соседние уровни ряда коррелируют между собой.

Частная автокорреляционная функция показывает корреляцию между двумя случайными переменными за вычетом влияния всех внутренних значений автокорреляции. Частная автокорреляция на данном лаге аналогична обычной автокорреляции, за исключением того, что при вычислении из неё исключается влияние автокорреляций с меньшими лагами. Таким образом, график ЧАФ показывает, что значения ряда зависят от предыдущих значений с лагом 1, поскольку только первый коэффициент является статистически значимым.

В целом анализ автокорреляционной функции показал, что значения стоимости акций «The Coca-Cola Company» не являются стационарным рядом в силу наличия трендовой составляющей.

1.2 Приведение исходного ряда к стационарному виду

Для дальнейшего моделирования временного ряда необходимо, чтобы он был стационарен, то есть необходимо постоянство математического ожидания и дисперсии [4, с.389]. Для приведения ряда к стационарному воспользуемся способом, называемым взятием конечных разностей, а именно возьмём первые разности $\Delta^1 y_t = y_t - y_{t-1}$.

График данного ряда имеет следующий вид:

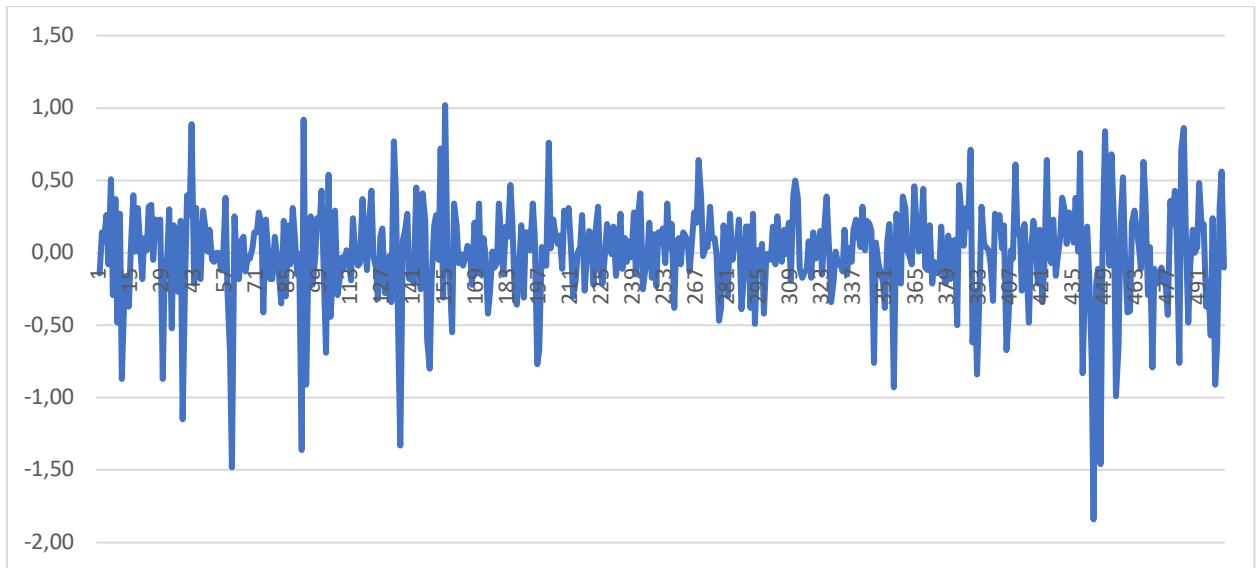


Рисунок 8 – Ряд первых разностей

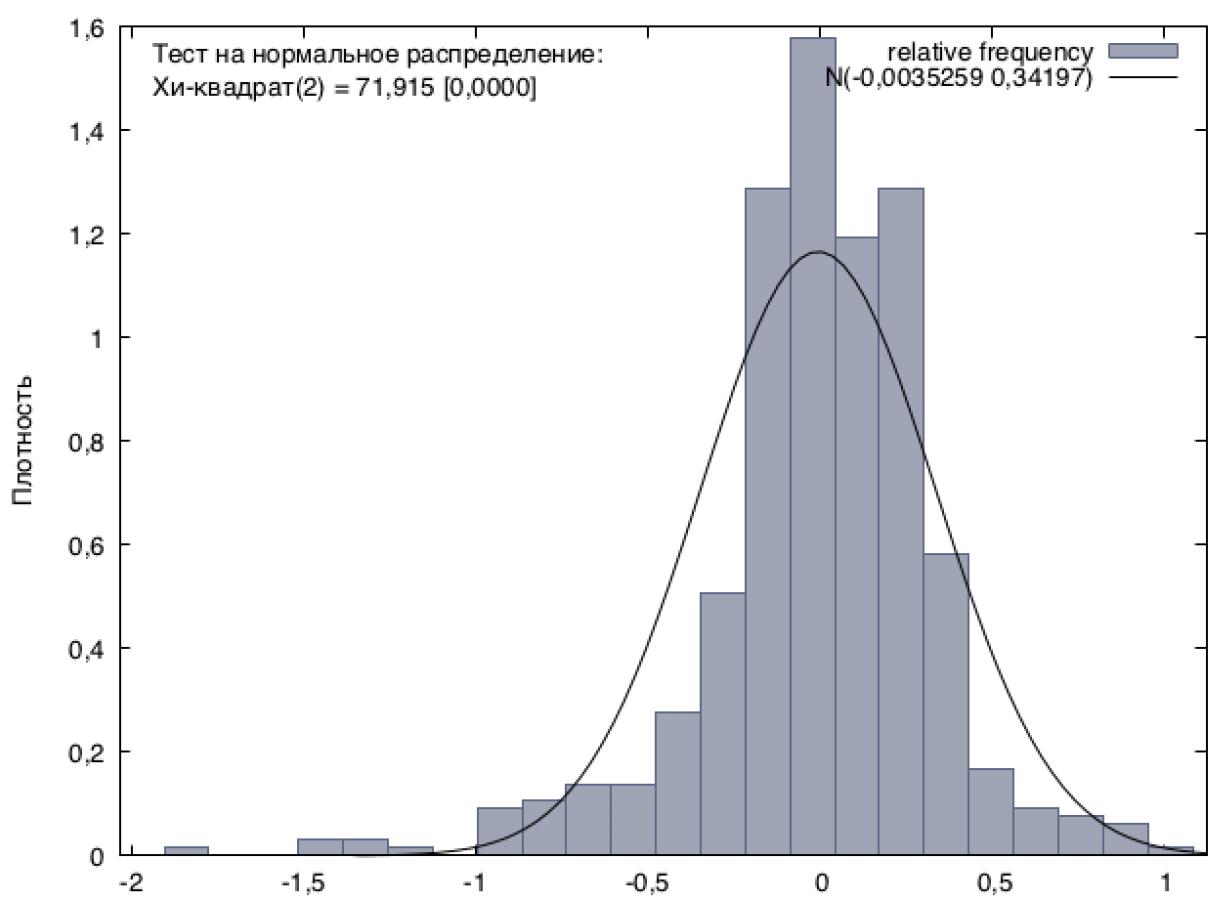
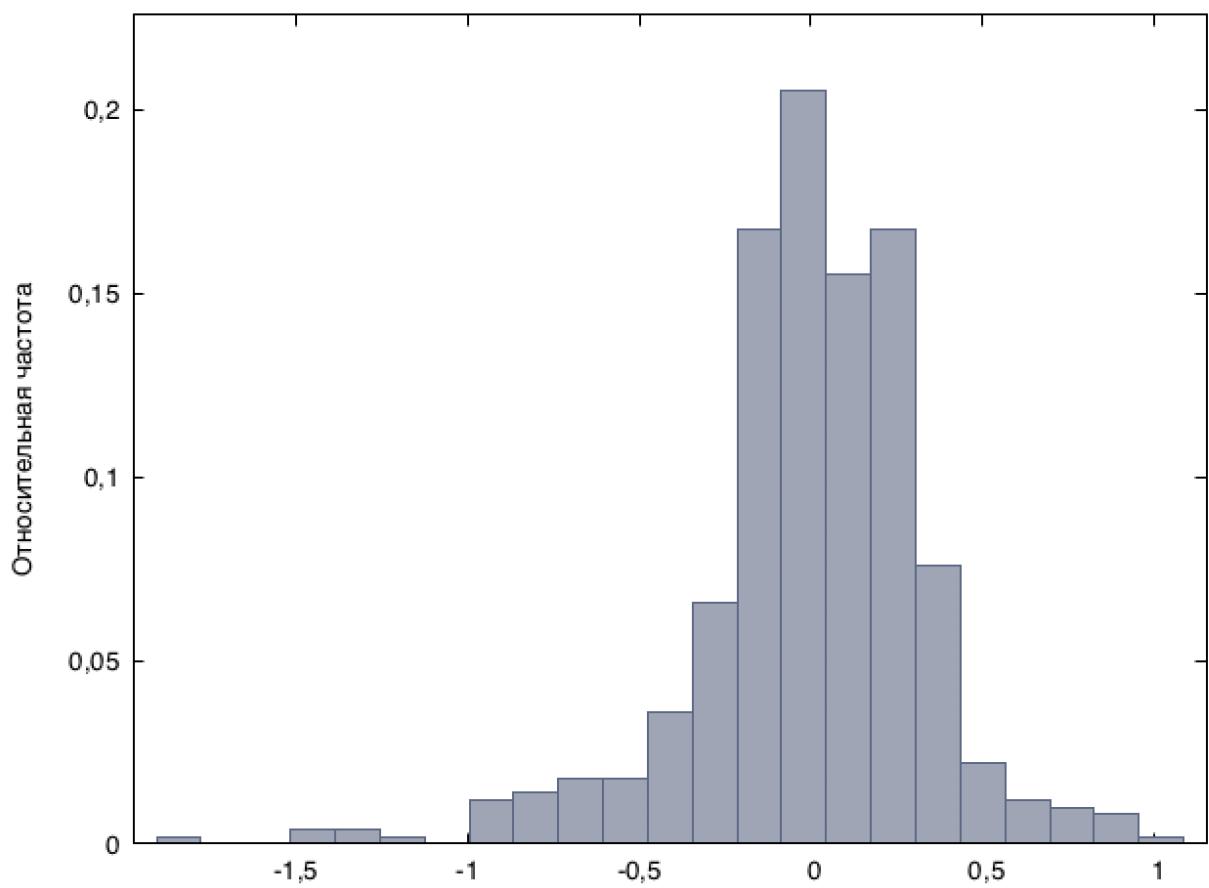
Визуально анализируя график данного ряда, можно предположить, что ряд не является стационарным. Для проверки этой гипотезы необходимо применить тесты на стационарность, выбор которых зависит от распределения ряда. Ряд имеет следующие характеристики:

Описательная статистика, наблюдения 2016-05-02 – 2018-04-30
для переменной 'd_Close' (использовано 502 наблюдений)

Среднее	-0,0035259
Медиана	0,019998
Минимум	-1,8400
Максимум	1,0200
Стандартное отклонение	0,34197
Вариация	96,989
Асимметрия	-1,0424
Эксцесс	3,8213
5% перцентиль	-0,64850
95%-перцентиль	0,45850
Interquartile range	0,35000
Пропущенные наблюдения	1

Рисунок 9 – Характеристики ряда первых разностей

Коэффициенты асимметрии, равный -1,04 не выходит за рамки допустимых значений [-2;2], в то время как коэффициент эксцесса, равный 3,82, выходит. Такое значение коэффициента эксцесса показывает, что в наблюдениях отмечаются более выраженная островершинность и более длинные хвосты, чем при нормальном распределении.



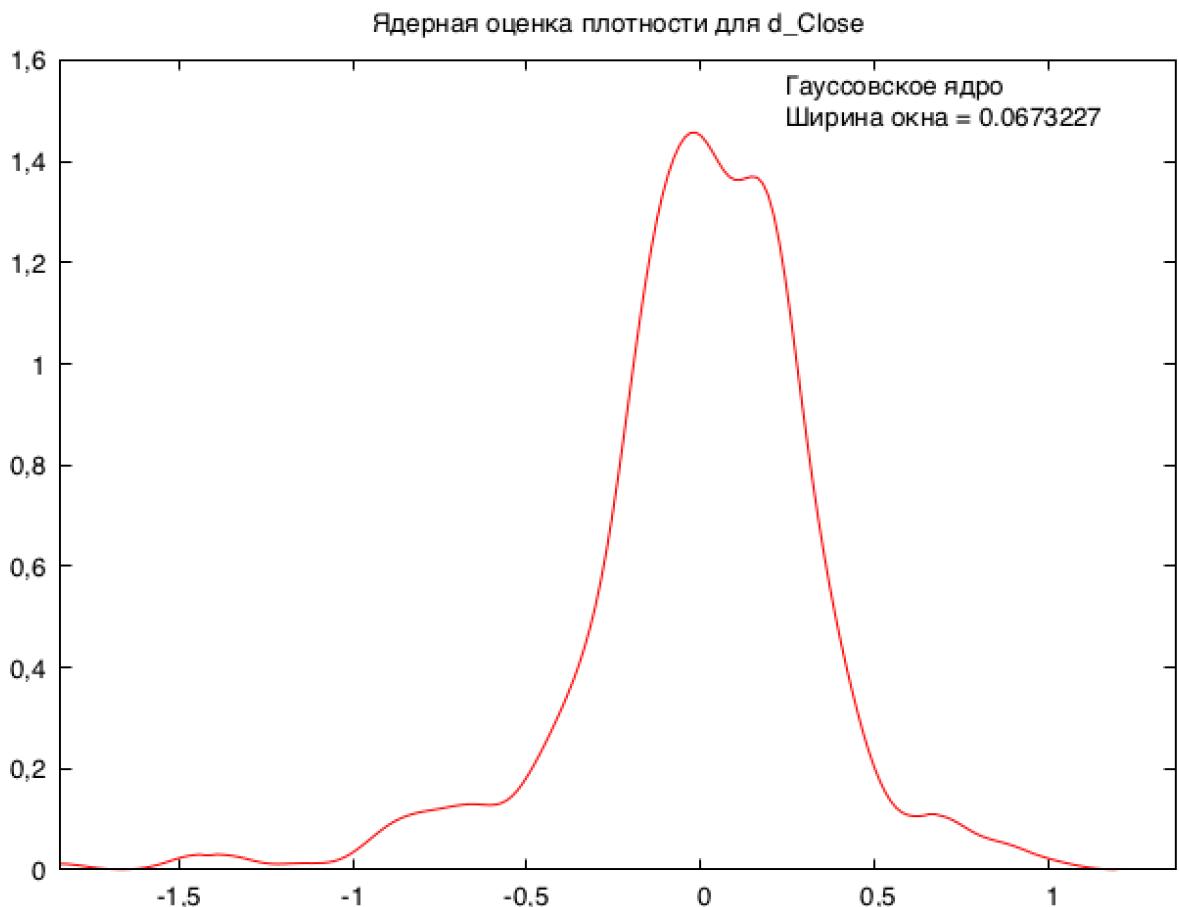


Рисунок 12 – Ядерная оценка плотности ряда первых разностей

Таким образом, закон распределения полученного ряда не является нормальным. Следовательно, для проверки данного ряда на стационарность целесообразно использовать непараметрические и полупараметрические тесты.

Для тестирование математического ожидания используем непараметрический тест **Манна-Уитни**.

Выдвигаются гипотезы:

Гипотеза H_0 : математическое ожидание ряда постоянно

Гипотеза H_1 : математическое ожидание не постоянно; вероятность ошибки первого рода $\alpha=0,05$

Разделим выборку на две части по $T_1 = 251$ и $T_2 = 251$ элементов. Упорядочим исходный ряд по возрастанию его значений. Значениям упорядоченного ряда ставятся в соответствие ранги, начиная с 1 (наименьшему значению), и заканчивая 502 (наибольшему значению). Затем

считается суммы рангов для первой и второй подвыборки R_1 и R_2 соответственно.

Далее проводим расчет статистики по формуле:

$$u^* = R_1 - \frac{T_1 * (T_1 + 1)}{2}$$

Эта статистика распределена по нормальному закону с математическим ожиданием $M[u^*] \approx \frac{(T_1 * T_2)}{2}$ и дисперсией $D[u^*] \approx \frac{T_1 * T_2 * (T_1 + T_2 + 1)}{12}$

Отсюда, случайная величина $z = \frac{u^* - M[u^*]}{\sqrt{D[u^*]}} \pm 0,5$ распределена по стандартному нормальному закону (поправка 0,5 вычитается, если значение z положительно и прибавляется, если отрицательно) [4, с.194].

Рассчитанное значение статистики Манна-Уитни для нашего ряда $Z = -0,529$ находится в пределах допустимых значений $[-1,96; 1,96]$, поэтому гипотеза о постоянстве математического ожидания принимается с вероятностью 95%.

Для проверки постоянства дисперсии используем **тест Сиджела-Тьюки.**

Выдвигаются гипотезы:

Гипотеза H_0 : дисперсия ряда постоянна

Гипотеза H_1 : дисперсия ряда не постоянна; вероятность ошибки первого рода $\alpha=0,05$

Для начала необходимо получить центрированный ряд. Далее разделим полученный ряд на две совокупности по $T_1 = 251$ и $T_2 = 251$ наблюдений. Упорядочим ряд и проранжируем его по следующей схеме: наименьшему элементу ставим в соответствие ранг 1, наибольшему – 2, значению, следующему за наименьшим – ранг 3, следующему за наибольшим – ранг 4 и так далее. Далее рассчитаем суммы рангов значений, приходящихся на 1-ю R_1 и 2-ю R_2 совокупности.

Тогда статистика $z = \frac{R_1 - \frac{T_1(T_1 + T_2 + 1)}{2} \pm 0,5}{\sqrt{\frac{T_1 T_2 (T_1 + T_2 + 1)}{12}}}$ распределена по стандартному нормальному закону [4, с.197]. Критическое значение равно $-1,96 < z < 1,96$. Вычисленное значение статистики в нашем случае составило -0.017 , следовательно, с вероятностью $0,95$ дисперсия является постоянной величиной.

Также используем полупараметрические тесты: тест Аббе

Тест Аббе

Расчетная статистика выглядит следующим образом:

$$\gamma^{(n)} = \frac{q^2}{\sigma^2}, \text{ где}$$

$$q^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2,$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Тогда критерий определяется следующим неравенством:

$$\gamma^{(n)} < \gamma_{(\alpha,n)}^{\min}, \text{ где} \quad (*)$$

$$\gamma_{(\alpha,n)}^{\min} = 1 + \frac{u_\alpha}{\sqrt{n + 0,5(1 + u_\alpha^2)}}$$

u_α - квантиль стандартного нормального распределения.

Если условие (*) выполняется, то гипотеза о стохастической независимости результатов наблюдения отвергается [5, с.78]. Для нашего ряда данное неравенство не выполняется ($\gamma^{(n)} = 0,9688 > 0,9267 = \gamma_{(\alpha,n)}^{\min}$).

Следовательно, гипотеза о стационарности ряда не отвергается.

Таблица 1- Результаты теста Аббе для ряда первых разностей

Тест Аббе	
n	502
σ^2	0,1169
q^2	0,1133
$\gamma^{(n)}$	0,9688
$\gamma_{(\alpha,n)}^{\min}$	0,9267

По результатам полупараметрических и непараметрических тестов можно сказать, что ряд $\Delta^1 y_t$ является стационарным.

Поведение стационарного ряда $\Delta^1 y_t$ характеризуется следующими выборочной и частной автокорреляционными функциями (ВАФ и ЧАФ):

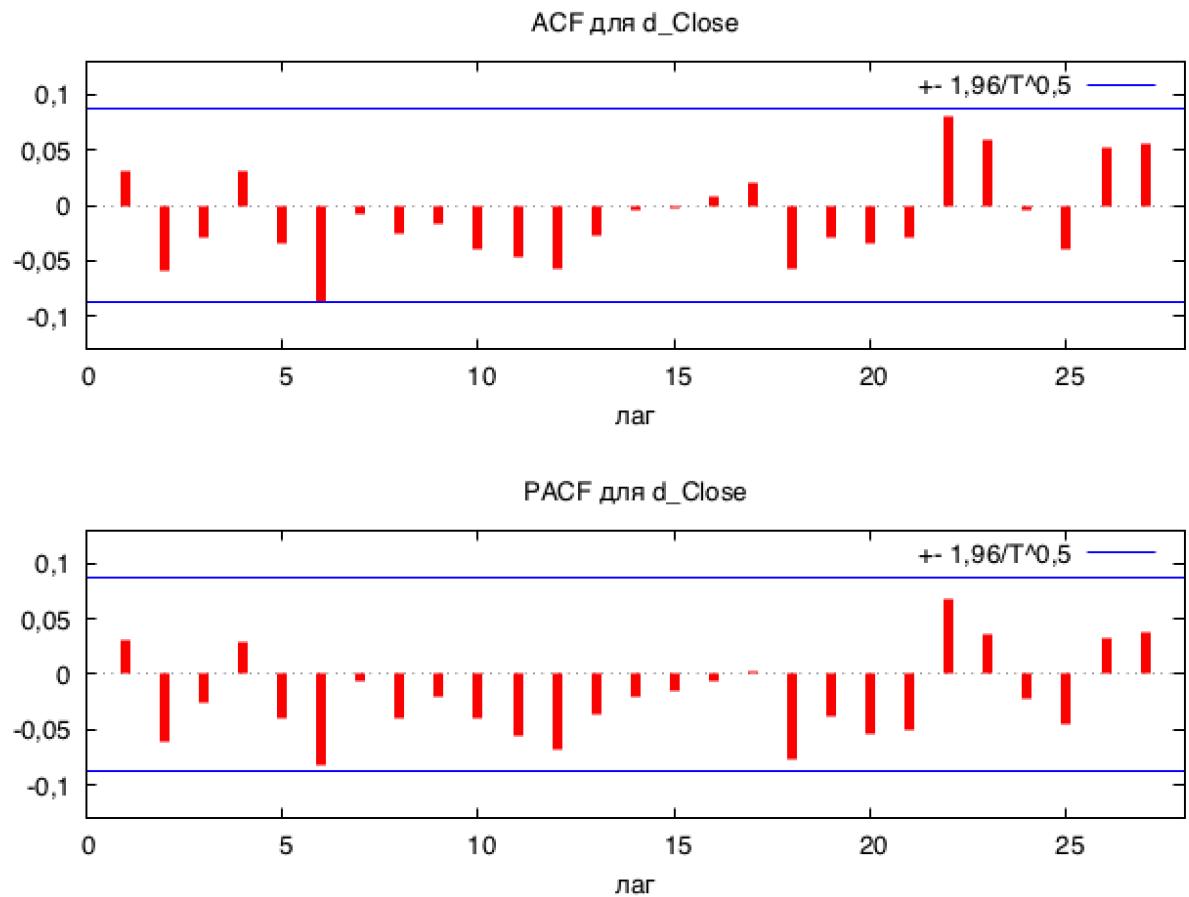


Рисунок 13 – ВАФ и ЧАФ ряда первых разностей

Так как все коэффициенты являются незначимыми, то данный ряд является «белым шумом», а значит его нельзя замоделировать с помощью ARIMA.

Таким образом, одна из предпосылок необходимых для принятия гипотезы о стационарности временного ряда, не выполняется. Взятие первых разностей не привело ряд к стационарному виду, поэтому попробуем другие способы приведения к стационарности: взятие вторых разностей, логарифмирование исходного ряда, взятие логдоходности, взятие вторых разностей от логарифмированного ряда.

Итак, среди вышеперечисленных преобразований эффективным оказалось взятие вторых разностей ряда. Таким образом, будем рассматривать ряд $\Delta^2 y_t = \Delta^1 y_t - \Delta^1 y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$.

График ряда вторых разностей имеет вид:

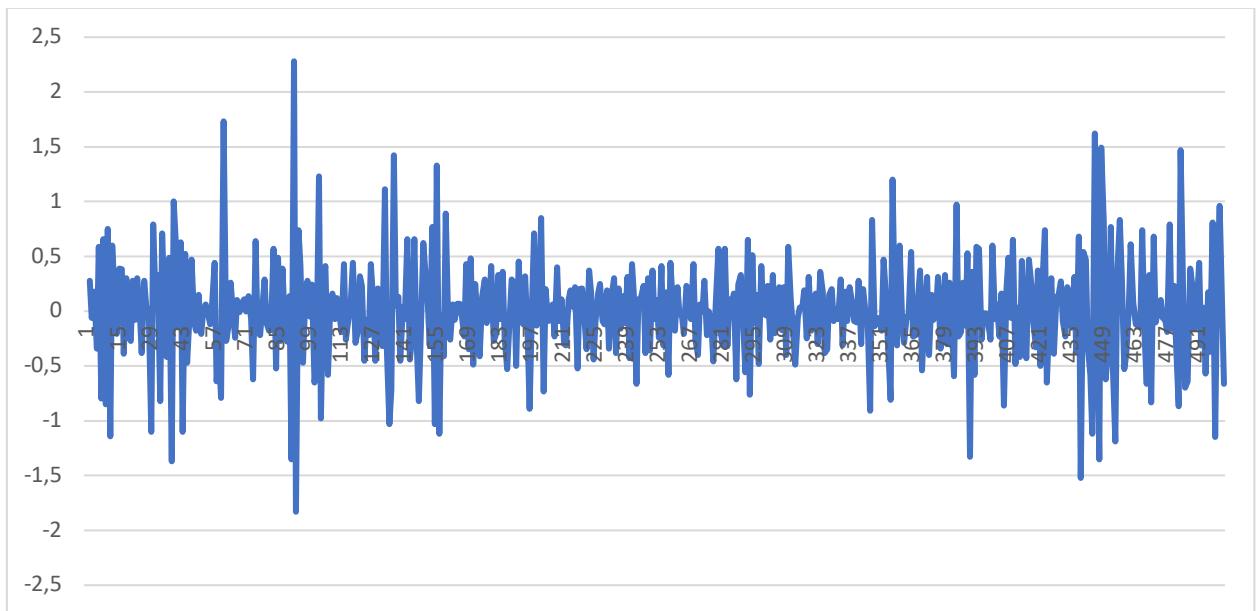


Рисунок 14 – Ряд вторых разностей

Проверим данный ряд на стационарность. Перед этим исследуем близость распределения ряда к нормальному.

Описательная статистика, наблюдения 2016-05-02 – 2018-04-30
для переменной 'd_d_Close' (использовано 501 наблюдений)

Среднее	7,9836e-05
Медиана	-0,0099950
Минимум	-1,8300
Максимум	2,2800
Стандартное отклонение	0,47651
Вариация	5968,6
Асимметрия	0,15914
Эксцесс	2,2680
5% перцентиль	-0,80900
95%-перцентиль	0,74000
Interquartile range	0,53000
Пропущенные наблюдения	2

Рисунок 15 – Характеристики ряда вторых разностей

Мы видим, что коэффициент эксцесса выходят за пределы, установленные для нормального закона распределения. Дополнительно рассмотрим распределение ряда:

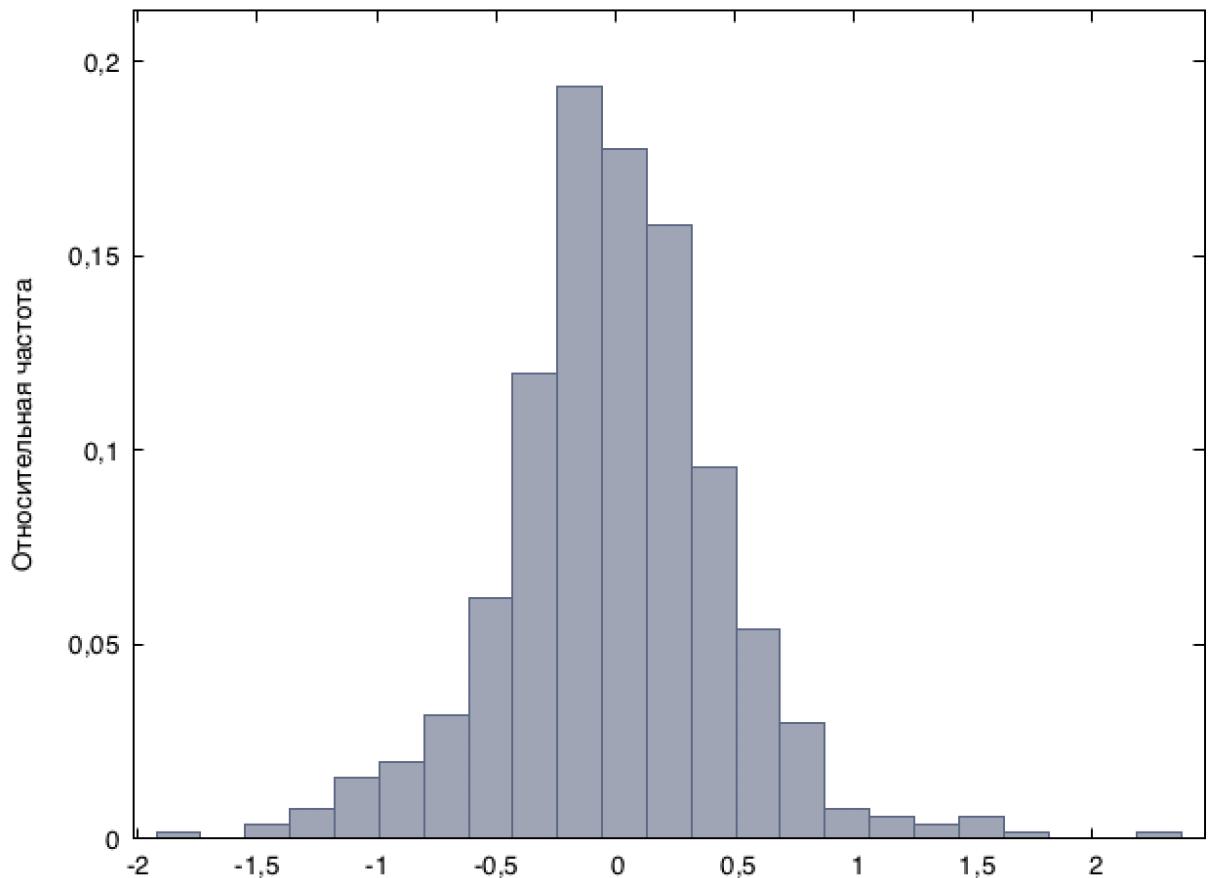
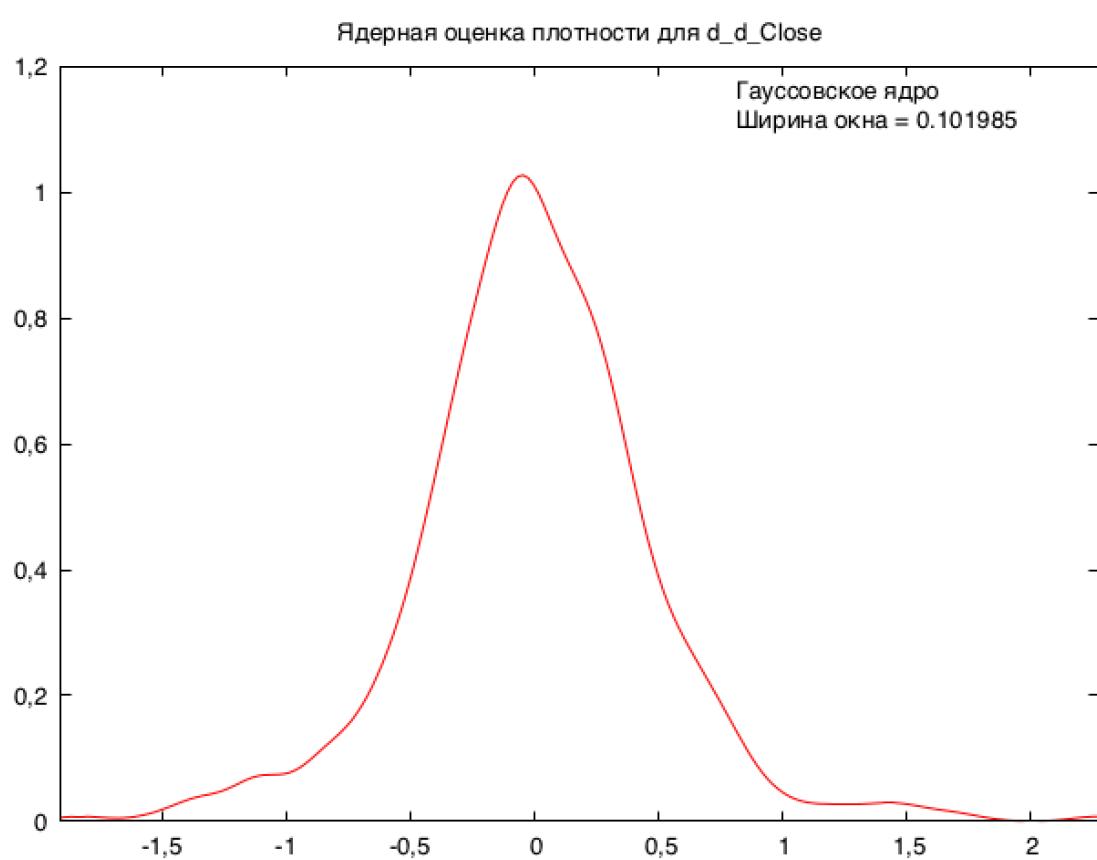
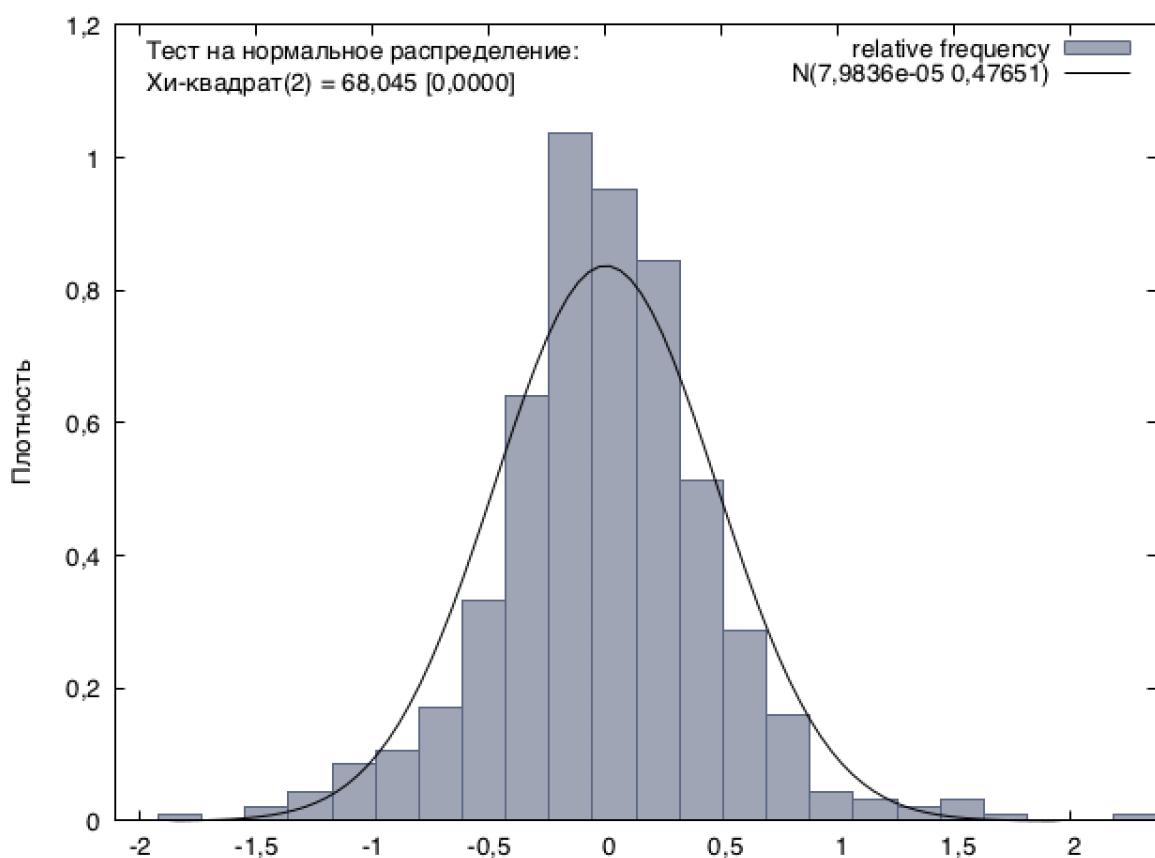


Рисунок 16 – Распределение ряда вторых разностей



Наш ряд не является нормально распределенным, поэтому следует использовать полупараметрические и непараметрические тесты.

Для начала используем непараметрические тесты Манна-Уитни и Сиджела-Тьюкки.

По описанному ранее алгоритму получаем расчётное значение для теста Манна-Уитни, которое составляет -0,148. Так как оно принадлежит интервалу [-1,96;1,96], то гипотеза о постоянстве математического ожидания принимается с вероятностью 95% [5, с.54].

Расчётное значение для теста Сиджела-Тьюкки составляет -0,113, данное значение не принадлежит отрезку [-1,96;1,96], следовательно, гипотеза о постоянстве дисперсии принимается с вероятностью 95%.

Теперь проведём полупараметрические тесты:

- тест Аббе,
- критерий восходящих и нисходящих серий,
- медианный критерий.

Тест Аббе

Таблица 2 - Тест Аббе для ряда вторых разностей

Тест Аббе	
n	501
σ^2	0,227
q^2	0,3293
$\gamma^{(n)}$	1,4505
$\gamma_{(\alpha,n)}^{\min}$	0,9266

Так как расчётное значение больше критического значения ($\gamma^{(n)} = 1,4505 > 0,9266 = \gamma_{(\alpha,n)}^{\min}$), то в рамках данного теста ряд признаётся стационарным.

Медианный критерий

Для начала определяем медиану для ряда $\Delta^2 y_t$, далее находим разницу между каждым элементом и медианой и выделяем серии следующим образом: если полученная разница больше нуля, элементу соответствует знак «+», если полученная разница меньше нуля, элементу соответствует знак «-», если разница равна нулю, элементу ничего не соответствует. Последовательности «+» и «-» образуют серии. Рассчитываем следующие показатели:

- общее количество серий: $v(n)$
- длина самой большой серии: τ_{\max}

Тогда критерий выглядит следующим образом:

$$\begin{cases} v(n) > 0,5(n + 2 - 1,96\sqrt{n-1}) \\ \tau_{\max} < 1,43 \ln(n+1) \end{cases}$$

Таблица 3- Медианный критерий для ряда вторых разностей

Медианный критерий			
Медиана	-0,01		
$v(n)$	295	>	229,587
τ_{\max}	5	<	8,893
n	501		

Если хоть одно из условий нарушено, то гипотеза о постоянстве математического ожидания отвергается. Однако в нашем случае оба условия выполняются, а значит, в рамках данного теста, мы делаем вывод о том что ряд $\Delta^2 y_t$ стационарен по математическому ожиданию.

Тест восходящих и нисходящих серий

Для проведения данного теста сначала необходимо построить ряд $\Delta^2 y_{t+1} - \Delta^2 y_t$. В этом ряду положительным элементам поставим в соответствие знак «+», а отрицательным - знак «-», если разница равна нулю,

элементу ничего не соответствует. Последовательности «+» и «-» образуют серии. Рассчитываем следующие показатели:

- общее количество серий: $v(n)$
- длина самой большой серии: τ_{\max}

Тогда критерий выглядит следующим образом:

$$\begin{cases} v(n) > \frac{1}{3}(2n - 1) - 1,96\sqrt{\frac{16n - 29}{90}} \\ \tau_{\max} < \tau_0 \end{cases}$$

В зависимости от объема выборки n определяется τ_0 :

Таблица 4 – Зависимость τ_0 от объема выборки

$n < 26$	$\tau_0 = 5$
$26 < n < 156$	$\tau_0 = 6$
$156 < n < 1170$	$\tau_0 = 7$
$n > 2000$	$\tau_0 = 8$

Таблица 5 – Критерий восходящих и нисходящих серий для ряда вторых разностей

Критерий восходящих и нисходящих серий			
$\vartheta(n)$	360	>	314,554
τ_{\max}	4	<	7
n	500		

Если хоть одно из условий нарушено, то гипотеза о постоянстве дисперсии отвергается. Однако в нашем случае оба условия выполняются, а значит, в рамках данного теста, мы делаем вывод о том ряд $\Delta^2 y_t$ имеет постоянную дисперсию.

По результатам полупараметрических и непараметрических тестов можно сказать, что ряд $\Delta^2 y_t$ является стационарным.

Поведение стационарного ряда $\Delta^2 y_t$ характеризуется следующими выборочной и частной автокорреляционными функциями (ВАФ и ЧАФ):

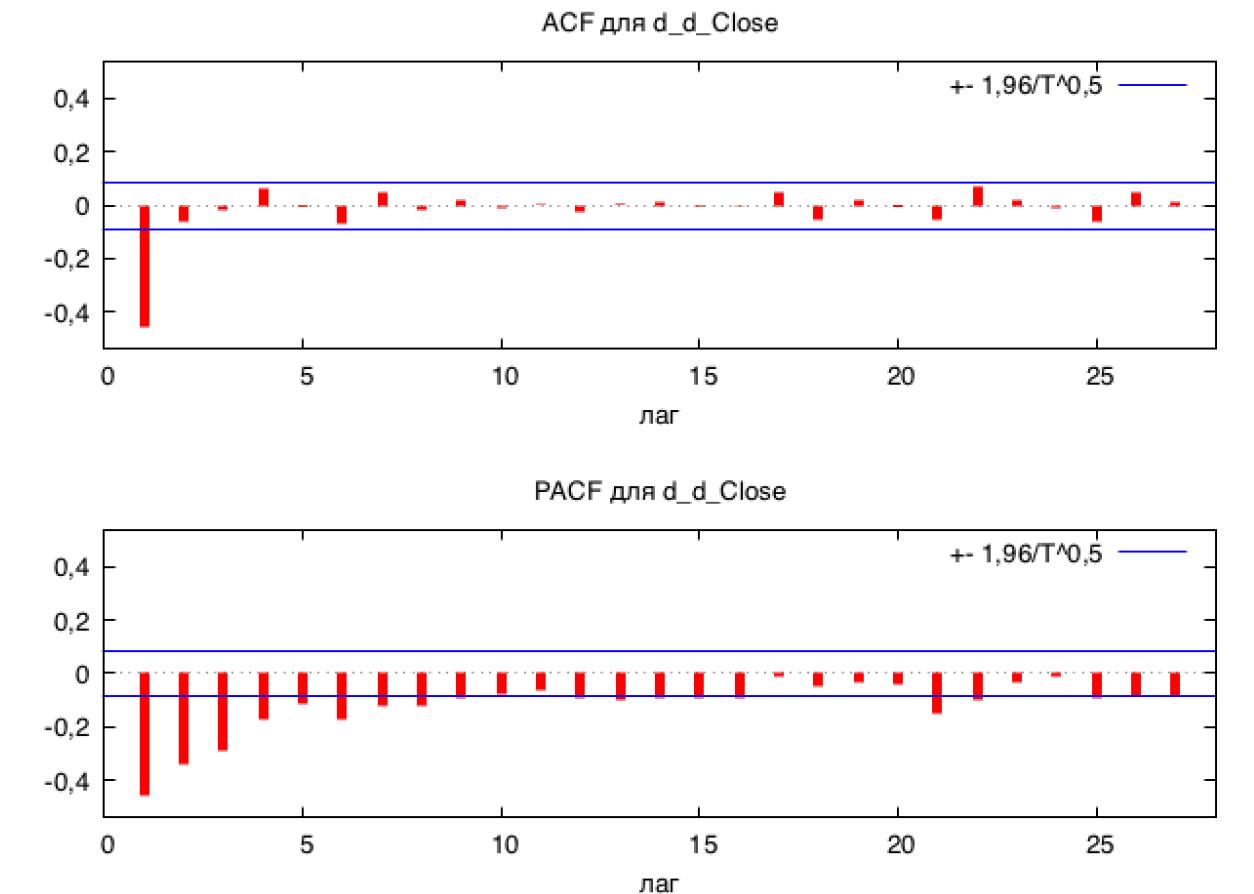


Рисунок 19 – ВАФ и ЧАФ ряда вторых разностей

Данный ряд не является «белым шумом», а значит его можно замоделировать с помощью ARIMA. График ВАФ и ЧАФ внешне напоминает MA(1) с положительным коэффициентом ($\beta_1 > 0$).

1.3 Построение модели ARIMA(0,2,1)

С помощью прикладного пакета Gretl получим оценки коэффициентов для модели скользящего среднего первого порядка.

	Коэффициент	Ст. ошибка	z	P-значение
const	0,000147583	0,000230774	0,6395	0,5225
theta_1	-0,987244	0,00695002	-142,0	0,0000 ***
Среднее зав. перемен	0,000080	Ст. откл. зав. перемен	0,476511	
Среднее инноваций	0,009105	Ст. откл. инноваций	0,345307	
Лог. правдоподобие	-178,1646	Крит. Акаике	362,3292	
Крит. Шварца	374,9790	Крит. Хеннана-Куинна	367,2925	
Действ. часть Мним. часть Модуль Частота				
MA				
Корень 1	1,0129	0,0000	1,0129	0,0000

Рисунок 20 – Оценка модели MA(1) для ряда вторых разностей

Первый коэффициент ВАФ является значимым и составляет -0,4528. Таким образом, основное ограничение для моделей скользящего среднего $|\rho_1| < 0,5$ не нарушено. Также важно отметить, что коэффициент при первом лаге статистически значим и составляет $\beta_1 = 0,9872$, а стандартная ошибка равна 0,007. Данное значение коэффициента удовлетворяет ограничению $|\beta_1| < 1$.

Для проверки обратимости процесса найдём корни характеристического уравнения:

$$1 - 0,9872z = 0$$

Решив уравнение получим: $z = 1,0129$.

Обратимость есть, так как значение находится за пределами единичного круга.

Таким образом, модель MA(1) имеет вид:

$$z_t = \beta_0 - \beta_1 * \varepsilon_{t-1} + \varepsilon_t,$$

$$\text{где } z_t = \Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2}$$

Рассчитаем ряд модельных значений:

$$\hat{z}_t = 0,00015 - 0,9872 * \varepsilon_{t-1} \quad (1)$$

Здесь мы сталкиваемся с проблемой того, что нам неизвестен ряд ошибки ε_{t-1} . Поэтому, прежде чем строить модель, восстановим сначала ряд ε_t согласно следующей схеме:

$$\begin{aligned}\varepsilon_1 &= z_1 \\ \varepsilon_2 &= z_2 - 0,00015 + 0,9872 * \varepsilon_1 \\ \varepsilon_3 &= z_3 - 0,00015 + 0,9872 * \varepsilon_2 \\ &\dots \\ \varepsilon_n &= z_n - 0,00015 + 0,9872 * \varepsilon_{n-1}\end{aligned}$$

Затем из восстановленного ряда ε_t мы получаем необходимые значения ε_{t-1} и переходим к построению \hat{z}_t по формуле (1).

После этого разынтегрируем ряд, сопоставим исходные значения с модельными и протестируем ошибки на свойства "белого шума". Опишем подробнее процесс разынтеграции модели MA(1):

$$\begin{aligned}z_t &= \beta_0 - \beta_1 * \varepsilon_{t-1} + \varepsilon_t \\ z_t &= \Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2} \\ y_t &= 2y_{t-1} - y_{t-2} + \beta_0 - \beta_1 * \varepsilon_{t-1} + \varepsilon_t \\ y_t &= 2y_{t-1} - y_{t-2} + 0,00015 - 0,9872 * \varepsilon_{t-1} + \varepsilon_t\end{aligned}$$

В итоге получен ряд модельных значений \hat{y}_t . Нанесём полученные значения на график исходных данных, чтобы визуально оценить насколько хорошо модель ARIMA(0,2,1) описывает реальные данные.



Рисунок 21 – Наблюдаемые и расчетные по ARIMA(0,2,1) значения исходного ряда

Визуально оценив график, можно сказать, что модельные значения повторяют исходный ряд. Коэффициент детерминации $R^2 = 0,9622$ говорит о том, что 96,22% вариации курса акций компании в течение исследуемого интервала времени объясняется построенной моделью ARIMA(0,2,1).

1.4 Тестирование остатков модели ARIMA(0,2,1)

Однако для того, чтобы уйти в краткосрочный прогноз с моделью ARIMA(0,2,1), необходимо в том, что остатки являются "белым шумом". То есть для ряда остатков должны выполняться следующие условия:

1. Равенство нулю математического ожидания.
2. Постоянство дисперсии.
3. Отсутствие автокорреляции.

Вычтем модельный значения из исходного ряда, чтобы найти остатки и нанесём их на график:

$$\varepsilon_t = y_t - \hat{y}_t$$

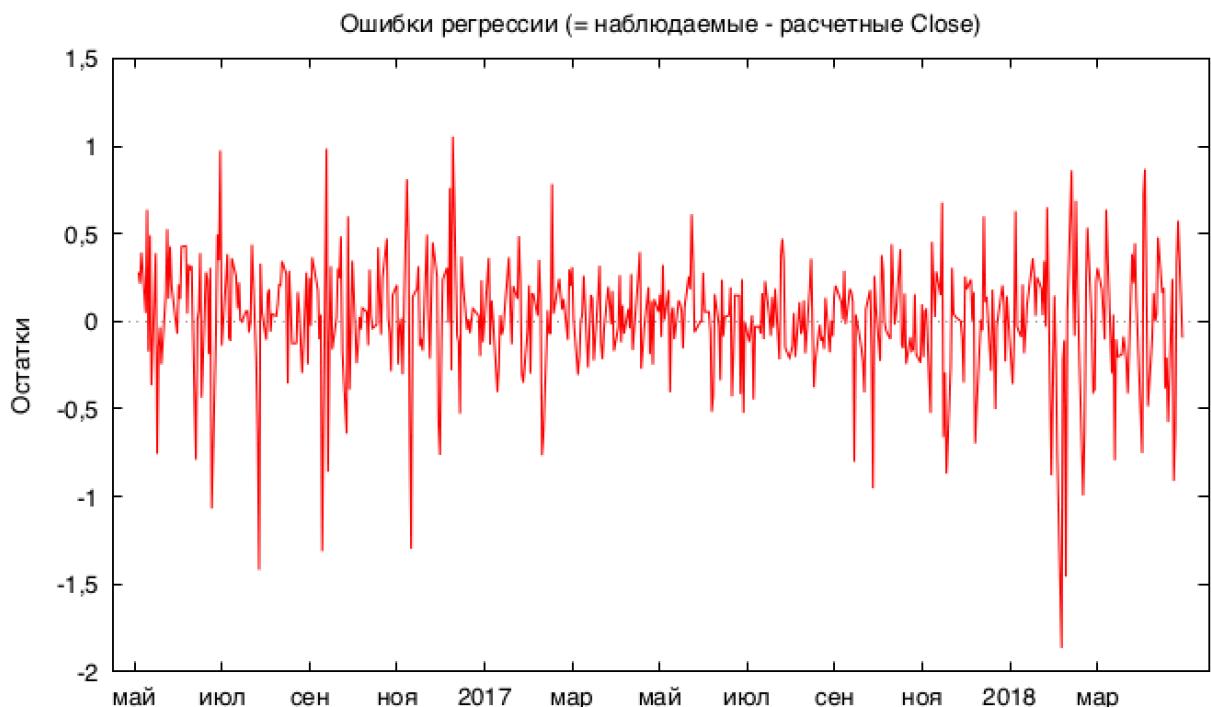


Рисунок 22 – Остатки модели ARIMA(0,2,1)

Визуально, судя по графику, можно выдвинуть гипотезу о том, что ряд ошибок является «белым шумом». Рассмотрим характеристики ряда остатков, а также его распределение.

Описательная статистика, наблюдения 2016–05–02 – 2018–04–30
для переменной 'err' (использовано 501 наблюдений)

Среднее	0,0091053
Медиана	0,032126
Минимум	-1,8637
Максимум	1,0551
Стандартное отклонение	0,34553
Вариация	37,948
Асимметрия	-0,96619
Эксцесс	3,6401
5% перцентиль	-0,63941
95%-перцентиль	0,48651
Interquartile range	0,34783
Пропущенные наблюдения	2

Рисунок 23 – Характеристики ряда ошибок модели ARIMA(0,2,1)

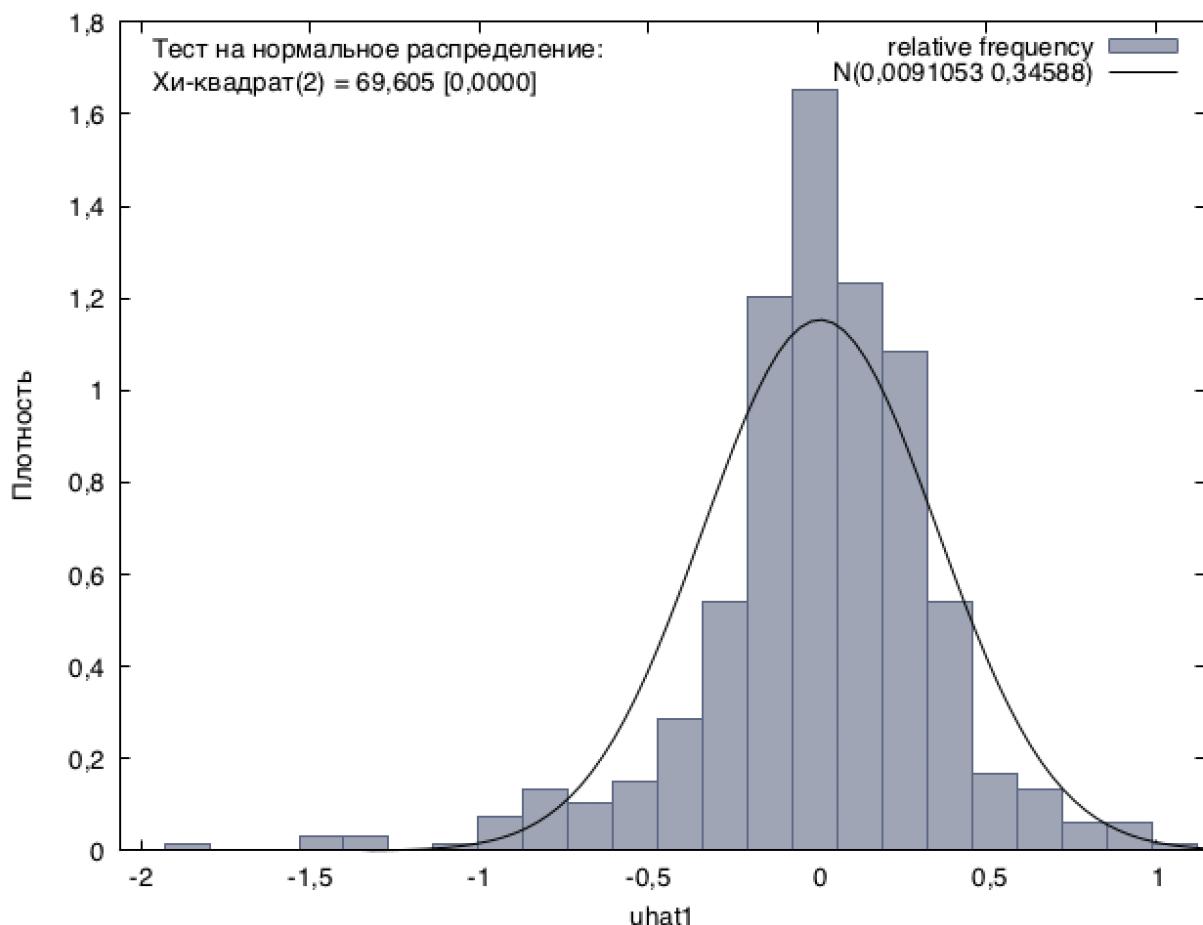


Рисунок 24 – Распределение ряда остатков модели ARIMA(0,2,1)

Математическое ожидание ряда остатков модели составляет 0,009, что статистически равно 0 по критерию Стьюдента.

Нулевая гипотеза: генеральное среднее = 0
 Размер выборки: $n = 501$
 Выборочное среднее = 0,00910535, ст. отклонение = 0,345532
 Тестовая статистика: $t(500) = (0,00910535 - 0)/0,0154372 = 0,58983$
 Двухстороннее р-значение = 0,5556
 (одностороннее = 0,2778)

Рисунок 25 – Проверка равенства нулю среднего ряда ошибок ARIMA(0,2,1)

Коэффициенты эксцесса не входит в диапазон допустимых значений для нормального закона распределения [-2;2]. Следовательно, используем тест Сиджела-Тьюки для проверки постоянства дисперсии ряда ошибок. Расчётное значение теста Сиджела-Тьюки составило -0,204, что входит в интервал [-1,96; 1,96]. Таким образом, гипотеза о постоянстве дисперсии не отвергается с вероятностью 95%.

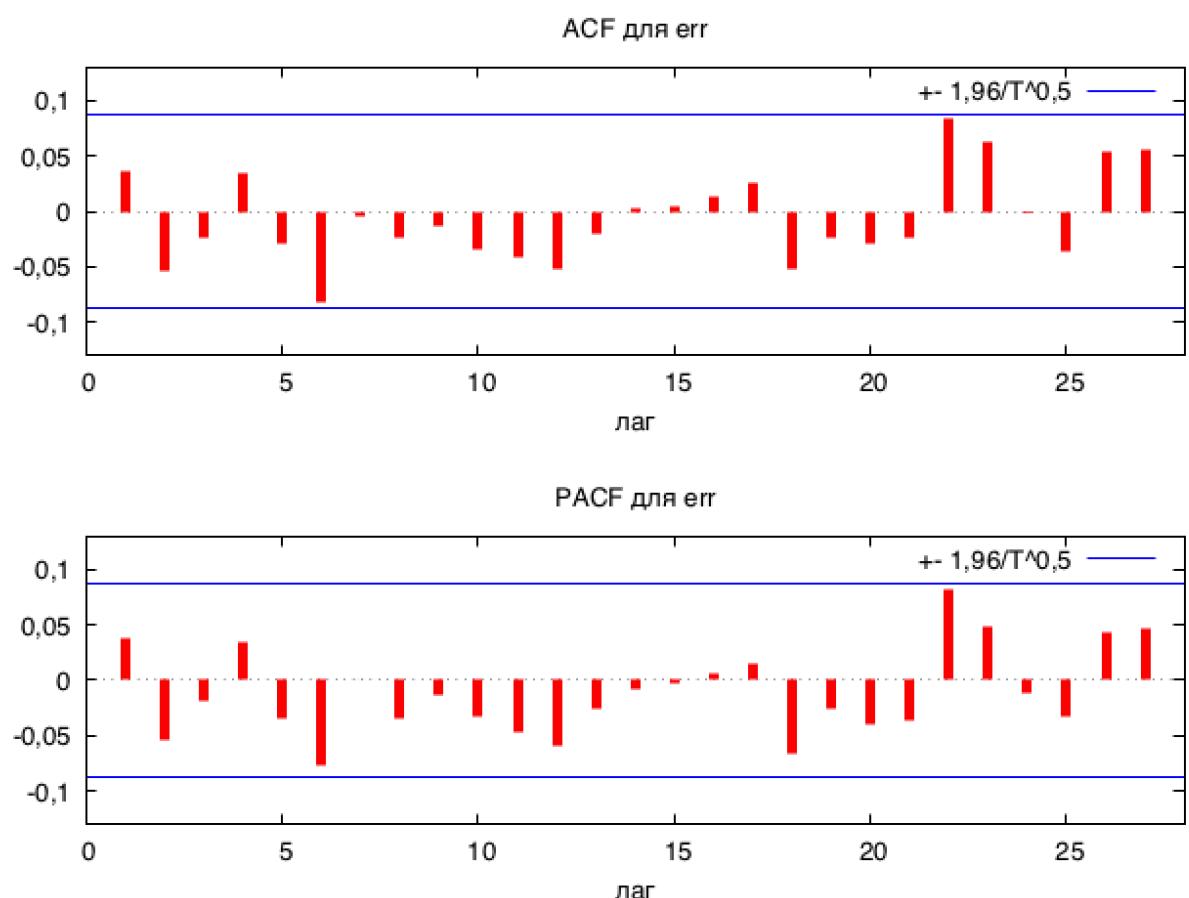


Рисунок 26 – ВАФ и ЧАФ ряда ошибок модели ARIMA(0,2,1)

На основании графиков ВАФ и ЧАФ можно сделать предположение, что остатки являются белым шумом. Однако для того, чтобы проверить данную гипотезу, необходимо провести ряд тестов.

Графики ЧАФ и ВАФ свидетельствуют о отсутствии автокорреляции, докажем это с помощью критерия Бокса-Пирса, который используется для проверки сразу всей последовательности выборочных коэффициентов автокорреляции: ρ_1, \dots, ρ_k [6,с.48].

Статистика Q рассчитывается по формуле:

$$Q = T \cdot \sum_{i=1}^{k=\max \text{ lag}} \rho_i^2,$$

где ρ_1, \dots, ρ_k - выборочные коэффициенты автокорреляции;

T - объем выборки;

k – максимальный лаг.

Если $Q < \chi_{\alpha,k}^2$, то в нашей последовательности отсутствует автокорреляция. В нашем случае, расчётное значение $Q = 71,11$ и $\chi_{0,05;100}^2 = 124,34$, следовательно, условие выполняется, а значит, в ряду ошибок автокорреляция отсутствует.

Также проверить автокорреляцию можно с помощью теста Льюнга-Бокса.

Статистика теста Льюнга-Бокса имеет следующий вид:

$$\tilde{Q} = \frac{T(T+2)}{T-k} \sum_{i=1}^k p_i^2$$

Расчётное значение статистики Льюнга-Бокса 83,93, что меньше табличного значения хи-квадрат распределения при уровне значимости 0,05 и со степенями свободы $k=100$. Следовательно, гипотеза о равенстве 0 коэффициентов автокорреляции не отвергается с вероятностью 95%.

Таким образом, очевидно, что ряд ошибок после разынтеграции модели ARIMA(0,2,1) обладает свойствами «белого шума».

1.5 Прогнозирование в краткосрочной перспективе

Построим прогноз на 2 дня: 01.05.2018 и 02.05.2018. Для прогнозирования будем использовать разынтегрированную модель МА(1).

Таблица 6 – Краткосрочный прогноз ARIMA(0,2,1)

	Границы 95% доверительного интервала прогноза			
	Прогноз	Нижняя	Верхняя	Факт
01.05.2018	43,20	42,53	43,88	42,59
02.05.2018	43,19	42,23	44,16	42,06

Для 95% доверительных интервалов, $z(0,025) = 1,96$

	Close	Предсказание	Ст. ошибка	95% доверительный интервал
2018-04-24	43,070000	43,980730		
2018-04-25	42,430000	43,059260		
2018-04-26	42,750000	42,411381		
2018-04-27	43,310001	42,735848		
2018-04-30	43,209999	43,303321		
2018-05-01	42,590000	43,202276	0,345307	42,525486 – 43,879066
2018-05-02	42,060000	43,194700	0,491463	42,231451 – 44,157949

Рисунок 27 – Прогнозы модели ARIMA(0,2,1)

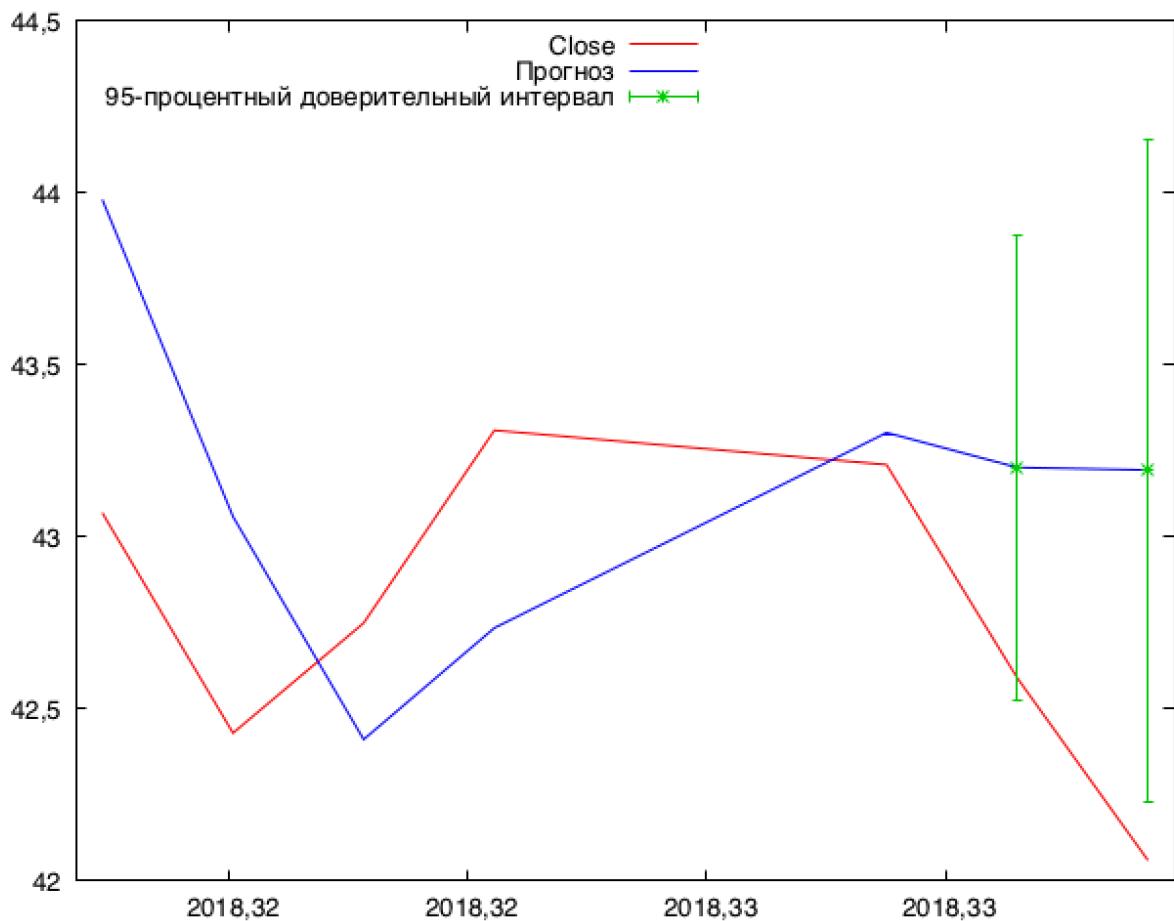


Рисунок 28 – Прогнозные и фактические значения

Несмотря на правильную спецификацию модели, мы можем увидеть, что фактическое значение от 02.05.2018 не попадает в 95% доверительный интервал.

Рассмотрим график ряда квадратов остатков и коррелограмму этого ряда.

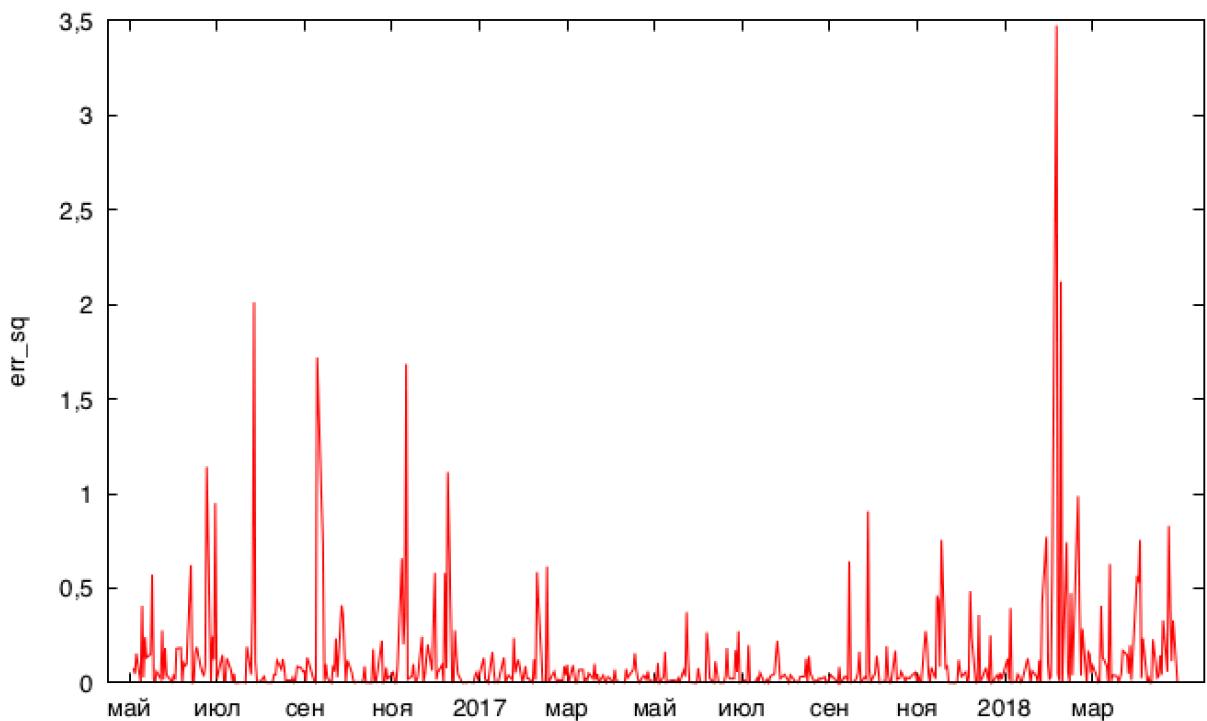


Рисунок 29 – Квадраты остатков модели ARIMA(0,2,1)

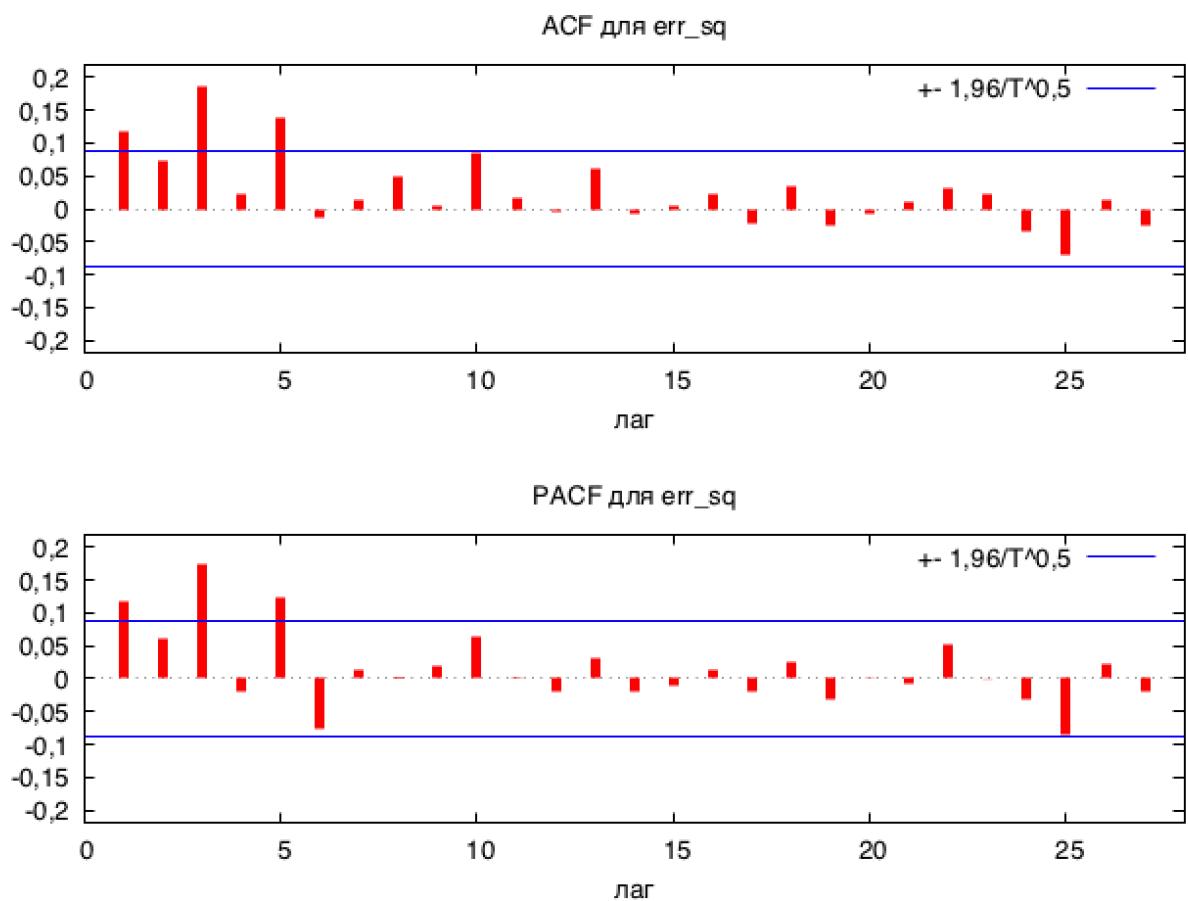


Рисунок 30 – ВАФ и ЧАФ ряда квадратов ошибок модели ARIMA(0,2,1)

Как мы можем увидеть, в ряде квадратов остатков есть автокорреляция. Следовательно, остатки модели можно домоделировать с помощью моделей ARCH/GARCH. Об этом также говорит следующий тест:

Тест на наличие ARCH процессов порядка 5

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
alpha(0)	0,0724617	0,0156395	4,633	4,62e-06	***
alpha(1)	0,105574	0,0446399	2,365	0,0184	**
alpha(2)	0,0207196	0,0448935	0,4615	0,6446	
alpha(3)	0,169843	0,0442684	3,837	0,0001	***
alpha(4)	-0,0314516	0,0449317	-0,7000	0,4843	
alpha(5)	0,131428	0,0447031	2,940	0,0034	***

Нулевая гипотеза: ARCH процессы отсутствуют

Тестовая статистика: LM = 31,9698

p-значение = P(Хи-квадрат(5) > 31,9698) = 6,02374e-06

Рисунок 31 – Тест на наличие ARCH процессов

Глава 2 Моделирование и прогнозирование курса акций «The Coca-Cola Company» в среднесрочной перспективе

2.1 Анализ исходных данных

Для моделирования и прогнозирования курса акций «The Coca-Cola Company» в среднесрочной перспективе необходимо использовать агрегированные по неделям данные за период с 29.04.2013 по 30.04.2018. В результате общий объём выборки агрегированного ряда составляет 262 значения.

Для начала рассмотрим график изменения курса акций в понедельном разрезе.

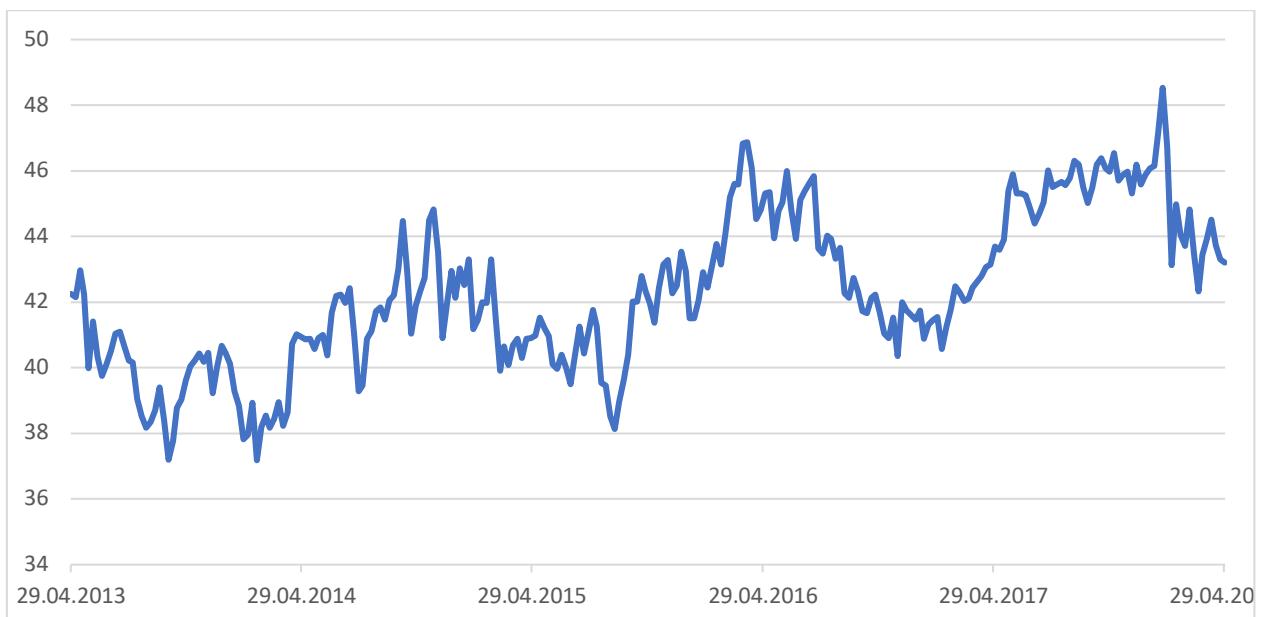


Рисунок 32 – Агрегированный по неделям ряд

На графике очень хорошо видна сезонность. На основе этих агрегированных значений построим несколько моделей и выберем лучшую.

Перейдём к первичной обработке данных и проанализируем исходный ряд.

Описательная статистика, наблюдения 2013-04-29 – 2018-04-30
для переменной 'Close' (использовано 262 наблюдений)

Среднее	42,292
Медиана	42,055
Минимум	37,180
Максимум	48,530
Стандартное отклонение	2,3382
Вариация	0,055286
Асимметрия	0,14879
Эксцесс	-0,69284
5% перцентиль	38,452
95%-перцентиль	46,095
Interquartile range	3,2875
Пропущенные наблюдения	0

Рисунок 33 – Характеристики агрегированного по неделям ряда

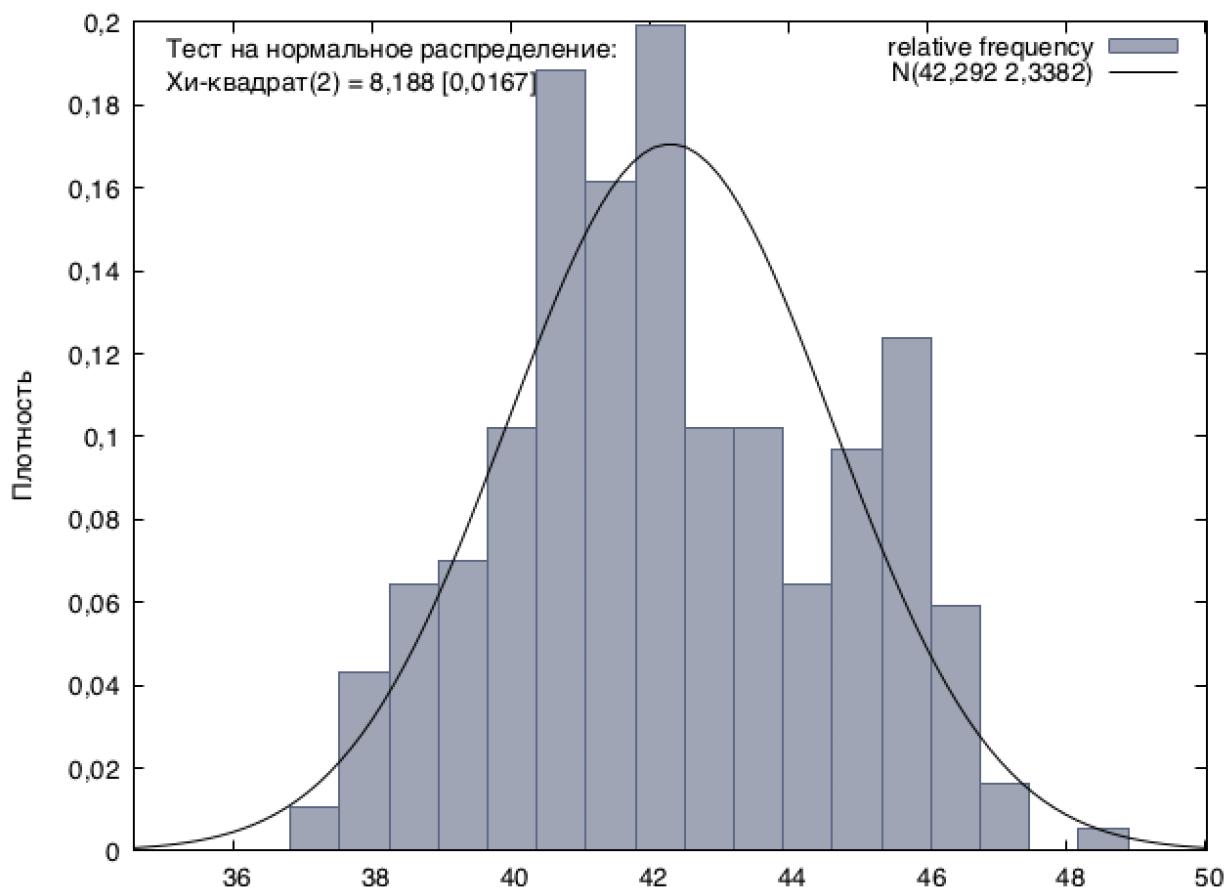


Рисунок 34 – Проверка нормальности агрегированного по неделям ряда

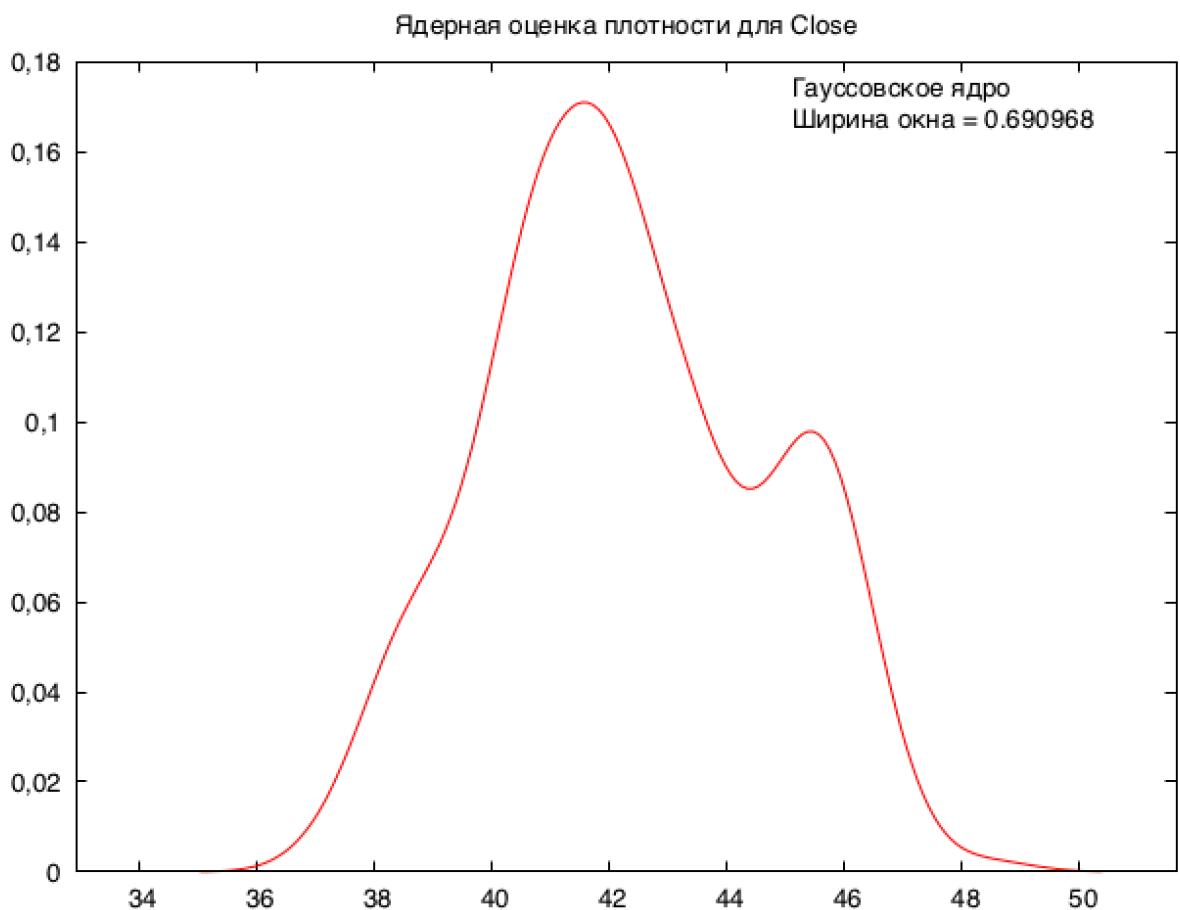


Рисунок 35 – Ядерная оценка плотности агрегированного по неделям ряда

График плотности говорит о наличии бимодальности. При этом коэффициенты асимметрии и эксцесса находится в допустимых пределах для нормального распределения от -2 до 2. Проанализируем выборочную и частную автокорреляционные функции рядов недельных значений.

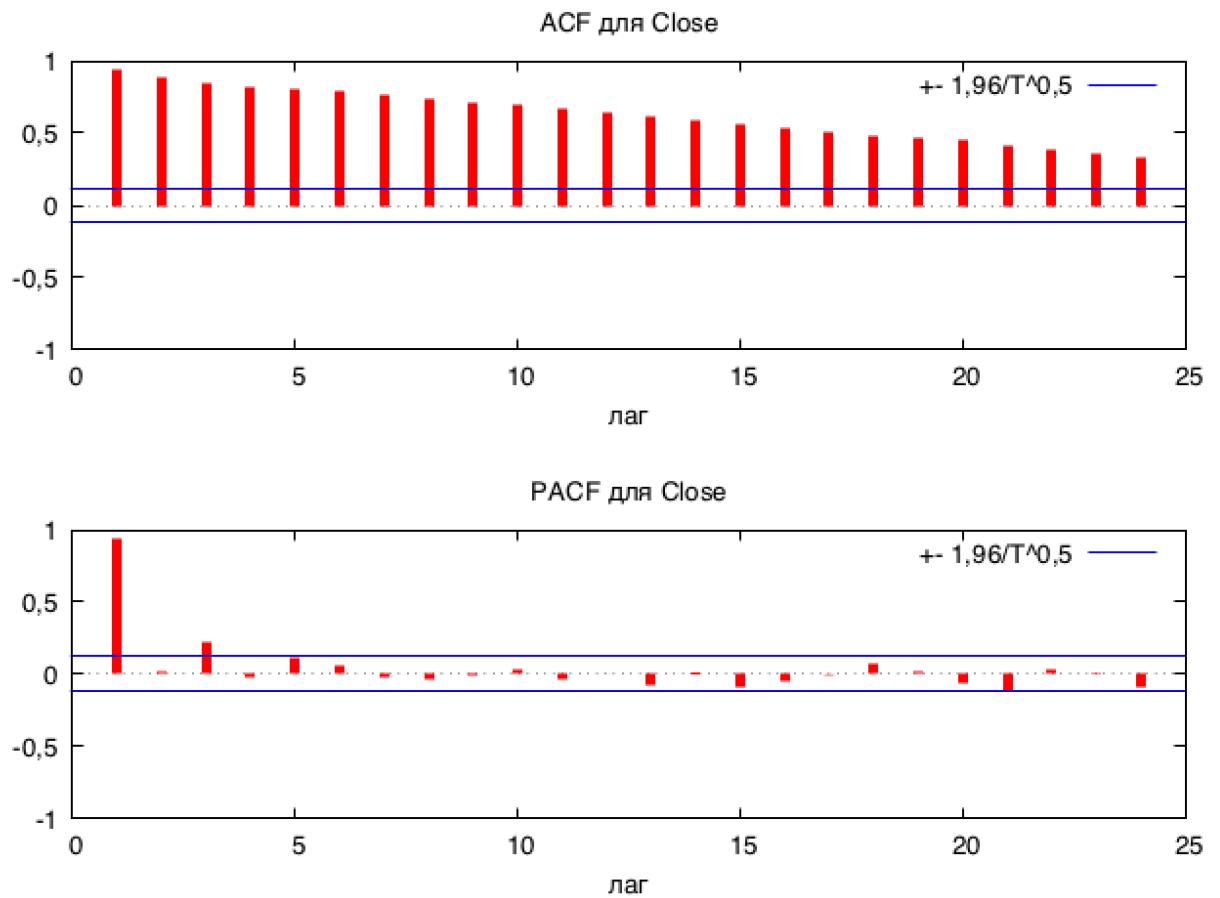


Рисунок 36 – ВАФ и ЧАФ агрегированного по неделям ряда

Значения выборочных коэффициентов автокорреляции имеют тенденцию к убыванию. Частная автокорреляционная функция показывает, что значения ряда зависят от предыдущих с лагом 1 и лагом 3, так как они являются значимыми.

2.2 Моделирование агрегированных недельных данных

Модель типа ARIMA можно использовать только в краткосрочной перспективе, поэтому альтернативой является модель с лаговой зависимой переменной.

На основании частной автокорреляционной функции можно предположить, что наш ряд зависит от предыдущих значений с лагом 3. Было рассмотрено несколько вариантов моделей с лаговыми переменными y_{t-1} , y_{t-2} и y_{t-3}

Было построено четыре модели:

- С одной лаговой переменной;
- С двумя лаговыми переменными;
- С тремя лаговыми переменными;
- С 1-м и 3-м лагом.

Из полученных моделей наиболее подходящей является модель с тремя лаговыми переменными.

Модель 1: МНК, использованы наблюдения 2013-05-20:2018-04-30 (T = 259)
Зависимая переменная: Close

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	2,03027	0,932179	2,178	0,0303	**
Close_1	0,919985	0,0609502	15,09	3,23e-37	***
Close_2	-0,189468	0,0829456	-2,284	0,0232	**
Close_3	0,221530	0,0610147	3,631	0,0003	***
Среднее зав. перемен	42,29027	Ст. откл. зав. перемен	2,351324		
Сумма кв. остатков	165,3907	Ст. ошибка модели	0,805351		
R-квадрат	0,884051	Испр. R-квадрат	0,882687		
F(3, 255)	648,0818	P-значение (F)	5,9e-119		
Лог. правдоподобие	-309,4221	Крит. Акаике	626,8441		
Крит. Шварца	641,0714	Крит. Хеннана-Куинна	632,5643		
Параметр rho	0,006305	h-статистика Дарбина	0,521686		

Рисунок 37 – Модель с тремя лаговыми переменными

Коэффициенты при факторах y_{t-1} , y_{t-2} и y_{t-3} являются значимыми на 95% уровне доверия, а значит, наша гипотеза, выдвинутая на основании

частной автокорреляционной функции, подтвердилась. Таким образом, полученная модель имеет следующий вид:

$$\hat{y}_t = 2,03 + 0,92y_{t-1} - 0,19y_{t-2} + 0,22y_{t-3}$$

Отметим также, что данная модель имеет достаточно высокий коэффициент детерминации $R^2 = 0,8841$ ($R_{adj}^2 = 0,8827$). То есть 88,41% вариации процесса объясняется построенной моделью.

Построив модельные и фактические значения на одном графике, мы также можем убедиться в высокой степени сходства процесса и построенной модели.



Рисунок 38 – Наблюдаемые и расчетные значения агрегированного по неделям ряда

Однако для того, чтобы уйти в среднесрочный прогноз, нам необходимо проверить остатки выбранной модели на свойства "белого шума".

2.3 Тестирование остатков модели с лаговыми переменными

График остатков, полученный после построения модели, выглядит следующим образом:



Рисунок 39 – Остатки модели с лаговыми переменными

Проведём статистическую обработку ряда остатков.

Описательная статистика, наблюдения 2013-04-29 – 2018-04-30
для переменной 'err' (использовано 259 наблюдений)

Среднее	-1,7064e-14
Медиана	0,12583
Минимум	-3,1436
Максимум	1,8994
Стандартное отклонение	0,80066
Вариация	4,6921e+13
Асимметрия	-0,59449
Эксцесс	0,90191
5% перцентиль	-1,4685
95%-перцентиль	1,2235
Interquartile range	0,87807
Пропущенные наблюдения	3

Рисунок 40 – Характеристики остатков модели с лаговыми переменными

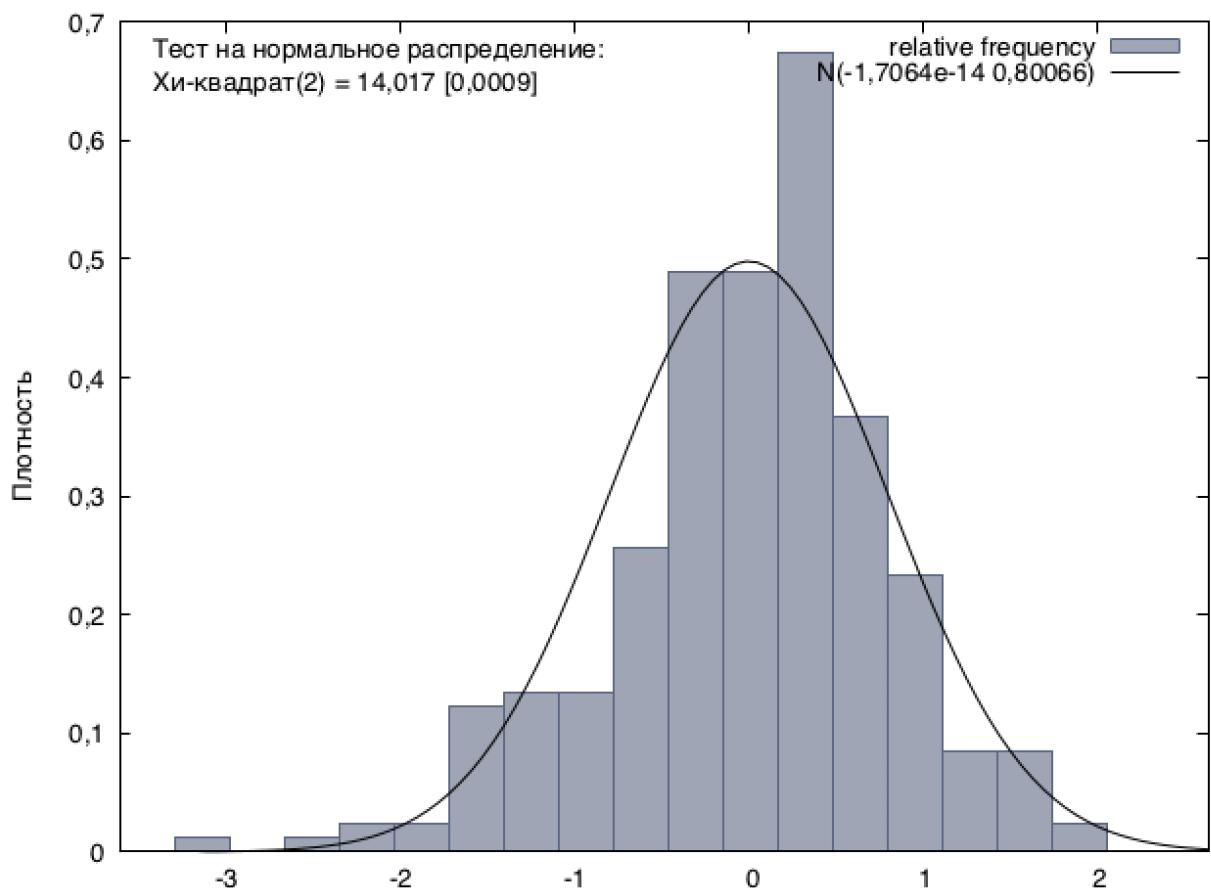


Рисунок 41 – Проверка нормальности ряда остатков модели с лаговыми переменными

Коэффициенты асимметрии и эксцесса попадают в интервал допустимых значений, но распределение ряда ошибок не является нормальным.

Тест на нормальное распределение err:

Тест Дурника–Хансена (Doornik–Hansen) = 14,0172, р-значение 0,000904059

Тест Шапиро–Уилка (Shapiro–Wilk W) = 0,975486, р-значение 0,000192717

Тест Лиллифорса (Lilliefors) = 0,072193, р-значение ~ 0

Тест Жака–Бера (Jarque–Bera) = 24,0342, р-значение 6,04001e–06

Рисунок 42 - Проверка нормальности ряда остатков модели с лаговыми переменными (Тесты)

Проверим ряд на постоянство математического ожидания и дисперсии непараметрическими тестами Манна–Уитни и Сиджела–Тьюки. Расчётные

значения статистик для данных тестов составили -1,866 и -1,255 соответственно. Данные значения попадают в интервал [-1,96;1,96]. Таким образом, по результатам данных тестов признаётся гипотеза о постоянстве математического ожидания и дисперсии. [4, с.197]

При этом отметим, что математическое ожидание ряда остатков статистически равно 0.

Нулевая гипотеза: генеральное среднее = 0

Размер выборки: n = 259

Выборочное среднее = -1,68003e-14, ст. отклонение = 0,800656

Тестовая статистика: t(258) = (-1,68003e-14 - 0)/0,0497503 = -3,37692e-13

Двухстороннее p-значение = 1

(одностороннее = 0,5)

Рисунок 43 – Проверка равенства нулю среднего ряда ошибок модели с лаговыми переменными

Далее протестируем ряд остатков на наличие автокорреляции с помощью статистических критериев Бокса-Пирса и Льюнга-Бокса. Расчётное значение статистики Бокса-Пирса равно 47,65, что меньше табличного значения хи-квадрат распределения при уровне значимости 0,05 и с k=65 степенями свободы (оно составляет 84,82). Расчётное значение статистики Льюнга-Бокса равно 64,11, что также меньше табличного значения хи-квадрат распределения. Это значит, что гипотеза о равенстве 0 коэффициентов автокорреляции принимается с вероятностью 95%. Об этом также свидетельствуют автокорреляционные функции.

Визуализируем выборочную и частную автокорреляционные функции ряда остатков модели с лаговыми переменными.

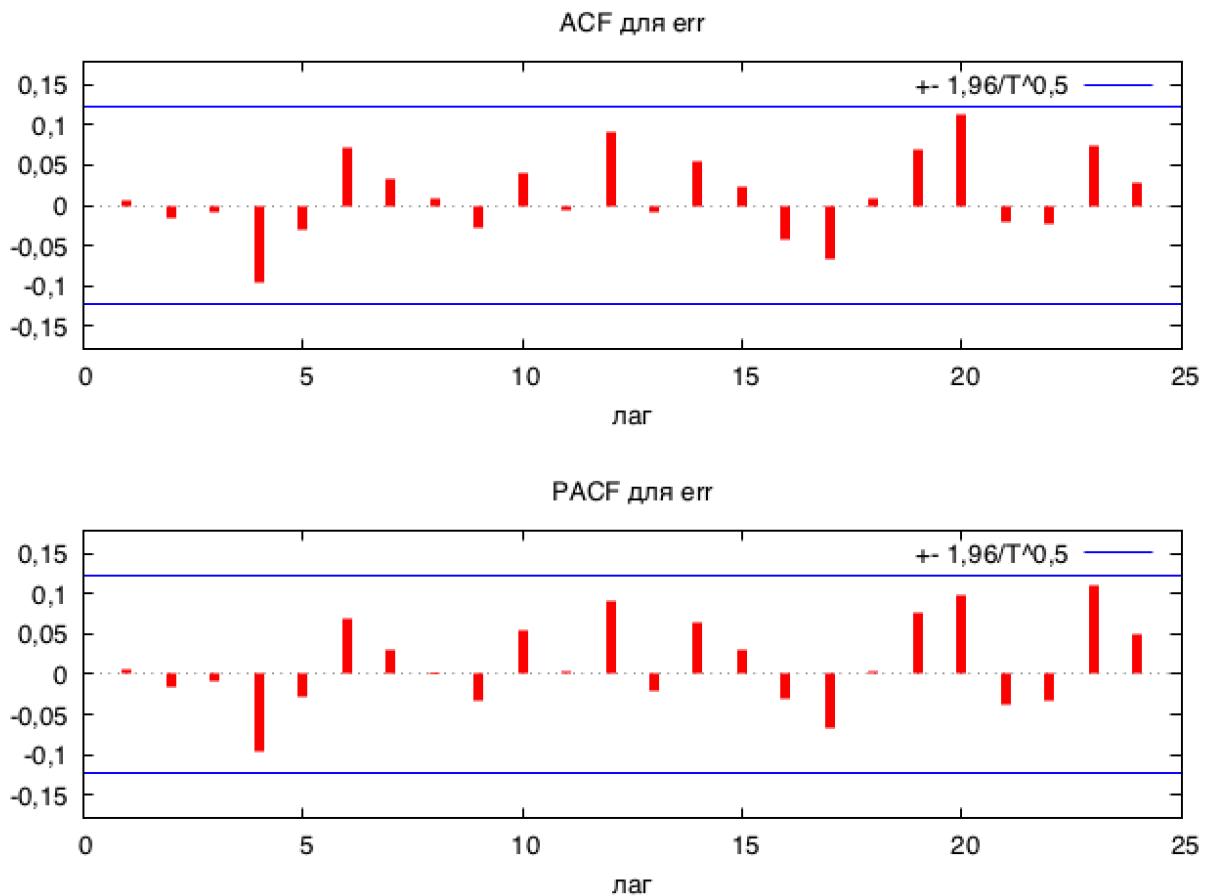


Рисунок 44 – ВАФ и ЧАФ ряда остатков модели с лаговыми переменными

Таким образом, проведя ряд тестов и проанализировав графики автокорреляционных функций, делаем вывод, что ряд остатков модели с тремя лаговыми переменными является «белым шумом», а это значит, что мы можем использовать данную модель в среднесрочном прогнозировании.

2.4 Прогнозирование в среднесрочной перспективе

Построим прогноз на 2 недели вперёд с помощью выбранной модели:

$$\hat{y}_t = 2,03 + 0,92y_{t-1} - 0,19y_{t-2} + 0,22y_{t-3}$$

Таблица 7 – Среднесрочный прогноз

	Границы 95% доверительного интервала прогноза			
	Прогноз	Нижняя	Верхняя	Факт
07.05.2018	43,27	41,68	44,85	42,14
14.05.2018	43,24	41,09	45,40	42,18

Для 95% доверительных интервалов, $t(255, 0,025) = 1,969$

	Close	Предсказание	Ст. ошибка	95% доверительный интервал
2018-04-02	43,919998	43,592715		
2018-04-09	44,509998	43,584764		
2018-04-16	43,740002	44,278398		
2018-04-23	43,310001	43,566776		
2018-04-30	43,209999	43,447774		
2018-05-07	42,140000	43,266668	0,805351	41,680681 – 44,852655
2018-05-14	42,180000	43,242492	1,094322	41,087431 – 45,397552

Рисунок 45 – Прогнозы модели с лаговыми переменными

В таблице представлен построен прогноз с доверительным интервалом на две недели вперед. Визуализируем данные, чтобы оценить точность прогноза.

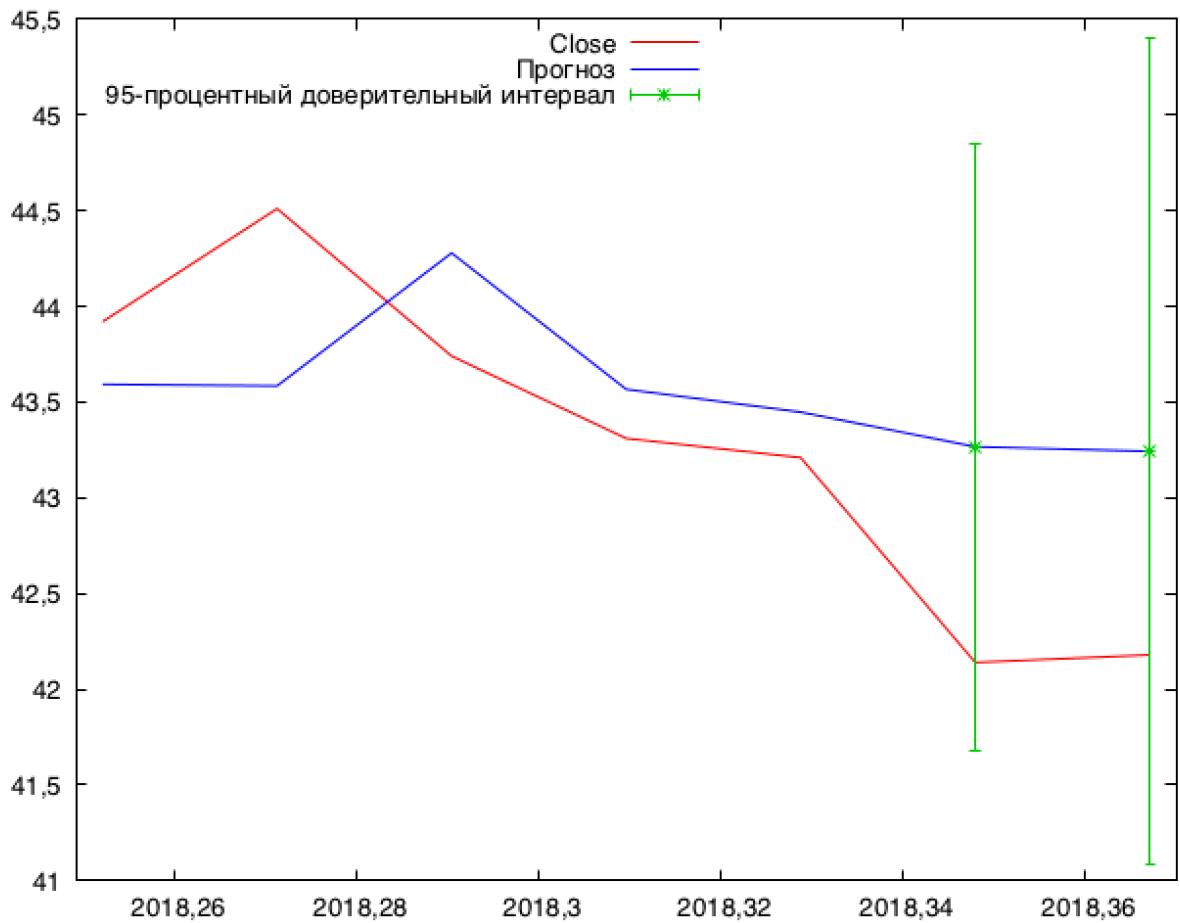


Рисунок 46 – Прогнозные и фактические значения

Различия между фактическими и прогнозными значениями есть, но фактические все же попадают в 95% доверительный интервал.

Глава 3 Моделирование и прогнозирование курса акций «The Coca-Cola Company» в долгосрочной перспективе

3.1 Анализ исходных данных

Для моделирования и прогнозирования динамики курса акций в долгосрочной перспективе были взяты агрегированные месячные данные за период с мая 2013 года по апрель 2018 года. Таким образом, ряд содержит 60 значения.

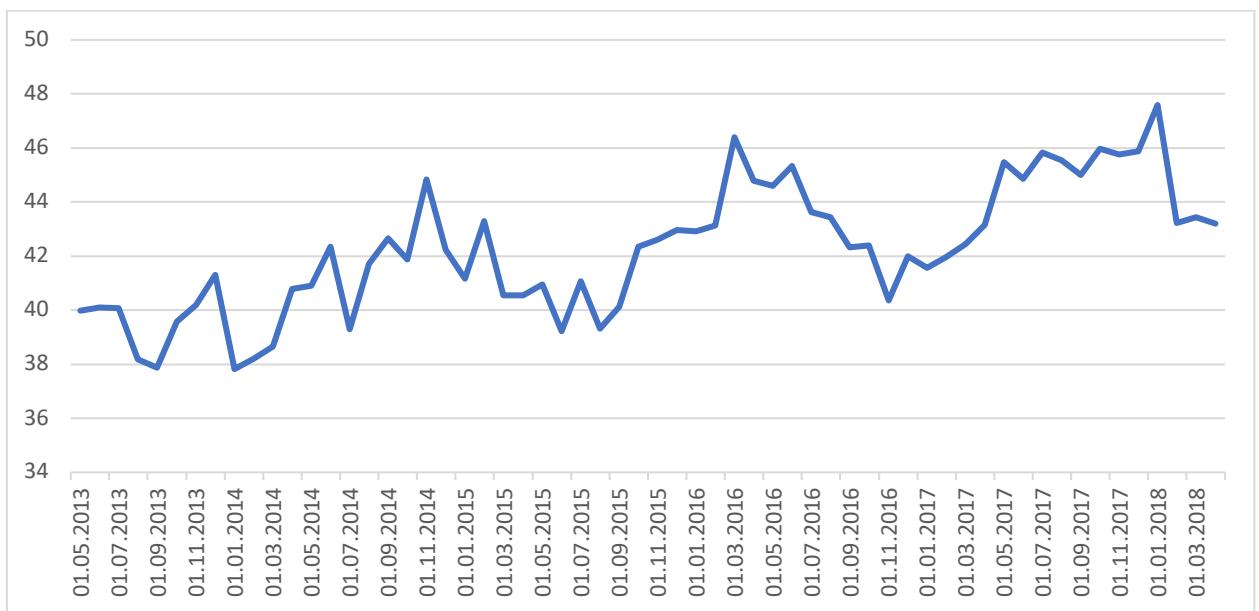


Рисунок 47 - Агрегированный по месяцам ряд

Из графика видно, что наш ряд имеет растущий тренд. Рассмотрим характеристики и распределение ряда.

Описательная статистика, наблюдения 2013:05 – 2018:04
для переменной 'Close' (использовано 60 наблюдений)

Среднее	42,252
Медиана	42,335
Минимум	37,820
Максимум	47,590
Стандартное отклонение	2,3765
Вариация	0,056246
Асимметрия	0,10114
Эксцесс	-0,70908
5% перцентиль	38,181
95%-перцентиль	45,975
Interquartile range	3,1800
Пропущенные наблюдения	0

Рисунок 48 – Характеристики агрегированного по месяцам ряда

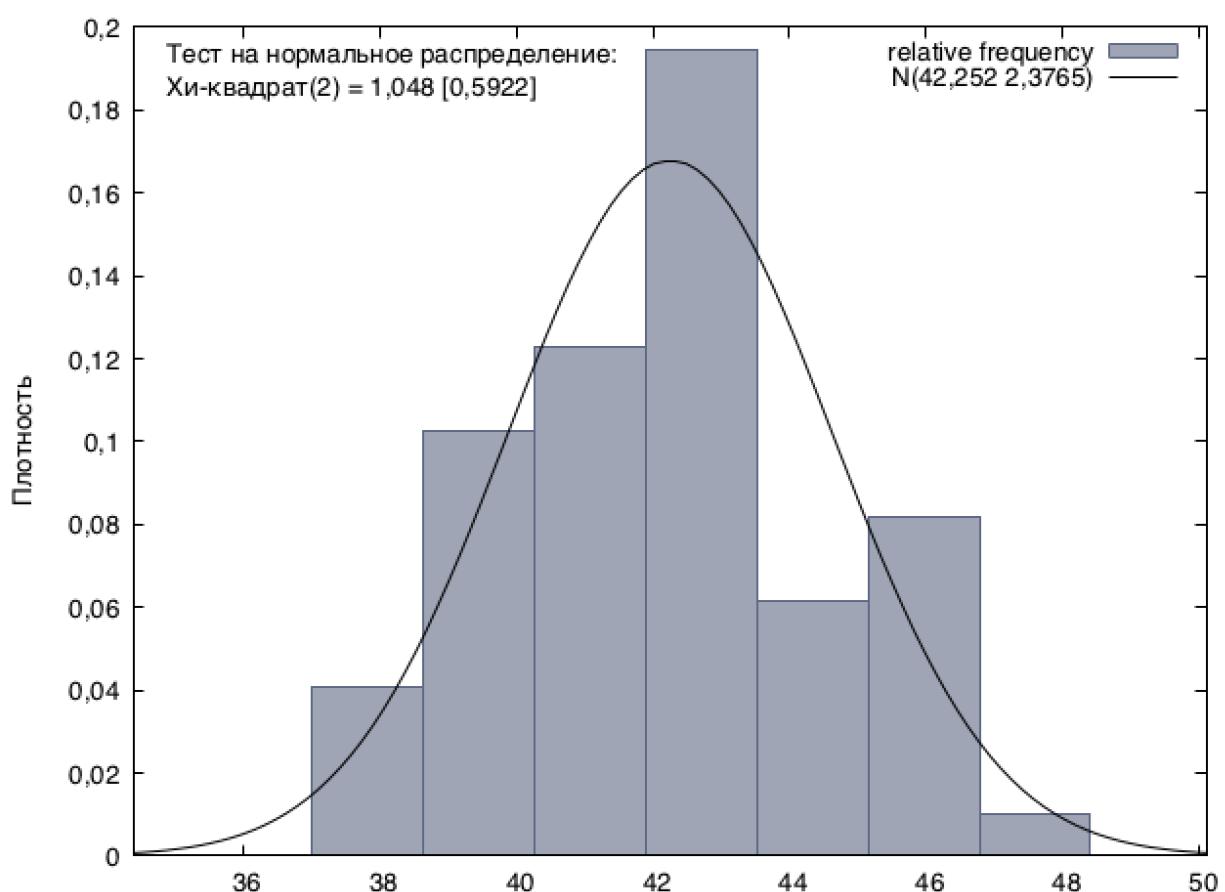


Рисунок 49 – Проверка нормальности агрегированного по месяцам ряда

Коэффициенты асимметрии и эксцесса попадают в интервал допустимых значений, и распределение нашего ряда является нормальным.

Тест на нормальное распределение Close:

Тест Дурника–Хансена (Doornik–Hansen) = 1,04775, р-значение 0,59222

Тест Шапиро–Уилка (Shapiro–Wilk W) = 0,979496, р-значение 0,4076

Тест Лиллифорса (Lilliefors) = 0,0748948, р-значение ≈ 0,54

Тест Жака–Бера (Jarque–Bera) = 1,3593, р-значение 0,506794

Рисунок 50 – Проверка нормальности агрегированного по месяцам ряда
(Тесты)

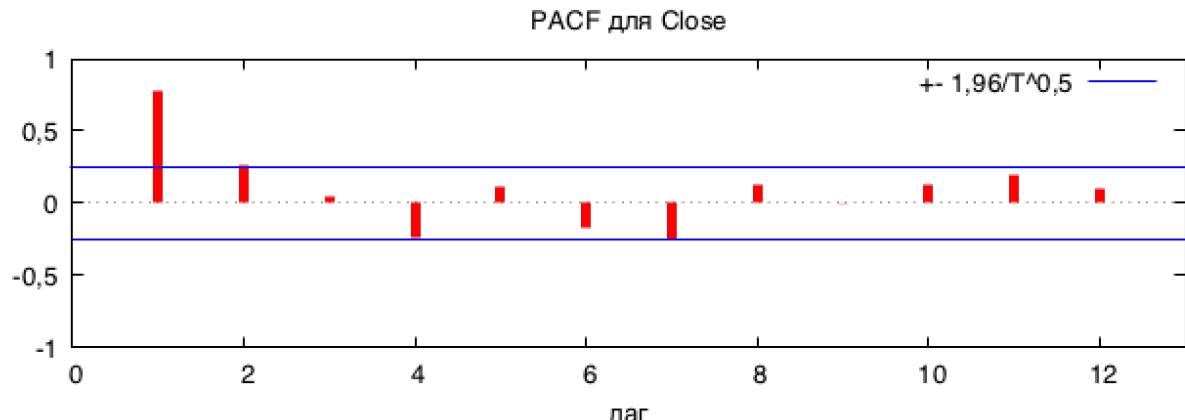
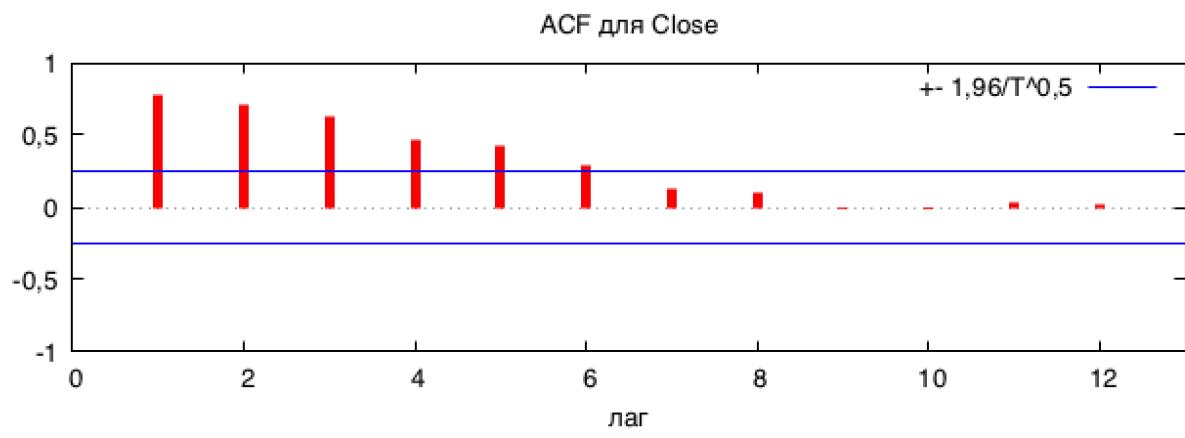


Рисунок 51 – ВАФ и ЧАФ агрегированного по месяцам ряда

3.2 Моделирование месячных данных с помощью сплайн функции

На основании частной автокорреляционной функции можно предположить, что наш ряд зависит от предыдущих значений с лагом 2. Было рассмотрено несколько вариантов моделей с лаговыми переменными y_{t-1} и y_{t-2} .

Было построено две модели:

- С одной лаговой переменной;
- С двумя лаговыми переменными;

Из полученных моделей наиболее подходящей является модель с двумя лаговыми переменными.

Модель 1: МНК, использованы наблюдения 2013:07–2018:04 (T = 58)

Зависимая переменная: Close

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	7,09747	3,61057	1,966	0,0544	*
Close_1	0,552135	0,129211	4,273	7,71e-05	***
Close_2	0,281632	0,128491	2,192	0,0326	**
Среднее зав. перемен	42,32741	Ст. откл. зав. перемен	2,381109		
Сумма кв. остатков	117,3322	Ст. ошибка модели	1,460587		
R-квадрат	0,636936	Испр. R-квадрат	0,623733		
F(2, 55)	48,24412	P-значение (F)	7,94e-13		
Лог. правдоподобие	-102,7309	Крит. Акаике	211,4617		
Крит. Шварца	217,6431	Крит. Хеннана–Куинна	213,8695		
Параметр rho	-0,019223	h-статистика Дарбина	-0,822734		

Рисунок 52 – Модель с двумя лаговыми переменными

Коэффициенты при факторах y_{t-1} и y_{t-2} являются значимыми на 95% уровне доверия, а значит, наша гипотеза, выдвинутая на основании частной автокорреляционной функции, подтвердилась. Таким образом, полученная модель имеет следующий вид:

$$\hat{y}_t = 7,097 + 0,552y_{t-1} + 0,281y_{t-2}$$

Коэффициент детерминации данной модели $R^2 = 0,6369$ ($R^2_{adj} = 0,6237$). То есть 63,69% вариации процесса объясняется построенной моделью.

Построим модельные и фактические значения на одном графике:



Рисунок 53 – Наблюдаемые и расчетные значения агрегированного по месяцам ряда

Чтобы уйти в среднесрочный прогноз, нам необходимо проверить остатки выбранной модели на свойства "белого шума".

3.3 Тестирование остатков сплайн функции

Изучим остатки модели на предмет "белого шума". Ряд остатков модели с двумя лаговыми переменными выглядит следующим образом:



Рисунок 54 – Остатки модели с лаговыми переменными

Определим основные характеристики ряда и посмотрим его распределение, а затем проведём ряд тестов на свойства "белого шума".

Описательная статистика, наблюдения 2013:05 – 2018:04 для переменной 'err' (использовано 58 наблюдений)

Среднее	3,0627e-15
Медиана	0,090683
Минимум	-3,4050
Максимум	3,3913
Стандартное отклонение	1,4347
Вариация	4,6846e+14
Асимметрия	-0,20546
Эксцесс	-0,096236
5% перцентиль	-2,7354
95%-перцентиль	2,5947
Interquartile range	1,8973
Пропущенные наблюдения	2

Рисунок 55 – Характеристики ряда остатков модели с лаговыми переменными

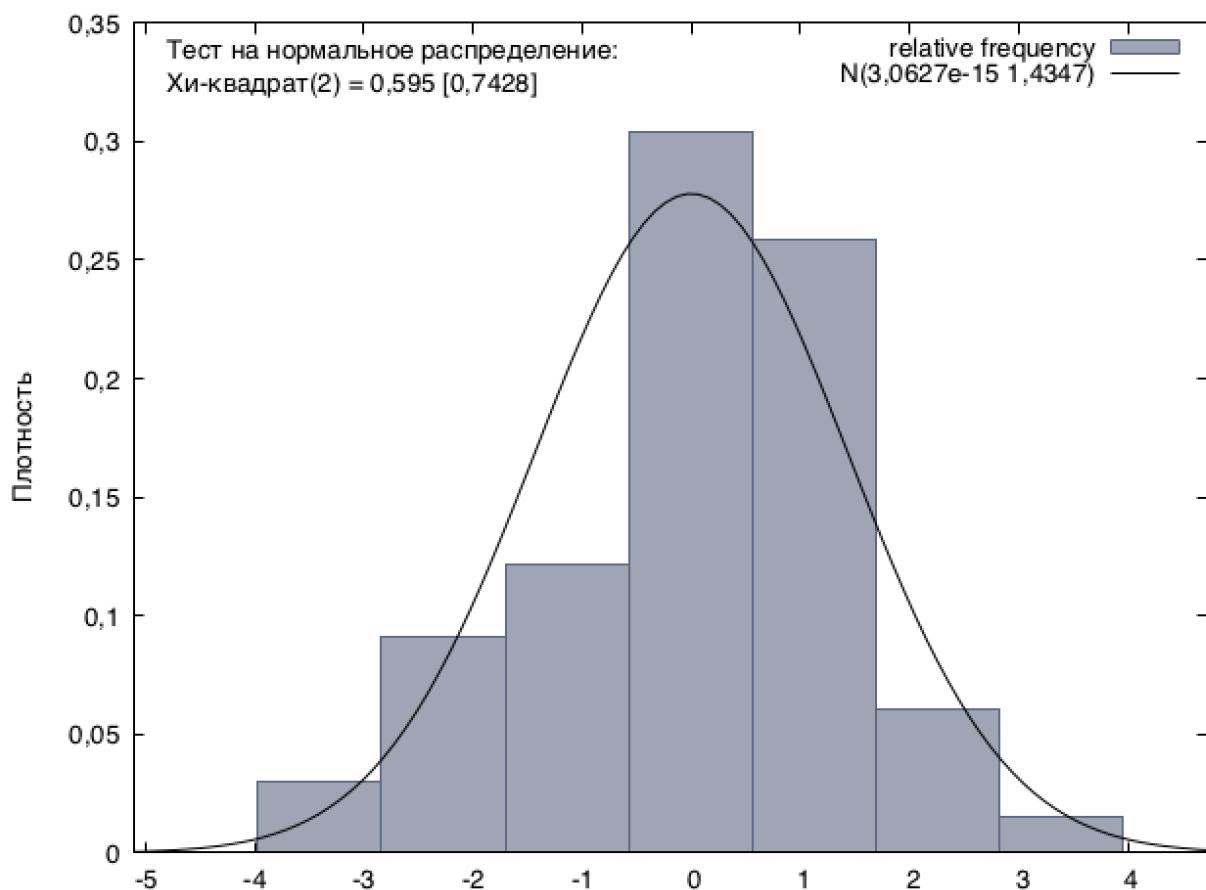


Рисунок 56 – Проверка нормальности ряда остатков модели с лаговыми переменными

Тест на нормальное распределение err:

Тест Дурника–Хансена (Doornik–Hansen) = 0,594572, р–значение 0,742832

Тест Шапиро–Уилка (Shapiro–Wilk W) = 0,985279, р–значение 0,704663

Тест Лиллифорса (Lilliefors) = 0,0901334, р–значение ~= 0,28

Тест Жака–Бера (Jarque–Bera) = 0,430444, р–значение 0,806363

Рисунок 57 - Проверка нормальности ряда остатков модели с лаговыми переменными (Тесты)

Ряд остатков распределён по нормальному закону, так как Коэффициенты асимметрии и эксцесса принадлежат интервалу допустимых для нормального закона распределения значений [-2;2]. Также нормальность наших остатков подтверждают тесты.

Проведём тестирование на постоянство дисперсии критерием Фишера:

$$F = \frac{S_1^2}{S_2^2} = 1,57 < F_{0,05;28;28} = 1,88$$

Так как расчётное значение меньше статистики Фишера, то дисперсия признаётся постоянной.

Математическое ожидание ряда остатков равно 3,0627E-15, что статистически равно 0.

Далее анализируется ряд остатков на наличие автокорреляции тестами Бокса–Пирса и **Льюнга–Бокса**. Расчётное значение статистики Бокса–Пирса равно 20,75, что меньше табличного значения хи-квадрат распределения (24,99). Это значит, что гипотеза о равенстве 0 коэффициентов автокорреляции не отвергается с вероятностью 95%.

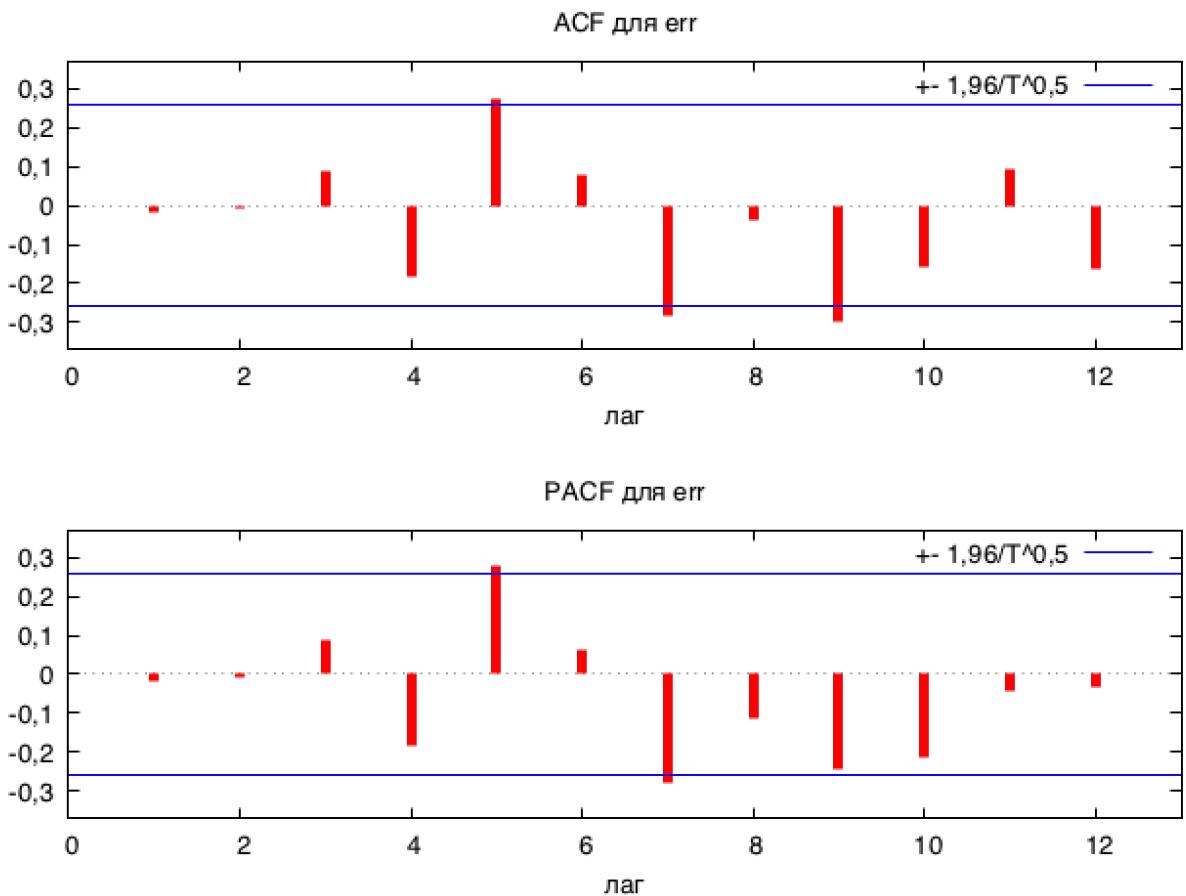


Рисунок 58 – ВАФ и ЧАФ ряда остатков модели с лаговыми переменными

Итак, остатки модели с двумя лаговыми переменными являются «белым шумом». Следовательно, с данной моделью может быть использована для долгосрочного прогнозирования.

3.4 Прогнозирование в долгосрочной перспективе

По выбранной модели построим прогноз на 2 месяца.

Таблица 8 – Долгосрочный прогноз

	Границы 95% доверительного интервала прогноза			
	Прогноз	Нижняя	Верхняя	Факт
Май 2018	43,19	40,26	46,11	43,00
Июнь 2018	43,11	39,77	46,46	43,86

Для 95% доверительных интервалов, $t(55, 0,025) = 2,004$

	Close	Предсказание	Ст. ошибка	95% доверительный интервал
2017:12	45,880001	45,318139		
2018:01	47,590000	45,319732		
2018:02	43,220001	46,294862		
2018:03	43,430000	44,363623		
2018:04	43,209999	43,248839		
2018:05	43,000000	43,186511	1,460587	40,259430 – 46,113592
2018:06	43,860000	43,111583	1,668431	39,767973 – 46,455193

Рисунок 59 – Прогнозы модели с лаговыми переменными

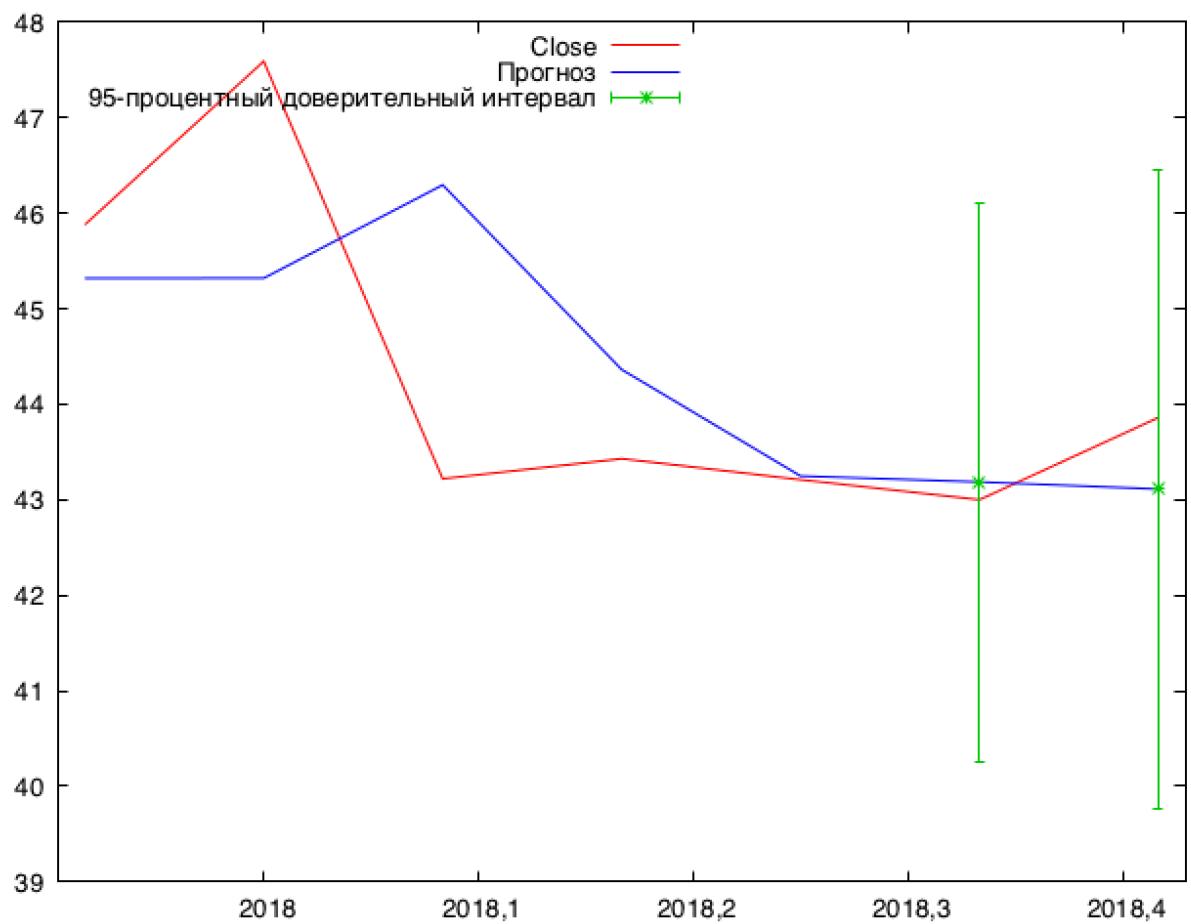


Рисунок 60 – Прогнозные и фактические значения

По графику видно, что прогнозные значения весьма близки к реальным, хотя и не полностью совпадают с ними. Модель верно описывает тенденцию процесса.

Заключение

В первом разделе в результате моделирования и прогнозирования курса акций компании «The Coca-Cola Company» были построены следующие модели:

1. Модель краткосрочного прогноза, построенная для ряда ежедневных данных за последние 5 лет. Была построена модель ARIMA(0,2,1) остатки который являлись «белым шумом». Прогноз по этой модели оказался весьма правдоподобным.

2. Модель с лаговыми зависимыми переменными, построенная на основе средненедельных данных. На основе данной модели был выполнен среднесрочный прогноз динамики курса акций на три недели вперед.

3. Модель регрессии с переменной структурой и зависимой лаговой переменной с лагом 1, построенная на основе среднемесячных данных за 5 лет. Данная модель позволяет проследить общую тенденцию в динамике показателя и сделать долгосрочный прогноз изменения среднемесячного курса акций на 3 месяца.

В ходе выполнения работы была применена как методология моделирования временных рядов, так и методология регрессионного анализа.

Таким образом, цель и задачи, поставленные в начале работы, были выполнены в полном объеме. На выходе получены модели, описывающие динамику курса акций за рассматриваемый период и позволяющие предсказать, как в будущем будет изменяться этот показатель.

Список литературы

1. Барбаумов В.Е. Сборник задач по финансовым инвестициям/ В.Е.Барбаумов, И.М.Гладких, А.С.Чуйко. - М.: Финансы и статистика, 2004. – 352 с.
2. Елисеева И.И. Эконометрика: Учебник / Елисеева И.И., Курышева С.В., Костеева Т.В. и др. – М.: Финансы и статистика, 2007 – 575 с.
3. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. – 8-е изд. – М.: Дело, 2007 – 368 с.
4. Тихомиров Н. П., Дорохина Е. Ю. Эконометрика: Учебник/ Н.П.Тихомиров, Е.Ю.Дорохина – М.: издательство «Экзамен», 2003 - 512 с.
5. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс: Учебник /– М.: Дело, 2007. – 504 с.
6. Лекции по эконометрике : учебное пособие для студентов высших учебных заведений, обучающихся по специальности «Прикладная информатика (в экономике)» / Н. И. Шанченко. – Ульяновск : УлГТУ, 2013. – 139 с
7. Акции - Акционерам и инвесторам - «ВТБ»/ Электронный ресурс:
[<http://www.vtb.ru/ir/shares/>] . – июнь 2017.
8. Банк «ВТБ» – Википедия/ Электронный ресурс:
[https://ru.wikipedia.org/wiki/Банк_«ВТБ»] . – июнь 2017.
9. Главная – Индексы и котировки - ММВБ/ Электронный ресурс:
[<http://www.micex.ru/marketdata/quotes>] . – июнь 2017.
10. Курс акций «ВТБ» - Финам/ Электронный ресурс:
[<http://www.finam.ru/analysis/profile04A6300007>] . – июнь 2017.
11. Ликвидные инструменты - БКС-экспресс – ао «ВТБ»/ Электронный ресурс:
[<http://bcs-express.ru/kotirovki-i-grafiki/vtbr>] . – июнь 2017.
12. Рынок акций – Инвестфандс/ Электронный ресурс:
[<http://stocks.investfunds.ru>] . – июнь 2017.
13. Аналитика и технический анализ - Портал Форекс/ Электронный ресурс:
[<https://www.fxteam.ru>] -июнь 2017.

14. Свободная энциклопедия Википедия/ Электронный ресурс:
[<http://ru.wikipedia.org>.] - июнь 2017.
15. Центр помощи Форекс- Полосы Боллинджера/ Электронный ресурс:
[[http://www.fxpro.ru/help-section/articles/fxpro-quant/bollinger-bands-\(bb\)](http://www.fxpro.ru/help-section/articles/fxpro-quant/bollinger-bands-(bb))] -
июнь 2017.

Список иллюстраций

Рисунок 1- Курс акций «The Coca-Cola Company»	9
Рисунок 2 – Последние 503 наблюдения ряда «The Coca-Cola Company».....	10
Рисунок 3 – Основные характеристики ряда «The Coca-Cola Company».....	10
Рисунок 4 – Распределение ряда «The Coca-Cola Company»	11
Рисунок 5 – Проверка нормальности ряда «The Coca-Cola Company»	12
Рисунок 6 – Ядерная оценка плотности ряда «The Coca-Cola Company»	12
Рисунок 7 – ВАФ и ЧАФ ряда «The Coca-Cola Company».....	14
Рисунок 8 – Ряд первых разностей.....	16
Рисунок 9 – Характеристики ряда первых разностей	17
Рисунок 10 – Распределение ряда первых разностей.....	18
Рисунок 11 – Проверка нормальности ряда первых разностей	18
Рисунок 12 – Ядерная оценка плотности ряда первых разностей	19
Рисунок 13 – ВАФ и ЧАФ ряда первых разностей.....	22
Рисунок 14 – Ряд вторых разностей.....	23
Рисунок 15 – Характеристики ряда вторых разностей.....	24
Рисунок 16 – Распределение ряда вторых разностей	24
Рисунок 17 – Проверка нормальности ряда вторых разностей.....	25
Рисунок 18 – Ядерная оценка плотности ряда вторых разностей	25
Рисунок 19 – ВАФ и ЧАФ ряда вторых разностей.....	29
Рисунок 20 – Оценка модели MA(1) для ряда вторых разностей	30
Рисунок 21 – Наблюдаемые и расчетные по ARIMA(0,2,1) значения исходного ряда	32
Рисунок 22 – Остатки модели ARIMA(0,2,1)	33
Рисунок 23 – Характеристики ряда ошибок модели ARIMA(0,2,1).....	34
Рисунок 24 – Распределение ряда остатков модели ARIMA(0,2,1).....	34
Рисунок 25 – Проверка равенства нулю среднего ряда ошибок ARIMA(0,2,1).....	35
Рисунок 26 – ВАФ и ЧАФ ряда ошибок модели ARIMA(0,2,1)	35
Рисунок 27 – Прогнозы модели ARIMA(0,2,1)	37
Рисунок 28 – Прогнозные и фактические значения	38
Рисунок 29 – Квадраты остатков модели ARIMA(0,2,1).....	39
Рисунок 30 – ВАФ и ЧАФ ряда квадратов ошибок модели ARIMA(0,2,1).....	39
Рисунок 31 – Тест на наличие ARCH процессов.....	40
Рисунок 32 – Агрегированный по неделям ряд.....	41
Рисунок 33 – Характеристики агрегированного по неделям ряда.....	42
Рисунок 34 – Проверка нормальности агрегированного по неделям ряда.....	42

Рисунок 35 – Ядерная оценка плотности агрегированного по неделям ряда.....	43
Рисунок 36 – ВАФ и ЧАФ агрегированного по неделям ряда	44
Рисунок 37 – Модель с тремя лаговыми переменными	45
Рисунок 38 – Наблюдаемые и расчетные значения агрегированного по неделям ряда	46
Рисунок 39 – Остатки модели с лаговыми переменными	47
Рисунок 40 – Характеристики остатков модели с лаговыми переменными	47
Рисунок 41 – Проверка нормальности ряда остатков модели с лаговыми переменными ...	48
Рисунок 42 - Проверка нормальности ряда остатков модели с лаговыми переменными (Тесты)	48
Рисунок 43 – Проверка равенства нулю среднего ряда ошибок модели с лаговыми переменными	49
Рисунок 44 – ВАФ и ЧАФ ряда остатков модели с лаговыми переменными	50
Рисунок 45 – Прогнозы модели с лаговыми переменными	51
Рисунок 46 – Прогнозные и фактические значения	52
Рисунок 47 - Агрегированный по месяцам ряд	53
Рисунок 48 – Характеристики агрегированного по месяцам ряда.....	54
Рисунок 49 – Проверка нормальности агрегированного по месяцам ряда.....	54
Рисунок 50 – Проверка нормальности агрегированного по месяцам ряда (Тесты).....	55
Рисунок 51 – ВАФ и ЧАФ агрегированного по месяцам ряда.....	55
Рисунок 52 – Модель с двумя лаговыми переменными	56
Рисунок 53 – Наблюдаемые и расчетные значения агрегированного по месяцам ряда	57
Рисунок 54 – Остатки модели с лаговыми переменными	58
Рисунок 55 – Характеристики ряда остатков модели с лаговыми переменными	59
Рисунок 56 – Проверка нормальности ряда остатков модели с лаговыми переменными ...	59
Рисунок 57 - Проверка нормальности ряда остатков модели с лаговыми переменными (Тесты)	60
Рисунок 58 – ВАФ и ЧАФ ряда остатков модели с лаговыми переменными	61
Рисунок 59 – Прогнозы модели с лаговыми переменными	62
Рисунок 60 – Прогнозные и фактические значения	63

Список таблиц

Таблица 1- Результаты теста Аббе для ряда первых разностей.....	22
Таблица 2 - Тест Аббе для ряда вторых разностей	26
Таблица 3- Медианный критерий для ряда вторых разностей	27
Таблица 4 – Зависимость τ_0 от объема выборки.....	28
Таблица 5 – Критерий восходящих и нисходящих серий для ряда вторых разностей.....	28
Таблица 6 – Краткосрочный прогноз ARIMA(0,2,1).....	37
Таблица 7 – Среднесрочный прогноз.....	51
Таблица 8 – Долгосрочный прогноз	62