

Данная работа посвящена анализу изменения стоимости акций компании 'Coca Cola' на продолжительном периоде наблюдения (1962 - 1921 г). Компания 'Coca Cola' является крупной международной компанией, работающей по всему миру на огромном рынке продуктов питания. В связи с этим стоимость ее акций должна быть достаточно стабильной и реагировать значительно только на глобальные изменения мировой экономики и отражать общие тенденции развития мирового рынка продовольствия. При построении достаточно точной модели изменения стоимости акций компании необходимо оценить возможность прогнозирования их стоимости на некоторый период времени, который зависит от общей ситуации в мировой экономике (неожиданный катаклизм практически устраняет возможность какого-либо разумного прогнозирования). В результате анализа тенденций изменения стоимости акций и возможности ее прогнозирования предлагается попробовать оценить минимальный разумный период учета данных для построения модели оценки стоимости акций в целях применения для прогноза текущей стоимости акций

#### Задачи:

- Провести анализ динамика изменения стоимости акций
- Определить факторы, влияющие на изменение динамики изменения стоимости акций.  
(выделить временные интервалы 'резких изменений тенденция' и попытаться найти причину в изменениях мировой конъюнктуры рынка)
- Построить модели оценки изменения стоимости акций с учетом всего представленного периода развития экономической ситуации
- Провести оценку качества прогнозов, получаемых с помощью построенных моделей

#### 1. Пример исходных данных

Date	Open	High	Low	Close	Adj Close	Volume
1962-01-31	0.248432	0.250784	0.245221	0.247692	0.048153	1.337018e+06
1962-02-28	0.243695	0.245546	0.241982	0.243318	0.047303	6.952421e+05

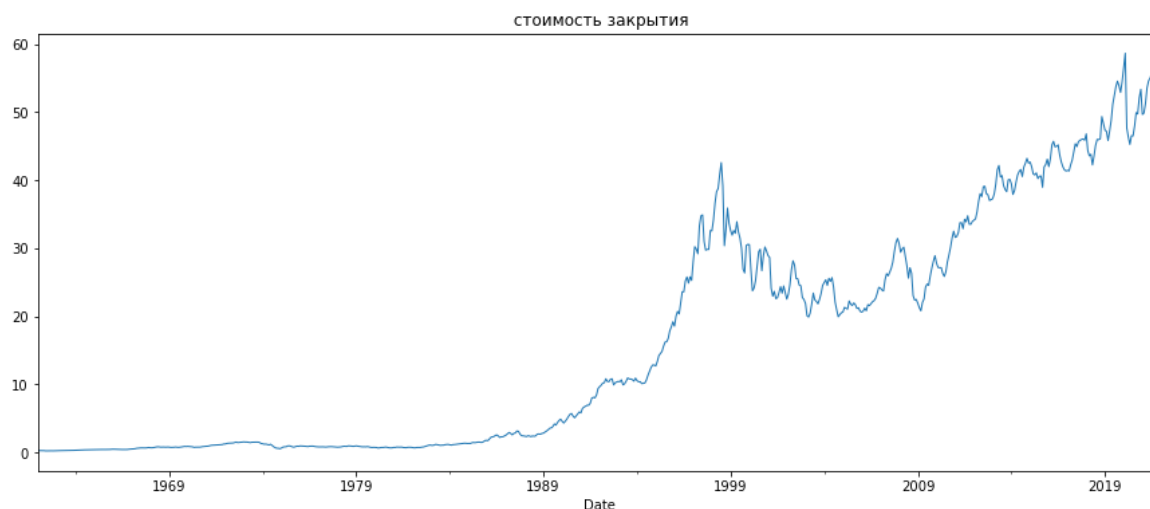
#### 2. Статистические характеристики исходных данных

	Open	High	Low	Close	Adj Close	Volume	
count	720.000000	720.000000	720.000000	720.000000	720.000000	720.000000	7.200000e+02
mean	16.067735	16.200015	15.933392	16.072319	11.384551	9.029490e+06	
std	16.961081	17.086733	16.832736	16.962937	14.287341	6.534544e+06	
min	0.204241	0.206814	0.200986	0.203233	0.039972	4.468364e+05	
25%	0.862740	0.868096	0.855472	0.861998	0.241078	3.170400e+06	
50%	9.479336	9.592989	9.386972	9.501573	4.717325	8.711218e+06	
75%	28.987626	29.382500	28.741837	29.052066	17.316536	1.334197e+07	
max	58.793158	59.192105	58.301053	58.652632	55.782308	4.038423e+07	

Из полученных данных можно сделать отдельные наблюдения:

- большой разброс объемов торгов ( 446 0000 - 40 000 000)
- наличие большой разницы средних значений Close и Adj Close
- (что указывает на наличие существенных влияний корпоративных мероприятий компании на стоимость акций на большом периоде времени)

В качестве метрики для прогнозирования выбрана стоимость закрытия Close (стоимость акции на конец торгового дня). График представлен ниже

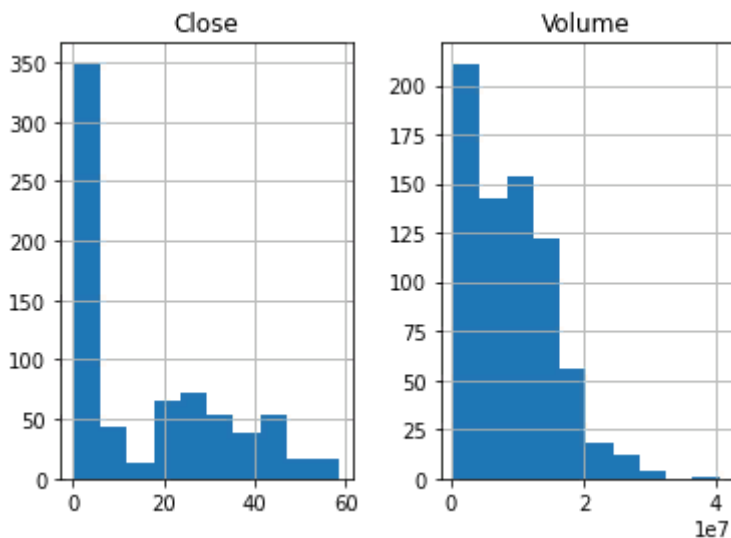


На графике видно что примерно в 1991 г. начинается стремительный рост стоимости акций (возможно связано с началом активного вывода средств из бывшего СССР и начало бурного роста финансового рынка (рост спекулятивного капитала), затем обвал стоимости в 1998 году (экономический кризис). Затем наблюдался период стабилизации (примерно до 2008г), затем опять обвал стоимости (кризис 2008 года) и затем плавный рост стоимости акций. В начале 2020 г. наблюдается резкий обвал стоимости акций (начало ковид-пандемии)

Из полученных графиков можно сделать следующие предположения:

- начиная с 1982 г. объем торговли акциями находится в среднем около 10 млн акций/месяц
- величина поправки стоимости акции была относительно стабильна и начала увеличиваться с моменты начала быстрого роста в 1991г. Возможно это было связано с нарастанием финансовых спекуляций (стремительный рост акций). Пик поправок пришел на начало кризиса 1998г. Затем, по всей видимости, произошла корректировка корпоративных процедур и постепенно к настоящему времени поправка стоимости акций (из-за корпоративных процедур) была приведена фактически к нулю и сейчас стоимость акций на момент закрытия дня совпадает со скорректированной стоимостью акций. Т. е. в настоящее время корпоративные процедуры не влияют на стоимость акций (доходность акций)

### 3. Гистограммы для анализа исходных данных



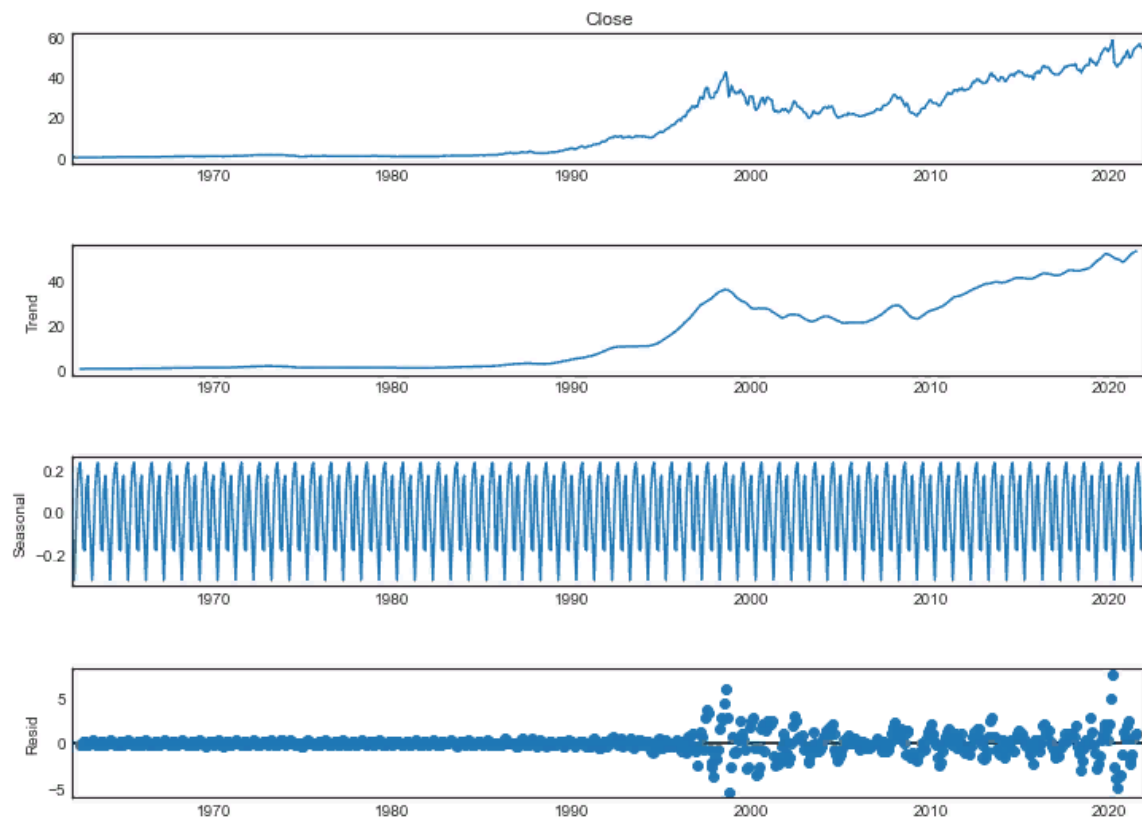
На гистограммах видно что среди значений стоимости акций преобладают значения из диапазона ( 0.2 - 10.0). Соответственно возможно предположить, что в создаваемой модели оценки стоимости акций может быть небольшим понижающий эффект (на текущий момент стоимость достигает 60 единиц)

Выводы вводной части:

- Полученные данные являются качественным набором данных без пропусков и пустых значений
- метрикой для построения модели прогнозирования стоимости акций выбирается цена закрытия Close т. к. значения Open, High, Low при выбранной дискретности представления данных близки значению Close и повторяют динамику значений Close. Параметр Adj Close, который считается существенным с точки зрения оценки доходности акций, стремится к нулю и практически совпадает с параметром Close на последнем временном интервале представленных данных

Для построения модели была произведена декомпозиция исходного временного ряда

Декомпозиция временного ряда (цена закрытия)



При декомпозиции наблюдается явный тренд в изменении параметра Close. Также присутствует сезонность. Период сезонности период явно не определен. Его будем определять при вычислении параметров модели

Построение моделей прогнозирования и их анализ

Задачи данного этапа:

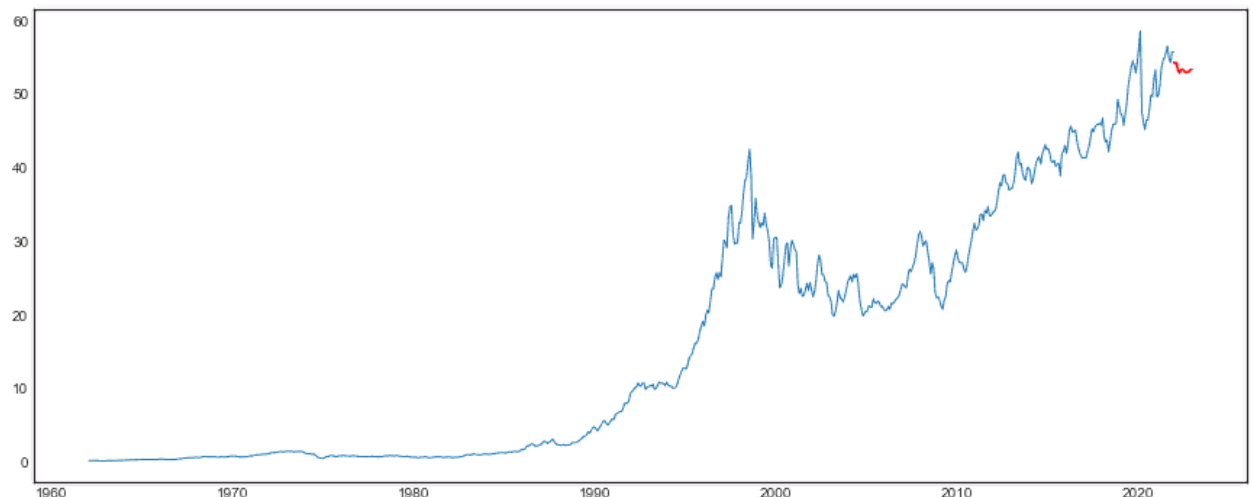
- описать модель
- подобрать оптимальные параметры с помощью метода ARIMA (необходимо достичь учета сезонности (на его наличие указывает декомпозиция))
- создать модель с помощью метода SARIMAX

- с помощью метода AR
- Далее для каждой модели
- обучить модель
    - на наборе обучающих данных (набор train)
  - сделать прогноз на период тестовой выборки
    - с помощью обученной модели на период равный размеру тестовой выборки (набор test)
  - сравнить прогноз с тестовой выборкой (построить график)
  - оценить качество прогноза
    - с помощью MSE (средняя квадратическая ошибка) и RMSE (среднеквадратическая ошибка)
  - сделать прогноз на определённый позднее период
    - будем ориентироваться на определённый при построении модели период сезонности
  - сделать выводы о работе данного метода прогнозирования

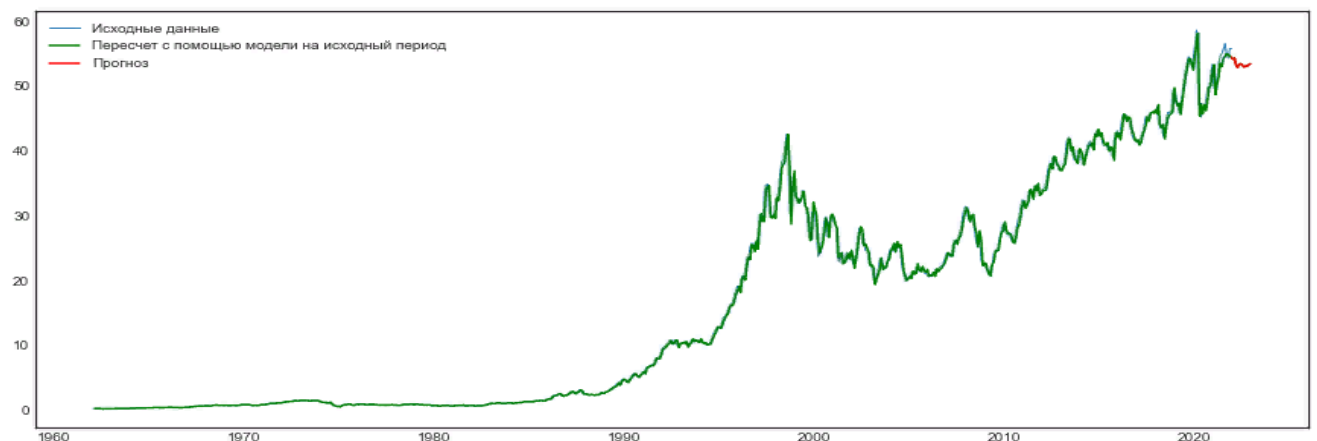
### Построение модели с помощью ARIMA

С использованием метода ARIMA были подобраны параметры нашей модели. Для определения сезонности в нашей модели был произведен подбор параметров модели с разными величинами  $m$  (величина сезонности). В результате было определено, что оптимальным является сезонность  $m = 8$  (8 месяцев)

Для визуализации 'соответствия' статистических характеристик прогнозной модели исходному набору данных построим общий график (исходные данные плюс период прогноза)



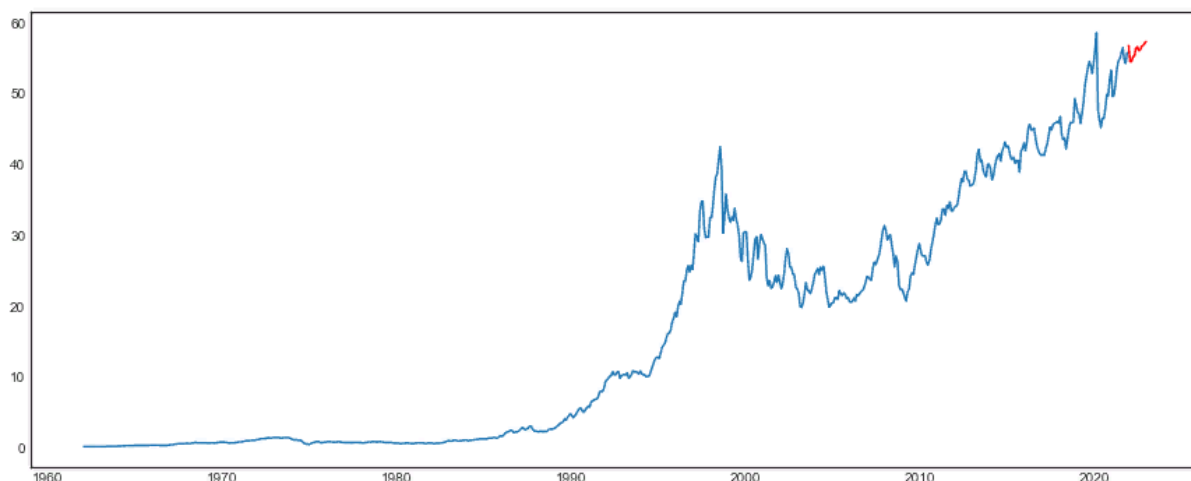
На общем графике, при сохранении тенденции изменения параметра Close, заметно некоторое 'понижение' прогнозного графика (предположение о таком эффекте было сделано ранее). Проверим как будет происходить 'стыковка' прогноза с графиком построенным моделью по всему исходному набору



На данном графике отражен пересчитанный график значений Close с использованием созданной модели на исходном интервале времени, который 'точно' переходит в прогнозный график. График исходных данных немного выше. (т. е. наблюдается 'проседание' данных модели). Возможно, это связано с большим интервалом относительно малых значений Close без явной тенденции к росту на раннем периоде.

Если применить модель на более коротком интервале (например период 2010 ...), то вероятно модель прогноза будет более точной

#### Построение моделей с помощью AR



На графике заметно что эффект 'проседания' прогноза, который наблюдался в предыдущей модели, отсутствует. Данный эффект, вероятно, объясняется тем, что в модели AR используются данные с лагом, вычисленным при построении модели (реализован шаг 20) а не за весь период. В этом случае на прогноз не влияют тенденции ранних периодов (тенденции, которые отличаются от тенденций ближайших к прогнозируемому периоду)

#### Заключение

В настоящей работе проведен анализ данных с использованием современных методов обработки статистической информации.

Рассчитаны основные статистические метрики, позволяющие судить о характере исследуемого явления. Анализ применения моделей показал степень влияния особенностей исходных данных на выбор метода построения модели прогнозирования.

Результат анализа представленных данных позволили сделать вывод о том, что для получения адекватных моделей прогнозирования для рынка акций компаний необходимо убирать дальнюю историчность данных, т.к. в течении нескольких лет общая экономическая ситуация и закономерности ее развития заметно изменяются и данные дальних периодов вносят недостоверные, для текущего периода, тенденции в модель, предназначенную для использования в текущем периоде.

Данный вывод базируется на т.н. плавности перехода от исходных данных к прогнозу и оценках точности построенных моделей.

Для использования рассмотренных методов прогнозирования в целях прогнозирования текущих стоимостей акций необходимо определить ближайший период относительно стабильного периода развития рынка и строить модель на основе данных этого периода.