ECO421 Fall 2021

University of Toronto

Predicting the Economic Recovery from the Covid-19 Pandemic using Previous Recessions

Shashwat Aggarwal

aggar174

s.aggarwal@mail.utoronto.ca

Kevin Xu

xuming11

kev.xu@mail.utoronto.ca

Poplar Wang

poplwang

poplar.wang@mail.utoronto.ca

December 12, 2021

1 Introduction

Predicting a recovery from the economic crisis as a result of Covid-19 is an area of concern for policymakers around the world. When the impact of the virus began to appear in early 2020, economies around the world shut down and went into lockdown to prevent the spread of the disease. Now with the availability of vaccines and widespread knowledge about the virus, economies have started to open up and things are returning to 'normal'. While things have started to open up, economy recovery patterns have been difficult to predict. This is similar to what has been observed in the past about recovery patterns. For example, Azis (2010) had predicted 'real sustainable recovery' two years after the apex of the 2008 recession but many indicators had not returned to the pr-crisis level until 2015.

Policymakers around the world are interested in predicting accurate information about fiscal and monetary decisions to get economies back to track, despite its inherent difficulty. While the magnitude of Covid-19's impact is greater, economic indicators generally exhibit similarities to other recent major recessions. The difference with this crisis however, lies with the accompanying health crisis and unemployment rates through the nature of lockdowns and social distancing measures.

Models trained using historical data can be used to understand the extent to which the Covid-19 recession is comparable to previous recessions. Incorporating pandemic specific data such as cases, deaths, and stringency of lockdown measures may help predict recovery and guide policy makers for the Covid-19 recovery plan compared to some traditional methods. Foroni et al. (2020) attempted to use methods such as Mixed-Frequency Bayesian VARS and Quantile Regressions but faced issues predicting previous recessions.

Keeping all of this in mind, our paper will aim to answer several different important questions. First we will see the similarity of the Covid-19 pandemic to previous recessions. We will analyse difference indicators such as unemployment, interest rates, inflation, and returns MA. Second, by using machine learning models, and algorithms containing economic indicators and Covid-19 statistics, we will measure how accurately a model predict future economic state and subsequent recovery. Lastly, assuming the health crisis abates and the aforementioned models are accurate, what will the economic recovery and recovery time look like.

2 Data and Methodology

2.1 Data

The datasets used for this project were obtained from various sources and are for the United States. This is due to Canadian data being less readily available/reliable. Upon joining into a single table and trimming null values, the monthly data ranged from 1999-04 to 2021-10 for a total of 274 data points.

Dataset	Description	Source
GDP	Monthly GDP index, consistent with real GPD in the NIPA	IHS Markit
Unemployment	Percentage of employees in labor force, but without a job	Bureau of Labor Statistics
Inflation	Estimations for the consumer price index	Eurostat
Market Volatility	CBOE Volatility Index (VIX)	CBOE Global Markets
Market Returns	Percent S&P500 daily returns using adjusted closing price	Yahoo! Finance
Covid-19 Related	Stringency index, total cases, total deaths	Vitenu-Sackey (2020)

2.2 Methodology

2.2.1 Stability of GDP, Redesign of Planned Experiment

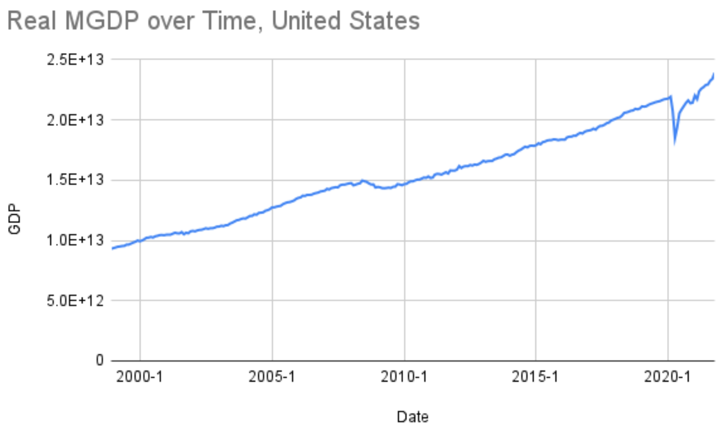
In Part 1 of this project, we tentatively stated the regression variable would likely be GDP to measure the macroeconomic effects of a recession. However, training and testing the model on GDP became an issue.

Since the dataset only has 274 monthly observational points, this value is quite low for our analysis. Using all of the indicators readily available with our dataset (inflation, unemployment, market returns, market volatility, months since) we trained the dataset using various regression models with GDP as the regression variable.[1] After running the model on the test data, we get the following RMSE and R^2 score for each regression:

Model	RMSE	R^2 Score
LinearRegression	244223609853.15073	0.9964202821671999
SGDRegressor	6878356301433.883	-1.839506323163195
ElasticNet	500761890646.55835	0.9849500158109488
BayesianRidge	4090330320404.0796	-0.004130747309760485
KernelRidge	2150605640476.7166	0.7224155947079207
GradientBoostingRegressor	106357258591.62572	0.9993210969307518
SVR	4164238106891.913	-0.04074566568292992

[1] To recreate, simply add desired indicators to the input DataFrame in Cell 3 of the attached notebook file

While such high scores are normally welcome news, this level of accuracy with such a small dataset is cause for concern and suggests a very basic model is sufficient, or that overfitting is occurring. Thus, all indicators are removed aside from date. This still provided an R^2 score of 0.993 using the gradient model[2]. This result can be understood by taking a look at the graph of GDP over months.



As seen here, GDP has a linear growth pattern that is resilient to economic events such as recessions in the long term. The Covid-19 recession sees a drastic dip due to the various lockdown measures, loss of service jobs, and the transition to work-from-home, but recovers soon after. Thus, when factoring in the small number of data points, GDP becomes an impractical choice for the dependent variable in regression, and suggests recovery well before other economic indicators.

2.2.2 Attempting GDP Delta

Unfortunately, attempting to use the monthly changes in GDP did not yield better results. Training various models using indicators as inputs and monthly change in GDP (GDP 'Delta') had inconsistent results[3], with high R^2 scores always being accompanied by unreasonable RMSE values (often greater than the mean monthly change itself). Calling `.corr()` on the DataFrame yielded no significant correlations between any of the variables with the exception of monthly change in unemployment.

Due to these results, the experimental design was ultimately changed to focus on the other economic indicators. These indicators have less volatility in the short term and a non-linear long term pattern, thus being more practical given the small number of data-points.

[2]See Cell 3 of the attached notebook file for a full table of results
[3]The results of one such training run can be seen in Cell 4 of the attached notebook file

2.2.3 Predicting Covid-19 Economic Recovery using the 2008 Financial Crisis

Various regression models were trained on the 2008 recession's data with "Months since Start of Recession" as the independent variable[4], and the various economic indicators as the dependent variables. Sampling was used on the training set after splitting data into training and test sets, in order to prevent intersections between the two while still having sufficiently large training sets for even the more demanding models.

A single independent variable was chosen due to the small number of data points available (not many months in a recession), thus being highly susceptible to over-fitting from the curse of dimensionality.

These models would be run a multitude of times. To run against Covid-19 economic data, the results would need to be normalized to compare[5]. Note as this model was trained and tested on 2008 Financial Crisis data, this Covid-19 economic data would be completely new in all regards.

This methodology was repeated using the 2000 recession data, but was ultimately dropped. As the 2000 tech bubble was relatively isolated to the tech sector with comparably minimal affects on GDP at just -0.3% (Kliesen, 2003), we did not expect it to be a comparable recession and was consequently not used in our project.[6]

2.2.4 Quantifying the Impact of Pandemic-Specific Variables on Economic Recovery

We followed a similar methodology as the previous section 2.2.3, but in the opposite direction. We trained our models in a similar manner to before, but added the various quantifiers of the pandemic's impact to the training data, and then attempted to map it to the recession data of 2008 with all of the indicators of the pandemic's severity being set to 0. These indicators were set to 0 since the 2008 recession was not accompanied by a global health crisis. These new pandemic indicators included new cases smoothed per million, new deaths smoothed per million, and stringency index (the intensity of public health guidelines implemented).

Similar to previous, these models would run against the 2008 recession data. Mapping to a recession in which all pandemic variables were 0 also serves as a sanity check against possible overfitting, due to the increased number of input variables as a result of the new independent variables.

[4] Labelled as "Months Since" in the DataFrames

[5] For implementation details of the normalization, view code or see corresponding section in Results

[6] Testing did support this; its models were completely inapplicable to the Financial Crisis and Covid-19 recessions. To recreate the test, simply change dates in Cell 5 of the attached notebook file

3 Results

3.1 Modelling Recovery of Economic Indicators

We began by training models to predict the recovery of the various economic indicators using data from the 2008 Financial Crisis. To increase training data size to accommodate all the models, after splitting the data into training and testing sets, the training set was sampled with $n = 500$.

The Gradient Boosting Regressor model consistently outperformed the other models when taking the means over multiple runs.

Economic Indicator	**RMSE**	R^2 **Score**
Inflation	0.005764612655885861	0.8914692498145762
Unemployment	0.0022583248431083188	0.8692299718717895
Market Volatility	3.892434346108393	0.8715599004068855
Market Returns	0.4657544517397555	0.9324668886301988

3.2 Predicting Covid-19 Economic Recovery

Before predicting recovery, data from the Covid-19 pandemic was first normalized to allow comparison to the Financial Crisis. This was accomplished via two steps:[7]

- "Shifting" all values by a constant amount, to account for different initial economic states immediately prior to the recessions
- "Scaling" all values by a constant factor, to account for different magnitudes of impact

As these normalization values can be calculated with minimal data from the first few months of a given recession, one does not need data on the recovery itself to determine them. Thus, this method does not impact our models' ability to predict recovery of future recessions. Instead, the models focus on the shape of the recovery.

Comparing the aforementioned models (trained using data from the Financial Crisis) to available economic data pertaining to the Covid-19 pandemic provides the following metrics:

Economic Indicator	**RMSE**	R^2 **Score**
Inflation	0.01172606292652328	0.7221861626472037
Unemployment	0.06513765450866668	-52.702418466249824
Market Volatility	5.926270473110732	0.683112831636121
Market Returns	2.114628481838211	-0.04027288487250935

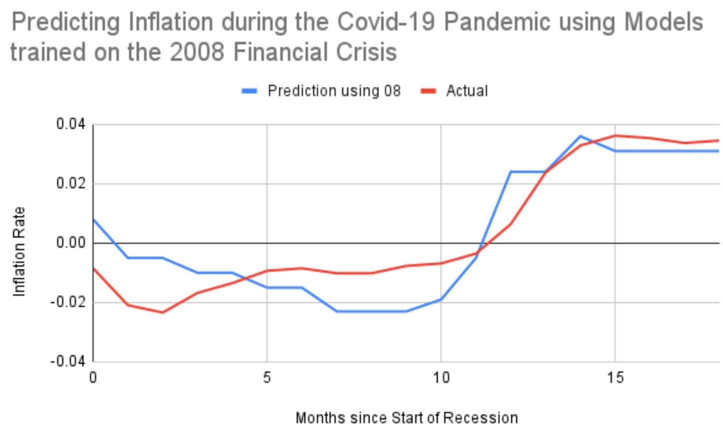
[7] For exact implementation, see `normalMean` and `normalRange` variables in cells 6 and 10 of the attached notebook file

3.2.1 Inflation

Fitting the 2008 model on Covid-19 proves to be worthwhile for inflation estimates, with a low RMSE of 0.011 and high R^2 score of 0.72. This can be interpreted that model is able to minimize the deviation of residuals and roughly 72% percent of the inflation movement during the Covid-19 recession can be explained by the model.

The predictability of inflation is rooted in governmental response to a recession. During a recession, the government will enact expansionary fiscal policy, to increase aggregate demand and stimulate economic growth. As experienced during Covid-19, the US government gave out stimulus packages for families in need. This expansionary policy on the aggregate demand will increase the inflation, but this increase is not experienced immediately. Firms need to adjust their expectations for inflation for the aggregate supply to have an effect. The increase in inflation is thus delayed until 10 months after the start of the recession, and reaches a new rate, hovering just below 0.04.

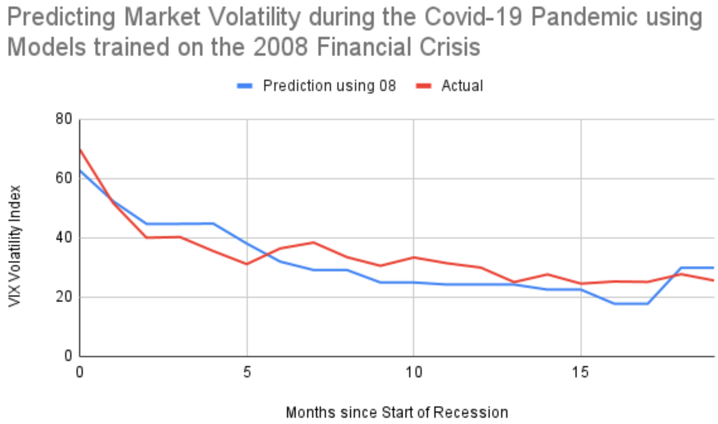
This relationship between model estimates and actual inflation data is represented in the following graph:



The implications of the model's ability to estimate inflation is that the response from the government is routine during recessions, and the model can be applied on future recessions as well.

3.2.2 Market Volatility

As with both the great recession of 2008 and Covid-19 recession, the market was very sensitive to changes and thus experienced many periods of unexpected and unpredictable movements. The model was able to capture the general trend of the market volatility with a R^2 score of 0.68, however, produced a large RMSE of 5.93. The RMSE value can be attributed to the unexpected nature of market volatility, with deviations of the residuals being extremely high. The trend of market volatility is reproduced across recessions, providing an estimate that market volatility will peak near the start of recession and eventually cool down to historic average values.

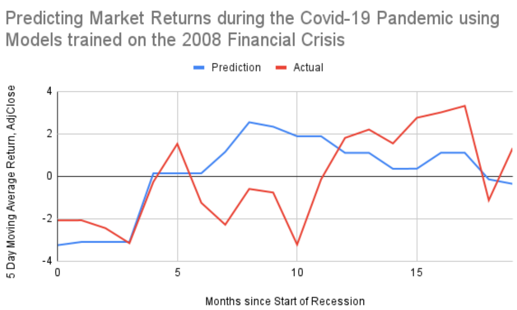


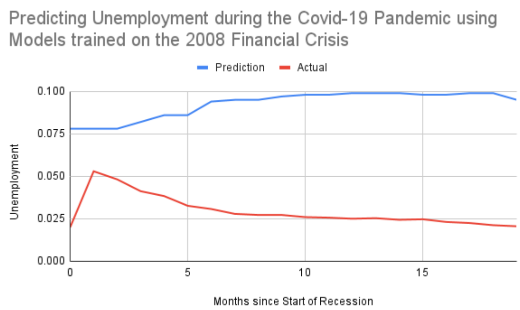
3.2.3 Market Returns and Unemployment

As seen with the extremely low R^2 scores for Market Returns and Unemployment, the model trained with 2008 data fails to predict the trend for the Covid-19 recession.

Unemployment during the financial crisis was often long term, and spread out across various industries. Thus, it had longer lasting impacts. By comparison, a considerable portion of the jobs lost in 2020 were service or physical labor related, and hence was able to recover much quicker. Instead, it was focused in the first few months of the recession, as seen in the graph's spike. For this reason, the model is unable to perform between these two recessions.

In the short run, market returns are inherently unpredictable. This can possibly be attributed to the efficient market hypothesis, where market prices reflect all information presently available to the public. As seen in the following "Predicting Market Returns" graph, the returns would rise and fall multiple times across the recovery of the recession to the news and updates of the pandemic impact on the economy. These peaks and valleys are not captured by the estimates of the model, resulting in a low R^2 score of -0.04.





3.3 Impact of Covid-19-Specific Factors on Economic Recovery

For quantifying the impact of pandemic specific variables on the economic recovery, we trained the model to predict the various economic indicators using the data from the Covid-19 recession. For mapping the pandemic specific to the recession data, the pandemic's indicators were set to 0. The results of the model are provided below:

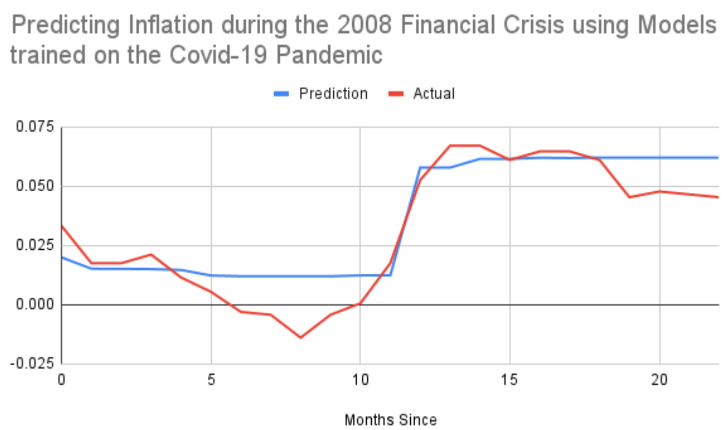
Economic Indicator	RMSE	R^2 Score
Inflation	0.011492602238661402	0.8140525676272509
Unemployment	0.19529939600995974	-42.03702341091725
Market Volatility	6.68696637508655	0.6452365981715584
Market Returns	2.035916254406029	-0.7233999804986149

The important takeaway that we see from an overview of these results is that our model stays viable for the inflation and market volatility data. This means that our model is able to predict these two indicators with a good accuracy after accounting for Covid-19 statistics. Hence, with some more manipulation as well as training the model with more Covid-19 data, we can try and see what these indicators will look like during future recessions and future Covid-19 waves.

3.3.1 Inflation

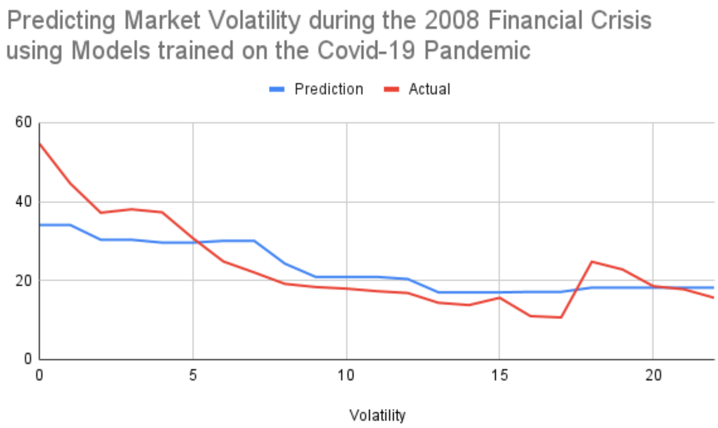
Fitting the Covid-19 model on 2008 recession indicators, we again see a low RMSE of approximately 0.011 and a high R^2 score of 0.81. Hence, the model was able to minimize the deviation of residuals and around 81% of the inflation movement during the 2008 recession could be explained by our Covid-19 trained model.

The relationship between the actual inflation data and our model estimates can be represented by the following graph:



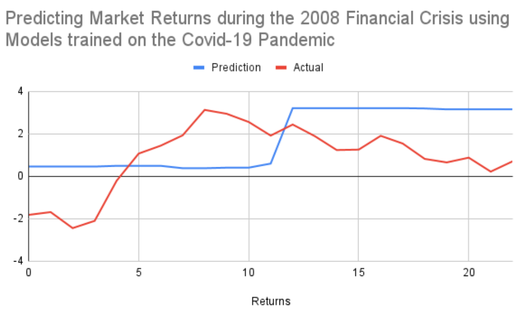
3.3.2 Market Volatility

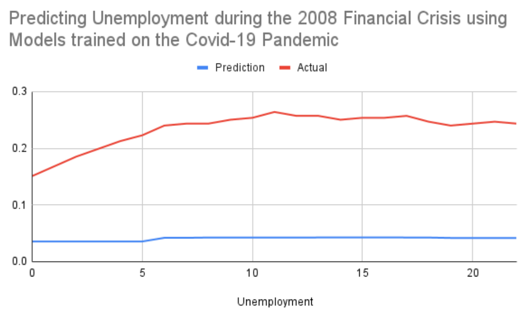
As mentioned and explained in section 3.2.2, the market was very sensitive to change during 2008 recession as well as the current Covid-19 recession. Our model was able to establish the general trend of market volatility with a R^2 score of 0.62 but produced a large RMSE of 6.69. Once again, this high RMSE could be traced back to the volatile nature of this indicator causing the deviations to be quite high.



3.3.3 Market Returns and Unemployment

Akin to section 3.2, we see the unemployment and market returns being poorly predicted for the 2008 recession using our Covid-19 training data. A R^2 score of -42.037 for unemployment and -0.72 for market returns tells us that the prediction quite poor from the training data. The unemployment data being predicted is very low but the actual unemployment observed was much higher. As for the market return, our model is unable to capture the variability of the actual data and this can be seen in the graph when it is unable to predict the peaks and troughs.





4 Conclusion

Our study looked at answering a multitude of questions relating to the Covid-19 pandemic and the recovery of the economic indicators. To do this, we used a panel dataset collected from multiple different data sources. This monthly data ranged from 1999-04 to 2021-10. Multiple different kinds of machine learning regression models were used to understand the recovery of indicators such as GDP, inflation, unemployment, market volatility, and market returns. In the end, it was observed that the Gradient Boosting Regression outperformed other models in multiple trials.

From the results of our models, we quickly learned that attempting to predict GDP from a dataset was difficult due to its resilience in the long term, and we had to exclude GDP from our analysis as a dependent variable. This, however, should not reflect on the nature of GDP as an economic predictor. As has been shown in the past by Asare Vitenu-Sackey & Barfi, (2021) and Foroni et al. (2020), GDP was used for economic recovery prediction but they employed methods such as OLS regression or nowcasting and forecasting which was considerably different from our machine learning models.

When we trained the 2008 recession data to predict the Covid-19 economic recovery, we saw success with inflation predictions, and the ability to capture general trend of market volatility, but poor trend predictions for market return and unemployment. This has good implications for our model. The most important result being the ability to expand our model to predict other recessions, as well as the economic recovery from Covid-19, for certain economic indicators.

Additionally, we saw success when using the same methodology with Covid-19 economic data as the training data. These models are more powerful in regards to our current recession, as they can be used to predict economic recovery while accounting for different possible pandemic recovery scenarios. They can also be used to estimate the exact impact on the economy of different aspects of the pandemic, such as deaths and governmental response (stringency index).

Overall, one can conclude that recessions which impact the entire economy are similar in recovery patterns for certain indicators. Thus, careful analysis of past recessions with tools such as machine learning can prove fruitful when determining fiscal and monetary policy subsequent recessions.

References

Asare Vitenu-Sackey, P., & Barfi, R. (2021). The impact of covid-19 pandemic on the global economy: Emphasis on poverty alleviation and economic growth

Athey, S. (2017). Beyond prediction: Using big data for policy problems. Science, 355(6324), 483–485. https://doi.org/10.1126/science.aal4321

Azis, I. J. (2010). Predicting a recovery date from the economic crisis of 2008. Socio-Economic Planning Sciences, 44(3), 122–129. https://doi.org/10.1016/j.seps.2010.03.003

Foroni, C., Marcellino, M., & Stevanovic, D. (2020). Forecasting the COVID-19 recession and recovery: Lessons from the financial crisis. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2020.12.005

Kliesen, K. L. (2003). The 2001 recession: How was it different and what developments may have caused it? ICPSR Data Holdings. https://doi.org/10.3886/icpsr01292

Mullainathan, S., &, Spiess, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives, 31(2), 87–106. https://doi.org/10.1257/jep.31.2.87