Econometrics Final Project

Jared Andreatta

April 2024

Introduction

The purpose of this analysis is to examine the effect of nonrenewable energy resource prices on end-use consumer energy costs. In order to measure these effects, the "Special Index for Energy", as denoted by the Bureau of Labor Statistics (BLS). The Special Index for Energy is composed by two different indices that measure consumer energy costs by indexing a "basket of goods" specific to energy consumption. These indices are the Household Energy Index, which measures prices of energy used for residential purposes (i.e. heating/cooling, cooking, electricity, etc.), and the Motor Fuels Index, which indexes the prices of energy resources used for fossil fuel-based resources used for motor vehicles.

The motivation for this analysis lies in the importance of using the energy-specific components to measure their effect on the overarching Consumer Price Index (CPI), which is an important macroeconomic indicator measuring differences in prices of baskets of goods for consumers. As an energy-specific index, it is an important market indicator to participants in the nonrenewable energy sector to forecast price movements. Additionally, it can be used by consumers for the purpose of budget planning by measuring the magnitude effect of nonrenewable price movements on their own end-use consumption.

Data

The data used in this analysis was collected from a variety of databases, namely the Federal Reserve Economic Database (FRED), Yahoo Finance, the Energy Information Administration (EIA), the BLS CPI Database, and the BLS PPI database.

Target (nominal monthly index Feb 2024 = 100)	**Regressors** (nominal monthly index Feb 2024 = 100)	**Econ Controls** (nominal monthly % change)	**Energy Control** (trillions BTU)
energy_index	ppi_coal ppi_crude ppi_nat_gas ppi_gasoline ppi_diesel ppi_heat_oil	effr_change cpi_change dxy	ren_cons_res

Table 1: Data Descriptions

Table 1 describes the data that was used for this regression analysis. The target variable for this model, denoted as energy_index, is the Special Energy Index. This data is a monthly nominal time-series index from 1994-2024, and is indexed for February 2024=100. Similarly, the regressor variables are Producer Price Indices (PPIs) for either the extraction or refining of an energy resource. They share common units and date-times as the target variable. For control variables, two types of controls were added to enhance the understanding of causal effects between the regressor variables. The first control captures macroeconomic changes by measuring nominal monthly % change of different economic indicators that may influence the target variable, namely effective federal fund rate change (effr_change), inflation rate change (cpi_change), and US Dollar Index change (dxy). The second control variable measures the residential renewable energy consumption (ren_cons_res), denoted in trillions BTU, to capture the effects of residential renewable energy consumption on end-use consumer energy costs.

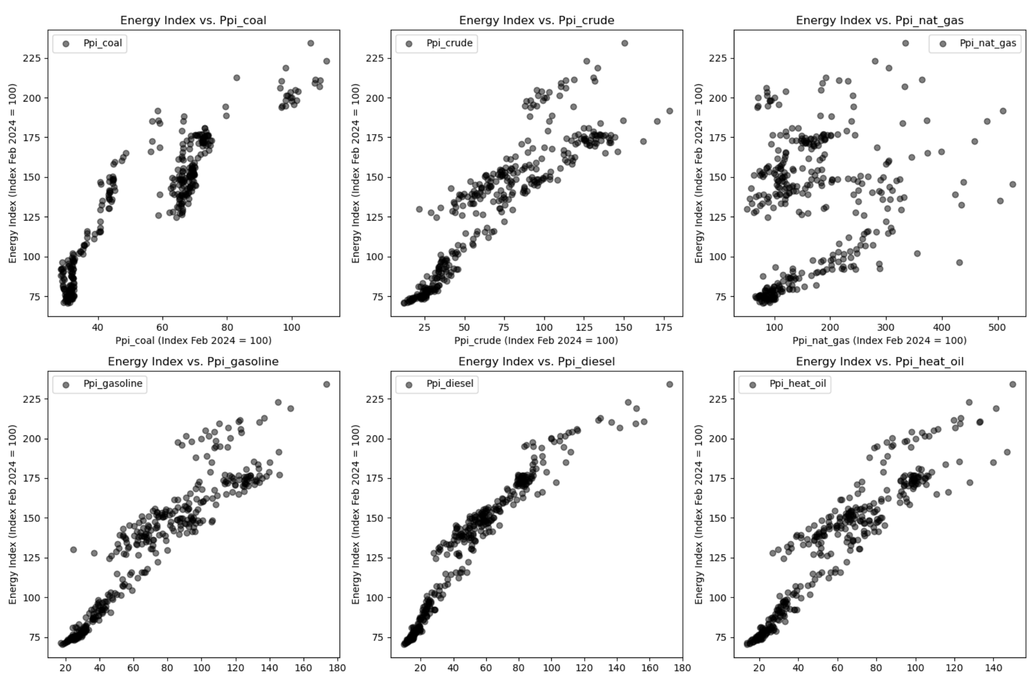

Figure 1: Special Energy Index vs Producer Price Indices

Methodology

While the original regression was purposed for an ordinary least squares (OLS) analysis. However, after some testing, the variables were severely multicollinear. Multicollinearity dramatically reduces the viability of regression results due to unreliable probabilitie values and confidence intervals of regression variables (Kim 2019) .Figure 2 depicts the correlation matrix between all variables included in the regression. Areas shaded in red are highly correlated variables. Note that in the upper left quadrant of the heat map, where the regressors (less natural gas) and target variables intersect, have a deep red shading, indicating a high amount of correlation between variables. Table 2 depicts a variance inflation factor (VIF) analysis of the included variables. With the consideration of a VIF of over 10 is accepted as multicollinear, there are clearly extreme amounts of multicollinearity within the data.

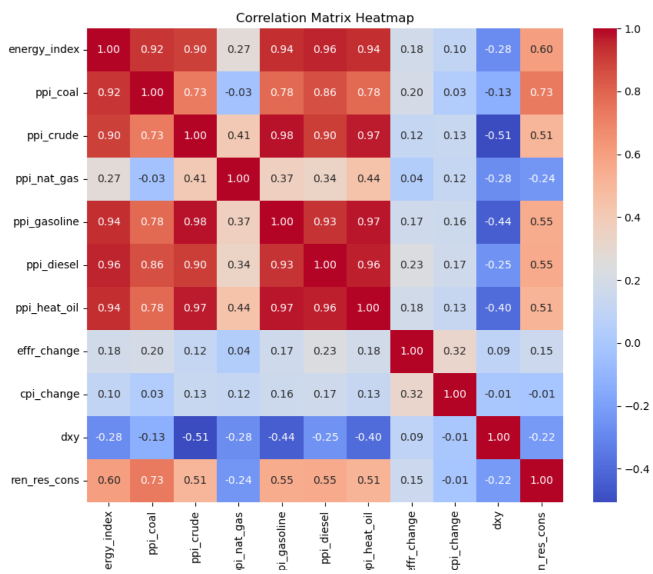

Figure 2: Correlation Matrix

Feature	**VIF**
energy_index	501.96
ppi_coal	168.87
ppi_crude	168.94
ppi_nat_gas	13.72
ppi_gasoline	213.85
ppi_diesel	91.23
ppi_heat_oil	283.96
effr_change	1.28
cpi_change	1.72
dxy	57.86
ren_cons_res	66.71

Table 2: VIF Table

To combat the problems of multicollinearity in the regression, principal component analysis (PCA) was used to reduce the degree of multicollinearity between variables. PCA is a statistical method used for dimensionality reduction by minimizing sum of squared distances (SSD) while maximizing the observed variance of the data (Jaadi 2023). The method works by decomposing the original data using eigendecomposition, which derives the eigenvalues and corresponding eigenvectors that are used to formulate principal components (Pankavich 2020). Principal components are essentially linear combinations of the original variables based on the weight they have on each principal components. Principal components are ordered by the amount of variance they explain; the weights (a.k.a principal component loadings) of the linear combinations describe how much influence the variable has on each principal component. This method corrects for multicollinearity because the decomposed eigenvectors are orthogonal to each other, meaning that they are statistically independent from one another.

$$\hat{Y} = \beta_0 + \sum_{i=1}^{n} (\beta_i PC_i) + \varepsilon \tag{1}$$

$\hat{Y}$ **-** Predicted value of Special Energy Index

β_0 **-** Intercept of the model

β_i **-** Coefficient of regressor variable i

PC_i **-** Principal component i

n **-** Number of principal components

ε **-** Random error term

(1) represents a modified version of an OLS regression which includes principal components as regressors, as opposed to the original variables. This method is also known as a principal component regression, where the aggregated effects of variables are measured on each principal component in the regression. This is advantageous in that the principal component loadings allows easier understanding of the extent of each independent variable's effect on the regression, however, it is disadvantageous in that the affect of each independent variable cannot be specifically described as it could be in a traditional OLS regression.

For the regression, the following hypothesis tests will be considered for the regression analysis:

$\mathbf{H}_0 =$ Changes in respective regressor have no significant impact on the energy index

$\mathbf{H}_1$ Changes in respective regressor have significant impact on the energy index

Results

PC Relations:

	PC1	PC2	PC3	PC4	PC5
ppi_coal	0.37	−0.35	−0.02	0.15	−0.09
ppi_crude	0.42	0.12	−0.07	−0.02	−0.03
ppi_nat_gas	0.15	0.68	0.07	0.32	0.11
ppi_gasoline	0.43	0.06	−0.03	0.02	−0.05
ppi_diesel	0.42	−0.03	0.06	0.21	−0.09
ppi_heat_oil	0.43	0.09	−0.02	0.13	−0.03
effr_change	0.10	−0.13	0.66	−0.02	0.72
cpi_change	0.07	0.14	0.66	−0.43	−0.59
dxy	−0.19	−0.33	0.32	0.74	−0.30
ren_res_cons	0.28	−0.49	−0.12	−0.27	0.04

Figure 3: Principal Component Loadings

Figure 3 depicts the PC loadings of the regression, that is, the weight of effect that each variable has on each principal component. The PPIs for heating oil and gasoline have the highest weights on PC1, implying that refined petroleum products have the greatest influence over this regressor. Raw crude comes in at a close second, with a weighting of.42. Thirdly, coal comes in with a weight of .37. Additionally, renewable energy consumption has a significant weighting of .28, indicating that as prices of oil, refined petroleum products, and coal rise, then renewable energy consumption tends to increase as well.

The effects of PC2 are mostly captured by the PPI for natural gas regressor. This interesting implication tells that natural gas has a less significant impact on the energy index than the oil counterparts. Additionally, PC2 has weightings for ppi_coal, dxy, and ren_res_cons of -.35, -.33, and -.49, respectively. This indicates that as prices for natural gas rise, coal prices fall, the USD index falls, and renewable energy consumption falls.

The effects of PC3 are primarily captured by changes in macroeconomic conditions (i.e. effective federal fund rate change and CPI change with an equal weighting of .66). Additionally, the dxy variable has a weighting of .32, suggesting that it correlates with the movement of effr_change and cpi_change. PC4 and PC5 can be neglected for this analysis, as they have little to no significance in terms of the amount of variance they observe from the original data.

```
Regressor: total
Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:           energy_index   R-squared:                       0.959
Model:                            OLS   Adj. R-squared:                  0.958
Method:                 Least Squares   F-statistic:                     1658.
Date:                Fri, 12 Apr 2024   Prob (F-statistic):         4.46e-244
Time:                        16:33:34   Log-Likelihood:                -1278.3
No. Observations:                 362   AIC:                             2569.
Df Residuals:                     356   BIC:                             2592.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        131.4360      0.438    299.952      0.000     130.574     132.298
x1            17.0022      0.191     89.115      0.000      16.627      17.377
x2            -3.5175      0.358     -9.819      0.000      -4.222      -2.813
x3            -0.6474      0.380     -1.702      0.090      -1.395       0.101
x4             6.9758      0.492     14.172      0.000       6.008       7.944
x5            -3.7388      0.548     -6.828      0.000      -4.816      -2.662
==============================================================================
Omnibus:                        6.880   Durbin-Watson:                   0.259
Prob(Omnibus):                  0.032   Jarque-Bera (JB):                6.896
Skew:                           0.338   Prob(JB):                       0.0318
Kurtosis:                       3.020   Cond. No.                         2.87
==============================================================================
```

Figure 4: Regression Results

Figure 4 depicts the results of running the principal component regression with the regressors (x1...x5) representing their respective principal components. The regression has an R^2 value of 0.959, indicating a high amount of correlation between predicted and actual outputs. Additionally, the P(F-statistic) value is extremely close to 0, indicating that the analysis is statistically significant with nearly 100% confidence.

The regression constant is 131.4360, which entails that when all regressor variables are equal to 0, then the special energy index will be the aforementioned value on average. For the relevant regressors, namely PC1-PC3, the original regressor variables cannot be observed independently, so the magnitude of importance will be observed by the loadings (weights) of each principal component. PC1 has a coefficient of 17.0022, entailing that per a 1 unit increase of the principal component aggregation, a 17.0022 increase of the special energy index on average. These coefficient and respective loadings of PC1 signify that refined petroleum product prices contribute most to the increase of the energy index. Raw crude comes in a close second, with coal prices coming in third, with a slight upward pressure attributed to renewable energy consumption. PC1 has a p-value of 0, indicating that the the null hypothesis can be rejected at with $\geq 99\%$ confidence.

In consideration of PC2's coefficient and loadings, the regression model expects a 3.5175 unit decrease of the energy index, on average, per one unit increase of PC2. PC2 has a p-value of 0, which means the the null hypothesis can be rejected at with $\geq 99\%$ confidence. This is an interesting, and mostly perplexing result. Intuitively, one would expect a natural gas price increase to increase end-use consumer costs, but the results indicate otherwise with definite statistical significance. A caveat in this is that PC2 has less statistical significance in relation to the variance it captures in the data, which indicates that natural gas prices have a lesser statistical impact on the special energy index.

PC3's coefficient implies that a one unit increase of PC3's aggregated variable combination will cause a decrease of .6474 the energy index. An interesting statistic to consider for this component is the p-value, which is .09. This means that the null hypothesis cannot be rejected with 95% confidence, and almost cannot be rejected with 90% confidence, in the context of how it affects the energy index. With the loadings of PC3 in mind, this weaker p-value indicates that macroeconomic changes (Effective Federal Fund Rate and CPI in this case) have a weaker statistical impact on the energy index. Additionally, PC3 captures less variance than PC1 and PC2, which indicates that this component (and thus the constituent macroeconomic indicators) are less significant to the impact on the index.

Conclusion

According to the analysis, it can be inferred that movements in refined petroleum product (gasoline, heating oil, diesel) prices have the greatest impact on end-use consumer costs. Crude oil is a close second, and coal follows as a third. The interesting result to note is the effects of natural gas price movements on the regression. When natural gas prices rise, the energy index seems to fall, indicating that consumer energy prices are negatively correlated with natural gas prices. It is puzzling, as natural gas in the United States is a prominent resource used by utility companies around America to produce thermal energy. It is another possible question to ask why this is. Macroeconomic indicators seem to have significantly less of a statistical impact on the regression as well. The effects of PC3, which is the principal component who's weightings are most explained by the effects of macroeconomic indicators, fail to reject the null hypothesis at a 95% confidence level. This indicates that changes in macroeconomic conditions have little effect on end-use consumer energy costs compared to other regressors. This is perhaps due to the lagged nature of monetary policy changes; inflationary changes tend to lag behind monetary policy changes, which follows that prices will have a lagged response to monetary policy changes as well. Additional observations to note are involved with renewable energy consumption. Renewable energy consumption has a slight positive correlation with refined petroleum, oil, and coal price movements, meaning that as the prices of these fuels increase, so will renewable energy consumption, which is an intuitive answer. On the other hand, renewable energy consumption has a strong inverse relationship, meaning that as natural gas prices rise, renewable energy will decrease, which is another interesting result. Further work on this topic could examine the dynamics of natural gas markets and end-use consumer energy costs. Additionally, other relevant factors could be included in this analysis to build relevance to the model, such as oil and gas production, average monthly electricity demand, meteorological conditions, and spatial analysis dependent on each state.

References

[1] Federal Reserve Bank of St. Louis. (n.d.). Welcome to fred, Federal Reserve Economic Data. your trusted source for economic data since 1991. FRED. https://fred.stlouisfed.org/

[2] Ice US dollar index - index - c (DX-Y.NYB) historical data - yahoo finance. (n.d.). https://finance.yahoo.com/qu
Y.NYB/history/

[3]Renewable Alternative Fuels - U.S. Energy Information Administration (EIA). (n.d.). https://www.eia.gov/rene

[4] A step-by-step explanation of principal component analysis (PCA). Built In. (n.d.). https://builtin.com/data-science/step-step-explanation-principal-component-analysis

[5] U.S. Bureau of Labor Statistics. (n.d.-a). Measures. U.S. Bureau of Labor Statistics. https://beta.bls.gov/dataQ

[6] U.S. Bureau of Labor Statistics. (n.d.-b). Measures. U.S. Bureau of Labor Statistics. https://beta.bls.gov/dataQ

[7] U.S. Bureau of Labor Statistics. (n.d.-c). Measuring price change in the CPI: Household energy. U.S. Bureau of Labor Statistics. https://www.bls.gov/cpi/factsheets/household-energy.htm

[8] Pankavich, S. (2020). Linear vector spaces & applications (MATH 500 Notes). Colorado School of Mines.