Predicting Uranium Prices with Linear Regression

Mattia Andrea Antinori, Leonard Ray Inciso, Egor Pantyukhin

November 2023

1 Introduction

The quest for sustainable and clean energy sources stands as a cornerstone in the global pursuit of environmental conservation and economic development. Within this landscape, uranium emerges as a critical player, and our analysis delves into the intricate dynamics of uranium prices. By comprehending and predicting the trajectory of uranium prices, we aim to contribute insights crucial for stakeholders navigating the evolving energy sector.

Recent market dynamics, accentuated by the uranium price hitting a 15-year high (Shaw, 2023), serve as a catalyst for our exploration. This surge is not merely a response to geopolitical events but is intricately woven into the fabric of broader global energy trends. Nations across the world are increasingly turning to nuclear power as a sustainable alternative to traditional fossil fuels, aligning with their commitments to meet sustainability goals (Shaw, 2023).

The impetus for our in-depth examination of uranium prices arises from the need to unravel the multifaceted influences on this vital energy commodity. The resurgence of interest in nuclear power, marked by a surge in global demand, coupled with concerns over potential supply disruptions in geopolitically sensitive regions, forms the crux of our exploration (Shaw, 2023).

To fortify our analytical approach, we incorporate the uranium price forecast by Fitri Wulandari (Capital.com), providing nuanced insights into the potential impacts of the protracted Russia-Ukraine conflict on uranium prices. This forecast highlights the challenges posed by global supply issues and underscores the growing significance of uranium in the context of nations contemplating nuclear power for electricity generation (Fitri, Capital.com).

Furthermore, insights from Australia's Department of Industry, Science Resources shed light on the intricate interplay between uranium supply and demand. The projection of potential supply shortfalls, amidst rising global demand and geopolitical factors, adds a layer of complexity to the uranium market, compelling us to delve deeper into the factors influencing uranium prices (Fitri, Capital.com; Australia's Department of Industry, Science Resources, 2022).

The dataset for the analysis of Uranium prices consists of the following variables:

- **fed_rate:** Federal Reserve interest rate. It represents the interest rate set by the Federal Reserve, which influences borrowing costs in the economy.
- **brent:** Brent crude oil prices. This variable reflects the market price of Brent crude oil, a major benchmark for worldwide oil prices.
- **coal:** Prices of coal. It denotes the cost of coal, an important energy resource used for various purposes, including electricity generation.
- **cpi:** Consumer Price Index (CPI). CPI measures the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services. It is an indicator of inflation.
- **electricity:** Average Price: Electricity per Kilowatt-Hour in U.S. City Average
- **usd_index:** Closing value of the U.S. Dollar Index. The U.S. Dollar Index measures the value of the U.S. dollar relative to a basket of foreign currencies.
- **sp_500:** Closing value of the S&P 500 Index. The S&P 500 is a stock market index that measures the performance of 500 large companies listed on stock exchanges in the United States.

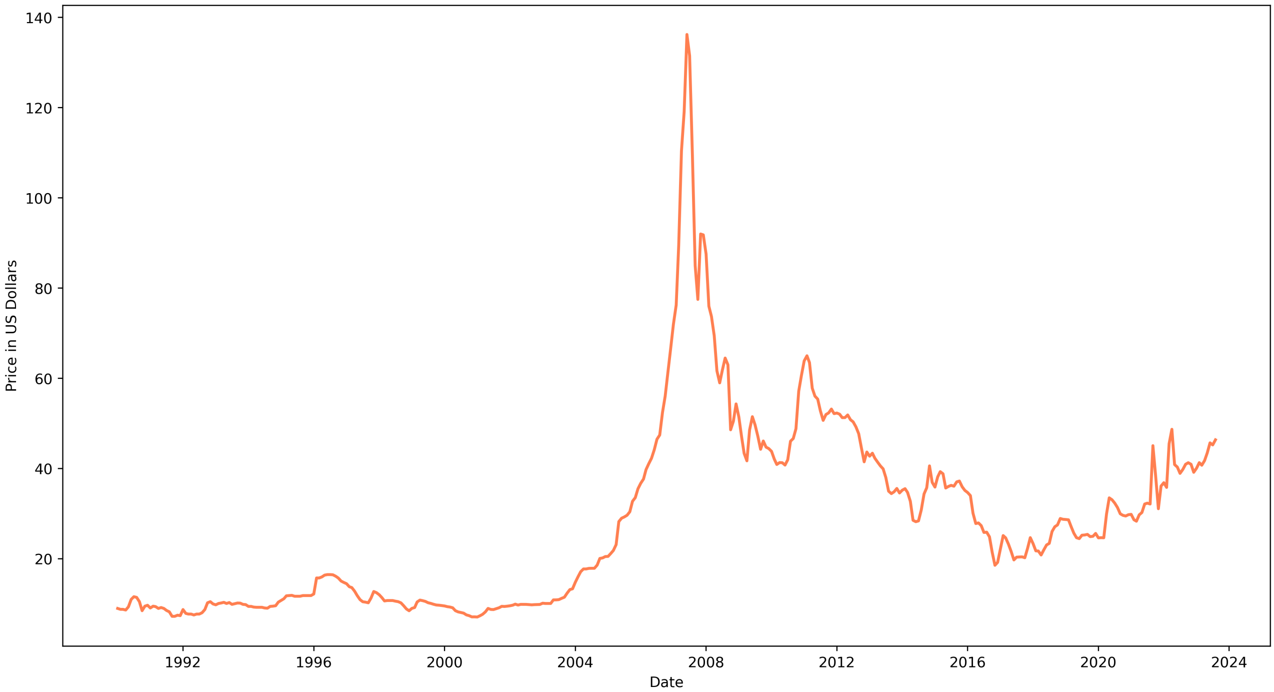

Figure 1: Uranium Prices 1990-2023

2 Data Preprocessing and Exploratory Analysis

Data manipulation was performed to merge separate datasets for each regressor and the regressand from FRED. The dataset was then narrowed down to observations from January 1, 2009, to August 1, 2023, resulting in 176 observations. No extensive cleaning was required, given the reliability of FRED data.

In this project, we decided to use a new practical correlation coefficient Phi_K, based on several refinements of the test of Pearson's hypothesis of independence of two variables. The combined features of Phi_K form an advantage over existing coefficients. First, it works consistently between categorical, ordinal and interval variables. Second, it captures non-linear dependency. Third, it reverts to the Pearson correlation coefficient in case of a bi-variate normal input distribution. These are useful features when studying the correlation matrix of variables with mixed types Baak, M., Koopman, R., Snoek, H., Klous, S. (2020).

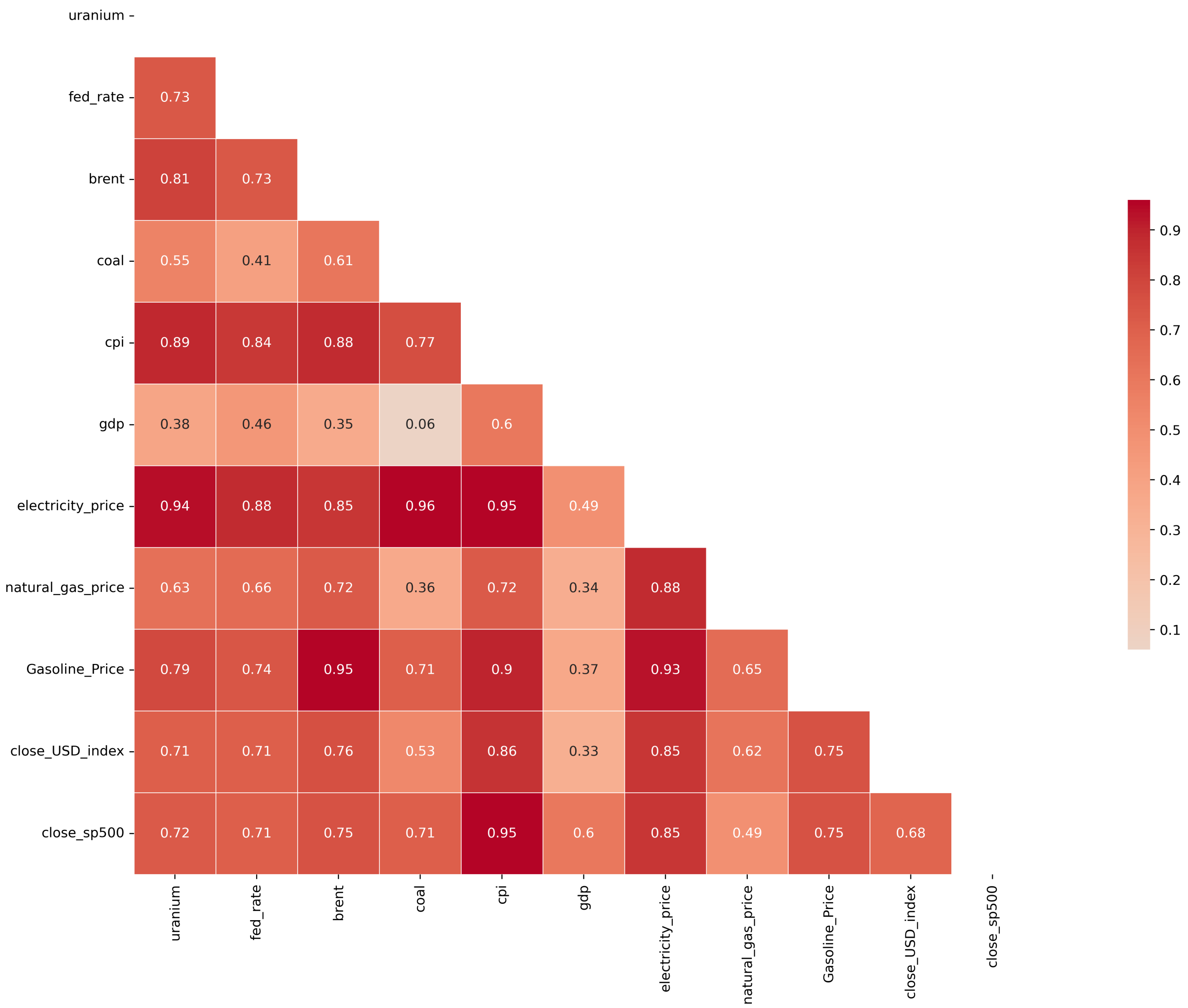

Figure 2: Correlation Matrix

High Correlations: Notable high correlations exist between:

- Uranium and CPI (0.864)
- Uranium and Electricity (0.862)

- CPI and Electricity (0.948)
- CPI and S&P 500 (0.947)

These relationships suggest potential multicollinearity issues, indicating that changes in one variable may be associated with changes in another.

Moderate Correlations:

- Fed Rate and Brent (0.689)
- Brent and Electricity (0.809)
- USD Index and CPI (0.852)

These correlations, while not extremely high, should still be considered in regression modeling to avoid biased coefficient estimates.

Low Correlations: Coal with Fed Rate (0.420) and USD Index (0.527) These variables exhibit relatively lower correlations, indicating less linear association.

The method of least squares (OLS) was used to select variables in the regression model. The t-statistic was used to assess the significance of the influence of individual variables on the dependent variable. The t-statistic provides information on the degree of significance of the regression coefficients, allowing us to identify variables that make a statistically significant contribution to the model. This method allows the statistical significance of the variables to be taken into account when forming the regression model.

Table 1: Regression Coefficients

Variable	Coef	Std Err	t	P> $\|t\|$	[0.025	0.975]
fed_rate	-1.4528	0.535	-2.714	0.007	-2.509	-0.396
brent	-0.0707	0.042	-1.665	0.098	-0.155	0.013
coal	0.0479	0.009	5.293	0.000	0.030	0.066
cpi	0.5938	0.112	5.296	0.000	0.372	0.815
usd_index	-0.9746	0.153	-6.380	0.000	-1.276	-0.673
electricity	56.1483	133.817	0.420	0.675	-208.021	320.317
sp_500	-0.0118	0.001	-8.210	0.000	-0.015	-0.009

Based on an analysis of the data using the least squares (OLS) method and focusing on the t-statistics of the variables, the following variables were selected for inclusion in the regression model: coal, cpi, usd_index and sp_500. This selection is based on the statistical significance of these variables in the context of their influence on the dependent variable, making them potentially important predictors in this analysis.

Analysis of outliers in the prices of Uranium (Figure 3) indicates their potential impact on linear regression models. Based on the constructed box plot, it was decided to use data from 2009 onwards. This decision was made to reduce the impact of outliers on the statistical estimates and to provide more stable modeling consistent with the data analysis.

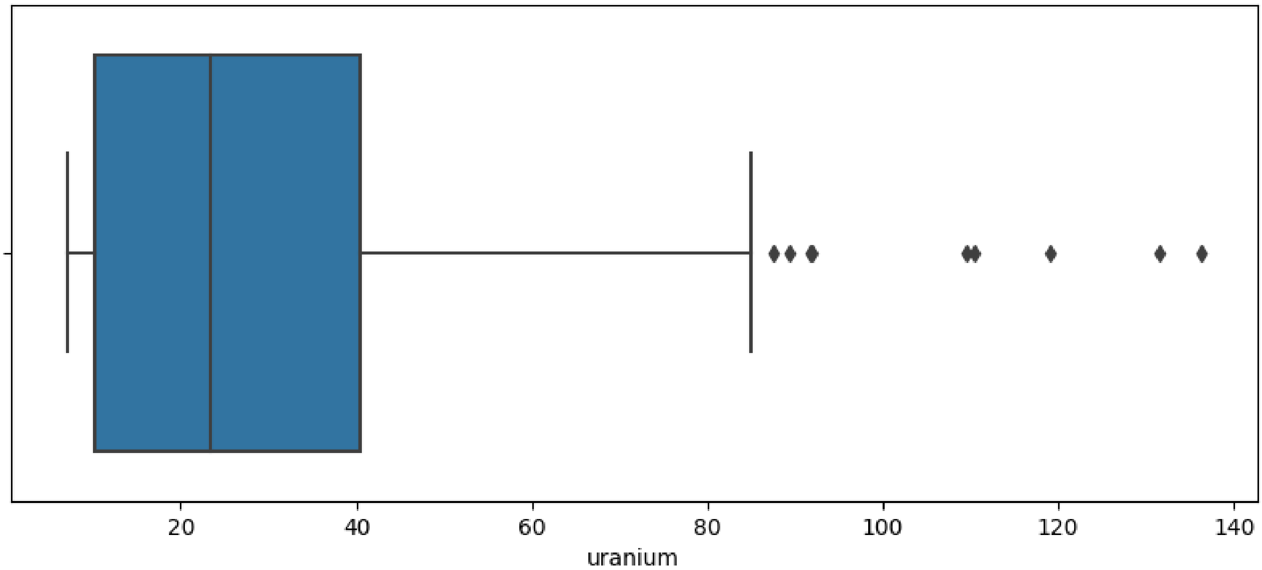

Figure 3: Box plot Uranium Price

3 Modeling

3.1 OLS

```
import statsmodels.api as sm
from sklearn.metrics import mean_squared_error

y = df.uranium
X = df[['coal', 'cpi', 'usd_index', 'sp_500']]

model = sm.OLS(y, X)
results = model.fit()
fitted_values = results.fittedvalues
```

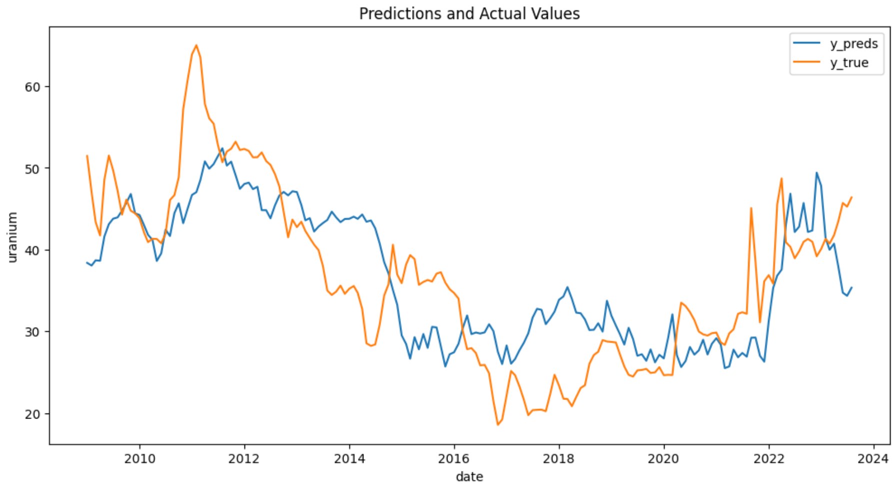

Figure 4: OLS: Predictions vs Actual Values

Table 2: Performance Metrics OLS

Metric	Value
MSE	49.6386
RMSE	7.0455
MAPE	0.17297
R-squared	0.966
Adj. R-squared	0.965
F-statistic	1216.
Prob (F-statistic)	6.40×10^{-125}
Log-Likelihood	-593.35
AIC	1195.
BIC	1207.

Model Specification:

$$Uranium = 0.037445 \times coal + 0.530842 \times cpi - 0.790399 \times usd\_index - 0.011327 \times sp\_500$$

3.2 LinearRegression

```
model_ols_constant = sm.OLS(y, X_with_constant)
results_constant = model_ols_constant.fit()
preds_constant = results_constant.predict(X_with_constant)
```

Table 3: LinearRegression Performance Metrics

Metric	Value
MSE	48.1443
RMSE	6.9386
MAPE	0.16852
R-squared	0.564
Adj. R-squared	0.553
F-statistic	55.22
Prob (F-statistic)	7.92×10^{-30}
Log-Likelihood	-590.66
AIC	1191.
BIC	1207

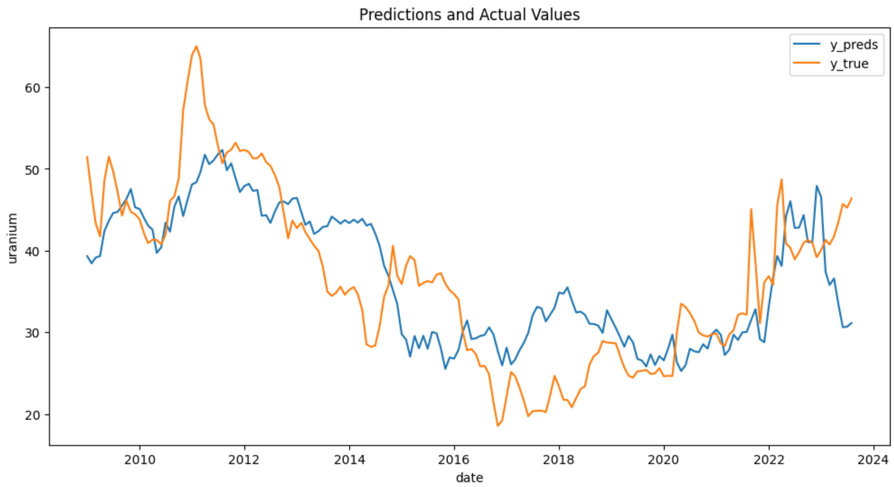

Figure 5: LinReg: Predictions vs Actual Values

Model Specification:

$$Uranium = 41.1517 + 0.0508 \times \text{coal} + 0.3186 \times \text{cpi} - 0.7842 \times \text{usd\_index} - 0.0076 \times \text{sp\_500}$$

3.3 Diebold-Mariano test

Based on the analysis of mean square error (MSE) and mean absolute percentage deviation (MAPE), the models showed almost identical results. For a more objective comparison and to determine the statistical significance of differences in their predictive performance, we decided to perform the Diebold-Mariano test. This test allows us to assess whether the differences in forecast accuracy are statistically significant and to emphasize the relative efficiency of each model.

```python
def diebold_mariano_test(actual, forecast1, forecast2, h=1):

    e1 = actual - forecast1
    e2 = actual - forecast2

    mse1 = np.mean(e1**2)
    mse2 = np.mean(e2**2)

    dm_stat = (mse1 - mse2) / np.sqrt((mse1**2 + mse2**2) / (len(actual) - h))

    p_value = 2 * (1 - t.cdf(np.abs(dm_stat), df=len(actual) - h))

    return dm_stat, p_value
```

Diebold-Mariano Test Results:

- Test Statistic: 0.2859
- P-value: 0.7753

Fail to reject the null hypothesis. There is no significant difference in forecast accuracy.

4 Predictions

As part of our study, we decided to conduct a forecast of uranium prices for the next 12 months. For this purpose, we created a new dataset including variables that are the basis for the specification of an ordinary least squares (OLS) model. This approach allows us to use a regression model to analyze and forecast future uranium prices, taking into account the influence of relevant factors.

```
np.random.seed(42)

X_new = pd.DataFrame(
    {
        "coal": np.random.uniform(120, 160, 12),
        'cpi': np.linspace(307, 315, 12),
        'usd_index': np.random.uniform(95, 110, 12),
        'sp_500': np.linspace(4500, 4900, 12)
    }
)
```

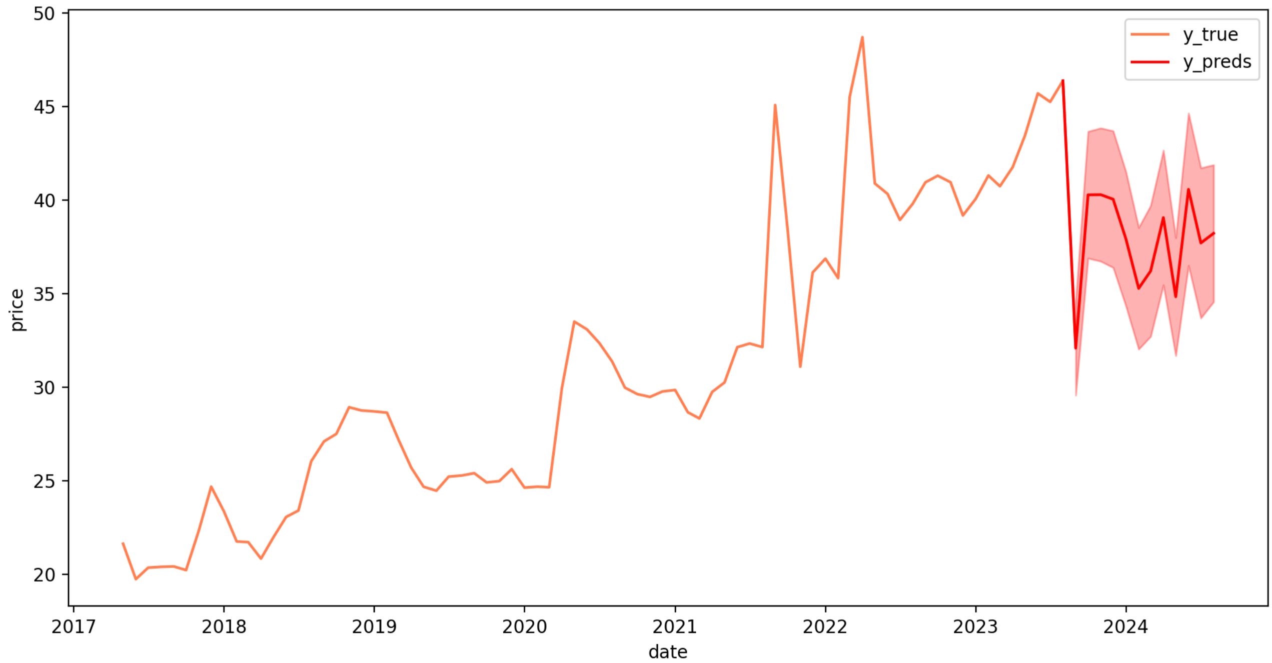

Figure 6: Uranium Prices Prediction for 12 month

Reference List

1. Australia's Department of Industry, Science & Resources. (2022, October 4). Potential supply shortfalls to support prices. *https://www.industry.gov.au/sites/default/files/minisite/static /8f915ab4-c06d-41cf-98ee-5bc628230bdc/resources-and-energy-quarterly-september-2022/ documents/ Resources-and-Energy-Quarterly-September-2022-Uranium.pdf*

2. Baak, M., Koopman, R., Snoek, H., & Klous, S. (2020). A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. *Computational Statistics & Data Analysis, 152*, 107043.

3. Federal Reserve Bank of St. Louis. (n.d.) FRED Economic Data. *https://fred.stlouisfed.org/*

4. Fitri, W. (Capital.com). Uranium price forecast: Can it continue to recoup loses? *https://capital.com/uranium-price-forecast*

5. NEA/IAEA. (2014). Uranium 2014: Resources, Production and Demand, A Joint Report by the OECD Nuclear Energy Agency and the International Atomic Energy Agency (NEA No. 7209). *Nuclear Energy Agency, Organisation for Economic Co-operation and Development. https://www.oecd-nea.org/upload/docs/application/pdf/2019-12/7209-uranium-2014.pdf*

6. Shaw, A. (2023, November 22). Uranium price hits 15-year high. *Mining Technology. https://www.mining-technology.com/news/uranium-price-hits-15-year-high/?cf-view*

7. World Nuclear Association. (2021). The Nuclear Fuel Report: Global Scenarios for Demand and Supply Availability 2021–2040. *https://www.world-nuclear-news.org/Articles/Nuclear-fuel-report-sees-positive-long-term-future*