



# GDP Growth Forecast Multivariate Time Series Analysis

HOYT LUI



# Steps

- ▶ Data Wrangling
- ▶ EDA – resampling, interpolation, differencing, transforming, missing values, visualization
- ▶ Stationarity test – Augmented Dickey-Fuller test, KPSS test
- ▶ Cointegration test – Johansen test
- ▶ Modeling – Vector Autoregression (VAR)
- ▶ Lag order selection
- ▶ Forecasting
- ▶ Impulse response analysis
- ▶ Evaluation – Forecast error variance decomposition (FEVD)
- ▶ Inverting



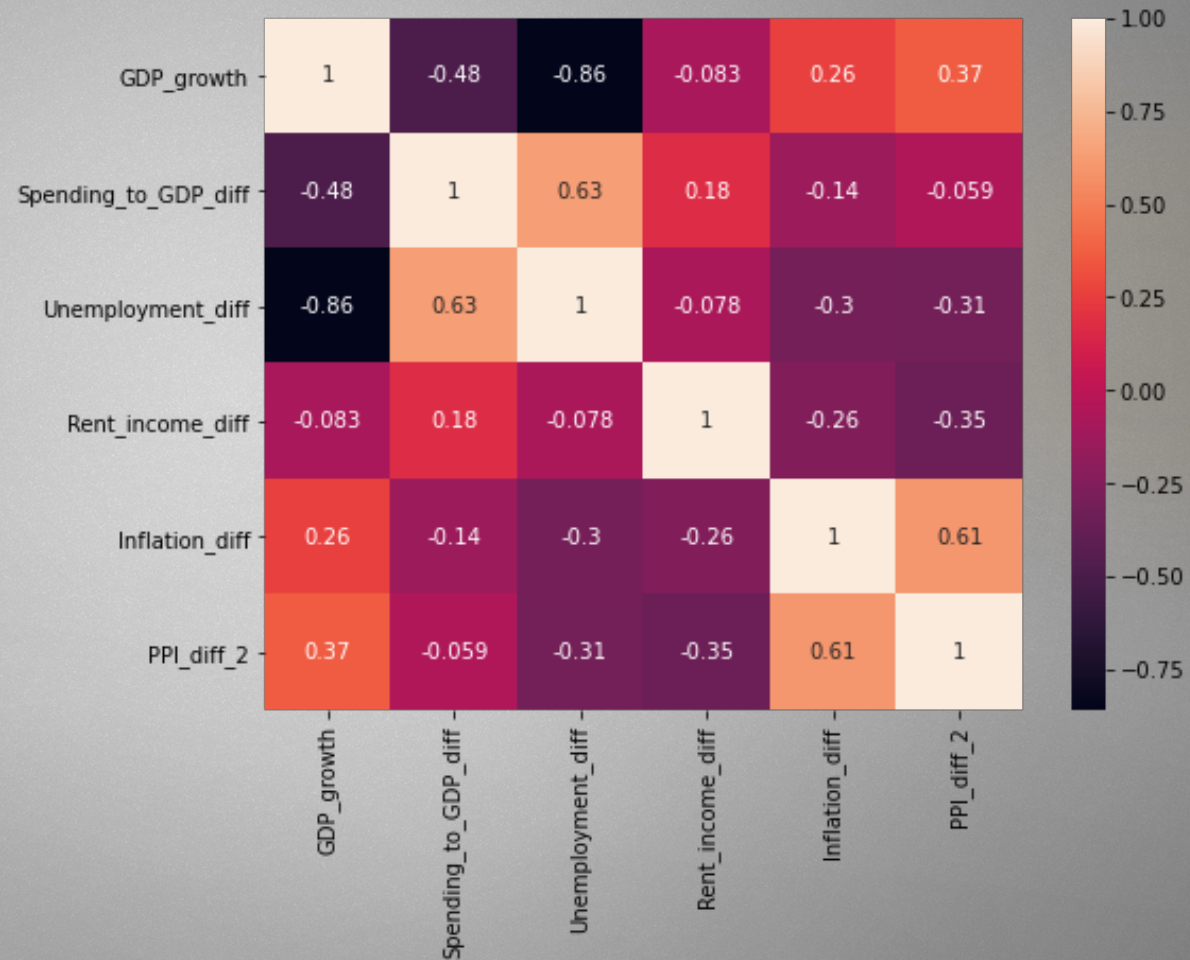
# Data Wrangling and EDA

- ▶ Preliminary plotting
- ▶ Resampling
- ▶ Interpolation
- ▶ Differencing and transforming
- ▶ Missing values imputation
- ▶ Merging dataframes
- ▶ Visualization



# Visualization

Which variables are highly and slightly correlated with GDP growth?





# Stationarity Test – Augmented Dickey-Fuller Test, KPSS Test

Critical value at 5%

```
=====
GDP_growth
Dickey-Fuller test (difference test):
Null hypothesis: series is not stationary
Test statistics      -3.394415
p-value             0.011151
Lags used           0.000000
Observations used   97.000000
Critical value (1%)  -3.499637
Critical value (5%)  -2.891831
Critical value (10%) -2.582928
dtype: float64
Fail to reject null hypothesis at critical value 1: series is not stationary
Reject null hypothesis at critical value 2: series is stationary
Reject null hypothesis at critical value 3: series is stationary

KPSS test (trend test):
Null hypothesis: series is stationary
Test statistics      0.401707
p-value             0.076420
Lags used           12.000000
Critical value (10%) 0.347000
Critical value (5%)  0.463000
Critical value (2.5%) 0.574000
Critical value (1%)  0.739000
dtype: float64
Reject null hypothesis at critical value 1: series is not stationary
Fail to reject null hypothesis at critical value 2: series is stationary
Fail to reject null hypothesis at critical value 3: series is stationary
Fail to reject null hypothesis at critical value 4: series is stationary
```



# Cointegration Test – Johansen Test

What is the ideal number of coefficients for cointegrated time series?

Trace statistics

	cv=90%	cv=95%	cv=99%	trace stat	r
<b>0</b>	91.1090	95.7542	104.9637	132.680942	r = 0
<b>1</b>	65.8202	69.8189	77.8202	83.916761	r <= 1
<b>2</b>	44.4929	47.8545	54.6815	47.276366	r <= 2
<b>3</b>	27.0669	29.7961	35.4628	30.106856	r <= 3
<b>4</b>	13.4294	15.4943	19.9349	18.158044	r <= 4
<b>5</b>	2.7055	3.8415	6.6349	8.801103	r <= 5

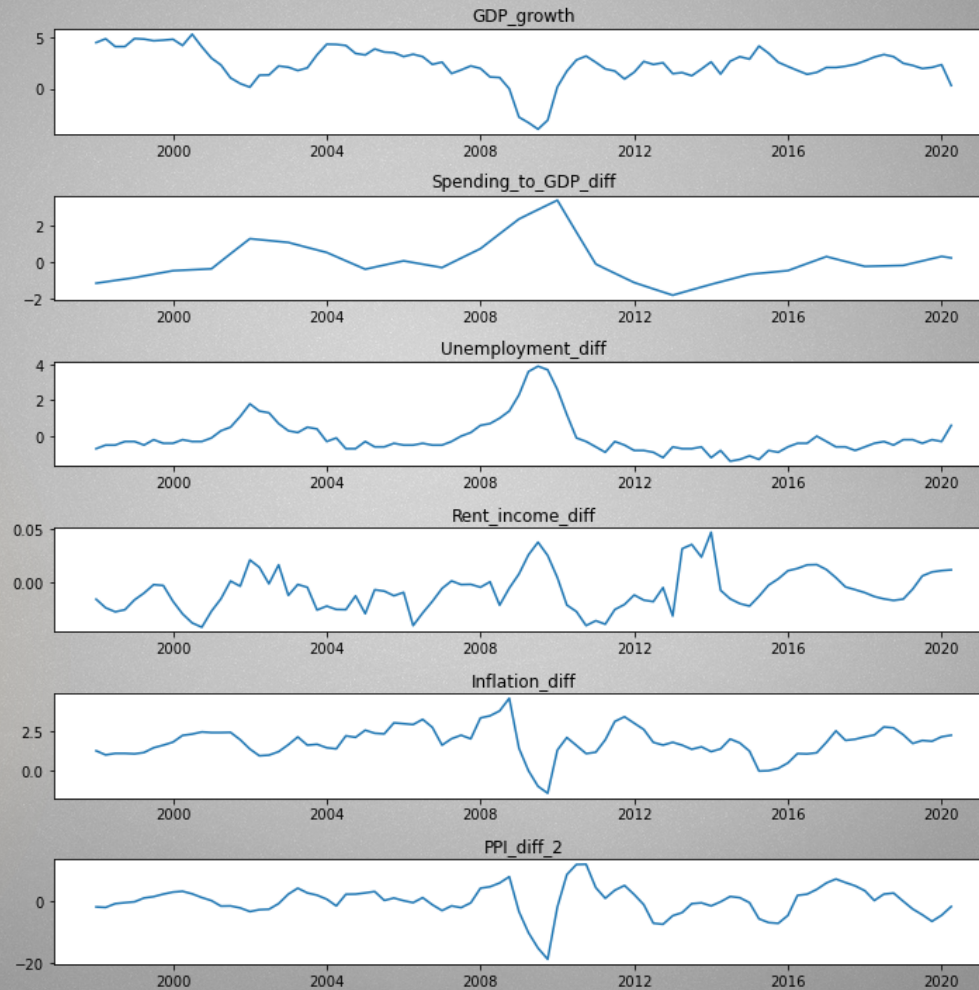
Eigen statistics

	cv=90%	cv=95%	cv=99%	eig stat	r
<b>0</b>	37.2786	40.0763	45.8662	48.764181	r = 0
<b>1</b>	31.2379	33.8777	39.3693	36.640395	r <= 1
<b>2</b>	25.1236	27.5858	32.7172	17.169509	r <= 2
<b>3</b>	18.8928	21.1314	25.8650	11.948812	r <= 3
<b>4</b>	12.2971	14.2639	18.5200	9.356941	r <= 4
<b>5</b>	2.7055	3.8415	6.6349	8.801103	r <= 5



# Modeling – Vector Autoregression (VAR)

VAR vs. ARIMA





# Lag Order Selection

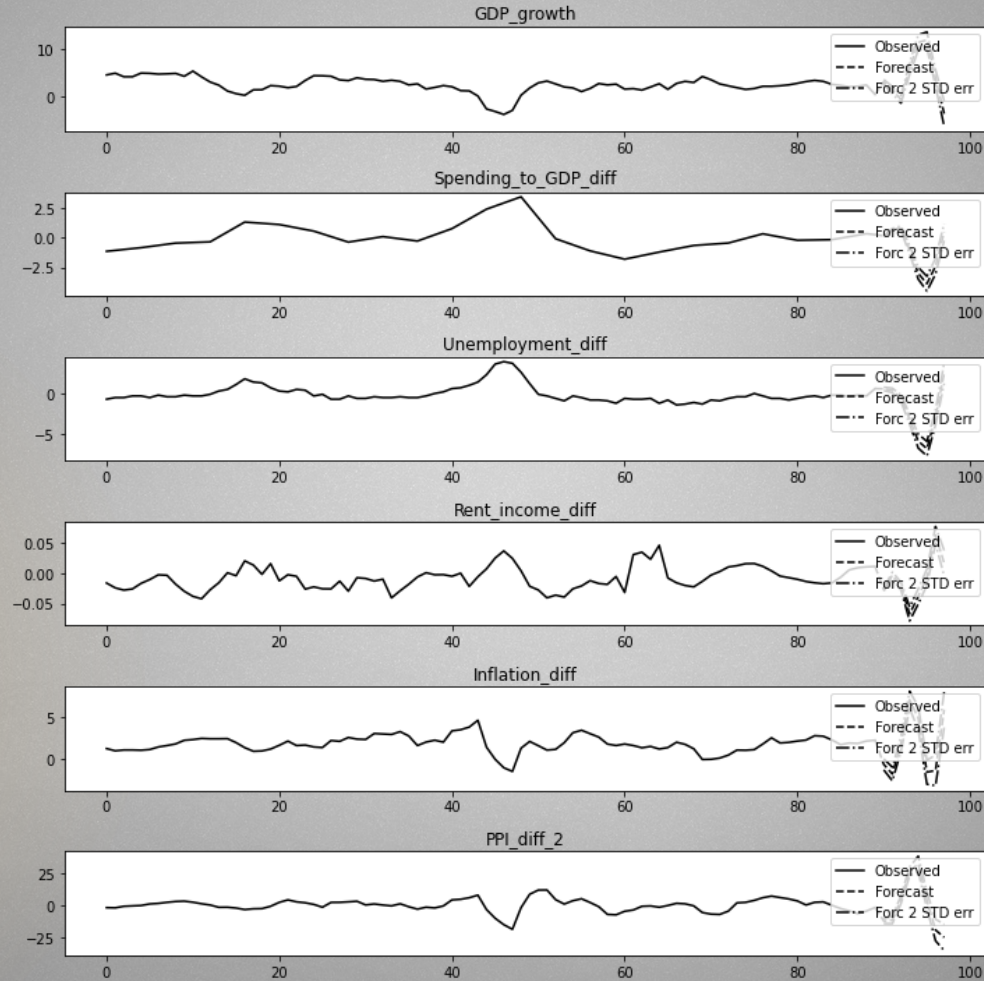
Selecting lag order based on AIC

AIC vs. BIC

Summary of Regression Results				
=====				
Model:	VAR			
Method:	OLS			
Date:	Mon, 12, Jul, 2021			
Time:	21:14:48			
-----				
No. of Equations:	6.00000	BIC:	-66.0466	
Nobs:	78.0000	HQIC:	-73.9827	
Log likelihood:	2865.87	FPE:	3.72855e-31	
AIC:	-79.2804	Det(Omega_mle):	1.73808e-33	
-----				
Results for equation GDP_growth				
=====				
	coefficient	std. error	t-stat	prob
const	1.358866	1.507464	0.901	0.367
L1.GDP_growth	-0.859822	0.410261	-2.096	0.036
L1.Spending_to_GDP_diff	-2.733166	1.520391	-1.798	0.072
L1.Unemployment_diff	-0.618433	1.037732	-0.596	0.551
L1.Rent_income_diff	26.931163	17.666132	1.524	0.127
L1.Inflation_diff	-2.021421	1.224211	-1.651	0.099
L1.PPI_diff_2	0.575526	0.372005	1.547	0.122
L2.GDP_growth	-0.402642	0.412703	-0.976	0.329
L2.Spending_to_GDP_diff	0.125916	2.693030	0.047	0.963
L2.Unemployment_diff	-0.592058	1.481904	-0.400	0.690
L2.Rent_income_diff	18.333273	13.118627	1.397	0.162
L2.Inflation_diff	-1.712199	1.181714	-1.449	0.147
L2.PPI_diff_2	0.298473	0.388722	0.768	0.443
L3.GDP_growth	0.036086	0.367740	0.098	0.922
L3.Spending_to_GDP_diff	4.107359	3.196769	1.285	0.199
L3.Unemployment_diff	-1.555735	1.467040	-1.060	0.289
L3.Rent_income_diff	-8.078692	25.395536	-0.318	0.750
L3.Inflation_diff	-2.154883	1.290066	-1.670	0.095
L3.PPI_diff_2	0.173975	0.347061	0.501	0.616
L4.GDP_growth	-0.270262	0.315999	-0.855	0.392
L4.Spending_to_GDP_diff	0.256761	1.977541	0.130	0.897
L4.Unemployment_diff	-3.460657	1.681094	-2.059	0.040
L4.Rent_income_diff	33.957997	17.001685	1.997	0.046
L4.Inflation_diff	-1.796284	1.563137	-1.149	0.250
L4.PPI_diff_2	0.062464	0.416148	0.150	0.881

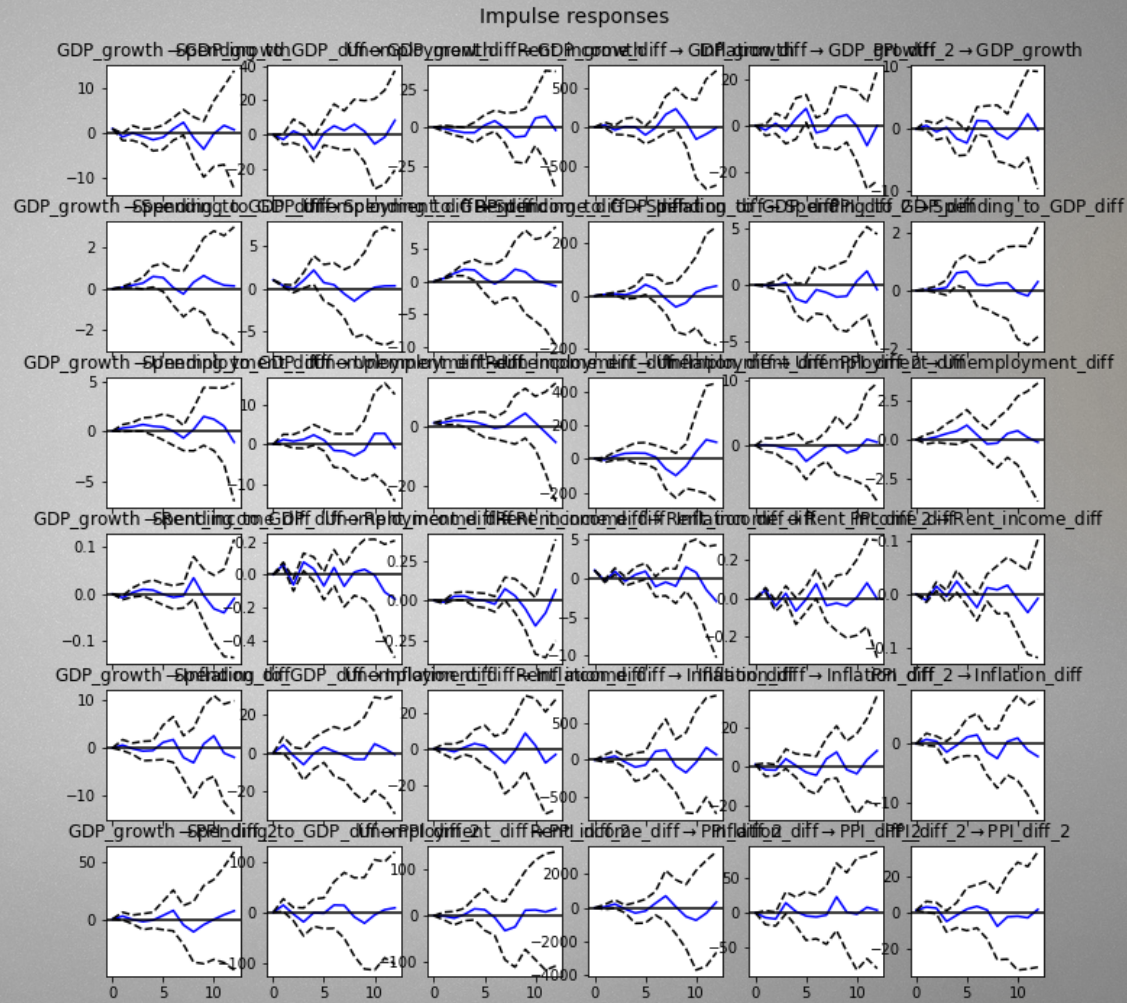
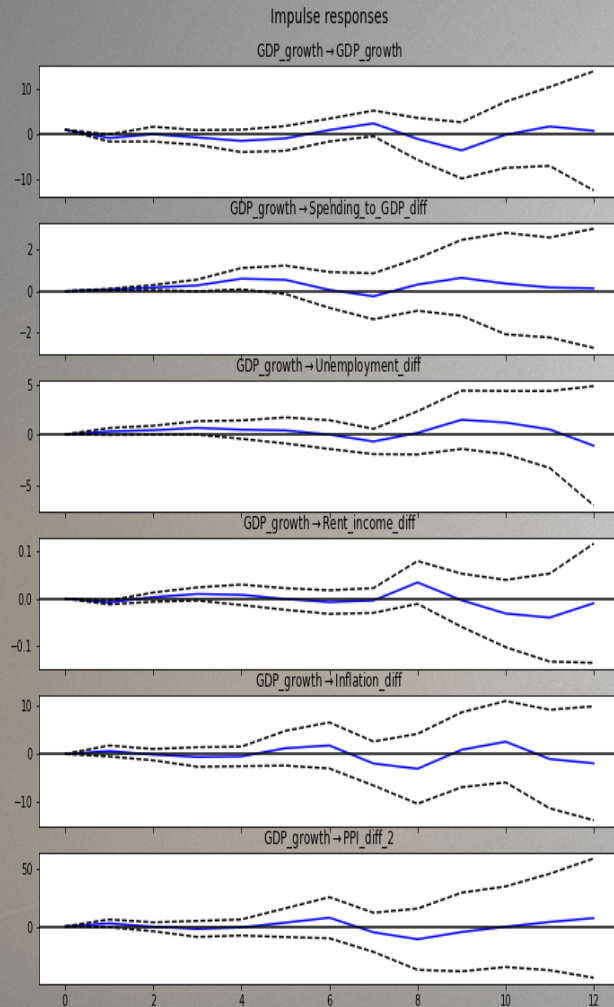


# Forecasting





# Impulse Response Analysis

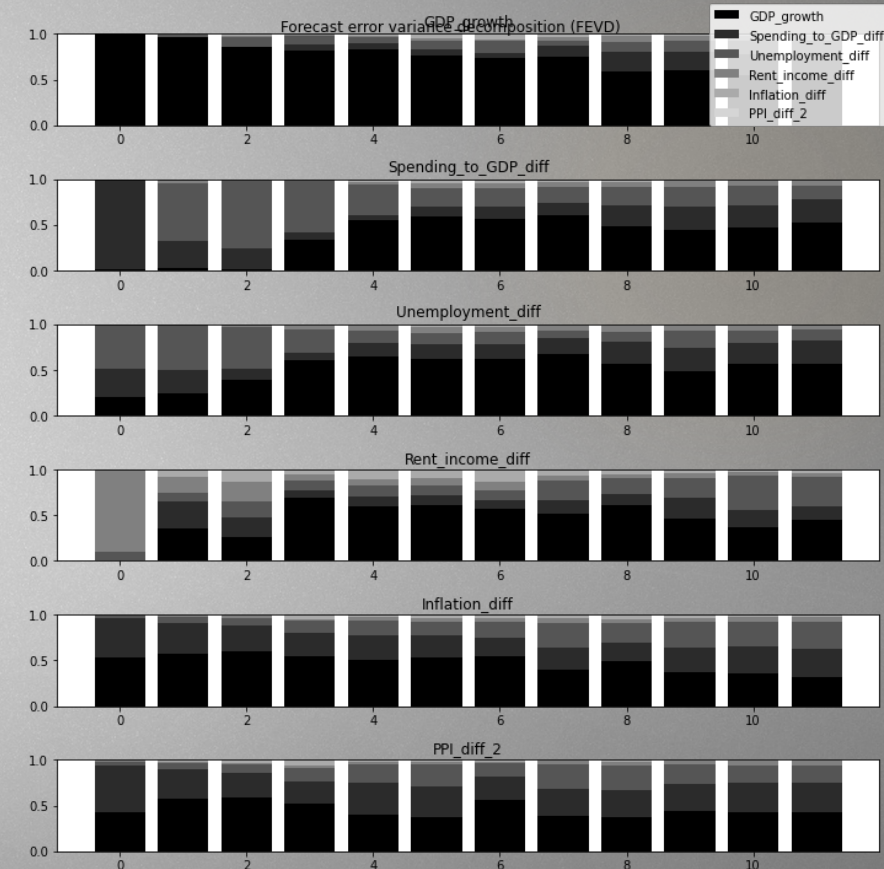




# Evaluation - Forecast error variance decomposition (FEVD)

	GDP_growth	Spending_to_GDP_diff	Unemployment_diff	Rent_income_diff	Inflation_diff	PPI_diff_2
0	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.968181	0.002702	0.016125	0.003038	0.009954	0.000000
2	0.858026	0.002989	0.102566	0.021371	0.015048	0.000000
3	0.817643	0.063195	0.099412	0.013462	0.006288	0.000000
4	0.830714	0.070067	0.063785	0.011868	0.023566	0.000000
5	0.769835	0.066011	0.089410	0.022449	0.052296	0.000000
6	0.732179	0.064685	0.121298	0.021144	0.060694	0.000000
7	0.743930	0.124775	0.061089	0.039123	0.031083	0.000000

Which variables contribute more/less error to GDP growth in each step?





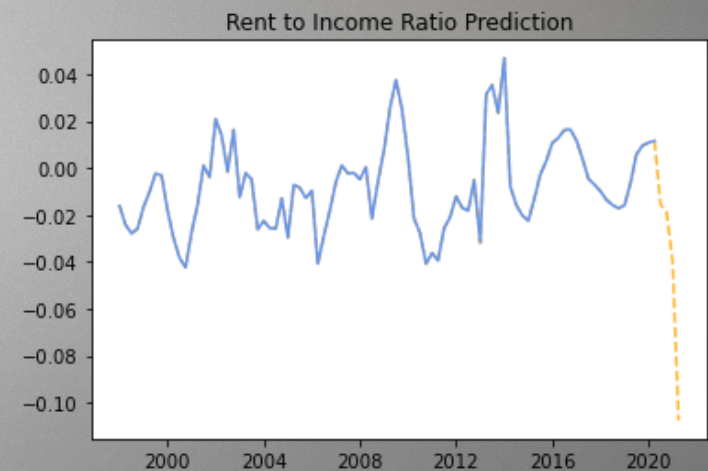
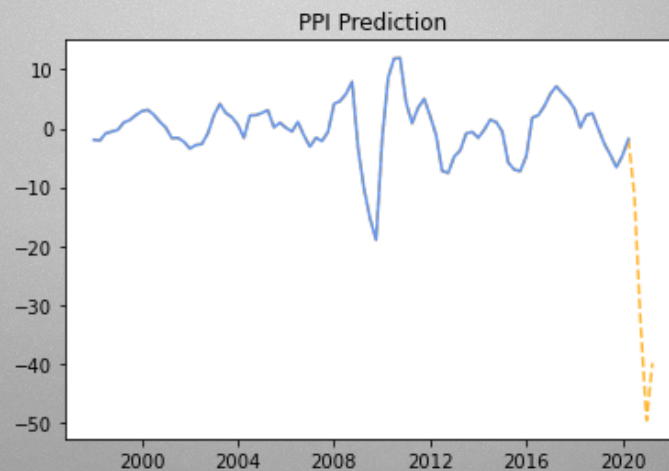
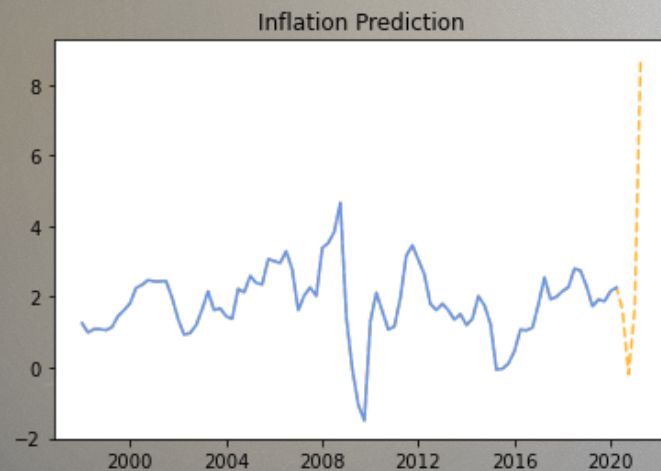
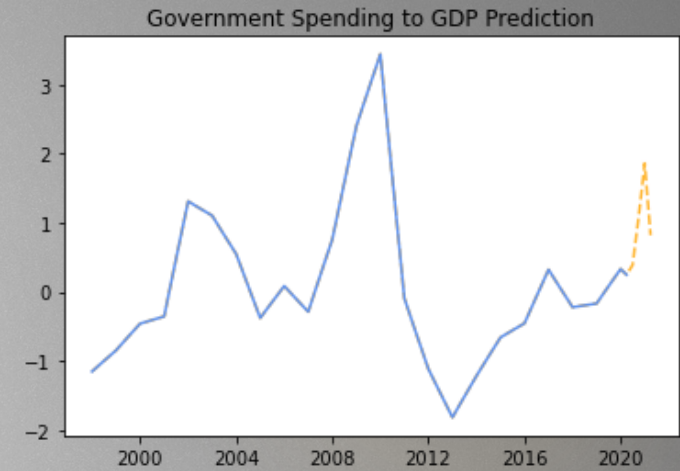
# Inverting

Variables to be inverted back to its original form

	GDP_growth_fc	Spending_to_GDP_diff_forecast	Unemployment_diff_forecast	Rent_income_diff_forecast	Inflation_diff_forecast	PPI_diff_2_forecast
Time						
2020-06-30	3.044324	0.386650	1.165243	-0.015498	1.613682	-11.429863
2020-09-30	1.236768	1.077823	1.539629	-0.018698	-0.195304	-32.753610
2020-12-31	-0.709559	1.858498	0.952496	-0.040002	1.642830	-49.655060
2021-03-31	5.891302	0.818498	-2.209186	-0.107554	8.752460	-39.896010



# Variables Forecast





# Takeaway

- ▶ Heatmap: unemployment rate vs. GDP growth, rent to income ratio vs. GDP growth
- ▶ Impulse response analysis: rent to income ratio vs. GDP growth
- ▶ Forecast error variance decomposition: GDP growth forecast vs. producer price indices (PPI)
- ▶ Johansen test: 2 variables only
- ▶ Unemployment rate is always in the same conversation with GDP growth. GDP growth may be explained by inflation, government spending and producer price indices, but not too relevant with rent to income ratio.



# Further Work

- ▶ Short-term quick fix:
- ▶ Johansen test for cointegration: eliminate all columns but 1-2 most significant ones along with the GDP growth
- ▶ Effect on GDP growth: replace the most insignificant variables with other variables to see the new predictive power
- ▶ Long-term work:
- ▶ Test other countries of our choice, may find surprising results and foretell an upcoming market crash.