Forecasting Real GDP

Matthew Burrell

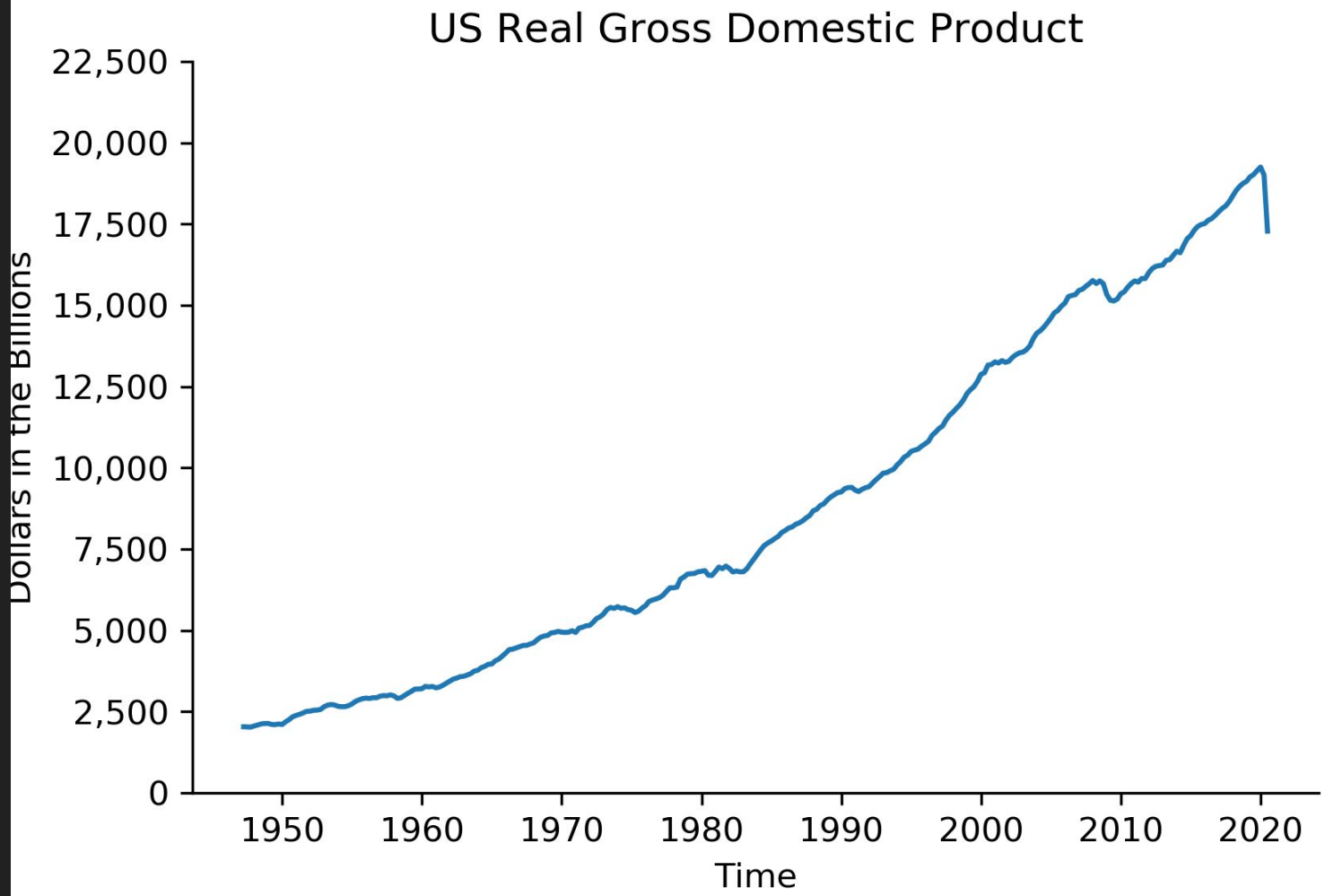
September 9th, 2020

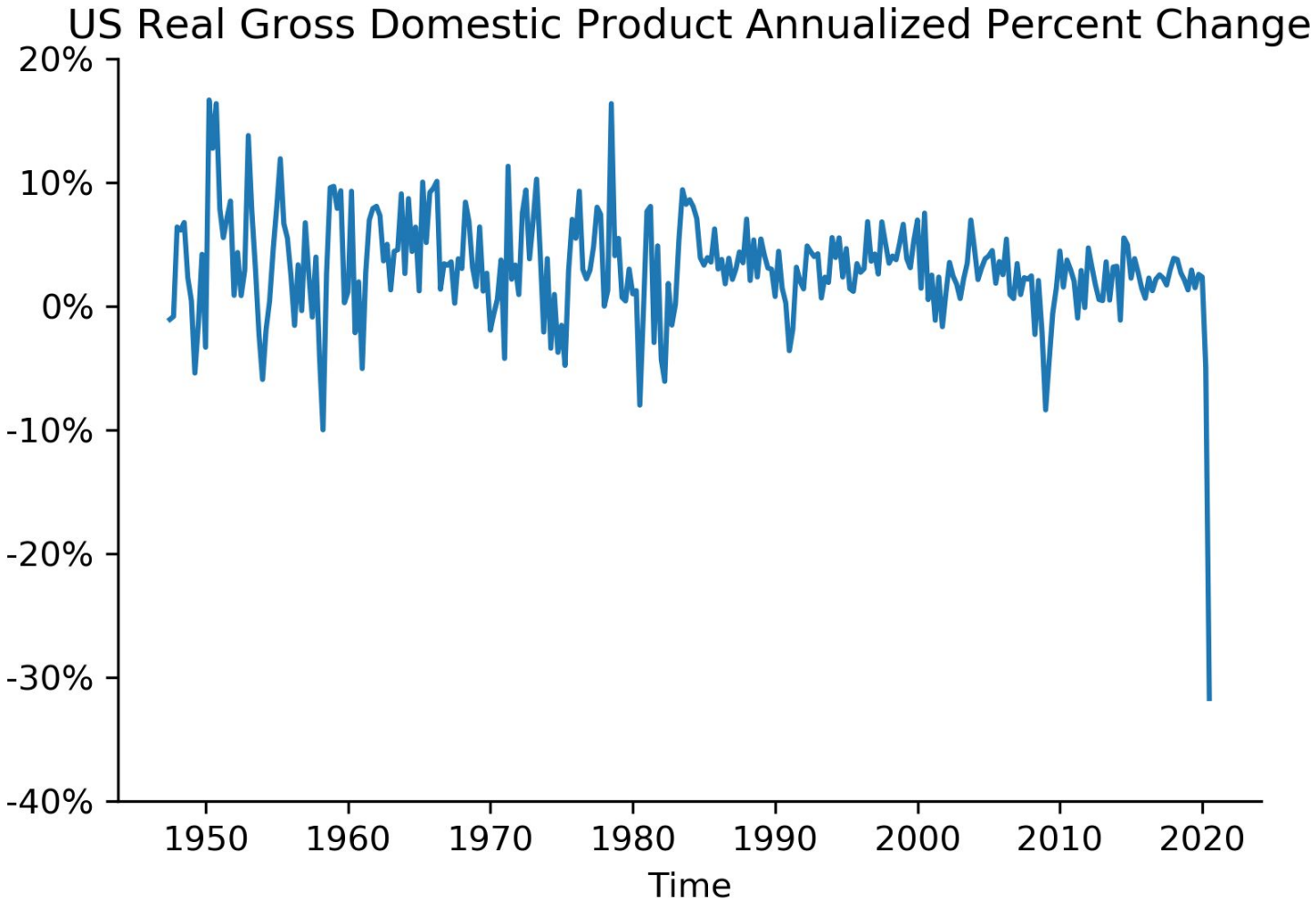
Problem Statement

Due to the 2020 Pandemic, the United States economy has plunged into a deep recession. With the economy ensambles, many business are uncertain about future demand. The goal of this project is to build a model that can forecast real gross domestic product percent change from previeces year. The models trained will be ARIMA, ARIMAX, CNN, LSTM and CNN-LSTM. The model that predicts the best will be the model implemented in production.

Data

- U.S. Bureau of Economic Analysis API - Table 1.1.6. Real Gross Domestic Product, Chained Dollars
  - [Billions of chained (2012) dollars] Seasonally adjusted at annual rate
  - 1947Q4 to present
  - Gross Domestic product, personal consumption expenditures, gross private domestic investment, exports, imports, government consumption expenditures and gross investment
- U.S. Bureau of labor Statistics API-Python library blsconnect
  - 1948M1 to 2020M07
  - U3 official unemployment rate, employment, unemployment, labor force, and labor force participation rate.


US Real Gross Domestic Product
Dollars in the Billions
22,500
20,000
17,500
15,000
12,500
10,000
7,500
5,000
2,500
0
1950
1960
1970
1980
1990
2000
2010
2020
Time


US Real Gross Domestic Product Annualized Percent Change
20%
10%
0%
-10%
-20%
-30%
-40%
1950
1960
1970
1980
1990
2000
2010
2020
Time

Model

Autoregressive Integrated Moving Average

- ARIMA is very flexible for modeling time series data
  - Taking information from long-term trends and sudden shocks
- Can be extended into more advanced models
  - For example: ARIMAX
- ARIMA performs well with moderate amounts of data.
  - Can be hard to gather a lot of time series data

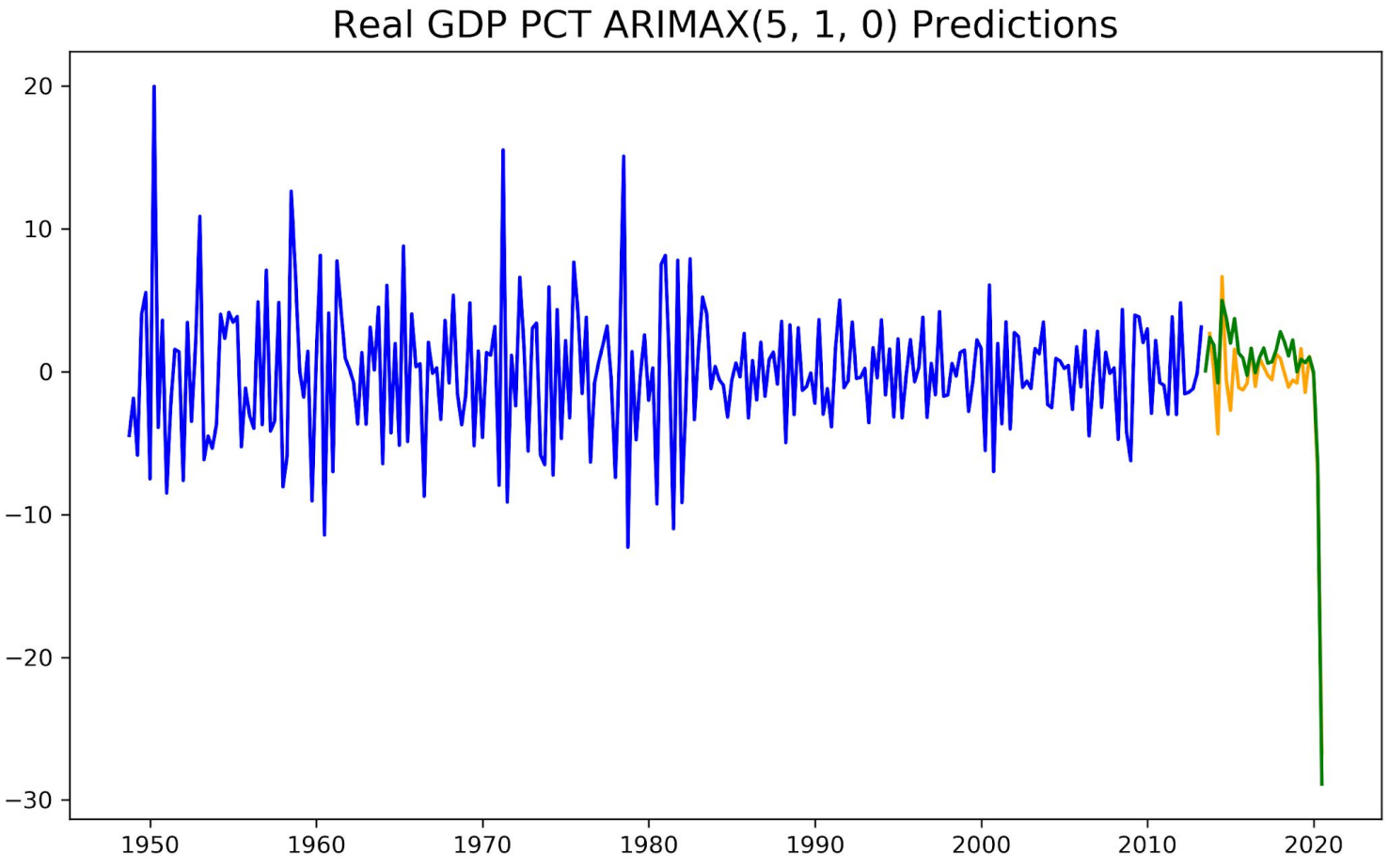
Model cont.

- Autoregressive
  - The regression of a variable on itself
  - explain long-term trends in our data
  - Hyperparameter p the number of previous values of Y to put into the model
- Moving Average Models (MA)
  - Takes previous error terms as inputs
  - The goal to predict future values based on forecasting errors
  - explain sudden shocks
  - Hyperparameter q is the number of previous errors to put into the model
- ARIMAX is a extension of ARIMA with
  - X being the exogenous variable
    - The variable can time-varying measurement, categorical or/and stand for a combination of several different external factors

Selecting Hyperparameter

Manual GridSearch to find the best parameters for the model

- Find the lowest AIC - 703.79
- The best p - 5
- The best q - 0


Real GDP PCT ARIMAX(5, 1, 0) Predictions
20
10
0
−10
−20
−30
1950
1960
1970
1980
1990
2000
2010
2020

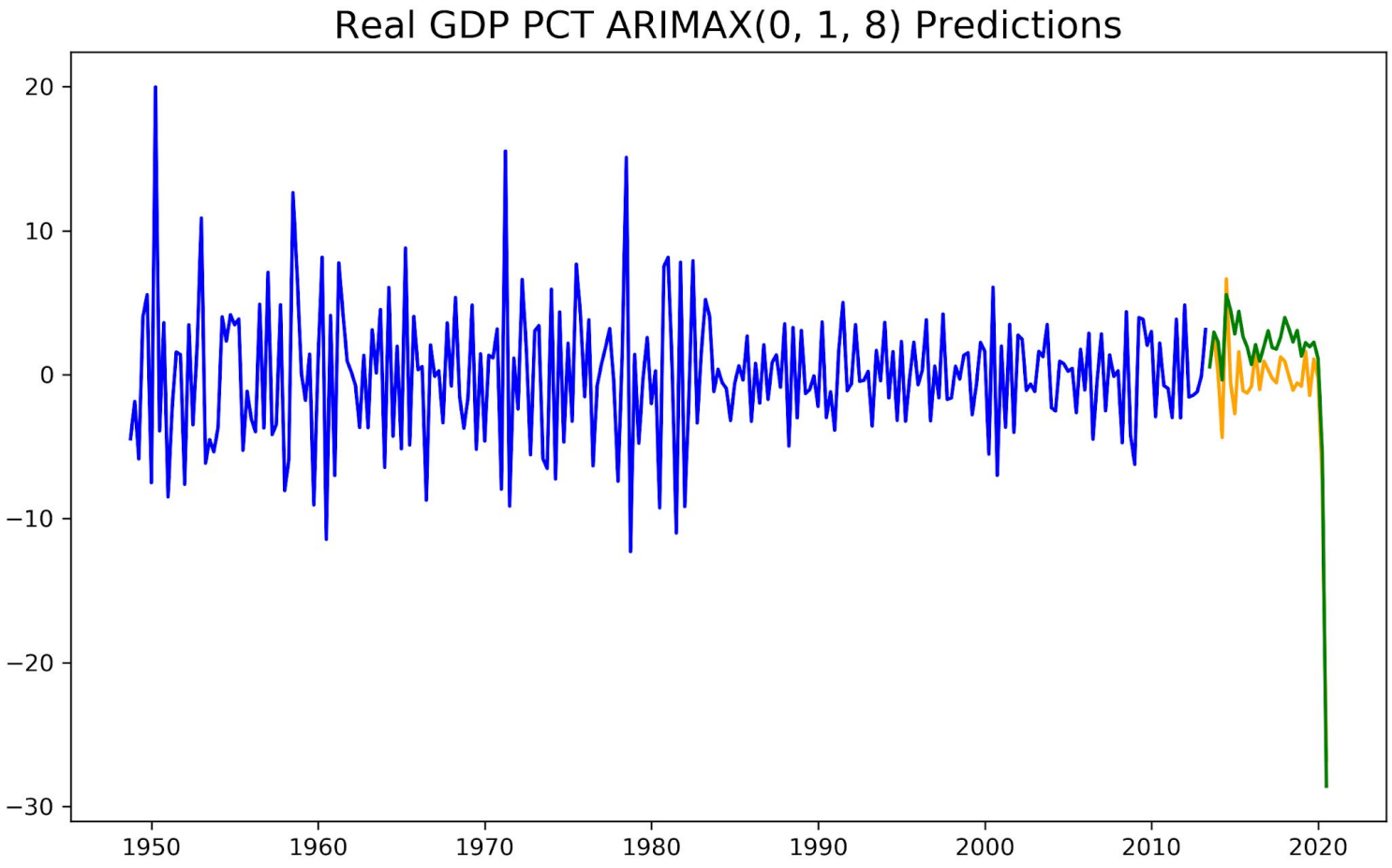
Model Summary Statistics

	coef	std err	z	P>|z|
const	-3.953e-11	0.007	-5.33e-09	1.000
consumption_pct_delta	0.5918	0.020	29.764	0.000
investment_pct_delta	0.1578	0.002	64.019	0.000
Exports_percent_delta	0.0390	0.003	12.205	0.000
Imports_percent_delta	-0.0316	0.003	-9.074	0.000
government_spending_percent_delta	0.1895	0.009	20.042	0.000
gdp_percent_2delta	0.0003	0.006	0.059	0.953
consumption_pct_2delta	0.0116	0.011	1.023	0.306
investment_pct_2delta	0.0001	0.000	0.341	0.733
Exports_pct_2delta	0.0082	0.014	0.573	0.566
government_spending_pct_2delta	0.0003	0.001	0.636	0.525
unemployment_rate	-0.1459	0.282	-0.518	0.605
employed	-0.3173	0.476	-0.666	0.505
Unemployed	-0.3173	0.476	-0.666	0.505
civilian_labor_force	0.3173	0.476	0.666	0.505
civilian_labor_force_participation_rate	0.6980	0.317	2.204	0.028
unemployment_rate_percent_delta	0.3611	9.031	0.040	0.968
employed_percent_delta	47.6664	38.828	1.228	0.220
Unemployed_percent_delta	1.0065	9.501	0.106	0.916
civilian_labor_force_percent_delta	9.1646	58.156	0.158	0.875
civilian_labor_force_participation_rate_percent_delta	-74.8989	44.618	-1.679	0.093
ar.L1	-0.7982	0.065	-12.212	0.000
ar.L2	-0.7733	0.089	-8.671	0.000
ar.L3	-0.6232	0.100	-6.218	0.000
ar.L4	-0.2806	0.092	-3.052	0.002
ar.L5	-0.1673	0.065	-2.589	0.010
sigma2	0.6682	0.066	10.129	0.000

Selecting Hyperparameter for drop insignificant Vars.

Manual GridSearch to find the best parameters for the model

- Find the lowest AIC - 661.76
- The best p - 0
- The best q - 8


Real GDP PCT ARIMAX(0, 1, 8) Predictions
20
10
0
−10
−20
−30
1950
1960
1970
1980
1990
2000
2010
2020

Model Summary Statistics

	coef	std err	z	P>|z|
const	-7.703e-09	1.41e-09	-5.474	0.000
consumption_pct_delta	0.6188	0.015	40.061	0.000
investment_pct_delta	0.1540	0.001	103.979	0.000
Exports_percent_delta	0.0411	0.003	15.968	0.000
Imports_percent_delta	-0.0341	0.003	-11.304	0.000
government_spending_percent_delta	0.1746	0.008	23.100	0.000
ma.L1	-0.9375	0.052	-18.075	0.000
ma.L2	-0.0802	0.066	-1.222	0.222
ma.L3	0.1290	0.095	1.354	0.176
ma.L4	0.1955	0.086	2.263	0.024
ma.L5	-0.1842	0.098	-1.885	0.059
ma.L6	0.0736	0.107	0.690	0.490
ma.L7	-0.0257	0.101	-0.255	0.799
ma.L8	-0.1358	0.067	-2.023	0.043
sigma2	0.6599	0.055	12.096	0.000

Conclusion

- Both Models have their strengths
  - First model over predicts closer to actual target
  - Second model has a better training AIC and significant variables
- Second model coefficients can be used to explain the movement
- The First model is better for predictions

Next steps

- Choose between the two models for production
- Try to add more engineered or gathered data
- Due to time constraints finish building CNN and/or LTSM