# Energy determinants of CO2 prices: results of exploratory data analysis

**Liudmila G. Reshetnikova** (0000-0001-9208-6005)[1(1)], **Natalia B. Boldyreva** (0000-0002-1610-6075)[1], **Anton P. Devyatkov** (0000-0002-4887-3797)[1], **Mariya S. Tsyganova** (0000-0002-9832-7882)[1], **Danila V. Ovechkin** (0000-0001-9356-6254)[1]

[1] University of Tyumen, Tyumen, Russia

**Abstract**. The key instrument for regulating the reduction of greenhouse gas (CO2) emissions is the emissions trading system (ETS), which Russia chose to create. One of the key issues for the ETS participants is the mechanism of CO2 pricing. Energy sources are the most important factors in the CO2 price. The study of the relationship between them and the CO2 price at foreign ETSs gives contradictory results. The paper investigates the relationship between energy variables and CO2 price using exploratory analysis tools as the first stage of machine learning. The univariate analysis showed that the laws of price distribution of futures contracts for CO2 and coal are further from normal in comparison with the prices of futures contracts for gas and Brent oil. Logarithmization improved the statistics of the data. Bivariate analysis showed a close relationship between the prices of CO2 and coal futures contracts. The price data for the other energy variables showed a weak to moderate relationship. Correlation analysis, taking into account the different time lags between the energy variables and CO2, indicates that it is appropriate to include past energy price information in the model. The close linear relationships of the energy variables suggest exploring opportunities to reduce the dimensionality of the data. Exploratory analysis revealed groups of data that would be appropriate to describe with different machine learning models. The presence of data groups in the resulting variable indicates the presence of other CO2 pricing factors that should be identified and taken into account in modeling.

**Keywords:** Greenhouse gas emissions · Emissions trading system · CO2 · Allowance pricing · Machine learning.

## 1. Introduction

The Decree of the President of the Russian Federation No. 474 dated July 21, 2020, as part of the national goal “Comfortable and safe environment for life”, sets the task of reducing emissions of hazardous pollutants. The point at issue is the reduction of greenhouse gas emissions in the process of economic activity of market economy agents.

The cap-and-trade system is a key instrument for regulating CO2 emission reductions worldwide (Schmalensee and Stavins, 2017). ETS allows to attract private capital to solve the problem of CO2 emissions reduction. The experience of foreign countries indicates that, under certain conditions, ETS contribute to the reduction of CO2 emissions (Zhang et al., 2019; Gao et al., 2020; Heiaas, 2021).

[1] Corresponding author: reshetnikova-l@yandex.ru

Russia is faced the task of developing the climate agenda at the national level. Federal Laws No. 296-FZ and No. 34-FZ made a choice in favor of creating the ETS for greenhouse gas emissions. One of the key issues for ETS participants is the pricing mechanism. It plays an important role in the dissemination of more energy efficient and low carbon technologies. Since the creation of the European Union Emissions Trading Scheme (EU ETS) in 2005, a lot of scientific literature has appeared on the identification of factors that affect the CO2 price. Researchers take the balance between supply and demand as the basis for pricing, and also recognize the influence of other factors related to market structure, institutional policy and climate (Chevallier, 2013). They show that energy sources are the most important determinants of the CO2 price and that the nature of this relationship varies depending on the period under consideration (Alberola et al., 2007).

An important determinant of CO2 price is the marginal cost of switching from carbon-intensive energy sources to less carbon-intensive sources for electricity and heat generation (Alberola et al., 2007; Aatola et al., 2013).

Researchers, using econometric methods, draw different conclusions about the nature and form of the relationship between energy variables and the CO2 price (Zhu et al., 2019; Liu and Jin, 2020; Batten et al., 2020; Chu et al., 2020; Duan et al., 2021). It should be noted that the conclusions of researchers are often contradictory for a number of reasons: data, research horizon, methodology, national characteristics of the ETS, climatic features of the region, the standard of living of the population and, in general, the level of economic development of the region.

A big gap in the Russian scientific literature is the lack of publications on the topic of CO2 pricing. Developing the environmental policy, the Russian government cannot rely on the estimates of foreign researchers, which sets the task of conducting research for Russia. The article contains an exploratory analysis of the impact of energy resource prices on CO2 price as a necessary step in machine learning. To achieve the goal, the following tasks are solved: on the basis of the literature review, the energy factors of emission allowance pricing are identified; a univariate analysis of variables and their bivariate analysis in accordance with the concept of exploratory analysis are carried out.

The paper has the following structure: section 2 discloses the research methodology and data, section 3 presents the results of exploratory analysis, section 4 critically discusses the results obtained and section 5 formulates the final conclusions and identifies directions for further research.

## 2. Materials and Methods

In the last decade, the researchers have noted a paradigm shift in data analysis – a shift to machine learning technologies due to the accumulation of large amounts of data after the “information explosion” and overcoming the limitations of computing power. The initial stage of machine learning is exploratory analysis. It reveals the most general dependencies, patterns and trends, as well as the nature and properties of the data being analyzed (Tukey, 1981).

The purpose of the study is to identify the relationships between energy variables and the CO2 price using exploratory analysis tools.

The methodological apparatus of the study is represented by traditional statistical methods. Within the exploratory analysis, univariate and bivariate analyses were performed. Visual histogram analysis, distribution analysis, and descriptive statistics of the variables were used for univariate analysis. Bivariate analysis involved the correlation analysis to identify relationships.

The information base of the study was the daily futures contracts prices for the period from March 25, 2013 to April 24, 2022 for Brent crude oil and natural gas traded on the Moscow Exchange, prices of futures contracts for coking coal, The prices of Brent crude oil and natural gas futures traded on the Moscow Exchange and the prices of European Union Allowances (EUA) futures traded on the European Energy Exchange (EEX). Since CO2 futures are quoted in euros, we converted the prices of coal futures contracts, Brent crude oil futures contracts and natural gas futures contracts into euros using the daily exchange rate data available from the Central Bank of Russia.

## 3. Results

The authors begin the exploratory data analysis with a visual analysis of the histograms. The histograms of gas and oil prices show a distribution closer to normal. The authors logarithmed the CO2 and energy price data, testing the hypothesis that this procedure could bring the distribution closer to normal (Fig. 1).

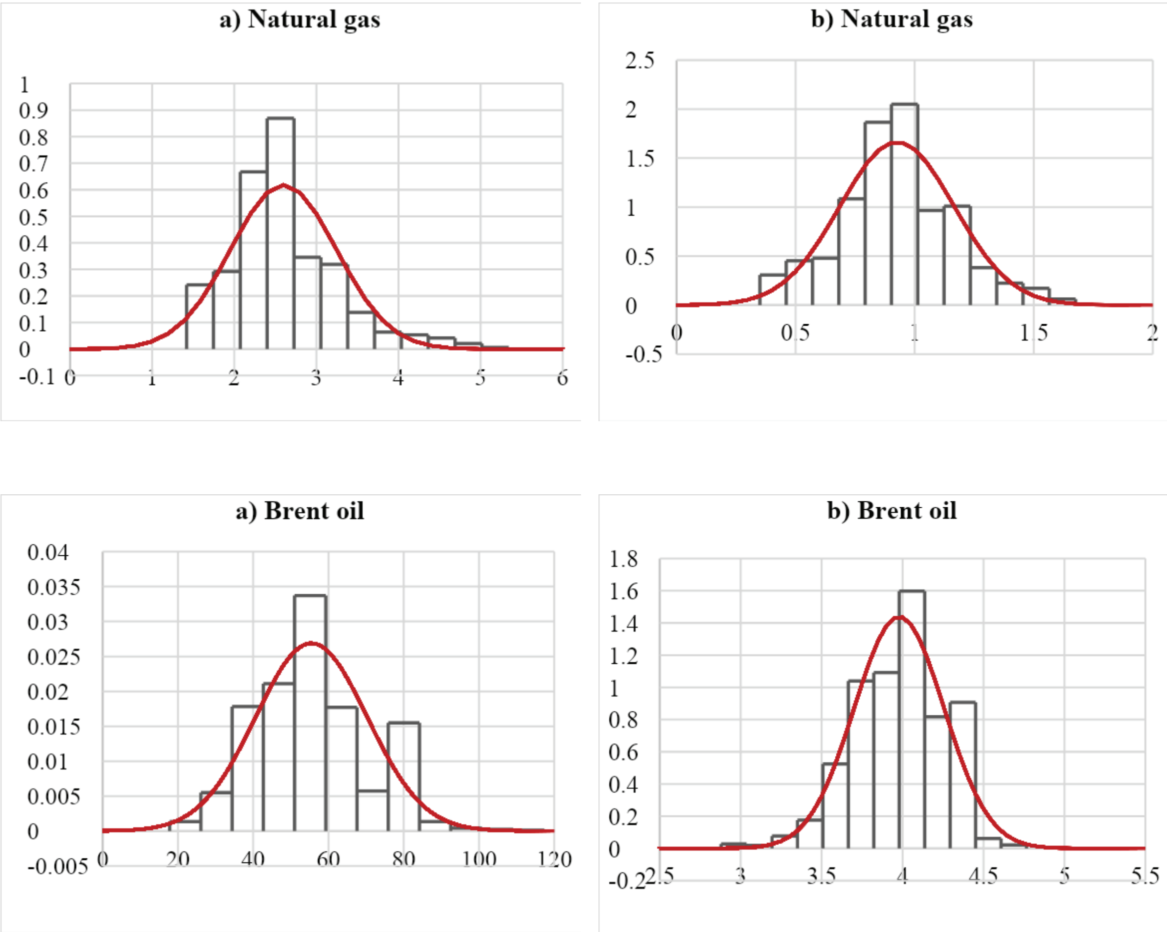


**Fig. 1.** Histograms of the empirical distribution and density plots of the normal distribution of a) prices and b) logarithmic prices.

The logarithm of prices had a positive effect on the distribution of gas and oil data.

The authors hereof analyzed the descriptive statistics (Table 1).

**Table 1.** Descriptive statistics for $CO_2$ and energy price data, euro.

| | $CO_2$ | Natural gas | Brent oil | Coking coal |
|---|---|---|---|---|
| Average | 19.33 | 2.60 | 55.46 | 163.54 |
| Median | 8.44 | 2.50 | 54.38 | 159.22 |
| Standard deviation | 18.99 | 0.65 | 14.84 | 78.47 |
| Skew | 1.79 | 0.96 | 0.37 | 2.05 |
| Kurtosis | 2.96 | 1.39 | -0.12 | 5.31 |

The values of the average prices for gas, oil and coal are close to their median values, which is typical for a normal distribution. A distribution close to normal is also characterized by skew and kurtosis. The normal distribution has a skew between -0.5 and 0.5. This criterion is met by oil prices.

The kurtosis value close to zero (-0.12), characteristic of a normal distribution, is shown by oil prices. A small negative kurtosis indicates a small amount of emissions. The prices for coal, CO2 and gas have peaked distributions, signaling a large number of values in the tails of the distributions. Requires pre-treatment of emissions.

Descriptive statistics for logarithmic prices for CO2 and energy prices are presented in Table 2.

**Table 2.** Descriptive statistics for logarithmic prices for CO2 and energy prices, euro

| | CO2 | Natural gas | Brent oil | Coking coal |
|---|---|---|---|---|
| Average | 2.56 | 0.93 | 3.98 | 5.01 |
| Median | 2.13 | 0.92 | 4.00 | 5.07 |
| Standard deviation | 0.87 | 0.24 | 0.28 | 0.41 |
| Skew | 0.47 | 0.15 | -0.40 | 0.59 |
| Kurtosis | -1.07 | 0.19 | 0.28 | 0.35 |

The values of the average logarithmic prices for gas, oil and coal almost equaled their median values, which is typical for a normal distribution. Mean and median values of log CO2 prices approached each other.

Skew, characteristic of a normal distribution, has logarithmic prices for almost all variables (from -0.40 for oil to 0.47 for CO2), only the skew for coal exceeds the threshold value (0.59). As a result of the logarithm, the kurtosis values approached zero (from -1.07 for CO2 to 0.35 for coal), which is typical for a normal distribution and a small amount of emissions.

Visual analysis and descriptive statistics show that the laws of distribution of CO2 and coal prices are further from normal in comparison with other variables. Logarithm improved the statistics of the data, bringing their distributions closer to normal.

As part of a two-dimensional analysis, we explored dependencies in the data. The scatterplots of both direct and logarithmic energy and CO2 prices show several groups of data, which may indicate the presence of other factors influencing the price for CO2. This is shown in fig. 2 for coal. It is advisable to single out these data groups and describe them with different machine learning models.

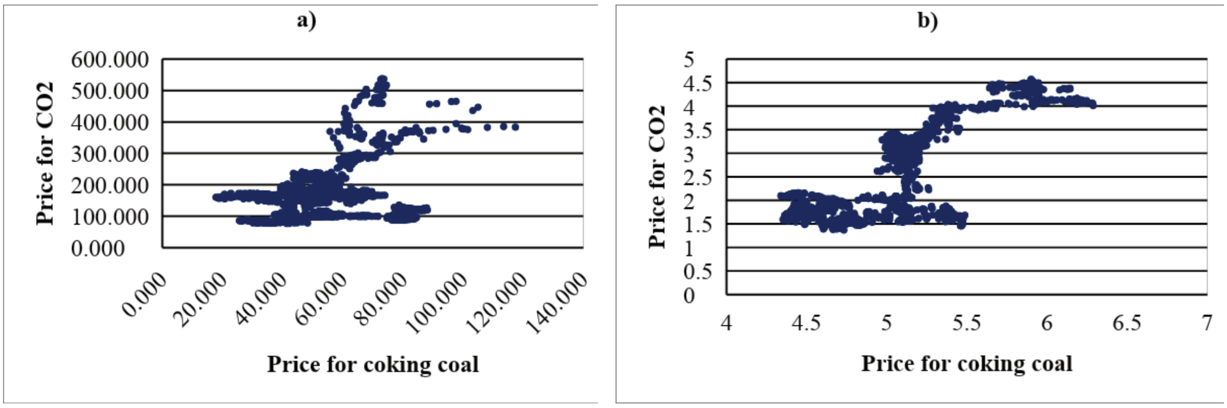


**Fig. 2.** Scatterplots of a) prices, b) logarithmic prices for coal and $CO_2$.

Graphs in fig. 3 show close linear relationships of energy variables. Dimensionality reduction, getting fewer main variables should be considered for machine learning.

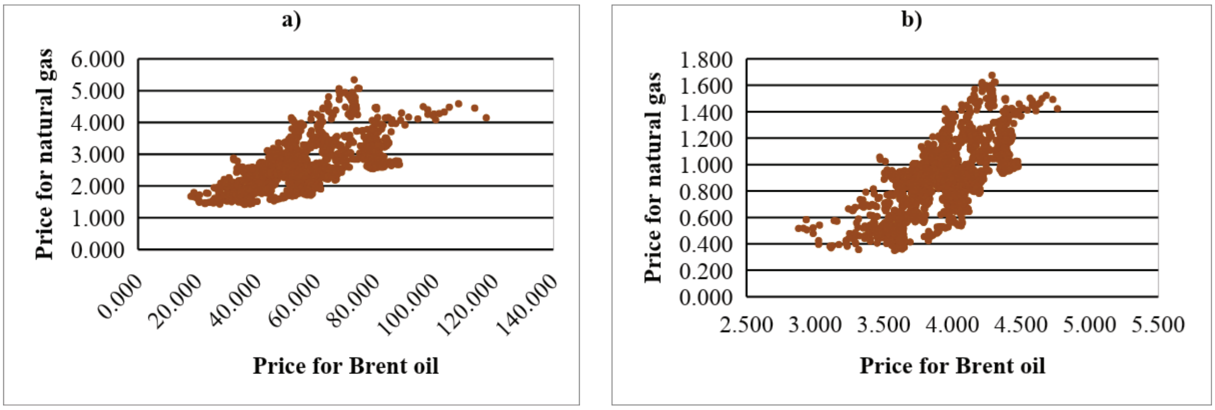


**Fig. 3.** Scatterplots of a) prices, b) logarithmic energy prices.

**Table 3.** The results of the correlation analysis.

| | CO2 | Natural gas | Brent oil | Coking coal |
|---|---|---|---|---|
| CO2 | 1.00 | | | |
| Natural gas | 0.33 | 1.00 | | |
| Brent oil | 0.22 | **0.65** | 1.00 | |
| Coking coal | **0.83** | **0.52** | 0.24 | 1.00 |

The correlation coefficients show that the CO2 price is closely related to the price for coal (0.83). At the same time, there is a significant linear relationship between gas prices and oil and coal prices.

The logarithm procedure had a positive effect on the results of the univariate analysis. The authors performed a correlation analysis for the logarithmic prices of all variables (Table 4).

**Table 4.** Correlation of logarithmic prices of CO2 and energy carriers

| | CO2 | Natural gas | Brent oil | Coking coal |
|---|---|---|---|---|
| CO2 | 1.00 | | | |
| Natural gas | 0.04 | 1.00 | | |
| Brent oil | 0.09 | **0.67** | 1.00 | |
| Coking coal | **0.76** | 0.34 | 0.20 | 1.00 |

The logarithm did not lead to an increase in the correlation coefficients between CO2 and energy variables.

Real markets are not perfectly efficient, so the authors performed a correlation analysis taking into account different time lags between logarithmic energy and CO2 prices (Table 5) and their logarithmic values (Table 6).

**Table 5**. Correlation of CO2 and energy prices with time lag

| Time lag | 0 months | 3 months | 6 months | 9 months |
|---|---|---|---|---|
| Natural gas | 0.33 | 0.30 | -0.29 | **-0.51** |
| Brent oil | 0.22 | 0.17 | -0.18 | **-0.34** |
| Coking coal | 0.83 | **0.84** | 0.65 | 0.55 |

The results of the correlation analysis show a close relationship between the prices for CO2 and coal with a time lag of 3 months, as well as significant and moderate negative relationships between the price of CO2 and the prices of gas and oil with a time lag of 9 months.

Table 6 presents the correlation coefficients between the logarithmic prices of energy variables with a delay of 3, 6 and 9 months and CO2 prices.

**Table 6.** Correlation of logarithmic CO2 futures prices and logarithmic energy futures prices with a time lag

| Time lag | 0 months | 3 months | 6 months | 9 months |
|---|---|---|---|---|
| Natural gas | 0.04 | 0.02 | -0.31 | **-0.40** |
| Brent oil | 0.09 | 0.02 | -0.13 | **-0.24** |
| Coking coal | 0.76 | 0.76 | 0.75 | **0.79** |

Statistically, a change in energy prices leads to a change in the price for CO2 in 9 months. These results indicate that the model for explaining CO2 prices should take into account not only current information, but also past information on energy prices.

## 4. Discussion

Researchers are unanimous in identifying energy variables as the main determinants of CO2 pricing. A review of the literature showed the absence of an unequivocal answer to the question about the nature, form and strength of the influence of energy variables on the price for CO2. One reason is related to the use of different econometric models. In machine learning, data comes first, not models. The role of data explains the importance of exploratory analysis.

The main result of the exploratory analysis substantiates the importance of considering the time lag when modeling the relationship between energy variables and CO2 prices. This result is a consequence of the insufficient maturity of the European CO2 market, which is consistent with the findings (Meraz et al., 2021; Ghazani and Jafari, 2021).

## 5. Conclusion

Market agents are interested in predicting CO2 prices to make investment and financial decisions. This study conducted an exploratory analysis of energy variables in order to incorporate the data into a machine learning model for CO2 price prediction.

The results show that it is reasonable to use logarithmic coal and gas prices with a time lag of 9 months in determining the CO2 price. Given the high linear relationships of the energy variables, we should consider reducing the dimensionality by moving to a smaller number of main variables.

Moreover, the results of the exploratory analysis allowed us to identify groups of data that are appropriate to describe with different machine learning models. The data groups in the resulting variable indicate the presence of other CO2 pricing factors that should be identified and considered in future studies.

## Acknowledgments

The study was funded by the Russian Science Foundation grant No. 22-28-02032. https://rscf.ru/project/22-28-02032/.

## References

1. J. Tukey, Analiz rezultatov nablyudenii [Analysis of observational results] (Mir, Moscow, 1981). Accessed on: October 26, 2022. [Online]. Available: https://bookree.org/reader?file=581991
2. P. Aatola, M. Ollikainen, A. Toppinen, Energy Econ. **36**, 380-395 (2013). https://doi.org/10.1016/j.eneco.2012.09.009
3. E. Alberola, J. Chevallier, B. Cheze, Energy Pol. **36**, 787-797 (2008). https://doi.org/10.1016/j.enpol.2007.10.029Get rights and content

4. J.A. Batten, G.E. Maddox, M.R. Young, Energy Econ. **96**, 105016 (2020). https://doi.org/10.1016/j.eneco.2020.105016
5. J. Chevallier, Int. J. Appl. Log. **4(4)**, 1-7 (2013). https://doi.org/10.4018/ijal.2013100101
6. W. Chu, S. Chai, X. Chen, M. Du, Sustainability **12(14)**, 5581 (2020). https://doi.org/10.3390/su12145581
7. K. Duan, X. Ren, Y. Shi, T. Mishra, C. Yan, Energy Econ. **95**, 105131 (2021). https://doi.org/10.1016/j.eneco.2021.105131
8. Y. Gao, M. Li, J. Xue, Y. Liu, Energy Econ. **90**, 104872 (2020). https://doi.org/10.1016/j.eneco.2020.104872
9. A.M. Heiaas, Rev. Bus. Econ. Stud. **9(1)**, 84-120 (2021). https://doi.org/10.26794/2308-944X-2021-9-1-84-120
10. X. Liu, Z. Jin, Energy Sustain. Dev. **55**, 82-94 (2020). https://doi.org/10.1016/j.esd.2020.01.008
11. M. Meraz, J. Alvarez-Ramirez, E. Rodriguez, Information Efficiency of the European CO2 Trading Market in the Period 2008-2021 (2021). https://doi.org/https://doi.org/10.21203/rs.3.rs-1050631/v1
12. M.M. Ghazani, M.A. Jafari, Environ. Sci. Pollut. Res., **28**, 61080-61095 (2021). https://doi.org/10.1007/s11356-021-15044-5
13. R. Schmalensee, R. Stavins, Oxford Rev. Econ. Pol. **33(4)**, 572-588 (2017). https://doi.org/10.1093/OXREP/GRX040
14. H. Zhang, M. Duan, Z. Deng, J. Clean. Prod. **234**, 912-924 (2019). https://doi.org/10.1016/j.jclepro.2019.06.247
15. B. Zhu, S.Ye, D. Han, P. Wang, K. He, Y. Wei, R. Xie, Energy Econ. **78**, 202-216 (2019). https://doi.org/10.1016/J.ENECO.2018.11.007