Test Exercise 2

Maria Korzun

12 March 2017

Introduction

This research is devoted to analysis of data provided by US Bureau of the Census.

The data is aggregate (for the whole of United States) annual observations for the period 1959 - 2003 on income, 20 categories of consumer expenditure and price index series for these categories. The income and expenditure variables are all measured in $ billion at 2000 constant prices. The price index series are all based with 2000 = 100.

We will provide detail analysis of expenditures on admissions to specified spectator amusements (ADM). It is very interesting to track changes in this type of data because expenditures on entertainment do not apply to essential goods. We can assume that such type of expenditures grows when income increases and vice versa.

Part 1

At the first part of the research we will analyse the series visually, comment on their basic statistics, characterise the period of time for which the series is fixed (what events occurred at various intervals of time and how they could affect the behavior of the series), put forward some hypotheses about what to expect from them.

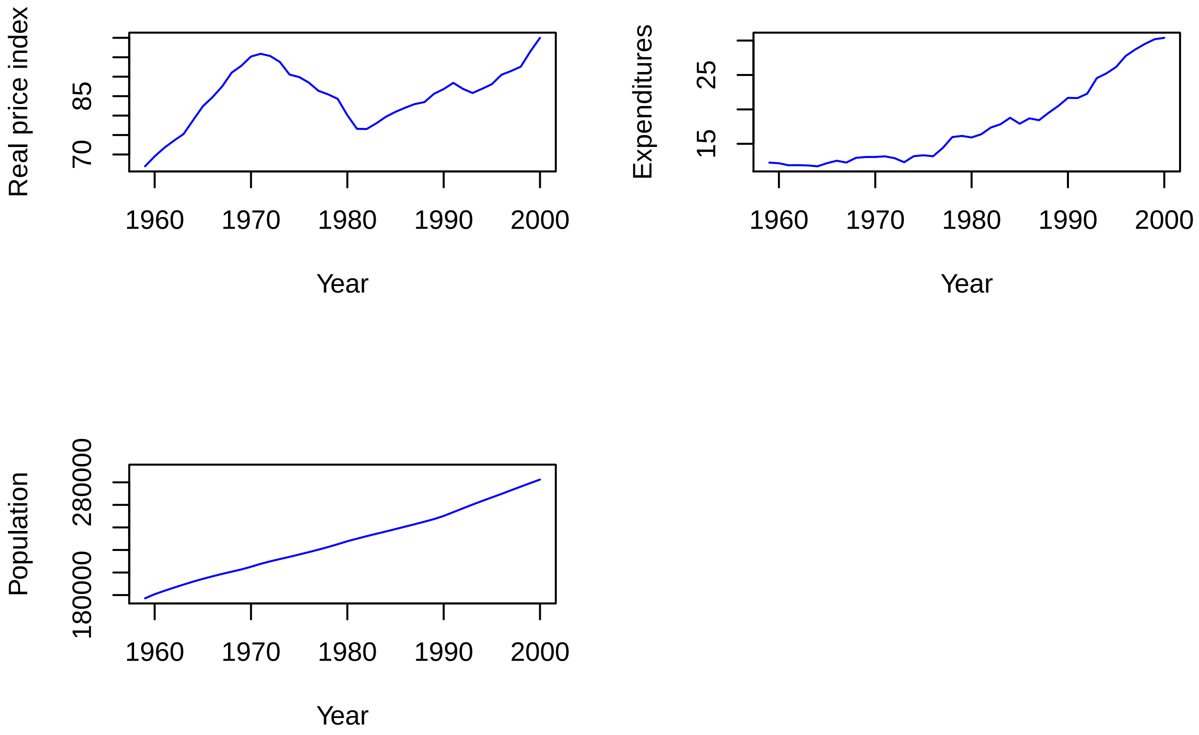
The nominal price index series for the admissions to specified spectator amusements have the name of the category $PADM$. The data set includes the nominal price index for total personal expenditure $PTPE$. In the regressions, economic theory (and common sence) suggests that one should use real price indices rather than nominal ones in regression analysis, where a real price index is defined relative to general inflation measured py $PTPE$. The real price index for admissions to specified spectator amusements, $PRELADM$, is defined as: $PRELADM$=100*($PADM/PTPE$).

Next step is visual analysis.Graphs below show that real price index does not have constant trend. The real price index grew till 1970 and then decreased from 1970 to 1980. It may be connected with 1973 oil crisis when the members of the Organization of Arab Petroleum Exporting Countries proclaimed an oil embargo.

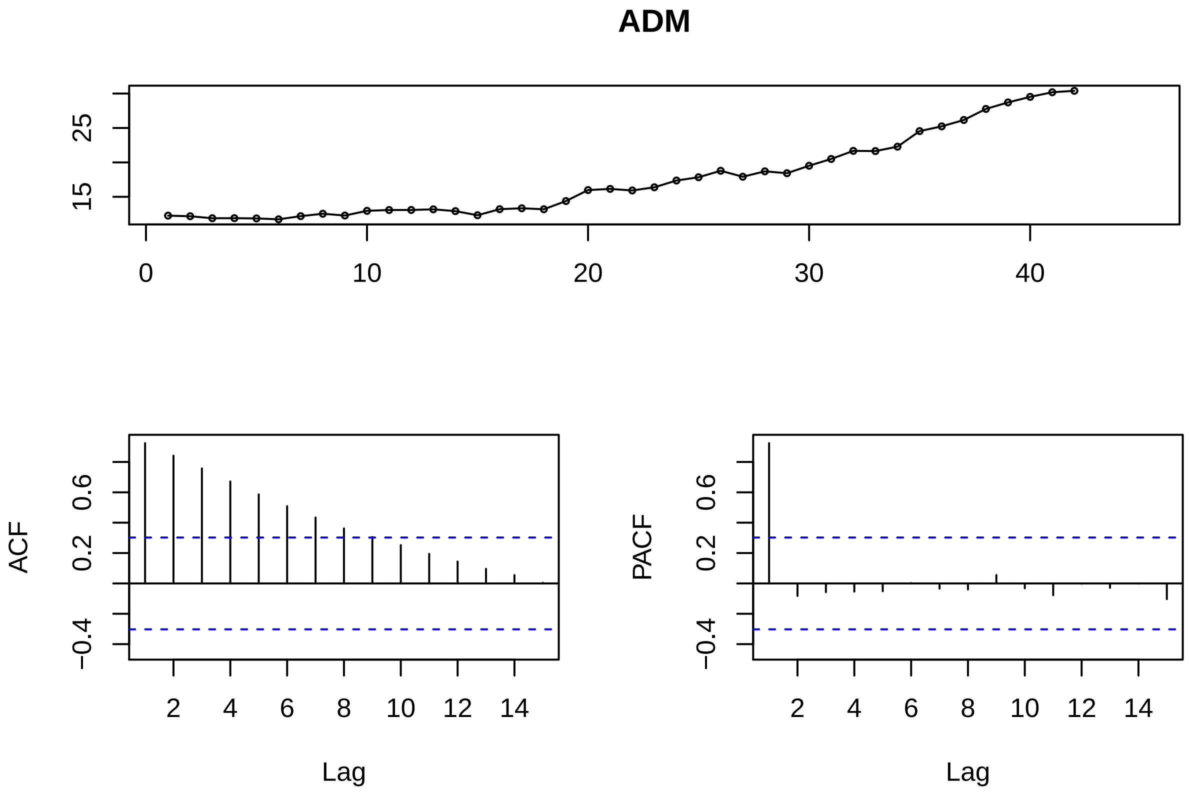
Amount of expenditures grows up over time (you also may pay attention that at the beginning of oil crisis expenditures slightly decreased. What is more, overall trend of expenditures growth may be related with growth of population.

Part 2

In this part we will make a dynamic model of demand for goods.



You can see three plots. First plot shows changes of ADF over time. This plot carries the same information as the graph higher (entertainment~year): the data set also includes trend variable $TIME$ that is defined to be 1 for 1959, 2 for 1960 and so on. The graphs of autoregressive function and the partial autocorrelation fuction show the non-stationarity: only first lag is significant in PACF, effects of other lags are close to zero.To sumup, ACF decreases extremely slowly, and PACF breaks off after the first partial correlation, then, probably, we have a random walk before us.

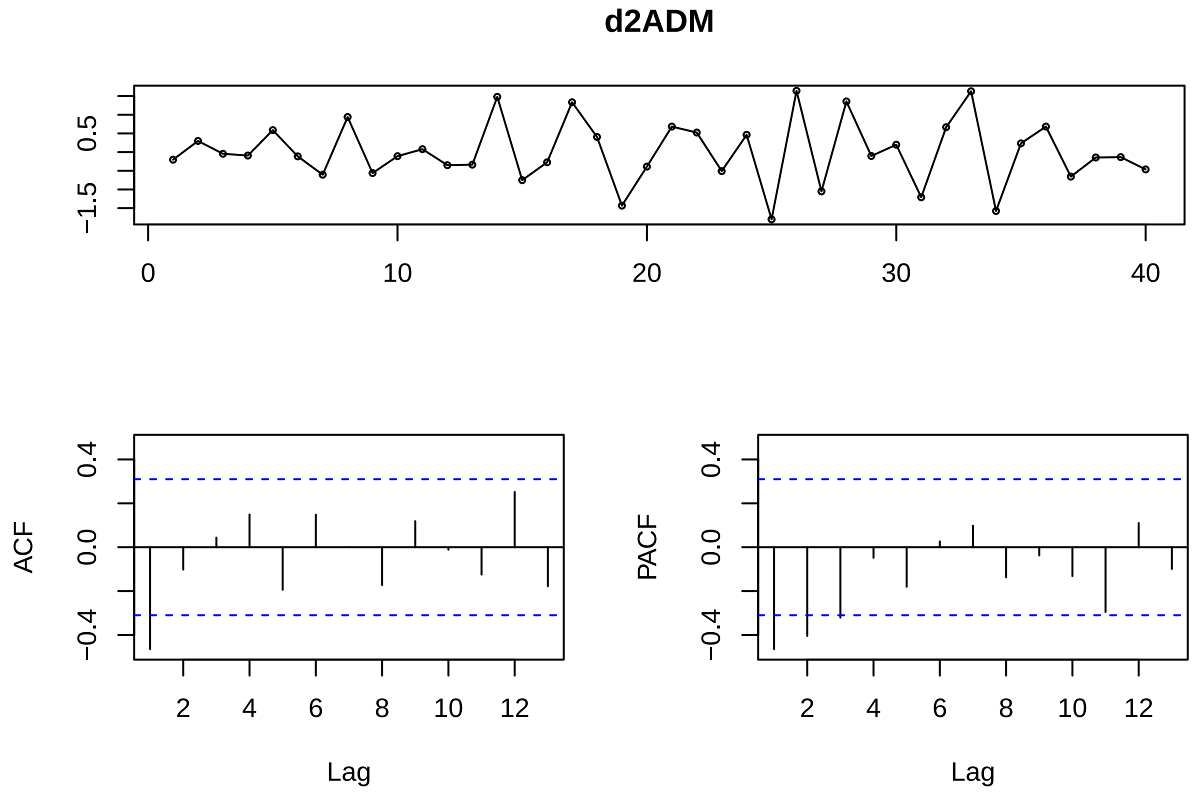


To proof the visual non-stationarity we need to make the Dickey-Fuller test. Choose 2 lags because we have only 42 observations.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ADM
## Dickey-Fuller = -0.66625, Lag order = 2, p-value = 0.9652
## alternative hypothesis: stationary
```

P-value = 0.9652 > 0.05 -> we cannot reject Ho: non-stationarity -> we proved our visual assumption.

Now we should make our time series stationary: differentiate it twice.



```
##
##  Augmented Dickey-Fuller Test
##
## data:  d2ADM
## Dickey-Fuller = -6.4534, Lag order = 2, p-value = 0.01
## alternative hypothesis: stationary
```

Graphs for differentiated time series show stationarity and the Dickey-Fuller test too (p-value = 0.01 < 0.05 -> we cannot accept Ho: non-stationarity).

We need to check if other variables are stationary. They are non-stationary so we make them stationary too. You can see the before and after results of the Dickey-Fuller test below.

```
##
##  Augmented Dickey-Fuller Test
##
## data:  TPE
## Dickey-Fuller = 1.4124, Lag order = 3, p-value = 0.99
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  d2TPE
## Dickey-Fuller = -4.6209, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  PRELADM
## Dickey-Fuller = -2.3795, Lag order = 3, p-value = 0.4241
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  d2PRELADM
## Dickey-Fuller = -4.0747, Lag order = 3, p-value = 0.01687
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  FLOW
## Dickey-Fuller = -0.29353, Lag order = 3, p-value = 0.9862
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  d2FLOW
## Dickey-Fuller = -4.1464, Lag order = 3, p-value = 0.01408
## alternative hypothesis: stationary
```

We made full ADL model and reject from it unsignificant variables. Final ADL model below: p-value < 0.05 -> we can reject the hypothesis of equality of the coefficients to zero.

```
##
## Call:
## lm(formula = d2ADM ~ lag(d2ADM, n = 1L) + lag(d2PRELADM, n = 1L) +
##     d2FLOW + lag(d2FLOW, n = 1L))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41624 -0.55925 -0.00197  0.37937  1.32462
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.02098    0.11293   0.186  0.85376
## lag(d2ADM, n = 1)     -0.57565    0.14526  -3.963  0.00036 ***
## lag(d2PRELADM, n = 1) -0.14266    0.07879  -1.811  0.07903 .
## d2FLOW                 0.59963    0.24465   2.451  0.01954 *
## lag(d2FLOW, n = 1)     0.25698    0.26620   0.965  0.34118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.704 on 34 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:   0.39,  Adjusted R-squared:  0.3183
## F-statistic: 5.436 on 4 and 34 DF,  p-value: 0.001708
```

Checking autocorrelation in residuals: p-value is very large so there are no autocorrelation.

```
##
## Call:
## lm(formula = res[-n] ~ res[-1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37926 -0.46672 -0.00032  0.32034  1.37355
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.00835    0.11015  -0.076    0.940
## res[-1]     -0.09304    0.16543  -0.562    0.577
##
## Residual standard error: 0.679 on 36 degrees of freedom
## Multiple R-squared:  0.008709,   Adjusted R-squared:  -0.01883
## F-statistic: 0.3163 on 1 and 36 DF,  p-value: 0.5773
```

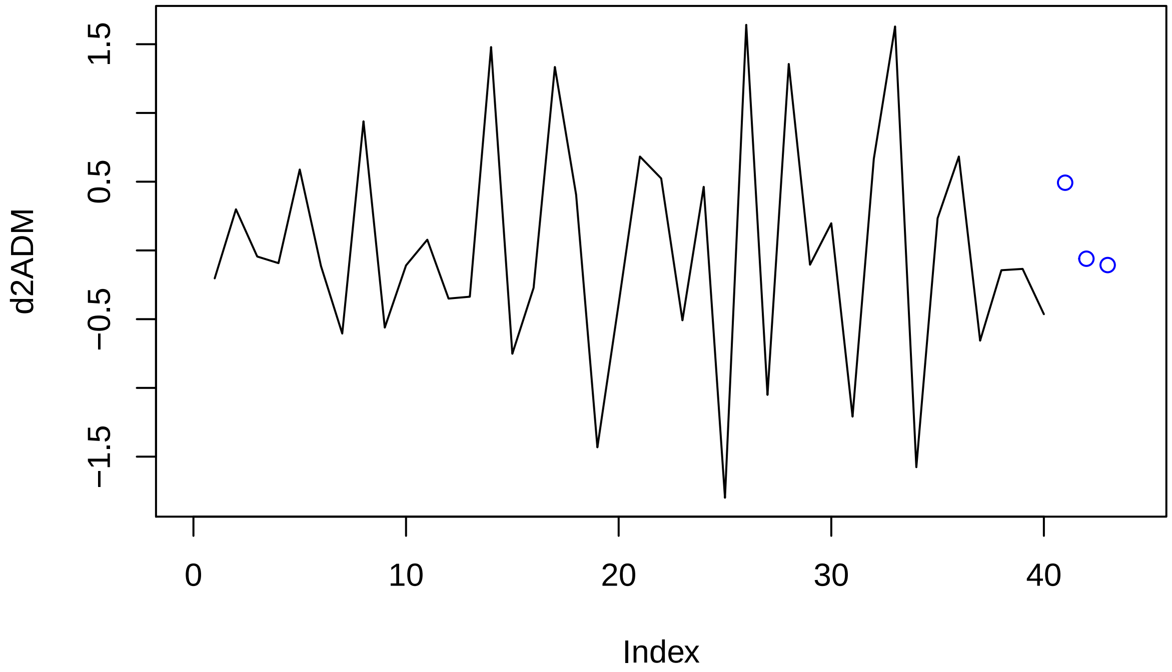
So our model will be: $d2ADM{=}lag(d2PRELADM,n{=}1L){+}d2FLOW{+}lag(d2FLOW,n{=}1L)$.

Part 3

In this part we build an ARIMA models for each time series and build three-step forward forecasts (missed observations for 2001-2003).

We have already carried out a series of stationary analysis and select a suitable stationary transformation in part 2.

```
## Series: d2ADM
## ARIMA(3,0,0) with zero mean
##
## Coefficients:
##           ar1      ar2      ar3
##       -0.7857  -0.6125  -0.3192
## s.e.   0.1493   0.1648   0.1445
##
## sigma^2 estimated as 0.4311:  log likelihood=-38.83
## AIC=85.66   AICc=86.8   BIC=92.42
##
## Training set error measures:
##                      ME     RMSE       MAE      MPE     MAPE     MASE
## Training set 0.04636958 0.631457 0.5044112 48.37579 118.6265 0.441773
##                     ACF1
## Training set -0.05738385
```



R-Studio chose automatically the best ARIMA(3,0,0) for *d*2*ADM* by the penalty criterion (the minimum value of the criterion of Akaike). This ARIMA model does not contains linear trend and constant.

So we got model for *d*2*ADM*: Yt=-0.7827Y$t-1$-0.6125Y$t-2$-0.3192Y$t-3$+Et.

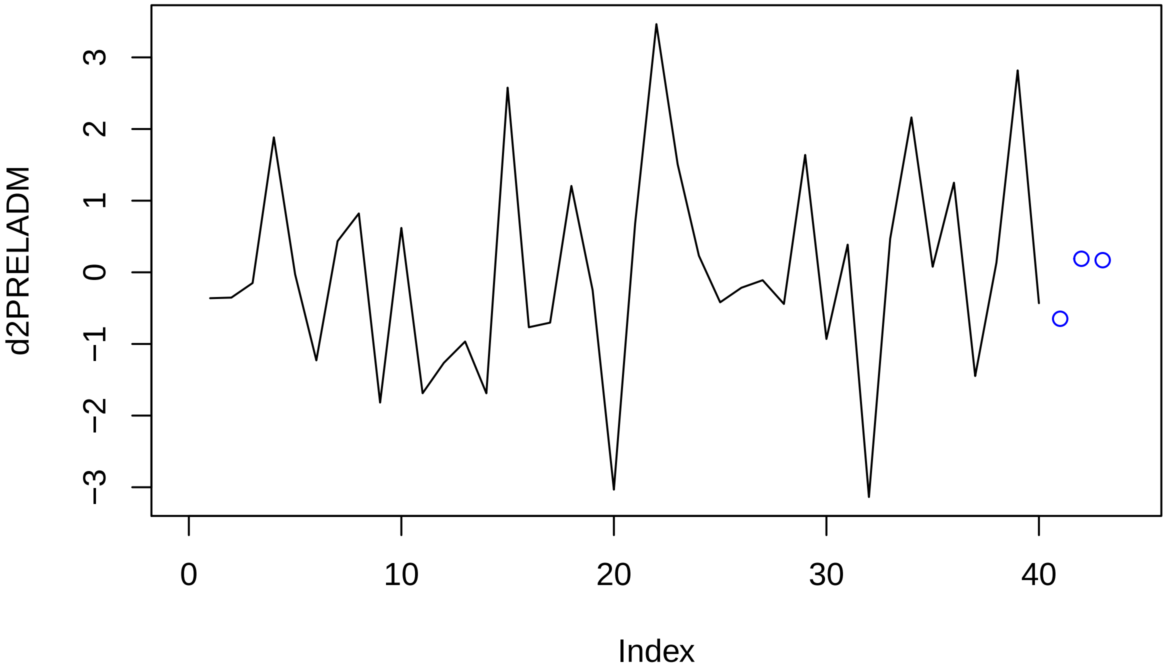
Now we can predict next three steps by two ways:

```
##    Point Forecast      Lo 80     Hi 80      Lo 95    Hi 95
## 41     0.49254432 -0.3488683 1.3339570 -0.7942853 1.779374
## 42    -0.06015051 -1.1302033 1.0099023 -1.6966550 1.576354
## 43    -0.10641127 -1.1764718 0.9636493 -1.7429277 1.530105
```

For *d*2*PRELADM auto.arima* gives (0,0,0) AIC = 144.25 so we will chose the best model ourselves. After long search - the best model was founded: ARIMA(2,0,0) AIC = 147.53.

```
##
## Call:
## arima(x = d2PRELADM, order = c(2, 0, 0))
##
## Coefficients:
##           ar1      ar2  intercept
##       -0.0888  -0.2518     0.0173
## s.e.   0.1521   0.1566     0.1649
##
## sigma^2 estimated as 1.91:  log likelihood = -69.77,  aic = 147.53
##
## Training set error measures:
```

```
##                         ME     RMSE      MAE      MPE     MAPE      MASE
## Training set -0.004740435 1.381961 1.075551 113.9184 127.8396 0.6049388
##                     ACF1
## Training set -0.01128554
```



Predictions:

```
##    Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 41     -0.6482034 -2.419258 1.122851 -3.356798 2.060391
## 42      0.1892303 -1.588796 1.967257 -2.530027 2.908488
## 43      0.1695339 -1.660211 1.999279 -2.628820 2.967888
```

Part 4

In this part we will make cointegration ratio between a number of consumption and a number of commodity prices, forecast the volume of consumption, depending on the price of the goods, three steps forward, Build an error correction model.

Making the cointegration test needs stationarity of time series and their same order.

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
```

```
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71079 -0.38618 -0.08895  0.49805  1.49118
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1     -2.0405     0.2618  -7.794 3.08e-09 ***
## z.diff.lag   0.4024     0.1533   2.625   0.0127 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7125 on 36 degrees of freedom
## Multiple R-squared:  0.7697, Adjusted R-squared:  0.7569
## F-statistic: 60.15 on 2 and 36 DF,  p-value: 3.331e-12
##
##
## Value of test-statistic is: -7.7935
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.62 -1.95 -1.61
```

P-value = -7.7935 < all critical values -> we cannot accept Ho: non-stationary residuals.

We proved cointegration of consumer expenditures and price index.

Conclusion

There are a lot of useful instruments fot time series data analysis. We studied to visualise time series data, make models for prediction.Gained skills will be very useful if we need to make time series data analysis in scientific researches or in business.