Univariate Time Series

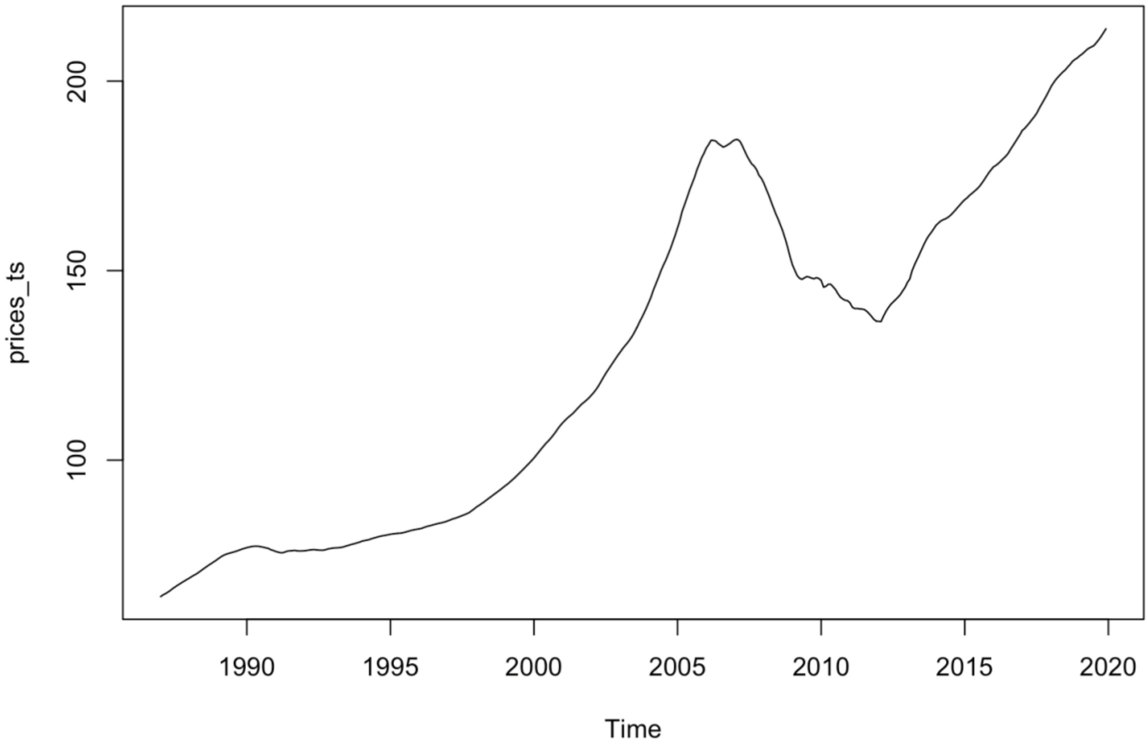
o Data source: https://fred.stlouisfed.org/series/CSUSHPISA
o Period considered in the analysis: January 1987 – latest data
o Frequency: monthly data

Using the data above

1. Forecast S&P/Case-Shiller U.S. National Home Price Index using an ARMA model.
2. Implement the Augmented Dickey-Fuller Test for checking the existence of a unit root in Case-Shiller Index series.
3. Implement an ARIMA(p, d, q) model. Determine p, d, q using Information Criterion or Box-Jenkins methodology. Comment the results.
4. Forecast the future evolution of Case-Shiller Index using the ARMA model. Test model using in-sample forecasts.
5. Research and suggest types of exogenous variables that can improve forecasts. In your references, indicate four (4) research articles or books at minimum.

Forecasting Home Price Index with ARIMA model

```
# reading data
data <- read_excel("CSUSHPISA.xls", col_types = c("date", "numeric"), skip = 10)

# constructing timeseries
prices_ts <- ts(data$CSUSHPISA, start = c(year(startd), month(startd)), frequency = 12)

# plotting time series shows a distinct trend present in the data
plot(prices_ts)
```



 # this is confirmed by the Augmented Dickey-Fuller Test
 # (for hypothesis of unit-root against alternative of stationarity)
 # result is -3.1079 with p-value = 0.1093:
 # cannot reject the hypothesis on unit root, the series is non-stationary
adf.test(prices_ts)

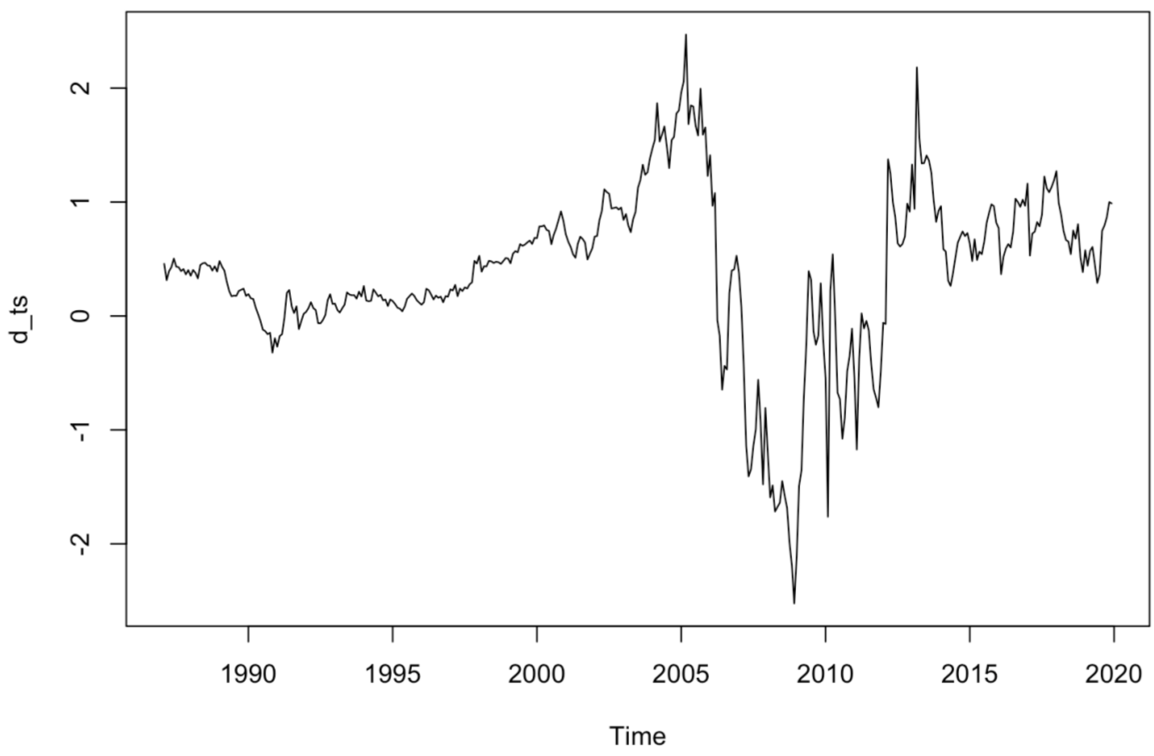
```
> adf.test(prices_ts)

        Augmented Dickey-Fuller Test

data:  prices_ts
Dickey-Fuller = -3.1079, Lag order = 7, p-value = 0.1093
alternative hypothesis: stationary
```

 # using kpss test the hypothesis is trend-stationarity
#
 # the test yields the following values:
 # KPSS Trend = 0.397, Truncation lag parameter = 5, p-value = 0.01
#
 # hence we cannot reject the hypothesis of trend-stationary
#
kpss.test(prices_ts, null = "Trend")

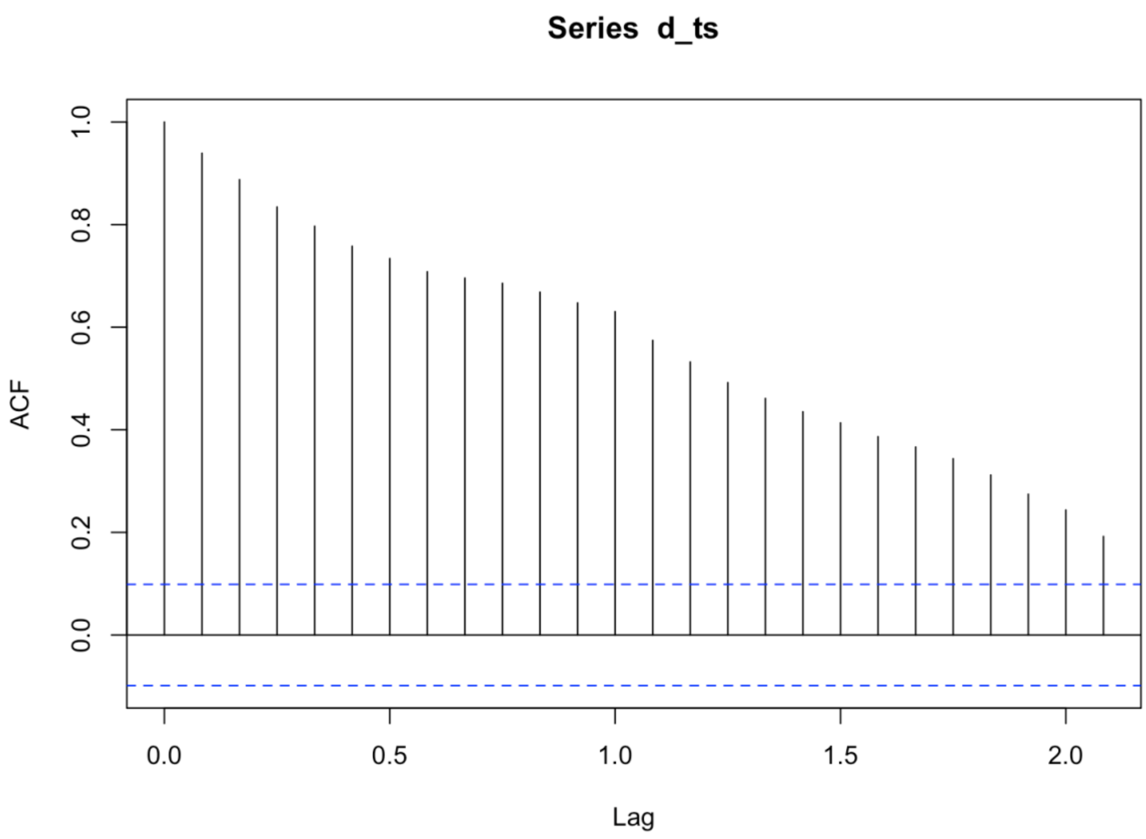
```
> kpss.test(prices_ts, null = "Trend")

	KPSS Test for Trend Stationarity

data:  prices_ts
KPSS Trend = 0.397, Truncation lag parameter = 5, p-value = 0.01
```

 # as neither unit root existence nor trend-stationarity is rejected
 # it means the data is not very informative or the tests are affected by outliers
 # so we elect to difference the data once
d_ts <- diff(prices_ts)

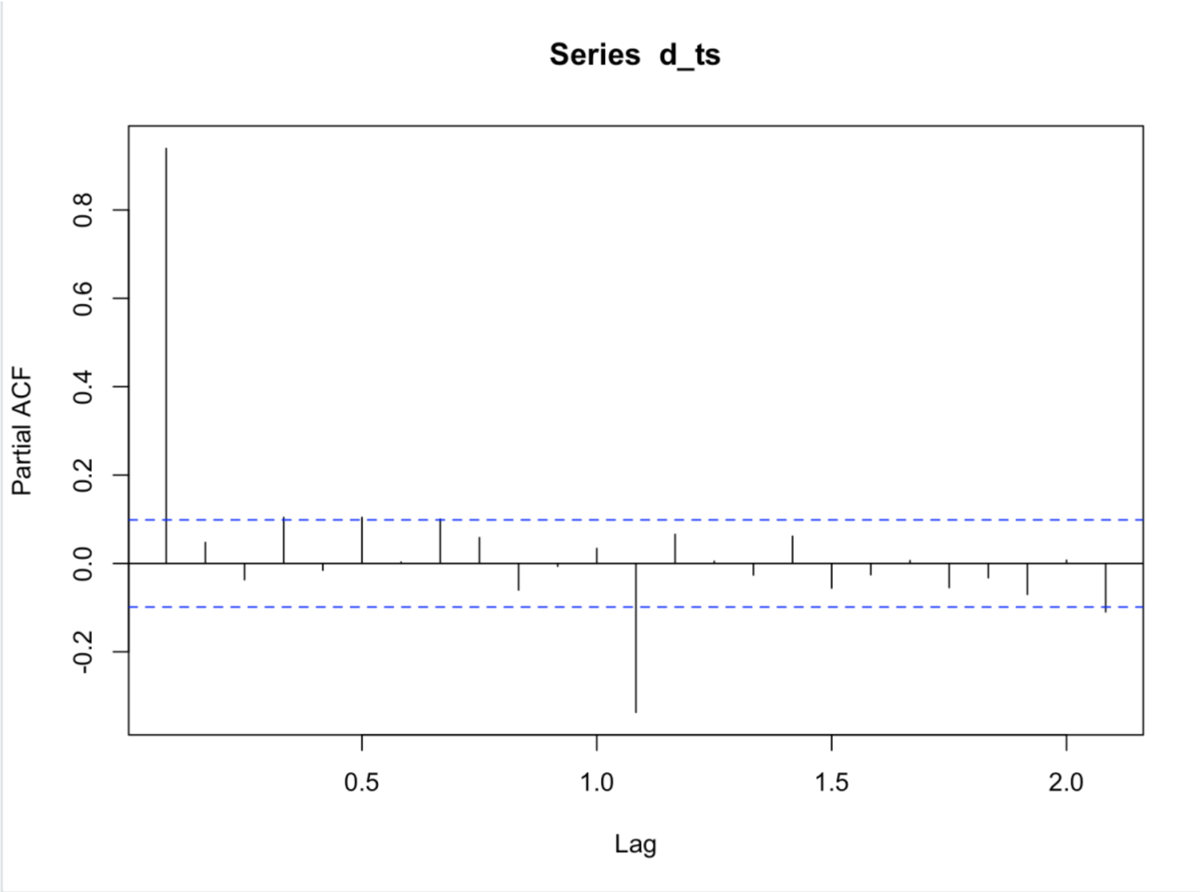
 # differenced time series still showing sign of being non-stationary
 # however, this time it is less so because of a trend in the data
 # and more so because of the heteroscedasticity (variable volatility) in the data.
plot(d_ts)



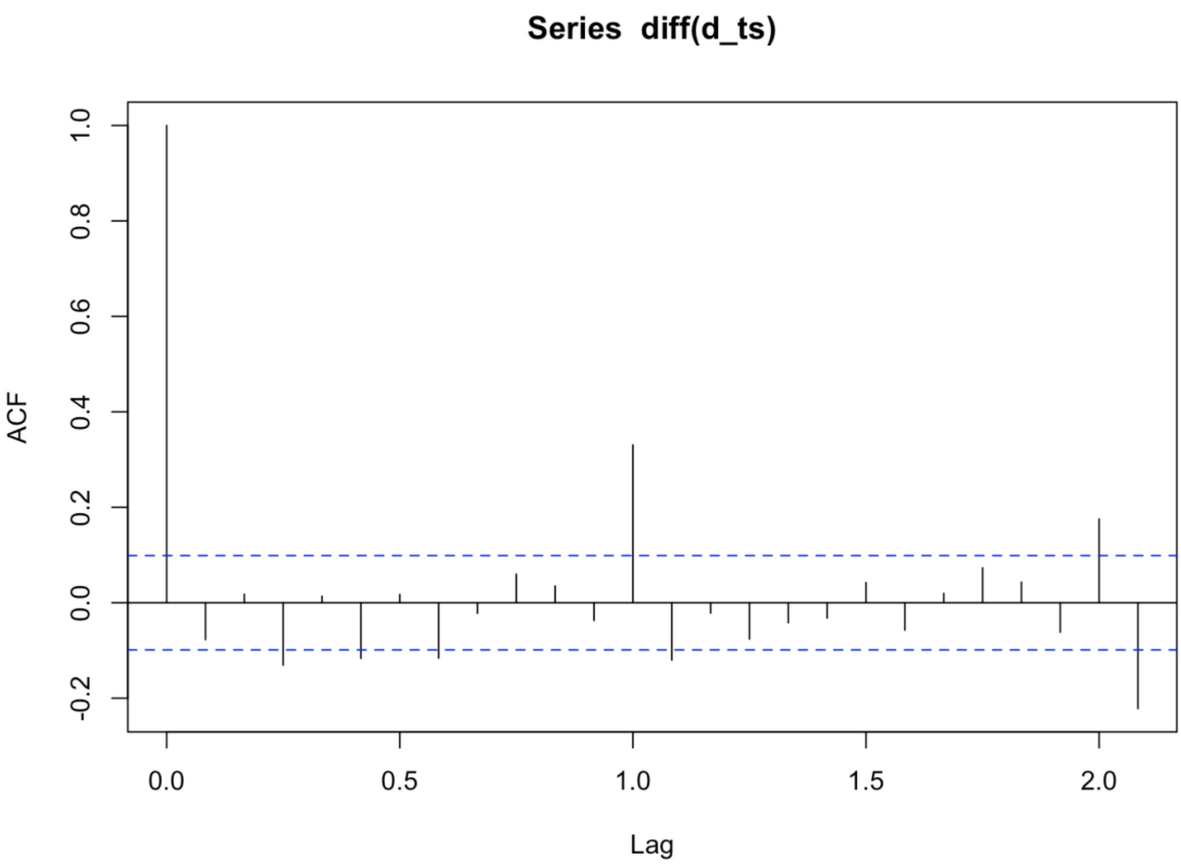
 # shows multiple gradually reducing self-correlations at all lags
acf(d_ts)




```
# pacf shows one significant spike at 1 and 13 months, suggesting seasonality and AR-Signature
# also shows less significant spikes at periods 3, 5 and 7
# this prompts an arima model with p = 7
pacf(d_ts)
```



 # There is a potential to differentiate second time but it will start to show signs of being over-differenced if done so (ACF changes sign back and forth a lot)
 # while those signs are mild, it's good to use minimum to use the minimum differencing that achieves mean-reversion-like properties
acf(diff(d_ts))



 # with above in mind the choice of the parameters leads us to a model (7,1,0) with seasonal order of 2
 # (taking into account two previous years for seasonality adjustment)

```
arima_710 <- arima(prices_ts,
              order = c(7,1,0),
              seasonal = list(order = c(2,0,0), period = 12))
```

 # Ljung-Box test output doesn't allow to reject the null hypothesis of autocorrelation in model residuals
 # Q* = 16.364, df = 15, p-value = 0.3583

```
checkresiduals(arima_710)
```

 # AIC information criterion value is -9.969872
 # experimenting with other model parameters shows this is a reasonably good value (verified by experiments with other model parameters)

```
AIC(arima_710)
```

```
> AIC(arima_710)
[1] -9.969872
```

 # with this we conclude the overall adequacy of the model built

 # the forecasts from the model
fc <- forecast(arima_710)

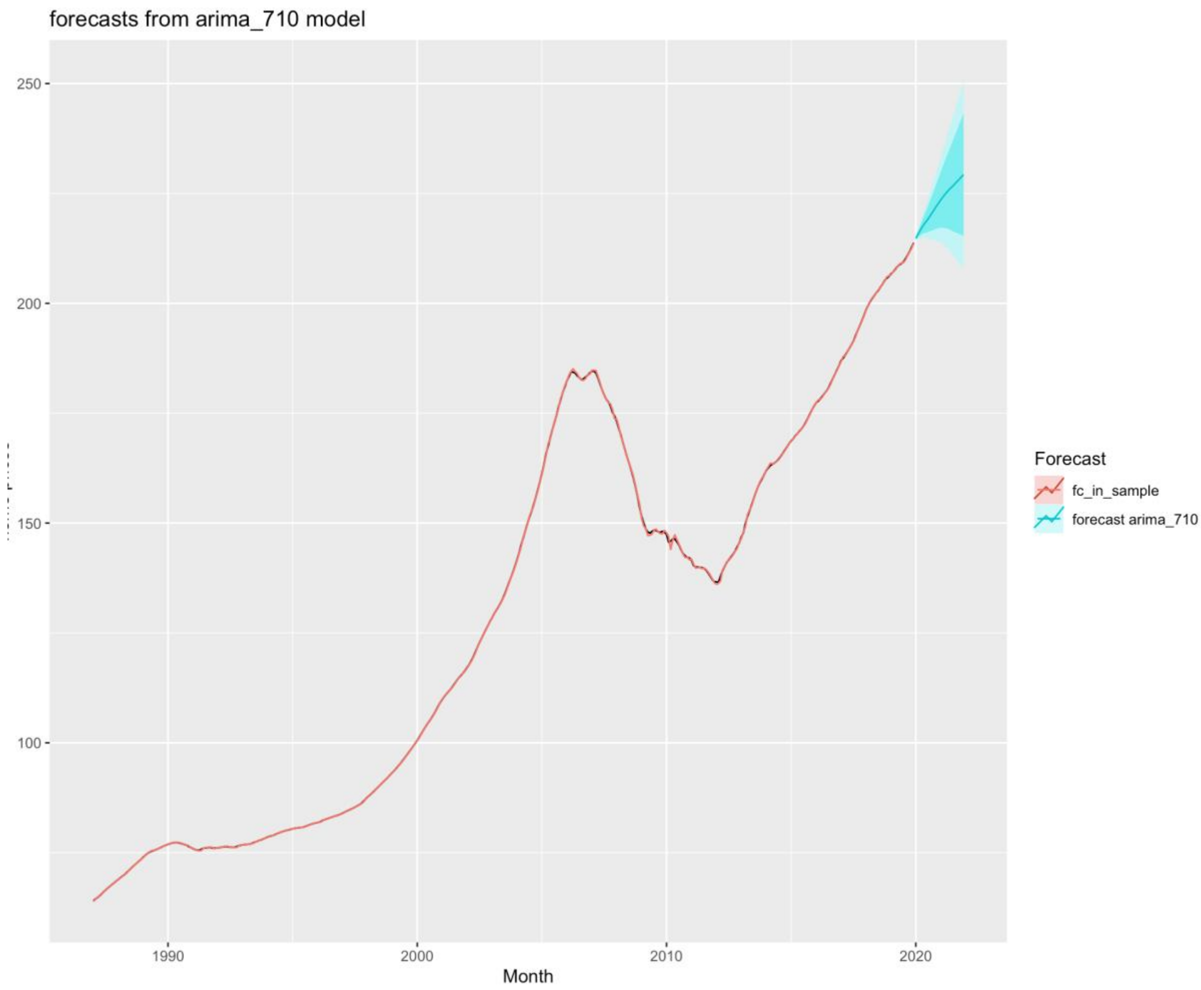
 # these are in-sample predictions
fc_in_sample <- fitted(arima_710)

 # in_sample accuracy of the model is
 # ME RMSE MAE MPE MAPE MASE ACF1
#0.01565473 0.2314025 0.1385282 0.01440401 0.1012788 0.2127796 -0.003648528
accuracy(arima_710)

```
> accuracy(arima_710)
                     ME      RMSE       MAE        MPE      MAPE      MASE         ACF1
Training set 0.01565473 0.2314025 0.1385282 0.01440401 0.1012788 0.2127796 -0.003648528
```

 # plot original series, fitted values (in-sample predictions)
 # and forecasted values with ranges

```
autoplot(prices_ts) +
  autolayer(fc, series="forecast arima_710") +
  autolayer(fc_in_sample) +
  ggtitle("forecasts from arima_710 model") +
  xlab("Month") + ylab("home prices") +
  guides(colour=guide_legend(title="Forecast"))
```



Exogenous variables to improve forecast

These are the exogenous variables which can add value to analysis (each of them need to be investigated empirically in the model):

- Political regimes
- Rent price ratio
- Inflation
- GDP growth
- Tax rate

References

1. Leblanc, Matthieu and Bokreta, Rachid, Analysis of the US Real Estate Market: Time-Varying Estimation and Forecast of the S&P Case-Shiller Composite 20 Cities (October 10, 2009. [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1486682 (accessed on 29.03.2020)
2. Huang, Daisy J. and Leung, Charles Ka Yui and Tse,Chung-Yi, What account for the differences in rent-price ratio and turnover rate? A search-and-matching approach

(2017). [Online] Available at: https://mpra.ub.uni-muenchen.de/76864/1/MPRA_paper_76864.pdf (accessed on 29.03.2020)

3. Banerjee, A., Marcellino, M., & Masten, I. Leading indicators for Euro-area inflation and GDP growth. Oxford Bulletin of Economics and Statistics (2005). [Online] Available at: http://didattica.unibocconi.it/mypage/dwload.php?nomefile=Leading_indicators_for_Euro-area_Inflation_and_GDP_growth20130917123939.pdf (accessed on 29.03.2020)
4. Brian Romanchuk, Primer: Exogenous Versus Endogenous Variables [Online] Available at: http://www.bondeconomics.com/2014/02/primer-exogenous-versus-endogenous.html (accessed on 29.03.2020)