A TERM PROJECT SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND NATURAL SCIENCES

OF

SABANCI UNIVERSITY

BY

ALİ KARAKUŞ, EZGİ BİLGE İŞLİ, ONUR AKBAŞ, TUFAN KESER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF MASTER OF SCIENCE

IN

THE DEPARTMENT OF DATA ANALYTICS

SEPTEMBER 2021

We hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. We also declare that, as required by these rules and conduct, we have fully cited and referenced all material and results that are not original to this work.

Ali Karakuş, Ezgi Bilge İşli, Onur Akbaş, Tufan Keser

Quarterly Gross Domestic Product (GDP) Forecast Model via Time series Methods Using Higher Frequency Data Provided by Official Resources

Ali Karakuş, Ezgi Bilge İşli, Onur Akbaş, Tufan Keser

M.S., The Department of Data Analytics

September 2021

Abstract

Gross domestic product (GDP) is a measure of the total value of all goods and services produced annually within the borders of a country. This metric can be calculated every three months or quarterly to estimate annual GDP. Gross domestic product can basically be calculated in three ways using expenditures, production or revenues. Adjustments for inflation and population can be made to provide more accurate forecasts. GDP is the most effective tool in the strategic decision making stages of politicians, investors and businesses. It is crucial in determining the risk profile, estimating unemployment, and credit growth. That is why, GDP forecast is one of the most important research topics in applied economics. Using higher frequency data provided by official resources such as CBT and BRSA, time series methods have also been used to estimate the GDP. In this research, we focus on time series forecasts on GDP via classic statistical economic models along with other machine learning classification methods to be able to provide an accurate forecast for the stakeholders of the economy.

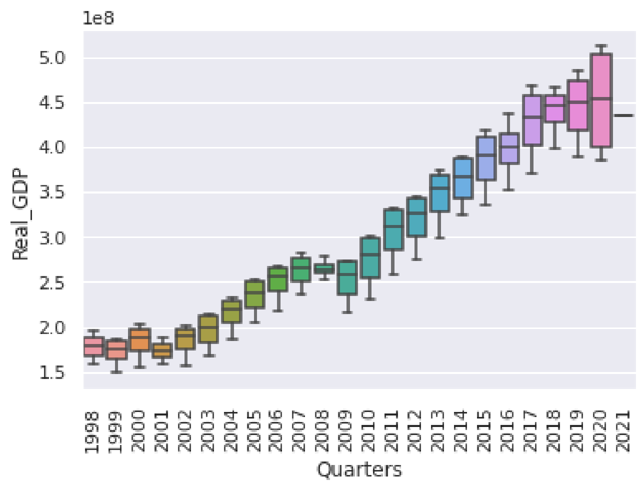
Keywords : GDP FORECAST, TIME SERIES, ARIMA, SARIMA, LSTM, DICKEY-FULLER TEST, GRANGER CAUSALITY, MACHINE LEARNING, CLASSIFICATION, XGB, BASE METHOD, IN SAMPLE WALK FORWARD, OUT OF SAMPLE WALK FORWARD, OUT OF SAMPLE PREDICTION, MLP, HOLT-WINTERS, PROPHET

Introduction

Gross domestic product forecasting is among the most comprehensive economic, financial and commercial analyses. Due to the complexity of the content knowledge and its compelling procedures and calculations, its applications in data science are relatively limited and have been the subject of applied economics to a large extent. However, after the Covid-19 epidemic, we see that the use of new and alternative models in GDP forecasting has gained importance, as the uncertainties arising from GDP have become unsettling for many companies and institutions in terms of sales growth and continuity. In this sense, time series can be useful in decision-making for those companies and business units by predicting future demand or future sales.

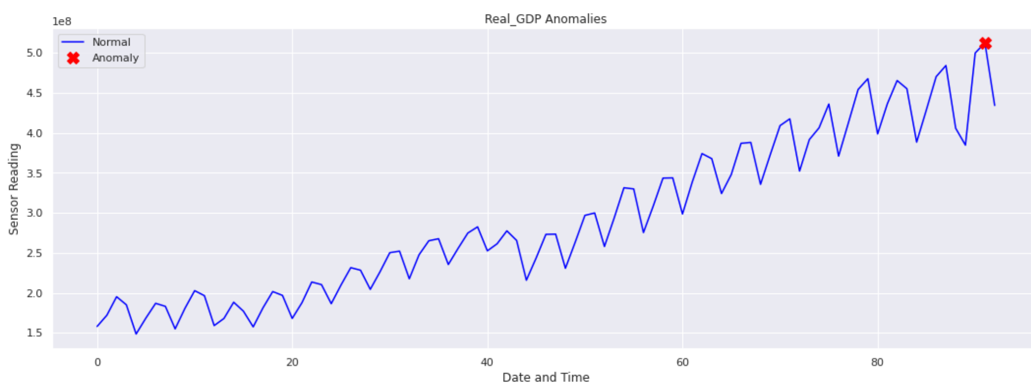
In light of this need, our purpose in this project is to forecast quarterly gross domestic product via time series methods using higher frequency data provided by official resources. Our data is from 1998 to the first quarter of 2021. These are only published quarterly on tcmb.gov.tr. That is why, the official website includes the biggest available official dataset. According to the Central Bank, BIST-100, USD/TL, credit default swap (CDS), 2Y –Bond Yield, 10Y – Bond Yield, private mortgage insurance–manufacturing, private mortgage insurance-service, employment statistics, confidence indices, retail sales index, consumption good imports, retail loans (including credit cards), retail non-performing loans (NPLs), industrial production of durable/non-durable consumer goods, special consumption tax collection, state personnel and goods & services expenditures, number of tourists, passenger car sales, white goods sales, house sales, industrial production index, state capital expenditures, commercial loans, capacity utilization rate, electricity production/consumption, retail sector inventory level index, investment good imports, exports, tourism revenues, imports,tourism expenditures are used as inputs in the calculation of GDP. As expected, an upward trend and seasonality were highly present in the data that we used. Besides, sharp shifts imply the economic crisis in 2008 and the effects of Covid-19 global pandemic in 2020 in the graph below which shows the GDP year by year.

Outlier Analysis

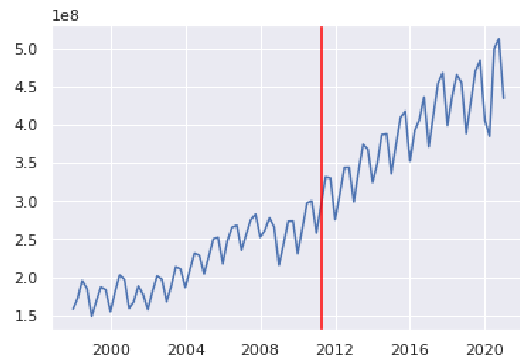


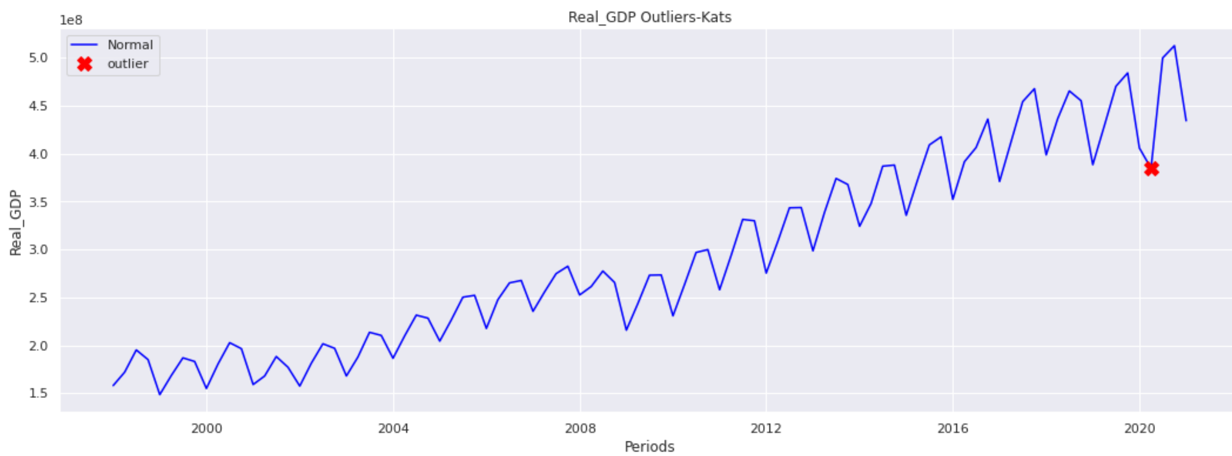
In the years 2008 and 2020 we can see the sudden fluctuations. To observe the effect of outliers we used an isolation forest algorithm. Taking the contamination parameter as 0.001, we detected the last quarter of the year 2020 as outlier. Nevertheless, there were not many outliers in the data and our data are not considerably huge at the same time, that is why we decided not to eliminate the outliers.

The results from isolation forest are below:

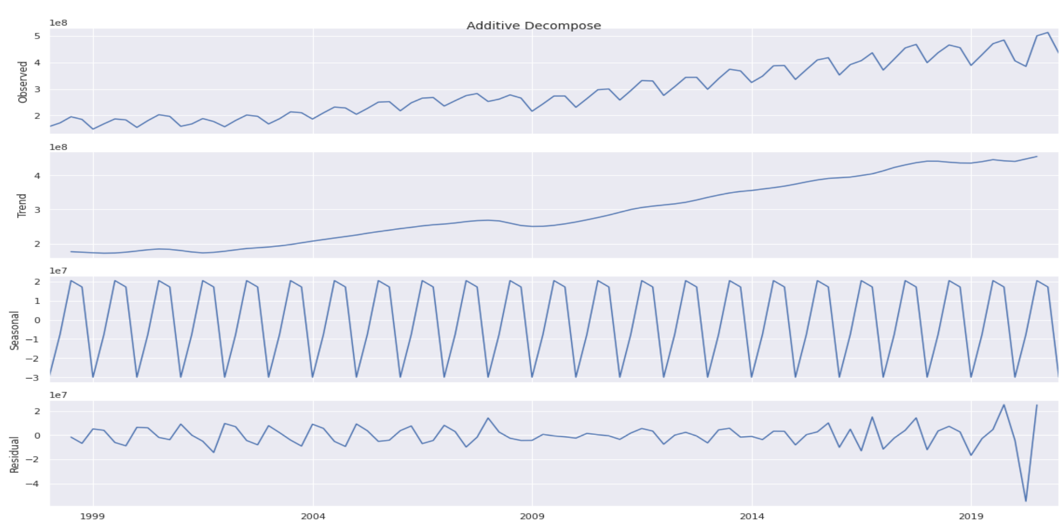


We also used Facebook Kats to find outliers, it gives us better results. In the year 2011, a shift was detected in the trend, and the effect of Covid-19 on the quarter of 2020 was detected as outlier.





TESTS
Non-Stationarity Check



When we look at the data, we see an upward trend and seasonality. And applied additive decomposition also shows trend and seasonal effect.

To quantitatively assess the stationarity of the data, statistical tests make strong assumptions about the data. They can only be used to inform the degree to which a null hypothesis can be accepted or rejected. They can provide a quick check and confirmatory evidence that your time series is stationary or non-stationary. For this data we apply Dickey-Fuller (DF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests.

Kwiatkowski-Phillips-Schmidt-Shin Test for Non-Stationarity

KPSS test is a statistical unit root test to check for stationarity of a series around a deterministic trend. The word 'deterministic' implies the slope of the trend in the series does not change permanently. That is, even if the series goes through a shock, it tends to regain its original path.

H0 : Stationarity (trend stationary: stationary around a deterministic trend, meaning that de-trended series is level-stationary like white noise)
HA : Non-stationary
If $p \leq 0.05$: Reject H0 (trend is stochastic, needs differencing)
If $p > 0.05$: Fail to reject H0 (NO evidence that the series is NOT trend stationary)

KPSS test says the p-value is significant with p-value = 0.02758690643282572
(p-value < 0.05) and hence, we can reject the null hypothesis (series is stationary) and derive that the series is NOT stationary. In the default option of the KPSS test, it tests for stationarity around a 'mean' only.For the stationarity testing around a trend, we need to explicitly pass the regression='ct' parameter to the KPSS.

Dickey-Fuller Test for Non-Stationarity

There are some differences between the Kwiatkowski-Phillips-Schmidt-ShinTest and Dickey-Fuller Test. The major difference between KPSS and DF tests is the capability of the KPSS test to check for stationarity in the 'presence of a deterministic trend'. What that effectively means to us is, the KPSS test may not necessarily reject the null hypothesis (that the series is stationary) even if a series is steadily increasing or decreasing.

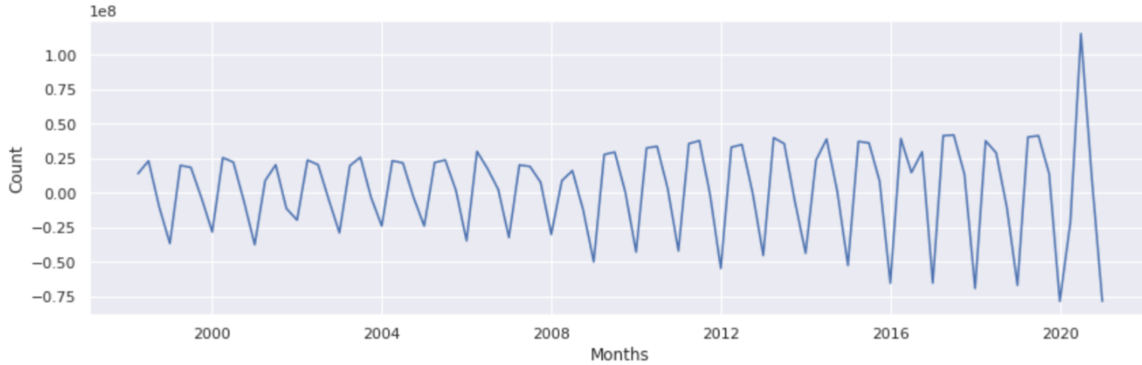
Stationarity implies that taking consecutive samples of data with the same size should have identical covariances regardless of the starting point. If the model satisfied the following conditions,

- The mean of the series is not a function of time
- The variance of the series is not a function of time (homoscedasticity)
- The covariance at different lags is not a function of time

It can be assumed that the model is stationary. Statistical tests, such as Dickey-Fuller, Augmented Dickey-Fuller and Phillips Perron, can be used to determine whether the model is stationary or non-stationary.. By taking the short-term period of the dataset, we used the Dickey-Fuller test in the project. Dickey-Fuller test is a statistical test for checking stationarity; the null hypothesis is that the TS is non-stationary.

```
Test Statistic           0.685057
p-value                  0.989535
# Lags                   4.000000
# Observations          88.000000
Critical Value (1%)     -3.506944
Critical Value (5%)     -2.894990
Critical Value (10%)    -2.584615
dtype: float64
 Series is Non-Stationary
```

If our test statistic is above an alpha value, we can reject the null hypothesis and say that the series is stationary. As seen in the output of the Dickey-Fuller Test, Series is Non-Stationary.



After differencing our data we can see from the above graph that our data becomes stationary and adf test below also confirms that our data become stationary.

```
Test Statistic            -4.837832
p-value                    0.000046
# Lags                     3.000000
# Observations            88.000000
Critical Value (1%)       -3.506944
Critical Value (5%)       -2.894990
Critical Value (10%)      -2.584615
dtype: float64
 Series is Stationary
```

METHODS

We generally use classical time series methods including in sample walk forward, out of sample walk forward, out of sample prediction, Holt-Winters and SARIMA. We also applied Facebook's Prophet, MLP, XGB and LSTM in addition to these time series approaches.

ARIMA MODEL INTRODUCTION- BOX-JENKINS METHOD

As an introduction before using the different time series methods we want to explain all of the components of the ARIMA model separately and elaborately.
The Autoregressive Integrated Moving Average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. ARMA model has two different components;

AR (Autoregressive Process)

** Autoregressive (AR) Process**

Let Yt represent GDP at time t. If we model Yt as

- $(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u$

where δ is the mean of Y and where it is an uncorrelated random error term with zero mean and constant variance σ 2 (i.e., it is white noise), then we say that Yt follows a first-order autoregressive, or AR(1), stochastic process Here the value of Y at time t depends on its value in the previous period and a random term; the Y values are expressed as deviations from their mean value. In other words, this model says that the forecast value of Y at time t is simply some proportion ($= \alpha 1$) of its value at a time ($t - 1$) plus a random shock or disturbance at time t; again the Y values are expressed around their mean values But if we consider this model,

- $(Yt - \delta) = \alpha 1(Yt-1 - \delta) + \alpha 3(Yt-2 - \delta) + ut$

then we say that Yt follows a second-order autoregressive, or AR(2), process. That is, the value of Y at time t depends on its value in the previous two time periods, the Y values being expressed around their mean value δ. In general, we can have

- $(Yt - \delta) = \alpha 1(Yt-1 - \delta) + \alpha 2(Yt-2 - \delta) +\cdots+ \alpha p(Yt-p - \delta) + u$

in which case Yt is a pth-order autoregressive, or AR(p), process.

Moving Average (MA) Process

** Moving Average (MA) Process**

The AR process just discussed is not the only mechanism that may have generated Y. Suppose we model Y as follows:

- ** $Yt = \mu + \beta 0ut + \beta 1ut-1$ **

where μ is a constant and u, as before, is the white noise stochastic error term. Here Y at time t is equal to a constant plus a moving average of the current and past error terms. Thus, in the present case, we say that Y follows a first-order moving average, or an MA(1), process. But if Y follows the expression;

- ** $Yt = \mu + \beta 0ut + \beta 1ut-1 + \beta 2ut-2$**

then it is an MA(2) process. More generally,

- ** $Yt = \mu + \beta 0ut + \beta 1ut-1 + \beta 2ut-2 +\cdots+ \beta qut-q$**

is an MA(q) process. In short, a moving average process is simply a linear combination of white noise error terms.

Autoregressive and Moving Average (ARMA) Process

It is quite likely that Y has characteristics of both AR and MA and is therefore ARMA. Thus, Yt follows an ARMA(1, 1) process if it can be written as:

- $Yt = \theta + \alpha 1Yt-1 + \beta 0ut + \beta 1ut-1$

Because there is one autoregressive and one moving average term. θ represents a constant term. In general, in an ARMA(p, q) process, there will be p autoregressive and q moving average terms.

Autoregressive Integrated Moving Average (ARIMA) Process

If we have to difference a time series d times to make it stationary and then apply the ARMA(p, q) model to it, we say that the original time series is ARIMA(p, d, q), that is, it is an autoregressive integrated moving average time series, where p denotes the number of autoregressive terms, d the number of times the series has to be differenced before it becomes stationary, and q the number of moving average terms.Thus, an ARIMA(2, 1, 2) time series has to be differenced once (d = 1) before it becomes stationary and the (first-differenced) stationary time series can be modeled as an ARMA(2, 2) process, that is, it has two AR and two MA terms.

In-Sample Prediction Walk-Forward

Random Walk is a special type of time series, the next value in the sequence is a modification of the previous value. This process is:

$X(t) = \alpha + X(t-1) + \varepsilon t$ where $\varepsilon t \sim WN(0,\sigma 2)$.

RW with a drift is a stochastic trend. Although X equals its previous value plus an additional increment in each step (analogous to deterministic trend), it's called stochastic because there is a non-stationary random component.

We fit the model on the training data and make a one-step-ahead (walk-forwarding) prediction on the test data using the true value from the previous time step to predict the next value in time. This is what we call an in-sample prediction.

We decide the first parameter by looking at the partial autocorrelation plot in first order differencing .After seeing 3 spikes in the partial autocorrelation graph, we use 3 as the first parameter, and number of differencing as 1. That is, our model will be (3,1,0).
Our prediction follows trend and catches increase points; however, decrease points cannot be captured by this model. Still, we improved the base method by significant margin with this model.

Out-of-sample prediction walk-forward

For the out-of-sample predictions, we make a one-step-ahead (walk-forwarding) prediction on the test data using the last predicted value from the previous time step to predict the next value in time. This is called out-of-sample prediction.

Out-of-sample prediction performs worse than in sample random walk prediction. Just like the previous in-sample prediction walk-forward model, this model cannot catch decreasing points.

Out-of-sample prediction (multi-step)

Out of sample multi step prediction automatize previous steps with forecast horizon which is chosen 11 according to our test size. As expected, our RMSE result is quite close to out of sample walk forward.

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                   81
Model:                 ARIMA(3, 1, 0)   Log Likelihood               -1444.044
Method:                       css-mle   S.D. of innovations       13019228.437
Date:                Wed, 14 Jul 2021   AIC                           2898.087
Time:                        10:50:27   BIC                           2910.060
Sample:                             1   HQIC                          2902.891

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       3.437e+06   4.23e+05      8.130      0.000    2.61e+06    4.27e+06
ar.L1.D.y     -0.8013      0.066    -12.137      0.000      -0.931      -0.672
ar.L2.D.y     -0.8792      0.050    -17.644      0.000      -0.977      -0.782
ar.L3.D.y     -0.8061      0.065    -12.403      0.000      -0.933      -0.679
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            0.0335           -1.0346j            1.0352           -0.2449
AR.2            0.0335           +1.0346j            1.0352            0.2449
AR.3           -1.1576           -0.0000j            1.1576           -0.5000
-----------------------------------------------------------------------------
```

According to the results of the ARIMA model above, we decided on the ARIMA (3,1,0) model for GDP forecast. Looking at p values, the 3th degree autoregressive model seems to be statistically significant. For further investigation, Granger Causality test and vector autoregression (VAR) model are used.

In-sample forecasting using Keras MLP

We created a function which will give us a dataframe for our machine learning model by using previous periods as features for GDP prediction. We used 8 previous periods as features, 20 nodes and 2000 epochs and 1 batch for our predictions. Results are considerably close to the observed results in the first 7 periods after Covid-19 global pandemic. Our model cannot predict a deep decrease in the second quarter of 2020 and a sharp increase in the third quarter of 2020 because of the adverse effects of the epidemic.

In-sample forecasting using XGBRegressor

We applied an XGB model along with an evaluation set for our model to be able to decide the best parameters via gridsearch. After the parameter tuning, the best parameters gave us a fair RMSE score, still the model cannot predict the trend of the dataset effectively.

Triple Exponential Smoothing (Holt-Winters) Method

In addition to the level and trend component (as in DES), we also update a season component.

Multiplicative model (for seasonality):

$L_t = \alpha\ (y_t / S_{t-T}) + (1- \alpha)\ (L_{t-1} + T_{t-1})$ (level smoothing)

$T_t = \beta\ (L_t – L_{t-1}) + (1- \beta)\ T_{t-1}$ (trend smoothing)

$S_t = \gamma\ (y_t / L_t) + (1- \gamma)\ S_{t-T}$ (seasonal smoothing)

RMSE score is not good, it is just more optimistic than the real results. That is why, we tried an alternative validation data to tune our parameters in our train dataset to see whether we can improve our results. After parameter tuning, causing an increase in our RMSE score, the best parameters catch the general trend much better, even though the model still does not seem to be the best model.

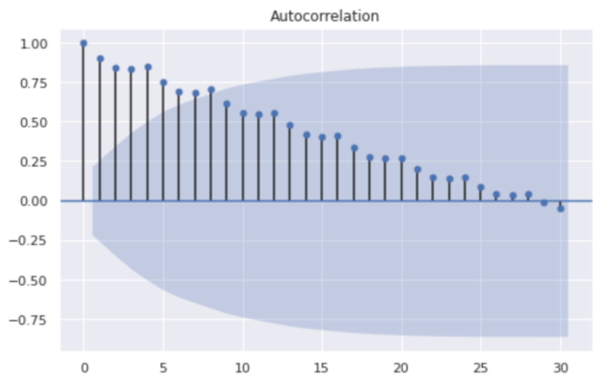
SARIMA

Basically, SARIMA models are ARIMA models with a seasonal component. In the formula SARIMA(p,d,q)x(P,D,Q,s), the parameters for these types of models are as follows:

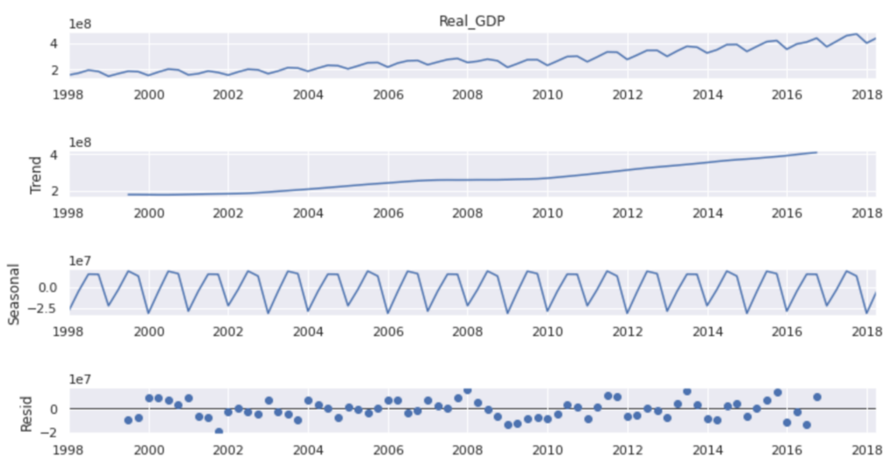
- p and seasonal P: indicate number of autoregressive terms (lags of the stationarized series)
- d and seasonal D: indicate differencing that must be done to stationarize series
- q and seasonal Q: indicate number of moving average terms (lags of the forecast errors)
- s: indicates seasonal length in the data

ACF (Autocorrelation Function)

Two approaches were taken to determine the ideal SARIMA parameters: ACF and PACF plots. The ACF plots were used as a starting point to narrow down to a few potential parameters



As above, there are significant positive spikes in the ACF plot between lag 0 to lag 28 (monthly seasonal component). These characteristics suggest a potential parameter starting point.



Seasonality and general trend are easily followed above, residual distribution looks quite good as well. Fairly constant standard deviation is good for our model, but rolling mean keeps increasing. That says, our model has a clear trend which we need to get rid of. According to results of the Dickey-Fuller Test, the series is not stationary as p value is bigger than 0.05. Different tests such as Augmented Dickey-Fuller or Phillips Perron can be used to determine whether the model is a stationary or non-stationary model. After differentiating the series once, the Dickey-Fuller Test shows us that the series is stationary as p value is under 0.05.

PACF (Partial Autocorrelation Function)

PACF plots, also, were used as a starting point to narrow down to a few potential parameters. In the PACF plot, there are significant negative and positive spikes at lags (i.e. lags are slower to decay)

ARIMA(1,1,1)(1,1,0)[12] is our best model for minimizing AIC

We tried models with different parameters in different time steps to find our best model by using cross validation. After deciding the best parameters, we tried them on our time steps to find the best working model to minimize RMSE. (0,1,0) (1,1,1,12) gives us the lowest RMSE score.

Long-term Forecasting

We fit the entire data, both train and test, using the parameters found earlier to be able to make a forecast for 5 and 10 years ahead. We saw in the graphs that the confidence interval grows as we make predictions that are further in time, so we cannot be confident in our long term forecasts.

Facebook's Prophet for Forecasting

We used Facebook's open-sourced forecasting tool Prophet by applying it to GDP forecasts. Unfortunately this model gave us the worst results.

LSTM

We created a dataframe with 8 previous periods to use as features for our LSTM model, by dropping rows with missing values, creating the same train and test data, min-max scaling to our data by using train values and reshaping the data.

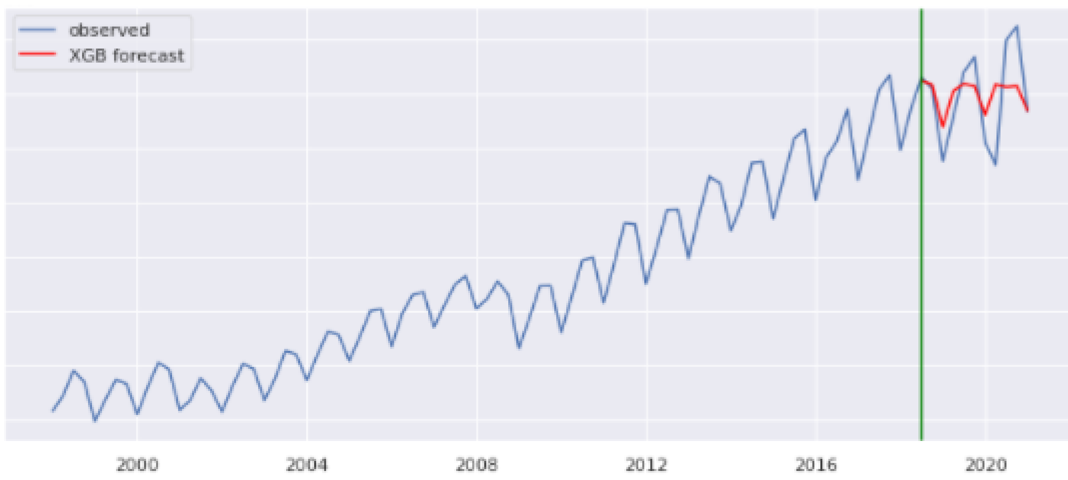
Results and Evaluation

According to the results table below, in sample walk forward seems to give the best result. The relatively small size of our dataset, which is from 1998 to the first quarter of 2021, makes estimating GDP with models like LSTM and prophet difficult. However, the best model captures the trend which is shaped by an uneven drop due to Covid-19 global pandemic in the second quarter of 2020 and the sharp increase after the effects of it. Last but not least, sample predictions seem to perform better than out-of-sample predictions. Not surprisingly, results are not that accurate in some models such as LSTM when compared to the previous models, due to the fact that data is considerably small. It is known that LSTM requires much more data to be able to give better results. Still, the model generally captures trends well, except the sharp drop in the second quarter of 2020 because of the Covid-19 global pandemic.

	method	RMSE
1	in sample walk forward	2.236547e+07
4	MLP	2.637476e+07
3	out of sample prediction	3.020529e+07
6	Holt-Winters	3.063410e+07
2	out of sample walk forward	3.117336e+07
5	XGB	3.286731e+07
7	SARIMA	3.436514e+07
9	ISTM	3.889423e+07
0	Base Method	5.648564e+07
8	Prophet	3.758095e+08

Using XGB, along with Trend as a Feature

We took the trend as a feature in the XGB model, but the results could not capture the trend.



Testing The Best Models Between 2015-2019

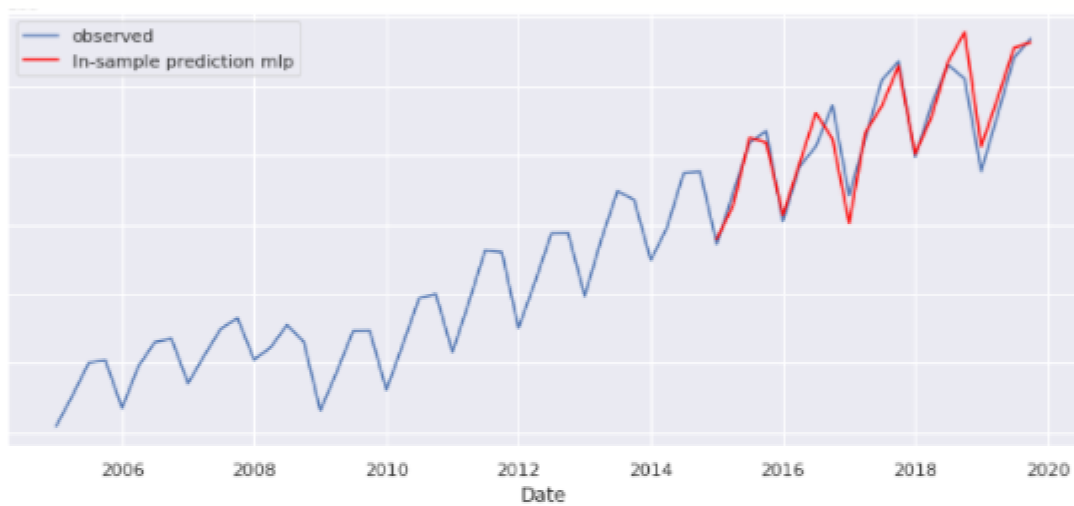
After the conversations with Akportföy, we decided to create a model with good results between 2015 and 2019. We took data before the year 2020, as the business side considered the effects of pandemics as a shock.

We considered 10 years as a length of our training data for the models we use two main approaches which are sliding and expanding. For expanding technique, at the beginning of every year, we took the previous ten years as our training data and expanded our history with these years results for in sample and expanded history with predictions for out of sample. After that, at the beginning of the new year this process is repeated. When it comes to sliding, we kept our history constant at ten years with every new period added to our history and the first period in our history removed from the history.

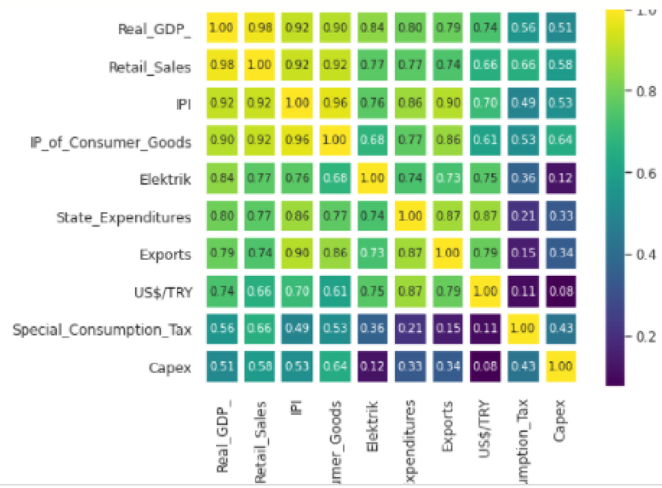
For the MLP model we basically use two approaches which are in sample and in sample expanding. With in sample technique, we train our model with all data before 2015 and the test data between 2015 and 2019. In sample expanding, we added every new period to our training data to be able to expand our training data.

	method	RMSE
3	MLP	1.394692e+07
7	sarima normal	1.492462e+07
6	sarima sliding	1.508496e+07
5	sarima sliding and expanding	1.723405e+07
1	in sample sliding	1.897315e+07
4	in sample mlp expanding	1.918179e+07
0	in sample sliding and expanding	2.359773e+07
2	out of sample prediction with uptading with r...	2.516972e+07

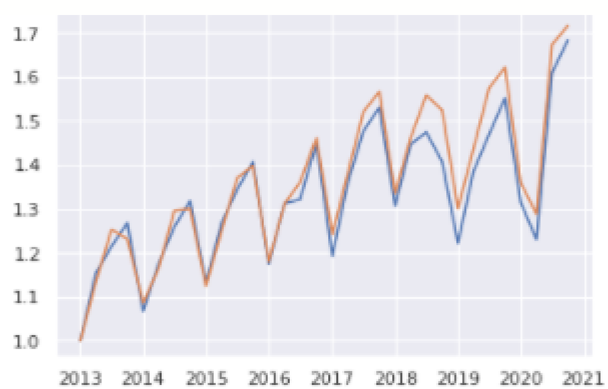
The MLP model gives us the best result among all our results, as we can see above. We can also see below that our best model captures trend and seasonality fairly well.



Exploratory Data Analysis For Features That are Correlated with GDP



These 10 parameters have a considerably high correlation with GDP.We can see from the graph below that retail sales and gdp have quite similar trends and seasonality.



Granger Causality

We perform granger causality test to find features which are useful(has casualty) to forecast GDP..These 11 features have a value below 0.01:

column x	column y	sinficance
Retail Sales_x	Real GDP _y	0
IP of Consumer Goods_x	Real GDP _y	0
Special Consumption Tax_x	Real GDP _y	0,0005
State Expenditures_x	Real GDP _y	0,0002
House sales_x	Real GDP _y	0
IPI_x	Real GDP _y	0,0001
KKO _x	Real GDP _y	0,0001
Elektrik_x	Real GDP _y	0,0024
export US$'000s_x	Real GDP _y	0,0001
import US$'000s_x	Real GDP _y	0
Exports_x	Real GDP _y	0,0061

References

- Composing High-Frequency Financial Conditions Index and Implications for Economic Activity, Abdullah Kazdal, Halil İbrahim Korkmaz, Muhammed Hasan Yılmaz, 2019 https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Paperss/2019/19-26 (Accessed 1 Sep 2021)

- Nowcasting Turkish GDP Growth withTargeted Predictors: Fill in the Blanks Mahmut Günay, 2020 https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Paperss/2020/20-06 (Accessed 1 Sep 2021)

- Nihai Yurt İçi Talep Kısa Dönemli Tahminleri, Mahmut Günay https://www.tcmb.gov.tr/wps/wcm/connect/24b30e1d-aecf-4161-9709-a247ba220d42/en1907eng.

pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-24b30e1d-aecf-4161-9709-a247ba220d42-mGmuRH0 (Accessed 1 Sep 2021)

- Nowcasting Turkish GDP with MIDAS: Role of Functional Form of the Lag Polynomial, Mahmut Günay, 2020 https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Paperss/2020/20-02 (Accessed 1 Sep 2021)
- Weekly Economic Conditions Index for Turkey, Aysu Çelgin, Mahmut Günay, 2020 https://www.tcmb.gov.tr/wps/wcm/connect/e35e43c4-45d6-4a89-88bf-2959cb524dbd/en2018eng.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-e35e43c4-45d6-4a89-88bf-2959cb524dbd-nnADSFk (Accessed 1 Sep 2021)

Libraries

- Pandas
- Numpy
- Seaborn
- Matplotlib
- Statsmodels
- Sklearn
- Keras