CASEPROJECT

2023-05-16

(a)

The table below summarizes the outcomes of four logit models to explain the direction of economic development (GDPIMPR) for the period 1951 to 2010. Perform three Likelihood Ratio tests to prove both the individual and the joint significance of the 1-quarter lags of li1 and li2, where the alternative hypothesis is always the model with both indicators included.

Dependent variable: GDPIMPR
Sample size: 240

Variable	Coeff	Coeff	Coeff	Coeff
Constant	0.693	0.812	0.636	0.729
li1(-1)	*x*	−0.340	*x*	−0.372
li2(-1)	*x*	*x*	−0.087	−0.120
Log likelihood	−152.763	−139.747	−149.521	−134.178

```
## Rows: 264 Columns: 8
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Date
## dbl (7): GDP, GDPIMPR, LOGGDP, GrowthRate, li1, li2, T
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
CG$Date <- as.yearqtr(CG$Date, format = "%Y/q%q")
CG <- mutate(CG, lagli1 = lag(CG$li1, n = 1))
CG <- mutate(CG, lagli2 = lag(CG$li2, n = 1))


CG1 <- CG["1951 Q1" <= CG$Date & CG$Date <= "2010 Q4",]
```

```
LOGIT1 <- glm(GDPIMPR~GDPIMPR, CG1, family = "binomial")

LOGIT2 <- glm(GDPIMPR~lagli1, CG1, family = "binomial")

LOGIT3 <- glm(GDPIMPR~lagli2, CG1, family = "binomial")

LOGIT4 <- glm(GDPIMPR~lagli1 + lagli2, CG1, family = "binomial")
```

```
LRT <- lrtest(LOGIT1, LOGIT2, LOGIT3, LOGIT4)
print(LRT)
```

```
## Likelihood ratio test
##
## Model 1: GDPIMPR ~ GDPIMPR
## Model 2: GDPIMPR ~ lagli1
## Model 3: GDPIMPR ~ lagli2
## Model 4: GDPIMPR ~ lagli1 + lagli2
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   1 -152.76
## 2   2 -139.75  1 26.034  3.355e-07 ***
## 3   2 -149.52  0 19.548  < 2.2e-16 ***
## 4   3 -134.18  1 30.684  3.036e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
LRT1<- lrtest(LOGIT3, LOGIT4)
LRT2<- lrtest(LOGIT2, LOGIT4)
LRT3<- lrtest(LOGIT1, LOGIT4)
print(LRT1)
```

```
## Likelihood ratio test
##
## Model 1: GDPIMPR ~ lagli2
## Model 2: GDPIMPR ~ lagli1 + lagli2
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   2 -149.52
## 2   3 -134.18  1 30.684  3.036e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(LRT2)
```

```
## Likelihood ratio test
##
## Model 1: GDPIMPR ~ lagli1
## Model 2: GDPIMPR ~ lagli1 + lagli2
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   2 -139.75
## 2   3 -134.18  1 11.137  0.0008464 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(LRT3)
```

```
## Likelihood ratio test
##
## Model 1: GDPIMPR ~ GDPIMPR
## Model 2: GDPIMPR ~ lagli1 + lagli2
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1   1 -152.76
## 2   3 -134.18  2 37.17  8.483e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

answer (a)

So, LR-test shows all models are significant at 1% level, but lagli2 model is significant at 5% level in this case.

(b)

It could be that the leading indicators lead the economy by more than 1 quarter. The table below summarizes outcomes of four logit models that differ in the lags of the indicators. For what reason can we use McFadden R2 to select the best lag structure among these four models? Compute the four values of McFadden R2 (with four decimals) and conclude which model is optimal according to this criterion.

Dependent variable: GDPIMPR
Sample size: 240

	1	2	3	4
Variable	Coeff	Coeff	Coeff	Coeff
Constant	0.729	0.731	0.746	0.749
li1(-1)	−0.372	−0.366	*x*	*x*
li1(-2)	*x*	*x*	−0.429	−0.421
li2(-1)	−0.120	*x*	−0.131	*x*
li2(-2)	*x*	−0.121	*x*	−0.129
Log likelihood	−134.178	−134.126	−130.346	−130.461

```
CG <- mutate(CG, lag2li1 = lag(CG$li1, n = 2))
CG <- mutate(CG, lag2li2 = lag(CG$li2, n = 2))
CG2 <- CG["1951 Q1" <= CG$Date & CG$Date <= "2010 Q4",]
llogit1 <- glm(GDPIMPR~lagli1+lagli2, CG2, family = "binomial")
llogit2 <- glm(GDPIMPR~lagli1+lag2li2, CG2, family = "binomial")
llogit3 <- glm(GDPIMPR~lag2li1+lagli2, CG2, family = "binomial")
llogit4 <- glm(GDPIMPR~lag2li1+lag2li2, CG2, family = "binomial")

LRT4 <- lrtest(llogit1, llogit2, llogit3, llogit4)
print(LRT4)
```

```
## Likelihood ratio test
##
## Model 1: GDPIMPR ~ lagli1 + lagli2
## Model 2: GDPIMPR ~ lagli1 + lag2li2
## Model 3: GDPIMPR ~ lag2li1 + lagli2
## Model 4: GDPIMPR ~ lag2li1 + lag2li2
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   3 -134.18
## 2   3 -134.13  0 0.1036  < 2.2e-16 ***
## 3   3 -130.34  0 7.5619  < 2.2e-16 ***
## 4   3 -130.46  0 0.2312  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
PseudoR2(llogit1, which = "McFadden")
```

```
##  McFadden
## 0.1216598
```

```
PseudoR2(llogit2, which = "McFadden")
```

```
##  McFadden
## 0.1219991
```

```
PseudoR2(llogit3, which = "McFadden")
```

```
##  McFadden
## 0.1467494
```

```
PseudoR2(llogit4, which = "McFadden")
```

```
##  McFadden
## 0.1459928
```

answer (b)

McFadden's Pseudo R^2 shows that $Constant + li1_{t-2} + li2_{t-1}$ model is the most optimal.

(c)

Use the logit model 3 of part (b) (with li1(-2) and li2(-1)) to calculate the predicted probability of economic growth for each of the 20 quarters of the evaluation sample. Assess the predictive performance by means of the prediction-realization table and the hit rate, using a cut-off value of 0.5. Evaluate the outcomes.

```
summary(llogit3)
```

```
##
## Call:
## glm(formula = GDPIMPR ~ lag2li1 + lagli2, family = "binomial",
##     data = CG2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2798  -0.9581   0.5610   0.7588   1.7679
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.74569    0.15733   4.740 2.14e-06 ***
## lag2li1     -0.42871    0.07632  -5.618 1.94e-08 ***
## lagli2      -0.13120    0.03859  -3.399 0.000675 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 305.53  on 239  degrees of freedom
## Residual deviance: 260.69  on 237  degrees of freedom
## AIC: 266.69
##
## Number of Fisher Scoring iterations: 4
```

```
CG <- mutate(CG, Pb = predict(llogit3, CG, type = "response"))
CG3 <- CG["2010 Q4" < CG$Date,]
CG3$Pb <- round(CG3$Pb, digits = 0)
print(data_frame(CG3$GDPIMPR, CG3$Pb))
```

```
## # A tibble: 20 × 2
##    `CG3$GDPIMPR` `CG3$Pb`
##            <dbl>    <dbl>
##  1             1        0
##  2             0        0
##  3             0        0
##  4             0        0
##  5             1        0
##  6             0        0
##  7             1        0
##  8             1        1
##  9             1        1
## 10             1        1
## 11             1        1
## 12             0        1
## 13             0        1
## 14             1        1
## 15             1        1
## 16             1        1
## 17             1        1
## 18             1        1
## 19             1        1
## 20             0        0
```

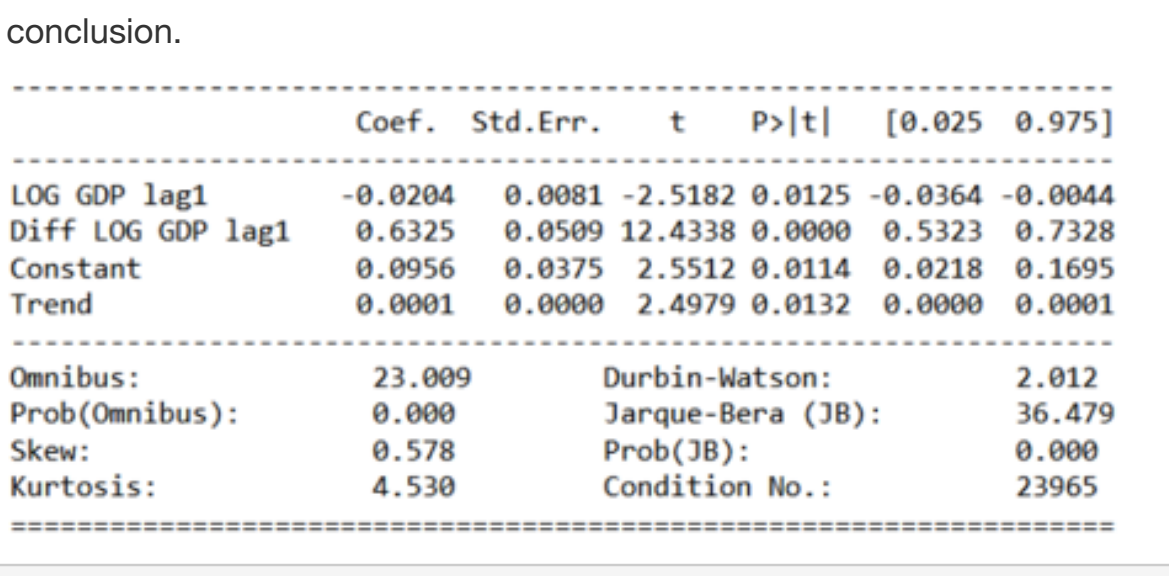
answer (c)

$h = \frac{1}{20}\Sigma_{i=1}^{20} w_i$. so, $h = \frac{1}{20}15 = 0.75$.

The hit rate of the model is 0.75 that was correct 75% of the time.

(d)

Perform the Augmented Dickey-Fuller test on LOGGDP to confirm that this variable is not stationary. Use only the data in the estimation sample and include constant, trend, and a single lag in the test equation (L = 1, see Lecture 6.4). Present the coefficients of the test regression and the relevant test statistic, and state your conclusion.



```
------------------------------------------------------------------
                    Coef.  Std.Err.    t    P>|t|   [0.025  0.975]
------------------------------------------------------------------
LOG GDP lag1       -0.0204   0.0081 -2.5182 0.0125 -0.0364 -0.0044
Diff LOG GDP lag1   0.6325   0.0509 12.4338 0.0000  0.5323  0.7328
Constant            0.0956   0.0375  2.5512 0.0114  0.0218  0.1695
Trend               0.0001   0.0000  2.4979 0.0132  0.0000  0.0001
------------------------------------------------------------------
Omnibus:             23.009       Durbin-Watson:          2.012
Prob(Omnibus):       0.000        Jarque-Bera (JB):       36.479
Skew:                0.578        Prob(JB):               0.000
Kurtosis:            4.530        Condition No.:          23965
==================================================================
```

```
tseries::adf.test(CG$LOGGDP, k = 1)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  CG$LOGGDP
## Dickey-Fuller = -3.0557, Lag order = 1, p-value = 0.1313
## alternative hypothesis: stationary
```

answer (d)

If T-value is smaller than 3.5, we can reject H_0 (in 5% level). It is same that y_t is non-stationary. In this case, T-value = −2.5182, so we can reject H_0. Therefore, this time series is non- stationary.

(e)

Consider the following model: $GrowthRate_t = \alpha + \rho GrowthRate_{t-1} + \beta_1 li1_{t-k_1} + \beta_2 li2_{t-k_2} + \varepsilon_t$. Here the numbers k_1 and k_2 denote the lag orders of the leading indicators. Estimate four versions of this model on the estimation sample from 1951 to 2010, by setting k_1 and k_2 equal to either 1 or 2. Show that the model with $k_1 = k_2 = 1$ gives the largest value for R^2 , and present the four coefficients of this model in six decimals.

answer (e)

```
CG <- mutate(CG, lagGR = lag(CG$GrowthRate, n = 1))
CG5 <- CG["1951 Q1" <= CG$Date & CG$Date <= "2010 Q4",]
```

If $k_1 = k_2 = 1$,

```
LM1 <- lm(CG5$GrowthRate~CG5$lagGR + CG5$lagli1 + CG5$lagli2)
summary(LM1)
```

```
##
## Call:
## lm(formula = CG5$GrowthRate ~ CG5$lagGR + CG5$lagli1 + CG5$lagli2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0097644 -0.0028559 -0.0001994  0.0026151  0.0158224
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.737e-03  3.197e-04   5.433 1.38e-07 ***
## CG5$lagGR    4.616e-01  4.830e-02   9.556  < 2e-16 ***
## CG5$lagli1  -1.023e-03  1.298e-04  -7.880 1.20e-13 ***
## CG5$lagli2  -1.494e-04  6.421e-05  -2.326   0.0209 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004106 on 236 degrees of freedom
## Multiple R-squared:  0.508,  Adjusted R-squared:  0.5017
## F-statistic: 81.22 on 3 and 236 DF,  p-value: < 2.2e-16
```

$k_1 = 2, k_2 = 1,$

```
LM2 <- lm(CG5$GrowthRate~CG5$lagGR + CG5$lag2li1 + CG5$lagli2)
summary(LM2)
```

```
##
## Call:
## lm(formula = CG5$GrowthRate ~ CG5$lagGR + CG5$lag2li1 + CG5$lagli2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.009053 -0.002661 -0.000229  0.002342  0.016491
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.894e-03  3.376e-04   5.611 5.60e-08 ***
## CG5$lagGR    4.109e-01  5.445e-02   7.546 9.70e-13 ***
## CG5$lag2li1 -9.871e-04  1.479e-04  -6.674 1.76e-10 ***
## CG5$lagli2  -1.585e-04  6.675e-05  -2.374   0.0184 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004233 on 236 degrees of freedom
## Multiple R-squared:  0.4772, Adjusted R-squared:  0.4705
## F-statistic:  71.8 on 3 and 236 DF,  p-value: < 2.2e-16
```

$k_1 = 1, k_2 = 2,$

```
LM3 <- lm(CG5$GrowthRate~CG5$lagGR + CG5$lagli1 + CG5$lag2li2)
summary(LM3)
```

```
##
## Call:
## lm(formula = CG5$GrowthRate ~ CG5$lagGR + CG5$lagli1 + CG5$lag2li2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.009785 -0.002792 -0.000221  0.002526  0.015915
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.754e-03  3.188e-04   5.501 9.79e-08 ***
## CG5$lagGR    4.578e-01  4.845e-02   9.449  < 2e-16 ***
## CG5$lagli1  -1.018e-03  1.296e-04  -7.860 1.37e-13 ***
## CG5$lag2li2 -1.471e-04  6.416e-05  -2.293   0.0227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004107 on 236 degrees of freedom
## Multiple R-squared:  0.5077, Adjusted R-squared:  0.5014
## F-statistic: 81.12 on 3 and 236 DF,  p-value: < 2.2e-16
```

$k_1 = k_2 = 2,$

```
LM4 <- lm(CG5$GrowthRate~CG5$lagGR + CG5$lag2li1 + CG5$lag2li2)
summary(LM4)
```

```
##
## Call:
## lm(formula = CG5$GrowthRate ~ CG5$lagGR + CG5$lag2li1 + CG5$lag2li2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0091209 -0.0026155 -0.0002767  0.0023941  0.0161471
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.911e-03  3.369e-04   5.673 4.10e-08 ***
## CG5$lagGR    4.069e-01  5.466e-02   7.444 1.82e-12 ***
## CG5$lag2li1 -9.833e-04  1.476e-04  -6.662 1.88e-10 ***
## CG5$lag2li2 -1.579e-04  6.667e-05  -2.368   0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004233 on 236 degrees of freedom
## Multiple R-squared:  0.4771, Adjusted R-squared:  0.4705
## F-statistic: 71.78 on 3 and 236 DF,  p-value: < 2.2e-16
```

The highest R^2 value Model is $k_1 = k_2 = 1$ model. It is 0.508.

(f)

Perform the Breusch-Godfrey test for first-order residual serial correlation for the model in part (e) with $k_1 = k_2 = 1$. Does the test outcome signal misspecification of the model?

answer (f)

```
bgtest(CG5$GrowthRate~CG5$lagGR + CG5$lagli1 + CG5$lagli2)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  CG5$GrowthRate ~ CG5$lagGR + CG5$lagli1 + CG5$lagli2
## LM test = 0.23037, df = 1, p-value = 0.6313
```

The P-value is 0.6613. That means it is not significant at a 5%-Level. So, we cannot reject the H_0 and can conclude that there is no no residual serial correlation. But, this P-value might be signal misspecification of the model cause it is so high.

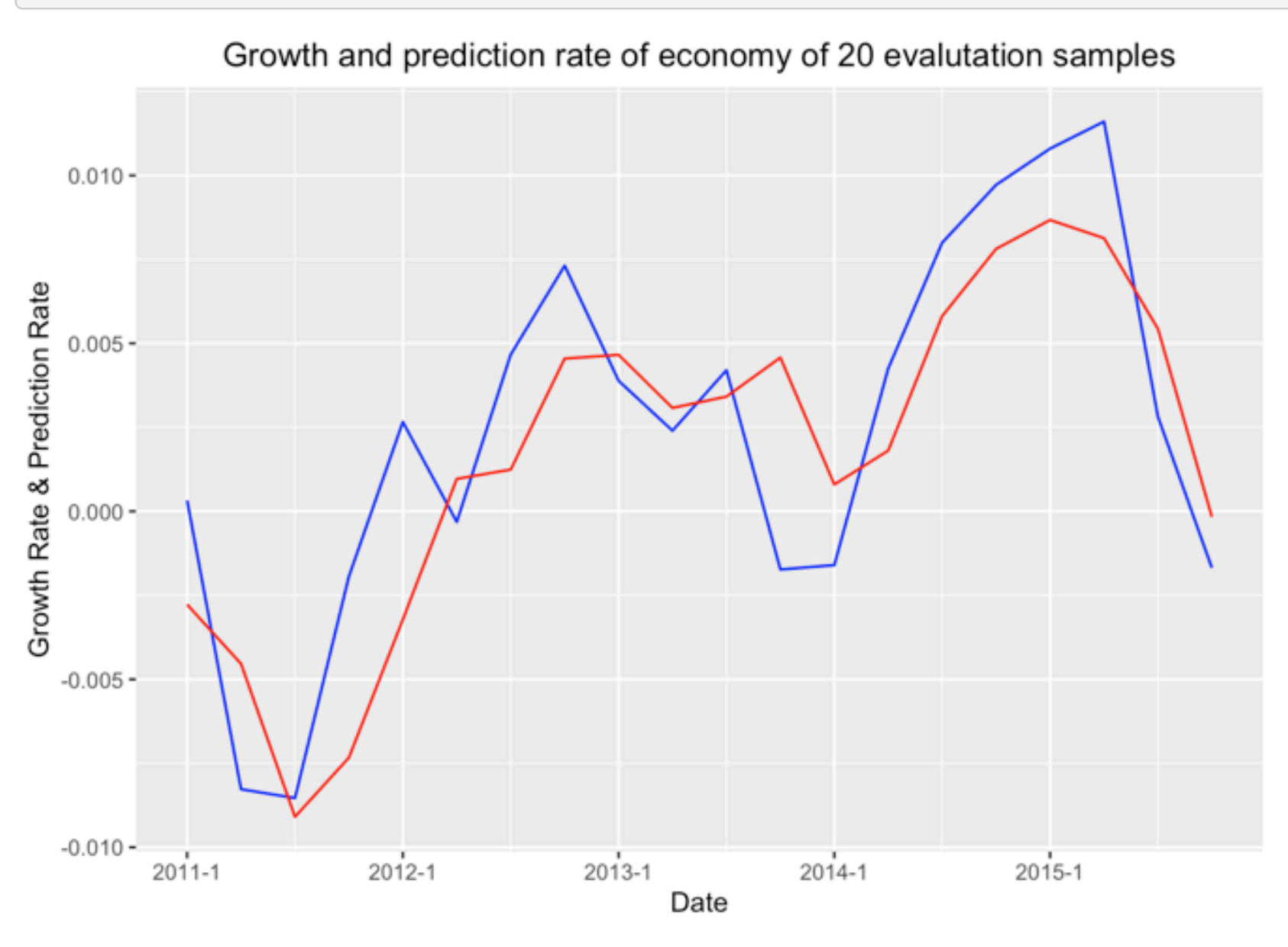
(g)

Use the model in part (e) with $k_1 = k_2 = 1$ to generate a set of twenty one-step-ahead predictions for the growth rates in each quarter of the period 2011 to 2015. Note that the required values of the lagged leading indicators are available for each of these forecasts. Calculate the root mean squared error of these forecasts and present a time series graph of the predictions and the actual growth rates.

answer (g)

```
LM2 <- lm(CG$GrowthRate~CG$lagGR + CG$lagli1 + CG$lagli2)
CG <- mutate(CG, Pb2 = predict(LM2, CG, type = "response"))
CG6 <- CG["2011 Q1" <= CG$Date,]

graph <- ggplot(CG6) +
    geom_line(aes(x = CG6$Date, y = CG6$GrowthRate), color = "blue") +
    ggtitle("Growth and prediction rate of economy of 20 evalutation samples") +
    xlab("Date") +
    ylab("Growth Rate & Prediction Rate") +
    theme(plot.title = element_text(hjust = 0.5))
graph + geom_line(aes(x = CG6$Date, y = CG6$Pb2), color = "red")
```



red = Prediction rate

blue = Growth rate

```
MAE(CG6$GrowthRate, CG6$Pb2)
```

```
## [1] 0.00266732
```

```
RMSE(CG6$GrowthRate, CG6$Pb2)
```

```
## [1] 0.003129364
```