The Economic Impact on HSR Merger Rates: Investigating Economic Indicators' Influence on Hart-Scott-Rodino Merger Filing Trends and future corporate consolidation behavior

Jeremiah A. Bloom

Slippery Rock University

Econometrics

Dr. Xintong Wang

April 24, 2024

Abstract

The Hart-Scott-Rodino (i.e. HSR) regulation is a key component of the United States antitrust efforts. This regulation provides additional safeguards in governing mergers and acquisitions to reduce monopolistic market behavior. Using the number of pre-merger notification transactions data, the implicit impact of market conditions on monopolistic market behavior can be used to predict future rates/amounts of HSR merger notifications preemptively.. The key macroeconomic indicators (reported monthly) used to analyze the encompassing trends in HSR transactions are producer price index (i.e. PPI), federal funds rate, m1 currency in billions, unemployment rates, population in thousands, consumer price index (i.e CPI), and a recession indicator. These macroeconomic predictor variables were used in two categories of data modeling. The first model is an ordinary least square (i.e. OLS) linear regression model. This model is used for feature engineering to better understand how the key indicators mentioned above impact the rates of HSR merger transactions. In addition to this regression methodology, a more powerful predictive model was aggregated to provide a forecast of the general trends for the future. This second model was produced using Azure Machine Learning Studio to iteratively find the best predictive model based on normalized mean squared error. The best model chosen for predictive power was an extra-tree regression model which works like a decision-tree model but with higher randomness. Both these models are implemented and analyzed using *SciKit-Learn's* (i.e. sklearn) application programming interface (i.e. API).

1.Introduction

Corporate monopolization is a phenomenon of markets and industries becoming increasingly concentrated through mergers and acquisitions. The impact of consolidated markets on consumers has large implications for consumer prices, product diversification, and relative market competitive health. Free market systems require all stakeholders to have an equal footing regarding the potential to be competitive. The Hart-Scott-Rodino Antitrust Improvements Act (i.e. HSR regulations) was enacted to govern mergers that fall under a specified threshold for transaction sizes and competitive reduction likelihoods. The three standard qualifications that require companies to notify the Federal Trade Commission (i.e. FTC) are transactions that significantly lessen competition, exceed the size-of-transaction threshold, or exceed the size-of-person threshold. The first requirement is more open ended regarding how the specific transaction may reduce the ability for related companies to compete in the given market, however, the other two requirements are cost parameters based on the Consumer Price Index (i.e. CPI). The size-of-transaction threshold, as of 2024, is any transaction that exceeds $119.5 million dollars. The size-of-person test is a situation in which two joining companies have a total asset/net-sales minimum of $239 million for the larger entity and $23.9 million for the smaller entity (Sidley, 2024).

1.1 Literature Review

Current literature on this topic are analytical studies on how specific indicators impact the merger rates in the United States. They delve into how macroeconomic indicators impact the number of mergers in a given business cycle or year. Some of the findings of this research find that macroeconomic indicators have a varied impact on merger rates in the country of interest. An example found by João Nuno Ferreira Antunes is a paper investigating the macroeconomic determinants of mergers and acquisitions in the United Kingdom. In this research, Antunes finds that gross domestic product has a significant short-term impact on merger and acquisition behavior. In addition to this, in the long run, he found that there is a cyclical relationship between merger and acquisition behavior and time. Antunes's conclusion is a bidirectional relationship between mergers and macroeconomic indicators. This means mergers influence performance of the indicators and vice versa (Antunes, 2017). One other notable paper was written by Ziran Ding and Benjamin Hemingway. In this, they examined the nature of merger rates using US firm-level data. This research found two key behaviors outlining acquirer behavior and the target company behavior. For the company acquiring the smaller company, it was found that the company buying another is typically larger than the company its purchasing. Moreover, their financial health is much better than that of the company they are interested in acquiring. For the target company, they have been found to have more innovative power, but are not able to keep up financially. This means that larger firms typically target struggling innovative firms, notably more so during recessions (Ding et al.).

These research papers show how the different macroeconomic indicators impact the overarching trends in mergers and acquisitions. While there are a lot of different papers looking into the impact of indicators on mergers, there is no modeling done to create a predictive machine-learning model to use as a guide to predict the rates and acquisitions. This research presents a novel approach to adding predictive capabilities to mergers. Using unseen 2019-2024 data, the model will be able to be seen its predictive capabilities. Then, using this model, these types of models can be used further to add more predictions into future mergers in the United States.

1.2 Motivation for Research

The data for these transactions presents a new opportunity to use machine learning modeling to better understand the trends in corporate consolidation efforts based on different macroeconomic indicators. The purpose for this research is to develop three different analytical predictive models to demonstrate the probabilistic changes and trends in large corporate merger behavior extended through 2024. The corporate behavior being modeled in this use case will be all mergers that fall under the HSR regulation mentioned previously. These robust predictive models being used for merger rate estimation are an ordinary least square (i.e. OLS) linear regression model and a model optimized using Azure Automated Machine Learning Studio that iteratively optimizes a regression model on normalized mean squared error to ultimately produce an extreme trees regression model (i.e. ETR). Both models will present a novel approach to modeling the behavior of U.S. company's efforts to grow more concentrated in the market based on economic indicators. By evaluating these three models, the best method can be selected based on the performance metrics of mean squared error (i.e. mse), mean absolute error (i.e. mae), R-squared (i.e. R2), and cross validation. In addition to these assessments, model visualization was performed to better understand the distribution and predictive power of the models. Using the best predictive model, future market merger trends can be outlined based on forecasted macroeconomic indicators.

2. Methodology

2.1 Data Aggregation and structure

The dataset in this paper was the combination of HSR merger filings by month (fiscal year 1990-2019) data from the FTC and financial indicators provided by the Federal Reserve Bank of St. Louis (i.e FRED). The FRED data that was retrieved for the equivalent years is producer price index (i.e. PPI), federal funds rate, m1 currency in billions, unemployment rates, population in thousands, and consumer price index (i.e CPI). This dataset also included the month and year of every reported transaction but was removed during training to avoid time as a predictor for the target variable. Summary statistics below.

The forecasted values for predictor variables leading to 2024 were extended from current values. The extension was done by inserting up-to-date FRED data to predict future trends of HSR merger rates. While this dataset did not include HSR transaction numbers leading up to 2024, they are available through annual reports from the FTC. This extended data will be compared to the modeled predicted data to evaluate the robustness of the model's predictions.

Metrics	PPI	fed_fund_rate	m1_in_billions	unemployment_rate	population_in_thousands	10_yr_mrkt_security_yield	CPI	recession	predictions
count	348	348	348	348	348	348	348	348	348
mean	155.3448	2.985316	1647.837644	5.929598	291466.0805	4.632902	190.3089	0.097701	239.336551
std	32.74569	2.458753	813.856758	1.559636	23679.74033	1.925761	36.6709	0.297338	130.316693
min	112.7	0.07	786.6	3.7	248174	1.5	125.4	0	-64.108789
25%	125.275	0.3675	1100.1	4.7	271414.75	2.87	159.325	0	145.324214
50%	143.95	2.985	1330.75	5.6	292542.5	4.55	187.25	0	217.816728
75%	187.75	5.25	1971.55	6.8	312203.25	6.0425	224.9533	0	355.446587
max	208.3	8.84	3702.2	10	329216	8.89	252.182	1	622.39359

2.2 Ordinary least Squares Regression

Ordinary least squares is a regression approach focused on uncovering the relationships between predictor variables explaining the target variable. The line of "best fit" is found using the sum of squared residuals (residual referring to the difference between actual and predicted value). This test of optimization is useful for analyzing data across time. In this case, it is considered the baseline model to be improved upon for predicted HSR merger amounts in recent and future years. The theoretical formula for this equation for is:

$$
\begin{aligned}
hsr\_merger\_filing\_count \\
&= \beta_0 + \beta_1 \times observation\_date + \beta_2 \times year\_num + \beta_3 \times month\_num + \beta_4 \times PPI \\
&+ \beta_5 \times fed\_fund\_rate + \beta_6 \times m1\_in\_billions + \beta_7 \times unemployment\_rate \\
&+ \beta_8 \times population\_in\_thousands + \beta_9 \times 10\_yr\_mrkt\_security\_yield + \beta_{10} \times CPI \\
&+ \beta_{11} \times recession
\end{aligned}
$$

The coefficients for all variables are calculated during training by the sum of residuals method described prior. While this method is extremely helpful for a baseline, it does not capture some of the randomness of variables and has difficulty with outlier data as demonstrated will be demonstrated in 2.6. The sum of all residuals is calculated by the following formula:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

2.4 Extreme Trees Regression

Pierre Geurts, Damien Ernst, and Louis Wehenkel proposed the extreme trees regression machine learning method in 2006 as an alternative method to Random Forest Classification. This proposed method is a new type of supervised classification/regression model that utilizes a tree-based structure in decision making. The birds eye functionality of the model is to build extremely randomized trees with the structure being "independent of the output values of the learning sample" (Geurts et al. 2006). The splitting algorithm for this machine learning model is outlined below in the table given from the Extremely Randomized Trees research paper.

Table 1 Extra-Trees splitting algorithm (for numerical attributes)

Split_a_node(S)
Input: the local learning subset S corresponding to the node we want to split
Output: a split $[a < a_c]$ or nothing
– If **Stop_split**(S) is TRUE then return nothing.
– Otherwise select K attributes $\{a_1, \ldots, a_K\}$ among all non constant (in S) candidate attributes;
– Draw K splits $\{s_1, \ldots, s_K\}$, where $s_i =$ **Pick_a_random_split**(S, a_i), $\forall i = 1, \ldots, K$;
– Return a split s_* such that $\text{Score}(s_*, S) = \max_{i=1,\ldots,K} \text{Score}(s_i, S)$.

Pick_a_random_split(S,a)
Inputs: a subset S and an attribute a
Output: a split
– Let $a^S_{\max}$ and $a^S_{\min}$ denote the maximal and minimal value of a in S;
– Draw a random cut-point a_c uniformly in $[a^S_{\min}, a^S_{\max}]$;
– Return the split $[a < a_c]$.

Stop_split(S)
Input: a subset S
Output: a boolean
– If $|S| < n_{\min}$, then return TRUE;
– If all attributes are constant in S, then return TRUE;
– If the output is constant in S, then return TRUE;
– Otherwise, return FALSE.

Source: Geurts et al.

This method was implemented by using sklearn's API interface to utilize and construct the model during testing, training, and validation phases. The parameters for this model are as follows; bootstrap set to false, criterion is squared error, max features 0.5, min leaf samples 0.005080937188890647, min samples split 0.0012814223889440828, and n estimators of 50. These tuned parameters were found to be at the optimal level using Azure's Automated Machine Learning feature in their Machine Learning Studio. This method optimized the algorithm and scaler based on normalized mean squared error (nmse). The scaler for this model was also chosen based on the nmse. More information on scalers can be found below.

2.5 Data Preprocessing and Model Development

Both of these models were passed through a grid search cross validator (i.e. GridSearchCV) to get an understanding of the best combination of data scaling and model selection. For model scaling, four distinct types were used through Sklearn: no scaler, min max scaler, standard scaler, and quantile transformer (see explanations for each below). The GridSearchCV handles training the data on five different folds of the data. A “fold” is a section of the data that represents one fifth of the dataset. Every model (OLS and ETR) is trained, scored, and then shown to the next fold. This method reduces the chance of overfitting by hiding at least one part of the data and outputs the performance metrics for the models in their entirety and across each fold. This is the initial test for the difference in predictive powers of these models.

While this is happening, the GridSearchCV applies the appropriate scaler to every model to uncover which preprocessing scaler best suits the model. To explain using one model as an example, the OLS model will be trained four separate times, each time the data will be scaled starting with no scaler and then trained on every scaler previously mentioned until it is finally trained on the quantile transformer. This aids in uncovering the scaler to use for preprocessing data upon training.

When referring to no scaler, this means the raw data in input into the model when training. Nothing regarding the data is altered or normalized on a specific scale. Min max scaler works by converting all data units to a scale of 0-1. 0 being the minimum value and the maximum value being 1. Every value in between takes a relative percent of the minimum and maximum. To demonstrate, a dataset with a maximum of 100 and minimum 0, a given value of 50 would be 0.5 using this scaler. The next scaler passed through the GridSearchCV is a standard scaler. Standard scaler alters the data by removing the mean and scaling unit variance. This method is also known as z-score. Regarding quantile transformer, this transformer works by calculating the different quantiles of the data and maps every datapoint to its respective quantile.

2.6 Performance Benchmarks

Each model was evaluated and compared to one another to find the best predictive power of each. While there are many different metrics to measure, these were chosen as key to understanding which is better for predictions and are outlined for their purpose below. The metrics below were used to evaluate the models in addition to the GridSearchCV mentioned previously.

(i) R-Squared is an indicator of how much variation in target variable (HSR mergers) is explained by the given model. Ultimately the best predictive model will have a larger R-Squared.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

(ii) Mean Squared Error is an accuracy metric to outline the size of error in the given model for predicting. This metric is used to add greater weight to larger errors in the model.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2$$

(iii) Mean Absolute Error is like the mse metric above. The difference between this metric and the one above is how error is treated. The mse above punishes large errors due to its squared nature while this metric treats all errors in the model the same way.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$$

3. Discussion of Results

The purpose of this study is to understand trends in HSR merger transactions to better model future corporate monopolization based on premerger notifications. While the data is available for HSR mergers going into recent years, they were not found in the given dataset and were left out intentionally as another measure of the model's accuracy. The relative scores for newer data will be presented in a section below. Nonetheless, all three phases (training, validation, and testing) presented new benchmarks to look at the model more thoroughly. Therefore, these sub-sections will outline the results from each phase and aggregate all results in the final sub-section. In addition, training and validation are handled together in the same phase and will therefore be described in the same sub-section below.

3.1 Training and validation

Both models were compared with each other during training using the previously mentioned GridSearchCV. In this training, all folds of the data were used to validate the model's general accuracy with the hidden fold while applying a different scaler to each individual training instance. After this was performed in code, the results were compiled to best demonstrate the results of the models performance during training and validation. The relative performance of these models in the testing phase are as follows:

param_model	param_scaler	split0_test_score	split1_test_score	split2_test_score	split3_test_score	split4_test_score	mean_test_score	rank_test_score
LinearRegression()	None	0.596041	0.361851	0.395499	0.534557	0.509164	0.479422	5
LinearRegression()	MinMaxScaler()	0.596041	0.361851	0.395499	0.534557	0.509164	0.479422	7
LinearRegression()	StandardScaler()	0.596041	0.361851	0.395499	0.534557	0.509164	0.479422	6
LinearRegression()	QuantileTransformer()	0.539007	0.20688	0.514353	0.557583	0.414333	0.446431	8

ExtraTreesRegressor(max_features=0.5, min_samp...	None	0.906456	0.781364	0.889053	0.937628	0.932345	0.889369	2
ExtraTreesRegressor(max_features=0.5, min_samp...	MinMaxScaler()	0.904257	0.795854	0.890639	0.926615	0.94293	0.892059	1
ExtraTreesRegressor(max_features=0.5, min_samp...	StandardScaler()	0.902146	0.786452	0.88954	0.92689	0.937337	0.888473	4
ExtraTreesRegressor(max_features=0.5, min_samp...	QuantileTransformer()	0.896716	0.797092	0.881317	0.94219	0.927539	0.888971	3

As demonstrated by the above chart, there are eight different training and validations of the models with four being linear regression attached to one of the three scalers scaler and four other models utilizing the ETR model. Looking at the best of each model type, a simple linear regression with no outside scaler performs the best. Regarding the ETR model, this value changes consistently through different trainings. When looking at R2 during these switches, it was found that R2 ranged from ~0.90-0.92. The random nature of this model changes its prediction ability consistently. Thus, the best model tends to be the ETR with a min max scaler. For reference, all below mentions of these models will be referencing the best performing model outlined in the training validation phase. Moreover, based solely on the training and validation phase, the best performance among all these models is a ETR model combined with a min max scaler.

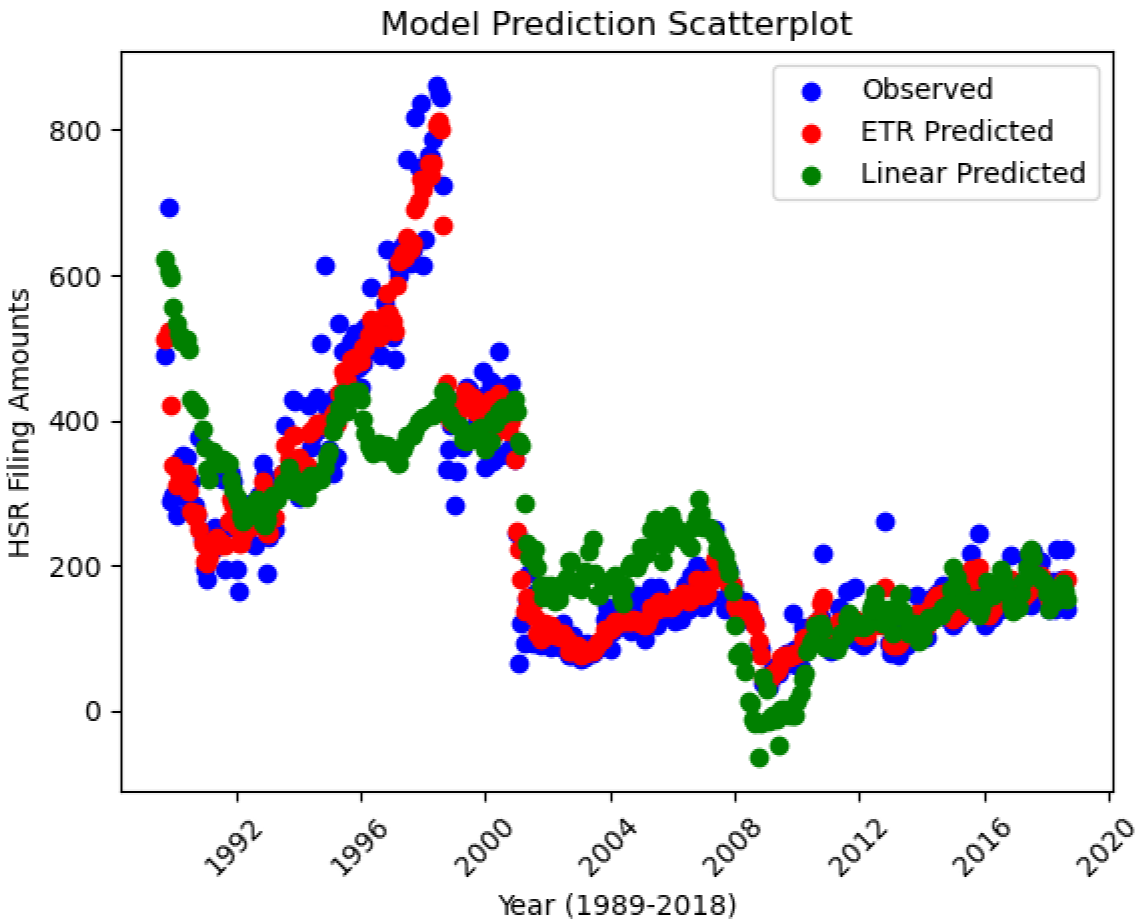
3.2 Testing Phase

The testing phase for both models was performed in the same manner. Prior to training and validation, 25 percent of the dataset was set aside to analyze its performance further on data lying inside the dataset unseen on creation. Starting with the performance of the linear regression model, it had the performance metrics of R2 ~ 0.5425815768824247, MSE ~ 17023.3726297243, and MAE ~ 93.09785003537758. Moving onto the ETR model, the performance metrics are as follows; R2 ~ 0.9270267097977019, MSE ~ 2715.7881019834244, and MAE ~ 34.14930288106455.

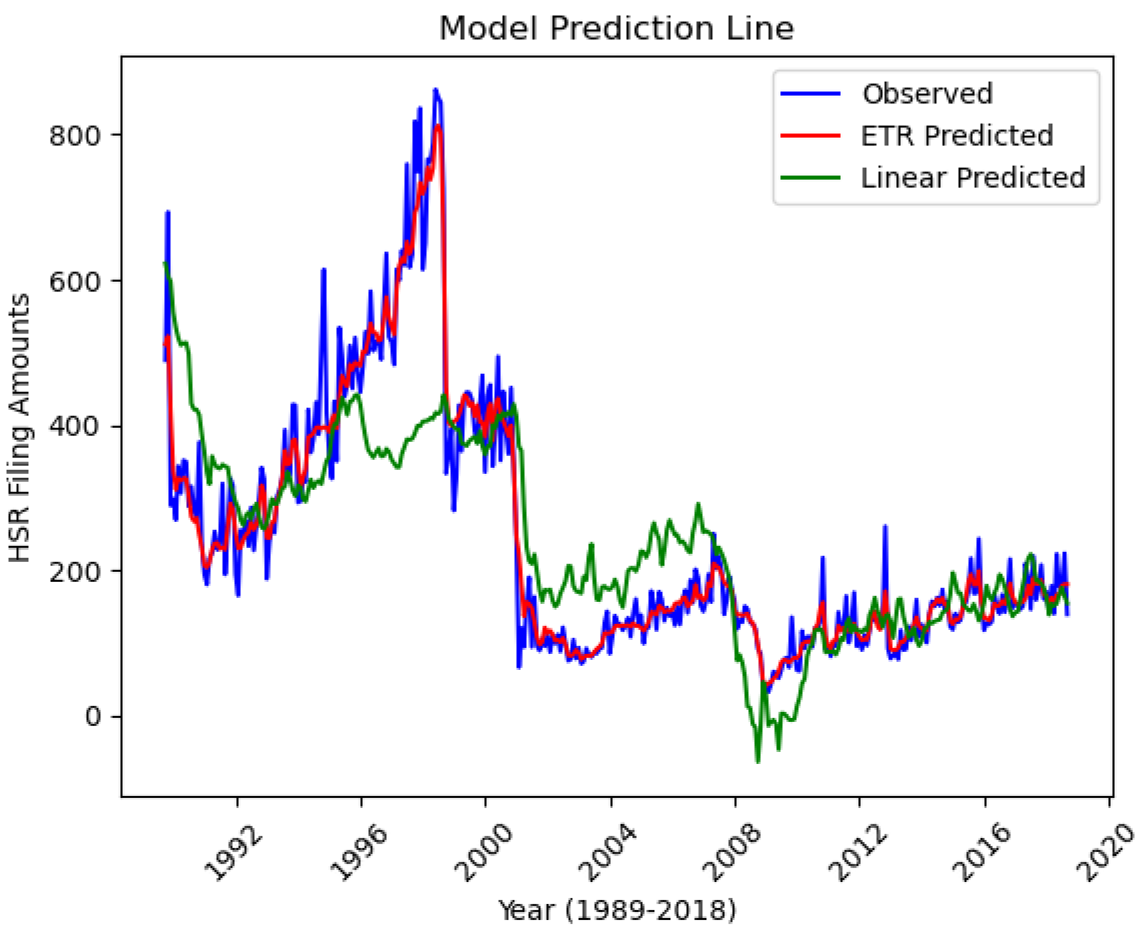
These metrics were retrieved by comparing the actual training HSR rates and the predicted HSR based on training parameters. Looking at these models based on this, the linear regression model is not performing well compared to the ETR model. Based on this model, it only explains ~54 percent of the variation in HSR merger transactions. This is relatively good for linear modeling but is vastly outperformed by the ETR which explains ~92 percent of all variation. One benchmark outlining the key problems with the regression model is MSE. The MSE of the linear regression model is ~14,308 higher than that of the ETR model. This means that the linear regression model is having hard times predicting outlier data when presented high rates of change in HSR mergers. ETR, however, does well at adapting to this rapid change in HSR mergers. Just to reiterate, these results are only based on 25% of the data given. The models were then passed through the entire dataset again.

3.3 Overall Performance Benchmark Evaluation

Now that the unseen data has been evaluated based on the performance benchmarks, both models went back and tried to predict every single datapoint in the dataset. For the ETR model, its benchmarks are as follows; R2 ~ 0.9589371086810032, MSE ~ 1284.389335217037, and MAE ~ 22.495948277053504. The linear regression model is as follows; R2 ~ 0.5667261054935889, MSE ~ 13552.196434704994, and MAE ~ 81.02357799093458. Looking at these, both models did well at improving when looking at the dataset in its entirety. Both their predictions were mapped along a time series to see the relative predictions compared to the actual value. The following are a scatter plot and then a line graph. Both present the same data with different presentations.



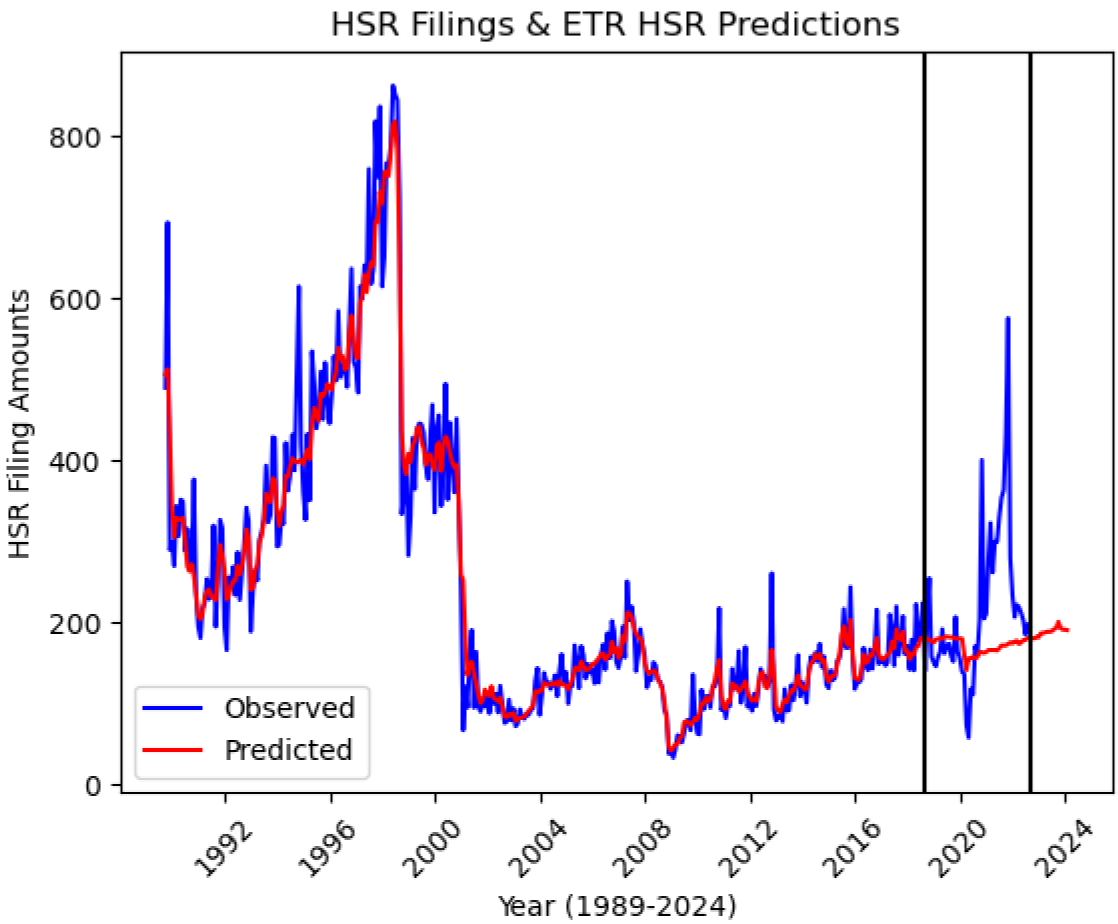
Model Prediction Scatterplot
Observed
ETR Predicted
Linear Predicted
HSR Filing Amounts
Year (1989-2018)
0
200
400
600
800
1992
1996
2000
2004
2008
2012
2016
2020



Based on simply looking at these graphs, it can easily be deduced that the ETR model outperforms traditional OLS regression methods. The ETR model will be used for the following predictive analysis.

4. Predictive Analysis

Now that the best model has been aggregated and found. Let's extend the data out to what is currently available on the FTC website which to September of 2022. This section of the dataset is peculiar in that it is the part of recent history with the introduction of the coronavirus. This was an unprecedented time where it impacts the performance metrics of our model. Nonetheless, the model was asked to extend its predictions out to 2024, with the only available data for HSR mergers reported up to September of 2022. Below is the same visualization as the graphs above but with two key differences. The first difference is the extended prediction and actual values; in addition, there are two horizontal lines. The first line represents the beginning of the model pre dicting values it did not see during training, validation, and testing phases. The second black line is a prediction of 2024 values that are not yet found on the FTC reports.



Now that we have seen an overview of the model predicting HSR merger notifications during the timeframe around covid, how did it perform based on the training metrics outlined above (R2, MSE, MAE)? Starting with R2, the model started to perform worse in this regard with only having a score of 0.9118858878658729 in the whole dataset. To add, MSE and MAE were 2523.6750157510396 and

28.271237120636936, respectively. While they have gone in the 'negative' direction with respect to each variable, the model still performs well given the new data.

Conclusion

The United States, like all other countries, must deal with the problem of corporate monopolization. Proper implementation of policy requires an encompassing view of the problem. There is a lot of research surrounding mergers and acquisitions that has various findings of the implications of different macroeconomic conditions and their impact on companies practicing this behavior. Established literature has foundationally discovered what causes this consolidation but does not seek to predict the future of it. While trying to model the future, it is important to implement robust algorithms alongside traditional techniques to find the best way of predicting. In this research, two models were used to outline two separate methods of predicting HSR mergers. An OLS linear regression does capture a lot of the variation in these rates (~55%), it is not robust enough to be considered a reliable predictor. This is where an ETR model comes in. This model provides a more robust methodology to consider. The ETR model vastly outperforms the OLS model and does well at describing variation. When presented with new data, however, its performance did drop. The reason for the drop in its accuracy could be due to the period it was attempting to predict. The new predictions were around pre and post Covid, and, as a result, offered a strange merger behavior the model did not predict. It is important to note that all machine learning modeling can only handle what is given. In this case, it was given only a few KPI's to attempt the predictive model. For the future, these kinds of methodologies need to be implemented on more found KPI's and other variables to increase the model's ability to handle occurrences like pandemics, climate, and wars. With future research adding onto this, better models can be constructed to effectively analyze the trends in corporate consolidation efforts.

References

Board of Governors of the Federal Reserve System (US). (n.d.). Federal Funds Effective Rate [FEDFUNDS]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/FEDFUNDS

Board of Governors of the Federal Reserve System (US). (n.d.). Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity, Quoted on an Investment Basis [DGS10]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/DGS10

Board of Governors of the Federal Reserve System (US). (n.d.). M1 [M1SL]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/M1SL

Ding, Z., & Hemingway, B. (n.d.). Mergers and acquisitions over the cycle - An empirical investigation.

Federal Reserve Bank of St. Louis. (2024, April 19). NBER based Recession Indicators for the United States from the Period following the Peak through the Trough [USREC]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/USREC

Federal Trade Commission. (2013, November 18). Data sets. Federal Trade Commission. https://www.ftc.gov/policy-notices/open-government/data-sets

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1

Nuno, J., et al. (n.d.). The macroeconomic determinants of M&A activity in the United Kingdom.

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011

Sidley. (2024, January 23). FTC releases 2024 thresholds for Hart-Scott-Rodino Act filings and raises HSR filing fees. Sidley. https://www.sidley.com/en/insights/newsupdates/2024/01/ftc-releases-2024-thresholds-for-hart-scott-rodino-act-filings

U.S. Bureau of Economic Analysis. (n.d.). Population [POPTHM]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/POPTHM

U.S. Bureau of Labor Statistics. (n.d.). Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/CPIAUCSL

U.S. Bureau of Labor Statistics. (n.d.). Producer Price Index by Commodity: All Commodities [PPIACO]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/PPIACO

U.S. Bureau of Labor Statistics. (n.d.). Unemployment Rate [UNRATE]. Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/UNRATE