

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО  
ОБРАЗОВАНИЯ "МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ  
УНИВЕРСИТЕТ имени М.В.ЛОМОНОСОВА"  
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ  
КАФЕДРА ОБЩИХ ПРОБЛЕМ УПРАВЛЕНИЯ**

**КУРСОВАЯ РАБОТА**

**"Исследование ценовых рядов на основании  
Математической Статистики и Теории  
Вероятностей"**

**Автор - Нарембеков Темирлан, студент 3-го курса кафедры  
Общих Проблем Управления, 312 группа**

**Научный руководитель - Заплетин Максим Петрович, доцент  
кафедры Общих Проблем Управления**

**Весна 2024**

# 1 СБОР И ПОДГОТОВКА ДАННЫХ

## 1.1 Сбор данных

Выберем временной период для анализа: май 2009г - апрель 2024г. со 180 ежемесячными показателями USD/KZT. Таким образом, интервал времени  $\Delta = 1$  месяц является регулярным.

|                                 |                  |                  |                  |                 |       |        |
|---------------------------------|------------------|------------------|------------------|-----------------|-------|--------|
| USD/KZT 446,435 +0,135 (+0,03%) |                  |                  |                  |                 |       |        |
| Дата                            | Цена             | Откр.            | Макс.            | Мин.            | Объём | Изм. % |
| 01.04.2024                      | 446,435          | 447,405          | 448,555          | 445,950         |       | +0.12% |
| 01.03.2024                      | 445,920          | 451,105          | 455,310          | 445,680         |       | -1.03% |
| 01.02.2024                      | 450,580          | 449,755          | 455,590          | 445,860         |       | +0.30% |
| 01.01.2024                      | 449,230          | 456,810          | 458,905          | 444,290         |       | -0.92% |
| 01.12.2023                      | 453,400          | 459,360          | 462,510          | 452,150         |       | -0.81% |
| 01.11.2023                      | 457,100          | 468,860          | 471,200          | 455,910         |       | -2.39% |
| 01.10.2023                      | 468,290          | 477,910          | 481,060          | 467,990         |       | -1.91% |
| 01.09.2023                      | 477,390          | 458,360          | 486,210          | 454,595         |       | +4.27% |
| 01.08.2023                      | 457,840          | 444,600          | 467,110          | 441,975         |       | +3.10% |
| 01.07.2023                      | 444,080          | 449,100          | 450,305          | 439,175         |       | -1.39% |
| 01.06.2023                      | 450,330          | 447,755          | 455,265          | 443,600         |       | +0.94% |
| 01.05.2023                      | 446,130          | 452,155          | 453,560          | 440,740         |       | -1.21% |
| 01.03.2010                      | 147,040          | 147,350          | 147,460          | 146,850         |       | -0.19% |
| 01.02.2010                      | 147,325          | 148,025          | 148,235          | 147,115         |       | -0.48% |
| 01.01.2010                      | 148,030          | 148,490          | 148,850          | 147,850         |       | -0.33% |
| 01.12.2009                      | 148,520          | 148,750          | 150,000          | 148,280         |       | -0.11% |
| 01.11.2009                      | 148,685          | 150,695          | 150,935          | 148,635         |       | -1.38% |
| 01.10.2009                      | 150,770          | 151,010          | 151,050          | 150,060         |       | -0.11% |
| 01.09.2009                      | 150,940          | 150,820          | 150,990          | 150,670         |       | +0.07% |
| 01.08.2009                      | 150,830          | 150,690          | 150,930          | 150,650         |       | +0.05% |
| 01.07.2009                      | 150,755          | 150,395          | 150,945          | 150,245         |       | +0.22% |
| 01.06.2009                      | 150,430          | 150,490          | 150,590          | 149,340         |       | -0.06% |
| 01.05.2009                      | 150,515          | 150,665          | 150,955          | 149,775         |       | -0.14% |
| Максимум: 527,125               | Минимум: 145,135 | Разница: 381,990 | Среднее: 294,247 | Изм. %: 196,202 |       |        |

## 1.2 Подготовка Данных

Мы имеем наблюдения ежемесячной стоимости валюты за последние 15 лет, но сами по себе эти данные интересуют нас лишь косвенно по следующей причине.

В середине 20 века появились работы, в которых было доказано, что в поведении цен акций и товаров нет ни ритмов, ни трендов, ни циклов, а суммы логарифмов цен - являются случайным блужданием, которые описывают всю эволюцию цен.

Поэтому при исследовании ценовых рядов используются логарифмы цен  $H_{tk} = \ln \left( \frac{S_{tk}}{S_{tk-1}} \right)$ .

## 2 ИССЛЕДОВАНИЕ ДОХОДНОСТИ ТЕНГЕ НА СЛУЧАЙНОСТЬ

Проверим логарифмическую доходность тенге разными статистическими методами  $H_{tk} = \ln \left( \frac{S_{tk}}{S_{tk-1}} \right)$  с общим числом наблюдений равным 179.

### 2.1 Ранговый критерий Вальда-Вольфовица

Проверка статистических гипотез заключается в том, чтобы решить какой из исходов влечет наибольшие риски, а затем ставят задачу отклонить этот исход.

Данный критерий использует понятие Ранга  $R_k$  наблюдения  $\ln \left( \frac{S_{tk}}{S_{tk-1}} \right)$ , которое является номером наблюдения в соответствующем вариационном ряду, для вычисления статистики  $R^* = \frac{R}{\sqrt{D[R]}}$  и улучшении ее до статистики  $R^{**} = R^* + 1.1216 \cdot n^{-0.523}$

1) Построим по известному ряду логарифмических доходностей вариационный ряд.

2) Введем гипотезу  $H_0$  и альтернативу  $H_1$ :

$H_0$  : Тренд отсутствует - т.е.случайность ряда,

$H_1$  : Тренд присутствует - неслучайность ряда.

3) В нашем случае ошибка I рода - это наиболее критическая. Значит, ошибка I рода - наиболее критическая

Таким образом наша (статистическая) задача сделать ошибку I рода  $\alpha$  как можно меньше, однако это увеличит ошибку II рода  $\beta$  (что не несет рисков кроме возможной потери прибыли).

4) Выберем уровень значимости  $\alpha$  (Ошибка I рода). Для объема выборки  $100 < n < 1000$  рекомендуется  $\alpha = 0.01$ .

Критерий двусторонний, гипотеза об отсутствии тренда принимается, если  $R^{**} \in [-2.57, 2.57]$ , где 2.57 - критическое значение нормального распределения, соответствующее  $\alpha = 0.01$ .

5) Построим статистику  $R = \sum_{i=1}^{n-1} \left(R_i - \frac{n+1}{2}\right) \left(R_{i+1} - \frac{n+1}{2}\right)$  для  $n = 179$ :  $R = 68436$

6) Вычислим  $D[R] = \frac{n^2 \cdot (n+1) \cdot (n-3) \cdot (5n+6)}{720}$ .

7)  $R^* = \frac{R}{\sqrt{D[R]}} = 1.9201840279563078$

8) улучшим  $R^*$  до  $R^{**}$ :  $R^{**} = 1.9945879562302098$

Таким образом принимаем гипотезу  $H_0$  об отсутствии тренда и получаем, что ряд логарифмических доходностей - случаен.

### 3 ПРОВЕРКА НА НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

#### 3.1 Коэффициент Эксцесса(вытянутости)

коэффициент эксцесса

$$K_N = \frac{\left( \frac{1}{n} \sum_{i=1}^n (r_i - \bar{r})^4 \right)}{\left( \frac{1}{n} \sum_{i=1}^n (r_i - \bar{r})^2 \right)^2} - 3,$$

где  $r_i$  - это логарифмическая доходность в момент времени  $i$ , где  $\bar{r}$  - средняя логарифмическая доходность, а  $n$  - количество наблюдений.

Для нормального распределения  $E = 0$ .

1) Для ряда логарифмических доходностей  $H_{tk} = \ln \left( \frac{S_{t_k}}{S_{t_{k-1}}} \right)$ . найдем величину  $\bar{r}$  как среднеарифметическое логарифмов цен:

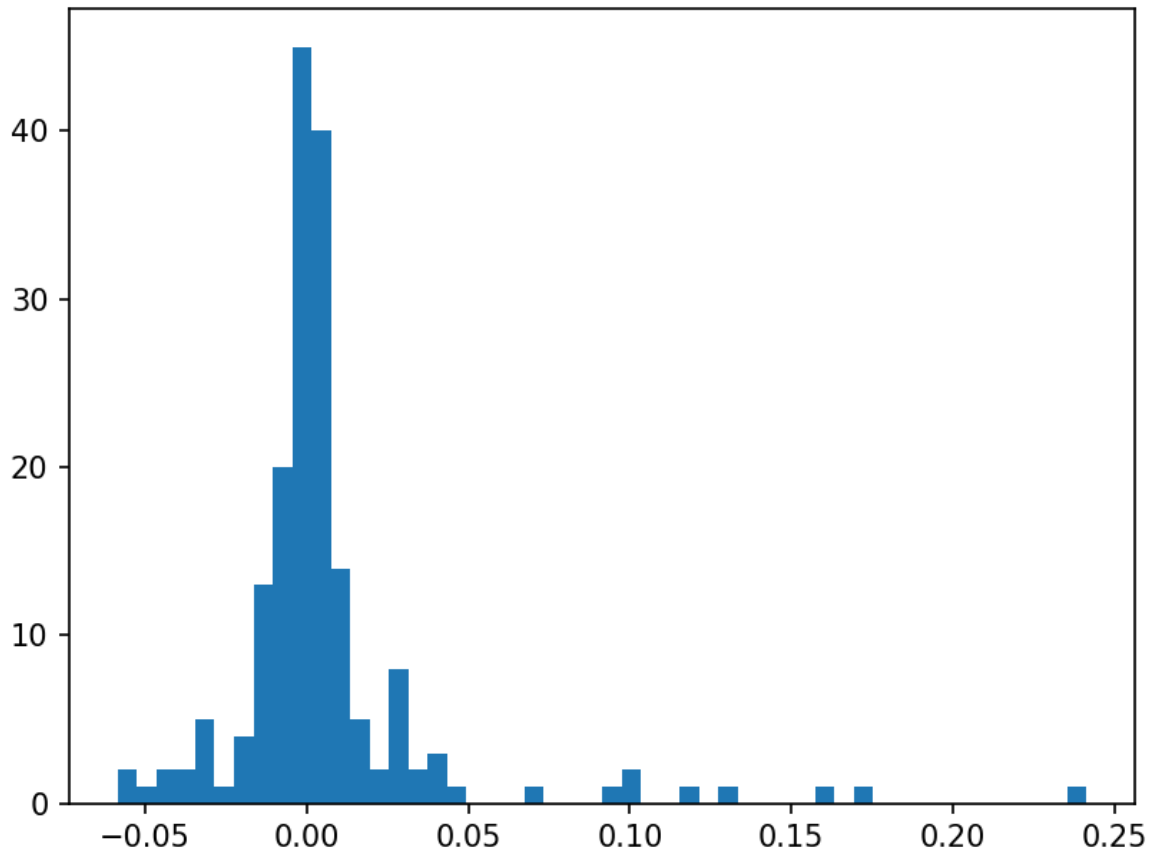
$$\bar{r} = 0.00607391656991618$$

2) Подставляем  $\bar{r}$  в формулу:

$$K_{179} = 16.38912221565546.$$

### 3.2 Гистограмма

Рассмотрим гистограмму: Наблюдаются выбросы справа - это сигнала-



лизирует о существенном отклонении от нормальности. Если от них избавиться, то на практике можно считать что распределение условно нормально, так как гистограмма имеет колоколообразную форму и небольшую ассиметрию. Однако нашем случае распределение нормальным не будет, убедимся в этом:

#### Тест Шапиро-Уилка

$H_0$  : Распределение нормальное,

$H_1$  : Распределение не нормальное..

PROBLEMS 10 OUTPUT DEBUG CONSOLE TERMINAL PORTS

## 4.2 НЕЗАВИСИМОСТЬ

Проведем тест Бройша-Годфри на наличие автокорреляции в остатках после внедрения модели линейной регрессии в структуру ряда.

Тест Бройша-Годфри

### 1. Модель регрессии

#### 1) Модель линейной регрессии

Введем модель линейной регрессии с известной числовой матрицей  $X$ , строками которой являются предшествующие 13 значений  $i$ -го элемента, в качестве предиктора (независимой переменной) и лог. доходностью  $Y$  в качестве зависимой переменной.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{13} X_{13i} + \epsilon_i$$

$$Y = X * \beta + \epsilon$$

Прежде чем использовать эту модель, выясним является ли она статистически значимой, т.е. описывает ли  $Y$ . Для этого найдем коэффициенты и исследуем их значимость.



| OLS Regression Results  |                  |                     |          |       |        |        |
|---|------------------|---------------------|----------|-------|--------|--------|
| =====   |                  |                     |          |       |        |        |
| Dep. Variable:  | Y                | R-squared:          | 0.097    |       |        |        |
| Model:  | OLS              | Adj. R-squared:     | 0.013    |       |        |        |
| Method:   | Least Squares    | F-statistic:        | 1.156    |       |        |        |
| Date:   | Wed, 30 Oct 2024 | Prob (F-statistic): | 0.315    |       |        |        |
| Time:   | 19:33:42         | Log-Likelihood:     | 322.50   |       |        |        |
| No. Observations:   | 165              | AIC:                | -615.0   |       |        |        |
| Df Residuals:   | 150              | BIC:                | -568.4   |       |        |        |
| Df Model:   | 14               |                     |          |       |        |        |
| Covariance Type:  | nonrobust        |                     |          |       |        |        |
| =====   |                  |                     |          |       |        |        |
|   | coef             | std err             | t        | P> t  | [0.025 | 0.975] |
| -----   |                  |                     |          |       |        |        |
| const   | 0.0062           | 0.003               | 1.913    | 0.058 | -0.000 | 0.013  |
| LogReturns  | 0.1797           | 0.081               | 2.206    | 0.029 | 0.019  | 0.341  |
| X1  | -0.0191          | 0.083               | -0.231   | 0.817 | -0.183 | 0.144  |
| X2  | 0.0191           | 0.083               | 0.231    | 0.817 | -0.144 | 0.182  |
| X3  | 0.2029           | 0.083               | 2.451    | 0.015 | 0.039  | 0.366  |
| X4  | -0.0608          | 0.084               | -0.721   | 0.472 | -0.228 | 0.106  |
| X5  | -0.0677          | 0.084               | -0.803   | 0.423 | -0.234 | 0.099  |
| X6  | -0.0317          | 0.084               | -0.377   | 0.707 | -0.198 | 0.135  |
| =====   |                  |                     |          |       |        |        |
| X7  | -0.1225          | 0.085               | -1.449   | 0.149 | -0.290 | 0.045  |
| X8  | 0.0612           | 0.085               | 0.720    | 0.473 | -0.107 | 0.229  |
| X9  | 0.0527           | 0.085               | 0.620    | 0.536 | -0.115 | 0.221  |
| X10   | 0.0224           | 0.083               | 0.269    | 0.788 | -0.142 | 0.187  |
| X11   | -0.0418          | 0.084               | -0.500   | 0.618 | -0.207 | 0.123  |
| X12   | -0.0486          | 0.084               | -0.581   | 0.562 | -0.214 | 0.117  |
| X13   | -0.0656          | 0.083               | -0.795   | 0.428 | -0.229 | 0.097  |
| =====   |                  |                     |          |       |        |        |
| Omnibus:  | 150.501          | Durbin-Watson:      | 2.001    |       |        |        |
| Prob(Omnibus):  | 0.000            | Jarque-Bera (JB):   | 2427.390 |       |        |        |
| Skew:   | 3.341            | Prob(JB):           | 0.00     |       |        |        |
| Kurtosis:   | 20.562           | Cond. No.           | 40.1     |       |        |        |
| =====   |                  |                     |          |       |        |        |
| Notes:  |                  |                     |          |       |        |        |
| [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. |                  |                     |          |       |        |        |

Обратим внимание на р-значение F-теста:

$H_0$  : Модель с одной только константой объясняет данные так же как и текущая модель, т.е. модель статитически незначима,

$H_1$  : Текущая модель значительно лучше модели с одной константой, т.е. модель статистически значима.

Для F-статистики р-значение = 0,315 > 0.01 и, следовательно, гипотеза  $H_0$  принимается и модель требуется доработать.

Для этого оставим в модели только те коэффициенты, р-значения которых достаточно малы ( $p < 0.01$ ) - именно они являются статистически значимыми. Согласно результатам, таких коэффициентов нет,

значит заключаем, что модель линейной регрессии не подходит для анализа ряда.

## 2) Модель регрессии Фурье

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{13} X_{13i} + \epsilon_i + \sum_{k=1}^6 \left( \gamma_k \sin \left( \frac{2\pi kt}{n} \right) + \delta_k \cos \left( \frac{2\pi kt}{n} \right) \right)$$

| OLS Regression Results |                  |                     |         |       |        |        |
|------------------------|------------------|---------------------|---------|-------|--------|--------|
| =====                  |                  |                     |         |       |        |        |
| Dep. Variable:         | Y                | R-squared:          | 0.278   |       |        |        |
| Model:                 | OLS              | Adj. R-squared:     | 0.148   |       |        |        |
| Method:                | Least Squares    | F-statistic:        | 2.139   |       |        |        |
| Date:                  | Tue, 03 Dec 2024 | Prob (F-statistic): | 0.00297 |       |        |        |
| Time:                  | 21:22:30         | Log-Likelihood:     | 340.89  |       |        |        |
| No. Observations:      | 165              | AIC:                | -629.8  |       |        |        |
| Df Residuals:          | 139              | BIC:                | -549.0  |       |        |        |
| Df Model:              | 25               |                     |         |       |        |        |
| Covariance Type:       | nonrobust        |                     |         |       |        |        |
| =====                  |                  |                     |         |       |        |        |
|                        | coef             | std err             | t       | P> t  | [0.025 | 0.975] |
| -----                  |                  |                     |         |       |        |        |
| const                  | 0.0201           | 0.004               | 5.135   | 0.000 | 0.012  | 0.028  |
| X1                     | -0.1946          | 0.083               | -2.356  | 0.020 | -0.358 | -0.031 |
| X2                     | -0.1599          | 0.082               | -1.949  | 0.053 | -0.322 | 0.002  |
| X3                     | 0.0214           | 0.083               | 0.258   | 0.797 | -0.143 | 0.186  |
| X4                     | -0.2096          | 0.084               | -2.507  | 0.013 | -0.375 | -0.044 |
| X5                     | -0.2236          | 0.083               | -2.679  | 0.008 | -0.389 | -0.059 |
| X6                     | -0.1784          | 0.083               | -2.158  | 0.033 | -0.342 | -0.015 |
| X7                     | -0.2702          | 0.084               | -3.227  | 0.002 | -0.436 | -0.105 |
| X8                     | -0.0968          | 0.084               | -1.159  | 0.248 | -0.262 | 0.068  |
| X9                     | -0.0788          | 0.084               | -0.936  | 0.351 | -0.245 | 0.088  |
| X10                    | -0.1112          | 0.085               | -1.315  | 0.191 | -0.278 | 0.056  |

|                |         |                   |           |       |        |        |
|----------------|---------|-------------------|-----------|-------|--------|--------|
| X11            | -0.1532 | 0.084             | -1.826    | 0.070 | -0.319 | 0.013  |
| X12            | -0.1461 | 0.083             | -1.761    | 0.080 | -0.310 | 0.018  |
| X13            | -0.1765 | 0.084             | -2.104    | 0.037 | -0.342 | -0.011 |
| sin_1          | 0.0104  | 0.004             | 2.681     | 0.008 | 0.003  | 0.018  |
| cos_1          | -0.0213 | 0.005             | -4.321    | 0.000 | -0.031 | -0.012 |
| sin_2          | -0.0172 | 0.005             | -3.789    | 0.000 | -0.026 | -0.008 |
| cos_2          | -0.0050 | 0.004             | -1.350    | 0.179 | -0.012 | 0.002  |
| sin_3          | 0.0111  | 0.004             | 2.603     | 0.010 | 0.003  | 0.019  |
| cos_3          | 0.0098  | 0.004             | 2.553     | 0.012 | 0.002  | 0.017  |
| sin_4          | -0.0057 | 0.004             | -1.360    | 0.176 | -0.014 | 0.003  |
| cos_4          | -0.0150 | 0.004             | -3.620    | 0.000 | -0.023 | -0.007 |
| sin_5          | 0.0031  | 0.004             | 0.762     | 0.447 | -0.005 | 0.011  |
| cos_5          | 0.0117  | 0.004             | 2.903     | 0.004 | 0.004  | 0.020  |
| sin_6          | 0.0046  | 0.004             | 1.166     | 0.246 | -0.003 | 0.013  |
| cos_6          | -0.0113 | 0.004             | -2.596    | 0.010 | -0.020 | -0.003 |
| =====          |         |                   |           |       |        |        |
| Omnibus:       | 106.316 | Durbin-Watson:    | 2.105     |       |        |        |
| Prob(Omnibus): | 0.000   | Jarque-Bera (JB): | 809.978   |       |        |        |
| Skew:          | 2.305   | Prob(JB):         | 1.30e-176 |       |        |        |
| Kurtosis:      | 12.827  | Cond. No.         | 46.6      |       |        |        |
| =====          |         |                   |           |       |        |        |

Р-значение F-теста меньше чем 0.01 и тогда нулевая гипотеза отклоняется и данная модель хорошо объясняет природу данных.

II.Применив модель регрессии Фурье к ряду, получим остатки и проверим тестом Бройша-Годфри наличие автокорреляции в них.

$H_0$  : Автокорреляция отсутствует в остатках, данные независимы,

$H_1$  : Автокорреляция присутствует, данные зависимы.

Вычисляем статистику:

```
bg_test = acorr_breusch_godfrey(initial_model, nlags=13)

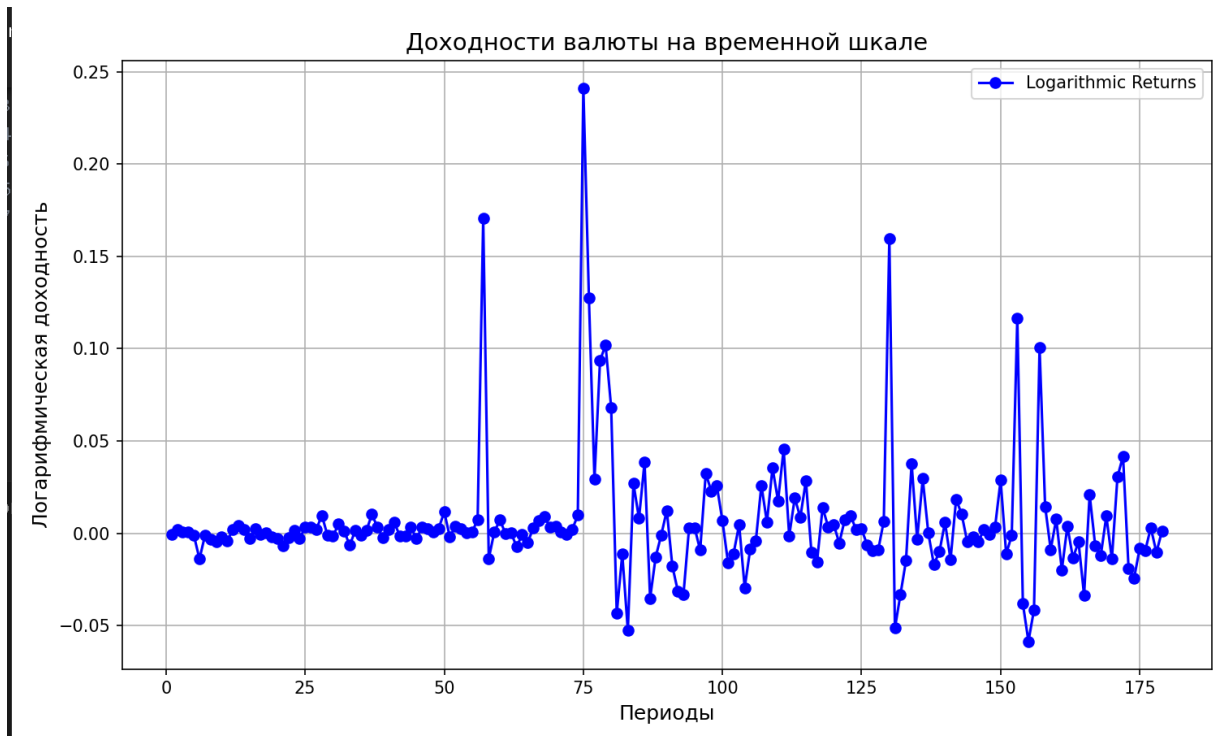
print(f"Статистика теста Бреуша-Годфри: {bg_test[0]}")
print(f"P-значение теста: {bg_test[1]}")
```

```
Статистика теста Бреуша-Годфри: 51.27151119893768
P-значение теста: 1.8013798830419523e-06
```

Так же значение p-value < 0.01 значит нужно отклонить  $H_0$  и тогда данные в выборке зависимы.

### 4.3 ОДИНАКОВАЯ РАСПРЕДЕЛЕННОСТЬ

Рассмотрим график ряда на временной шкале:



Видно явное различие между первыми 80 и оставшимися величинами. Проверим эти 2 подвыборки на однородность.

#### Тест Колмогорова-Смирнова

Пусть эмпирическая функция распределения (ЭФР)  $F_n$  построенная по выборке  $X = (X_1, \dots, X_n)$ , имеет вид:  $F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x}$ , где  $I_{X_i \leq x}$  указывает, попало ли наблюдение  $X_i$  в область  $(-\infty, x]$ :

$$I_{X_i \leq x} = \begin{cases} 1, & X_i \leq x; \\ 0, & X_i > x. \end{cases}$$

Таким образом имеем 2 эмпирические функции распределения  $F_{80}$  и  $F_{99}$

$H_0$  : Распределения подвыборок совпадают  
и данные одинаково распределены.

$H_1$  : Распределения отличаются  
и данные не являются одинаково распределёнными.

## Теорема Смирнова

Пусть  $F_{1,n}(x), F_{2,m}(x)$  — эмпирические функции распределения, построенные по независимым выборкам объёмом  $n$  и  $m$  случайной величины  $\xi$ . Тогда, если  $F(x) \in C^1(\mathbb{X})$ , то

$$\forall t > 0 : \lim_{n,m \rightarrow \infty} P \left( \sqrt{\frac{nm}{n+m}} D_{n,m} \leq t \right) = K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 t^2},$$

где  $D_{n,m} = \sup_x |F_{1,n} - F_{2,m}|$ .

Построим критическое множество  $R = D \mid D \geq K_{0.01} = 0.121$

Вычислим статистику:  $D = \sqrt{\frac{80 \cdot 99}{80+99}} \cdot D_{80,99}$  Результат:  $D = 0.36$

Таким образом выборки не одинаково распределены.

Хоть распределения и разные, может оказаться так, что они имеют одинаковый закон, но разные параметры, либо же вообще разные законы.

Стандартизируем, т.е. перейдем к

$$\frac{X - \bar{X}}{\sigma^{1/2}}$$

в каждой подвыборке, и выясним какой случай у нас, снова применив Тест:

```
80 standartized_subsample1 = [(x - stat.mean(sample_sub1))/stat.stdev(sample_sub1) for x in sample_sub1]
81 standartized_subsample2 = [(x - stat.mean(sample_sub2))/stat.stdev(sample_sub2) for x in sample_sub2]
82
83 ks_stat, ks_p_value = stats.ks_2samp(standartized_subsample1, standartized_subsample2)
84 print("Тест Колмогорова-Смирнова для двух выборок:")
85 print(f"Статистика K-S: {ks_stat}")
86 print(f"P-значение: {ks_p_value}")
87
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

powershell +

```
Тест Колмогорова-Смирнова для двух выборок:
Статистика K-S: 0.36464646464646466
P-значение: 9.145329456492403e-06
```

Выходит, что законы распределений различны.

## 5 ВЕРОЯТНОСТНЫЕ МОДЕЛИ ДЛЯ РЯДА

### 5.1 ОЧИСТКА ДАННЫХ

Исключим выбросы основанные, например, на отклонении данных от среднего значения на более чем 3 стандартных отклонения:

|     | LogReturns |
|-----|------------|
| 56  | 0.170719   |
| 74  | 0.241325   |
| 75  | 0.127618   |
| 129 | 0.159723   |
| 152 | 0.116695   |

Теперь протестируем очищенный ряд на стационарность тестом  
Дики-Фуллера  
гипотеза  $H_0$  : существует единичный корень, ряд нестационарный.

```
df_cleaned = df[(df['LogReturns'] >= mean - 3 * std_dev) & (df['LogReturns'] <=  
# Выполняем тест Дики-Фуллера на стационарность  
result = adfuller(df['LogReturns'])
```

```
ADF Statistic: -10.986895194268861  
p-value: 7.221386125274187e-20  
Critical Values: {'1%': -3.467631519151906,  
Ряд стационарен (отклоняем нулевую гипотезу)
```

Таким образом ряд доходностей стационарен, не имеет пропусков и выбросов, и не распределен нормально.

## ЛИТЕРАТУРА

1. "Основы Стохастической Финансовой Математики Том 1, А.Н. Ширяев.
2. "Критерии проверки гипотез о случайности и отсутствии тренда Б.Ю. Лемешко, И.В. Веретельникова.