An Inquiry into the Impact of Environmental Factors on Financial Markets

Shuya Zhao, Yi Tao, Hao Yu

Henan University of Economics and Law, Henan, China

ABSTRACT

The purpose of this paper is to study the impact of environmental factors on financial markets in depth by constructing mathematical models. We study and analyze the following four aspects.

Pearson correlation analysis was used to examine the relationship between environmental factors and the overall performance of the stock market. A scatter plot was created to examine the overall correlation, revealing a very weak correlation. Calculating the correlation coefficients between the variables and plotting a heatmap further confirms that the correlation is weak. The results are deemed insignificant by the significance test. In conclusion, there is no significant correlation between environmental factors and overall stock market performance.

The linear regression method is used to analyze the volatility of share prices in the specific energy industry and its influencing factors. The line graph illustrates the fluctuation in stock prices, showing short-term volatility in oil and gas stocks. Data refinement indicates that extreme weather has a minor impact on the oil and gas sector, implying that extreme weather has a limited effect on the energy sector.

The Ordinary Least Squares (OLS) linear regression model was utilized to analyze the linear relationship and correlation between AQI and Oil_Price, as well as between AQI and Gas_Price. The results indicated that the coefficients were insignificant, suggesting no significant linear correlation. The conclusion indicates that the Air Quality Index (AQI) has a lesser impact on the stock price of the energy sector.

A multivariate time series model is utilized to forecast oil and gas prices for the upcoming 20 months, relying on the initial 60 months of data. Firstly, the correlation between variables is analyzed, a smoothness test is performed, the model order is determined, and then tested. The results show that the model is ideal. Forecasting commodity prices for the next 20 months, the results show that oil and natural gas prices are stable with minor fluctuations. The overall prediction is accurate with minor deviations from historical data.

Keywords: Pearson Correlation Analysis; Regression Analysis; Multivariate Time Series Modeling; Python; Eviews

1 INTRODUCTION

In recent years, global climate change has become an increasingly serious issue with significant implications for the economy and financial markets. With rising temperatures, more extreme weather events, and declining air quality, environmental factors are becoming crucial variables in economic analysis. To address this issue, we will analyze the short-term impact of extreme weather events on sector stock prices, with a special focus on energy stock price movements, and explore the potential influence of changes in the Air Quality Index (AQI) on energy stocks [1]. To shed light on the impact of environmental factors on financial markets, we will construct a forecasting model for oil and gas prices that incorporates environmental factors. This will provide a scientific basis for investors and policymakers.

2 MODEL ESTABLISHMENT AND SOLUTION

2.1 Correlation Modeling

2.1.1 The analysis process of the model

In the financial market analysis, variables examined for correlation between environmental factors and overall stock market performance include stock indices, extreme weather events, air quality indices, and climate change. After analysis, it is not necessary to exclude any factors. Therefore, simple correlation analysis can be used to directly calculate the correlation degree of variables [2]. The correlation coefficient, significance, and thermal map are utilized to describe the correlation between variables.

2.1.2 Establishment of model

(1) Plot scatter plots

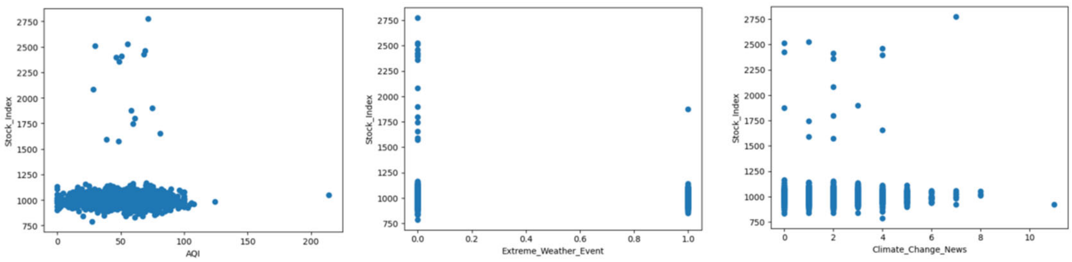

Figure 1: Scatterplot showing the relationship between environmental factors and the overall stock market

By visualizing and analyzing the data in Python, the scatterplot shows that environmental factors are weakly correlated with the stock market and have a small impact, and the correlation coefficient is further calculated to quantify the correlation [3].

(2) Calculate correlation

The correlation matrix was calculated using Python's corrcoef () function, and the presence of a statistically correlated relationship between the variables was tested by the value of r. The decision was made to utilize the Pearson correlation coefficient for the analysis [4].

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

After analysis, It was initially determined that Stock_Index is positively correlated with AQI and Climate_Change_News, and that Stock_Index is negatively correlated with Extreme_Weather_Event [5]. However, the correlation is very weak. The correlation matrix was visualized using the heatmap () function in Python.

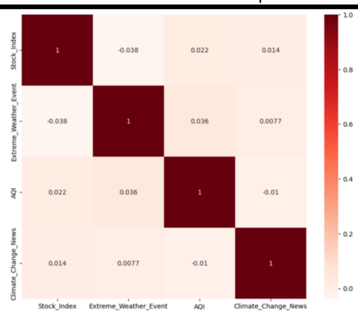

Figure 2: Heat Map of Environmental Factors and the Overall Stock Market

The figure above displays the correlation coefficient values in a heat map format, using color shades to represent the magnitude of the values. It is evident from the visualization that the correlation between Stock_Index and AQI, Climate_Change_News, and Extreme_Weather_Event is very weak [6]. Due to the presence of randomness and small sample sizes, the sample correlation coefficient cannot be directly used to determine whether the variables have a significant correlation. Therefore, it is necessary to conduct a significance test.

(3) significance test

The existence of a statistically significant relationship between the variables was tested to determine whether the p-value indicated significance ($p < 0.05$). First, assuming zero correlation between the variables, construct a new statistic, t, which follows a t-distribution with n-2 degrees of freedom when the variables are normally distributed [7]. Calculate the t-statistic, determine the corresponding probability P-value from the t-distribution, and then make a judgment.

$$t = r\sqrt{\frac{n-2}{1-r^2}} \tag{2}$$

After determination, the following P-value is obtained:

Table 1: Significance of each variable with Stock_Index

	Extreme_Weather_Event	AQI	Climate_Change_News
P	0.10376677165914357	0.3505678883925943	0.5567446342157089

The significance results are shown in the table above: obviously all the p-values are greater than 0.05, so the results are insignificant [8]. Therefore according to previous data environmental factors do not have a significant correlation with overall stock market performance.

2.2 Multivariable linear regression model

By screening and organizing the provided data, we obtain information on extreme weather events and energy industry stock prices for visualization. First, we conducted preprocessing on the dataset, including appropriate treatment of outliers and irrelevant values, to ensure accurate and reliable analysis results. We processed and visualized the data by utilizing the Python programming language and relevant libraries [9]. After screening the data, we examined the basic statistical information, such as the frequency of extreme weather events and the trends of specific stock indices. We also analyzed the maximum and minimum values, standard deviation, and mean values of stock prices to ensure data accuracy and the reliability of the analysis results.

After completing the processing of the dataset, we focused on the changes in oil stock values and natural gas stock values following extreme weather events. After integrating and analyzing the data, we obtained the results shown in Figures 3, respectively.

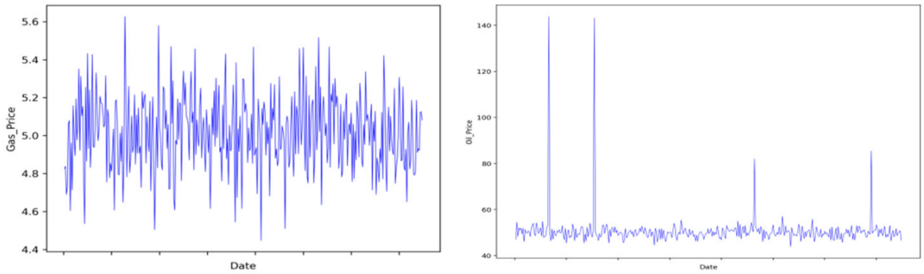

Figure 3: Stock Prices of Oil and Natural Gas Sectors Following Extreme Weather Events

It can be analyzed from Figures 3:

In the aftermath of extreme weather events, oil stocks typically remain stable, with oil prices mostly staying within a certain range. However, there are occasional brief periods of significant volatility. In contrast, natural gas stock prices have a relatively concentrated distribution of extreme weather impacts without any significant spikes [10]. Analyzing changes in oil and gas stock prices provides critical information for mathematical modeling and impact assessment.

In addition, to gain more insight into the impact of extreme weather events on stock prices in the oil and gas industry, we further processed the data. We selected 100 days for in-depth analysis, and as a result, we obtained the findings presented in Figures 4:

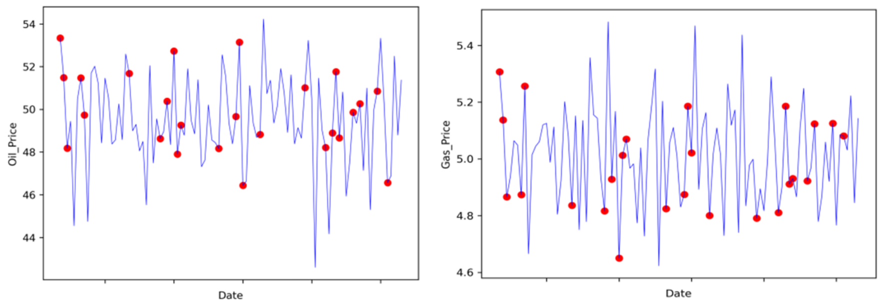

Figure 5: Curve of Oil and Natural Gas Stock Prices Subject to Extreme Weather Events Over a Given Time Period

From Figures 4 we observe the trend of stock prices in the oil and gas industry being affected by extreme weather events. This provides crucial insights for a more profound understanding of the impact of extreme weather events on specific energy sectors. By comparing Figures 3, it is evident that the fluctuations in the stock prices of the oil and gas industry in Figures 4 are more pronounced. This suggests that extreme weather events may have an impact on specific energy industries in the short term. This observation also offers valuable insights for future research and mathematical modeling.

To further illustrate the impact of extreme weather events, we integrate and compare dates, extreme weather events, and oil and gas stock prices to provide insight into the influence of extreme weather on the volatility of oil and gas stock prices. This analysis helps to assess the extent to which weather events affect stock prices and serves as a valuable reference for the industry.

2.3 Simple Linear Regression Mode

2.3.1The analysis process of the model

By examining the influence of fluctuations in the Air Quality Index (AQI) on stock prices in the energy sector, we also delve into the effects of AQI variations on oil and natural gas prices. We investigate the correlation between changes in Air Quality Index (AQI) and energy sector stock prices through regression analysis. The results of the regression analysis are used to determine the extent and direction of the impact of Air Quality Index (AQI) changes on energy industry stock prices.

Data pre-processing: To simplify the analysis, wvce calculate the monthly average data as the representative monthly value.

2.3.2 Modeling and Testing

（1）Modeling the Impact of AQI on Oil_Price

Setting Oil_Price as the explanatory variable and AQI as the explanatory variable, the one-way linear regression model is:

$$Y_1 = \alpha + \beta_1 X + \mu \tag{3}$$

where X is the AQI, Y_1 is Oil_Price, β_1 is the parameter to be estimated, α is the intercept term, μ is the random error term, and here is a representation of the others as considerations.

A trend plot of the two was produced using software EViews, Figure 5 below:

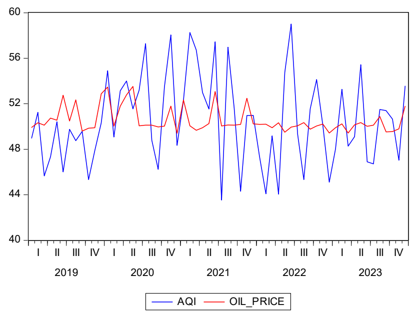

Figure 5: Trend plot of AQI and Oil_Price

From the analysis of Figure 5, it can be seen that the fluctuation trend between the two is roughly the same, when there is a large fluctuation in AQI, there is a small range of fluctuations in Oil_Price, and the trend of development and change between the two variables is relatively smooth. The regression equation obtained by the least squares method is:

$$Y_1 = 47.35 + 0.063X \tag{4}$$

$$R^2 = 0.052 \quad \overline{R^2} = 0.035 \quad F = 3.16 \quad n = 60$$

Based on the results of the model, it can be concluded that when all other factors are held constant, when AQI increases by one unit, Oil_Price increases by 0.063 units.

Test of goodness of fit: The decidable coefficient $R^2 = 0.052$, and when the decidable coefficient is closer to 1 indicates that the model's goodness of fit is better, so the model performs poorly in fitting the sample.

In summary, we can conclude that: there is a certain linear correlation between AQI and Oil_Price, but AQI has less influence on Qil_Price.

（2）Modeling the Impact of AQI on Gas_Price

Setting Gas_Price as the explanatory variable and AQI as the explanatory variable, the one-way linear regression model is:

$$Y_2 = \alpha + \beta_2 X + \mu \tag{5}$$

where X is the AQI, Y_2 is Gas_Price, β_2 is the parameter to be estimated, α is the intercept term, μ is the random error term, and here is a representation of the others as considerations.

The EViews was utilized to produce a trend graph for both, Figure 6 below:

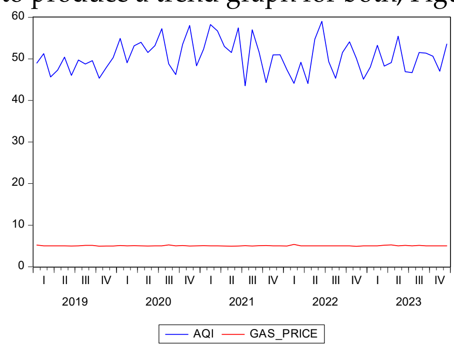

Figure 6: Trend plot of AQI and Gas_Price

As analyzed in Figure 6, Gas_Price remains relatively smooth when there are large fluctuations in AQI.

The regression equation obtained by the least squares method is:

$$Y_2 = 5.31 - 0.005X \tag{6}$$

$$R^2 = 0.069 \quad \overline{R^2} = 0.053 \quad F = 4.28 \quad n = 60$$

Based on the results of the model, it can be concluded that when all other factors are held constant, Gas_Price decreases by 0.005 units when AQI increases by one unit.

Test of goodness of fit: The decidable coefficient $R^2 = 0.069$,so the model performs poorly in fitting the sample.

In summary, we can conclude that: there is a certain linear correlation between AQI and Gas_Price, but AQI has less influence on Gas_Price.

2.4 Time series modeling

2.4.1 Preparation of the model

This question necessitates the creation of a forecasting model for commodity prices to anticipate the future prices of oil and natural gas by utilizing pertinent data, including environmental factors. After analyzing the data, including environmental factors such as Extreme Weather Events, AQI, Policy Changes, and Climate Change News, we have decided to utilize the time series ARIMA model to predict the average commodity prices for the next 20 months.

We utilized Python to preprocess the big data and compute monthly averages for each variable to eliminate random anomalies.

2.4.2 Modeling

（1）Correlation analysis

To conduct multi-factor time series forecasting, the initial step involves calculating the correlation between each variable. We utilize Python to generate a heatmap of each attribute. The heatmap reveals that there is no significant correlation between the price of oil and natural gas with each factor. The highest positive correlation is 0.23 with Oil_Price, and a positive correlation of 0.31 with Gas_Price.

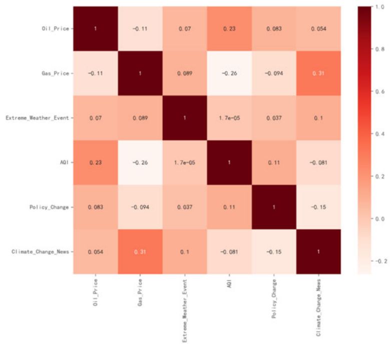

Figure 7: Heat Map of Factors and Commodity Prices

（2）Smooth Sequence Analysis

Since the ARIMA model requires the time series to be smooth, the first step is to conduct a smoothness test on the original data. If the series is not smooth, it needs to be converted from a differenced series to a smooth series before analysis. Using the LB statistic, if the p-value is less than the significance level of α=0.05, the original series is considered to be a non-white noise series, which is significant for the study. Draw a time series chart and subjectively assess the smoothness of the commodity price time series. If a clear trend is evident, it is considered a non-smooth series. If there is no obvious trend or uncertainty regarding smoothness, the ADF unit root statistic should be used for additional judgment on the smoothness of the series.

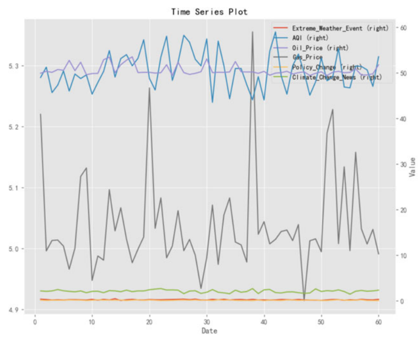

Figure 8: Timing diagram

As can be seen from the above figure, the original Oil_Price sequence is fairly smooth, while the Gas_Price sequence is unstable, especially in the 20th and 38th months. The price of natural gas suddenly rises and falls significantly. Therefore, it is necessary to use the Augmented Dickey-Fuller (ADF) test to assist in determining whether the sequence is smooth or not.

Autocorrelation coefficient ACF:

$$\rho_k = \frac{corr(X_{t+k}, X_t)}{\gamma_0} = \frac{\gamma_k}{\gamma_0}, k \in Z \tag{7}$$

Partial autocorrelation coefficient :

$$pacf(k,p) = \frac{\sum_{j=1}^{p}(\varphi_j \rho_{j-k}) - \sum_{j=1}^{p-1}(\varphi_j \rho_{j-k})}{\rho_{p-k}} = \varphi_p \tag{8}$$

k denotes the order of the interval and φ_p denotes the partial correlation coefficient after subtracting the effect of $\rho_1$to ρ_{p-1}.

For the ADF statistics, as shown in the table below, the ADF statistics of Oil_Price and Gas_Price are -6.4334 and -4.7911, respectively, which are both less than 0.05, and the p-values of Oil_Price and Gas_Price are both less than 0.05, which indicates that the corresponding time series of Oil_Price and Gas_Price are smooth series.

（3）Determine the model order

Determining the model order is initially established by examining autocorrelation and partial autocorrelation plots. These plots are generated using the autocorrelation coefficient (ACF) and the partial autocorrelation coefficient (PACF) through the ACF and PACF method. This process helps in obtaining autocorrelation and partial autocorrelation plots for the smoothed serie. The last lag value of the autocorrelation coefficient outside the threshold represents the q value, while the last lag value of the partial autocorrelation coefficient outside the threshold indicates the p value. From the figure below, we can see that the values of p and q for Oil_Price are both 18. However, it is clearly inaccurate to state that the values of p and q for Gas_Price are also both 18.

Figure 9: Autocorrelation and partial autocorrelation plots for Oil_Price

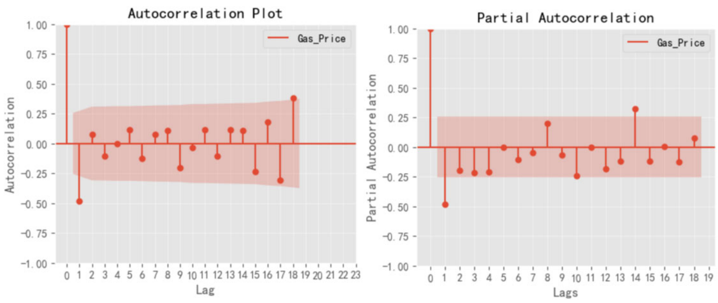

Figure 10: Autocorrelation and partial autocorrelation plots for Gas_Price

After the above initial judgment, the order of the model still cannot be determined. The next step is to solve for the p and q taking values by using the Akaike Informative Criterion AIC and Bayesian Informative Criterion BIC.

Bare Pool Informational Capacity AIC Formula：

$$AIC = 2k - 2\ln(L) \tag{9}$$

Bayesian Information Capacity BIC formula：

$$BIC = -2\ln(L) + k\ln(n) \tag{10}$$

L denotes the likelihood function of the model, k denotes the number of models

We implement this in python, we start the order from 0, and after fitting the model q*p times, we compare the values of AIC and BIC by the results of the model fitting, the smaller the value of BIC, the model is relatively optimal, and then we use the number of orders of this model as the number of orders of the final model fitting. The final optimal fixed order for Oil_Price is p=0, q=1 and for Gas_Price is p=0, q=1.

（4）model checking

Through the analysis of the above operation, the value of each parameter in ARIMA is obtained, in order to make the final prediction result more accurate, we also need to carry out the model parameter test and the model significance test. The model significance test is also known as the white noise test of the residuals, if p<0.05, the residuals are non-white noise series, otherwise the residuals are white noise series. The parametric test of the model is to test whether each unknown parameter is significantly zero, test whether the model is the most compact, if the parameter is not significant non-zero, can be removed from the fitted model from the elimination, look at the t-statistic.

To ensure accuracy, a residual test plot is also made, as can be seen in the following figure: the time series plot of the residuals is smoothly fluctuating, the data nodes on the Q-Q plot are near the diagonal line, and whether or not the residuals conform to a normal distribution indicates that useful signals have been extracted into the time series model, and all of the above indicates that the desired results are obtained.

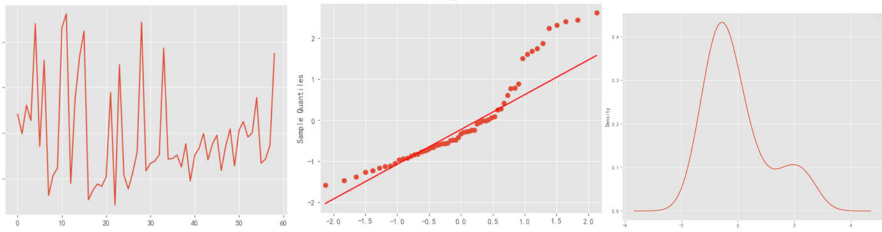

Figure 11: Residual test plot for Oil_Price

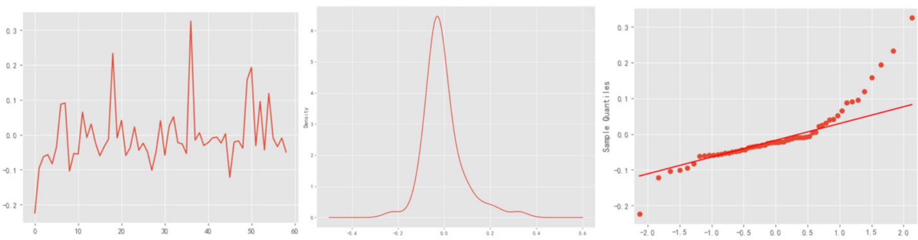

Figure 12: Residual test plot for Gas_Price

（6）model prediction

After the above analysis, the obtained relative optimal model is used to make predictions using the predict function in python to get the average price of oil and natural gas for each month of the next 20 months.

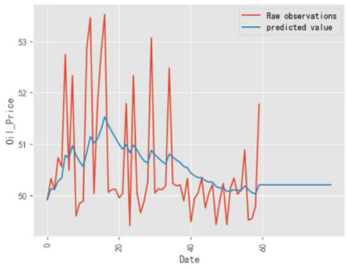

Figure 13: Time series plot of Oil_Price

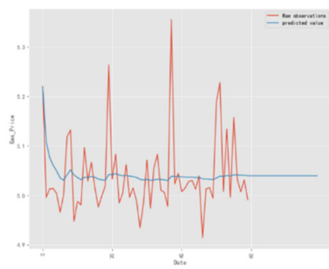

Figure 14: Time series plot of Gas_Price

As can be seen from the graphs in the chart above, the trend in oil and natural gas prices over the next 20 months is relatively stable, with very little volatility. If there are no sudden environmental, human, or other factors influencing the price of oil and natural gas, then the price of oil and natural gas will actually be more stable. Overall, our forecast results are quite accurate and not significantly different from those of previous months. In addition, the forecast results we have obtained are only based on oil and gas prices and environmental factors from the first 60 months. Without considering the impact of other factors, the real oil and gas prices will differ from our results. The actual situation will be influenced by the economic environment, unexpected accidents, and various other factors.

3 CONCLUSION

Based on the research on the impact of environmental factors on financial markets in this article, the following conclusions are drawn:

Environmental factors have a certain impact on the financial market, but the influence is relatively minor. A multivariate time series model is utilized to forecast oil and gas prices for the upcoming 20 months, relying on data from the initial 60 months. The model is ideal, indicating that the multivariate time series model is superior for studying the relationship between environmental factors and the price of a specific commodity in the financial market.

REFERENCES

[1] Chien, F., Sadiq, M., Kamran, H. W., Nawaz, M. A., Hussain, M. S., & Raza, M. (2021). Co-movement of energy prices and stock market return: environmental wa velet nexus of COVID-19 pandemic from the USA, Europe, and China. *Environment al Science and Pollution Research*, *28*(25), 32359-32373.

[2] Khan, H., Weili, L., & Khan, I. (2022). Institutional quality, financial development and the influence of environmental factors on carbon emissions: evidence from a global perspective. *Environmental Science and Pollution Research*, 1-13.

[3] Alessi, L., Ossola, E., & Panzica, R. (2021). What greenium matters in the stock m arket? The role of greenhouse gas emissions and environmental disclosures. *Journal of Financial Stability*, *54*, 100869.

[4] Nasir, M. A., Canh, N. P., & Le, T. N. L. (2021). Environmental degradation & ro le of financialisation, economic development, industrialisation and trade liberalisati on. *Journal of environmental management*, *277*, 111471.

[5] Wang, S., Wang, H., Wang, J., & Yang, F. (2020). Does environmental information disclosure contribute to improve firm financial performance? An examination of th e underlying mechanism. *Science of the Total Environment*, *714*, 136855.

[6] Ramelli, S., & Wagner, A. (2020). What the stock market tells us about the conseq uences of COVID-19. *Mitigating the COVID Economic Crisis: Act Fast and Do Whatev er*, *63*(7), 63-76.

[7] Acheampong, A. O., Amponsah, M., & Boateng, E. (2020). Does financial develop ment mitigate carbon emissions? Evidence from heterogeneous financial economie s. *Energy Economics*, *88*, 104768.

[8] Wong, J. B., & Zhang, Q. (2022). Stock market reactions to adverse ESG disclosure via media channels. *The British Accounting Review*, *54*(1), 101045.

[9] Guo, M., Kuai, Y., & Liu, X. (2020). Stock market response to environmental polici es: Evidence from heavily polluting firms in China. *Economic Modelling*, *86*, 306-316.

[10] Rasheed, R., Rizwan, A., Javed, H., Sharif, F., & Zaidi, A. (2021). Socio-economic a nd environmental impacts of COVID-19 pandemic in Pakistan—an integrated analy sis. *Environmental Science and Pollution Research*, *28*(16), 19926-19943.