Assignment 4: MARIMA models

Arturo Arranz Mateo

April 28, 2016

During the writing of the assignment I have been working with Stefan Rethmeier.

1 Question 1: Presenting the data

The following plots show the time series of the prices at zip2000, zip2800, zip7400, zip8900 and the consumer price index(cpi), discount and 30 real estate bond rate(real30yr).

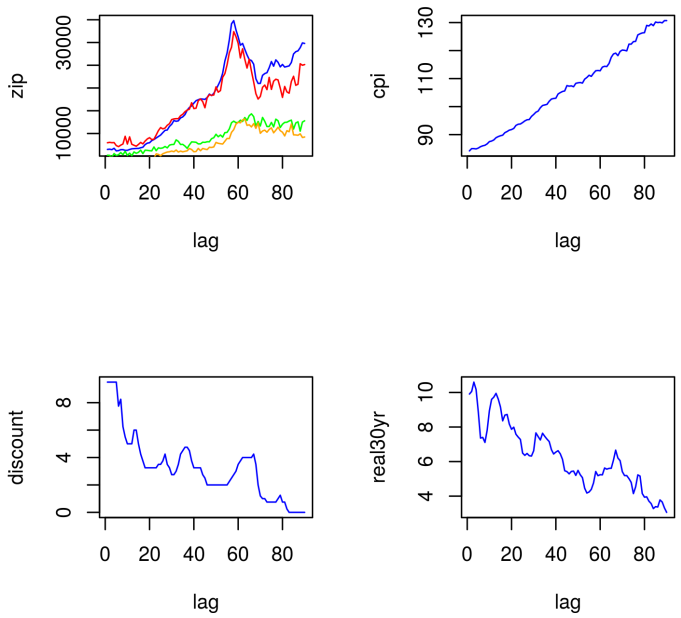

Figure 1: Prices, cpi, discount and real30yr without transformations.

As we can see the series are far from being stationary, so transformations are needed. We try with the first difference at each series.

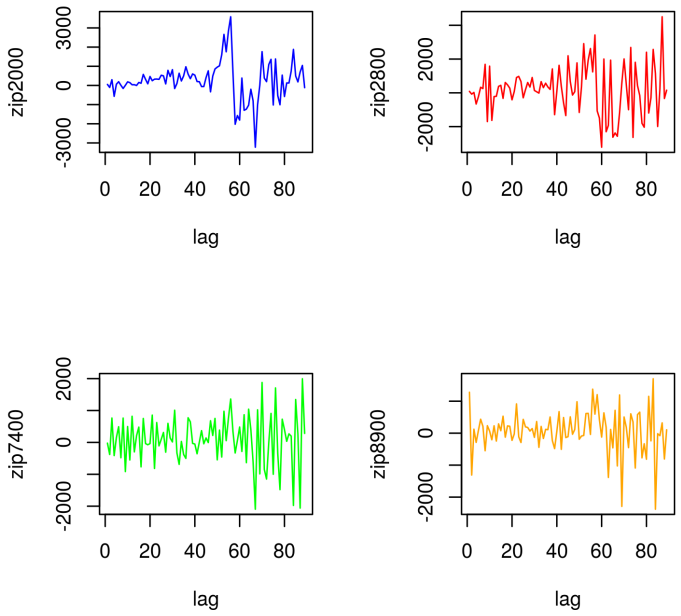

Figure 2: First difference of prices.

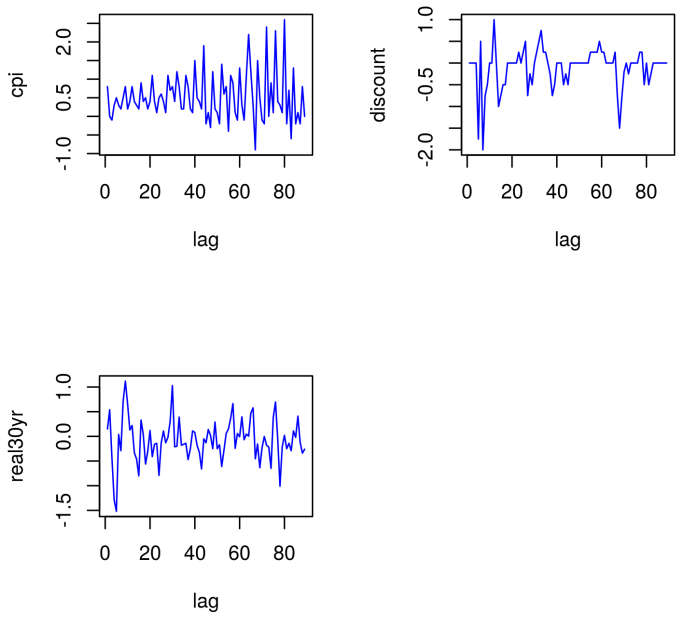

Figure 3: First difference of cpi, discount and real30yr.

Now the values oscillate around 0. However, we can see how in the variance increase over time in the prices series. That can be fixed with a logarithmic transformation.

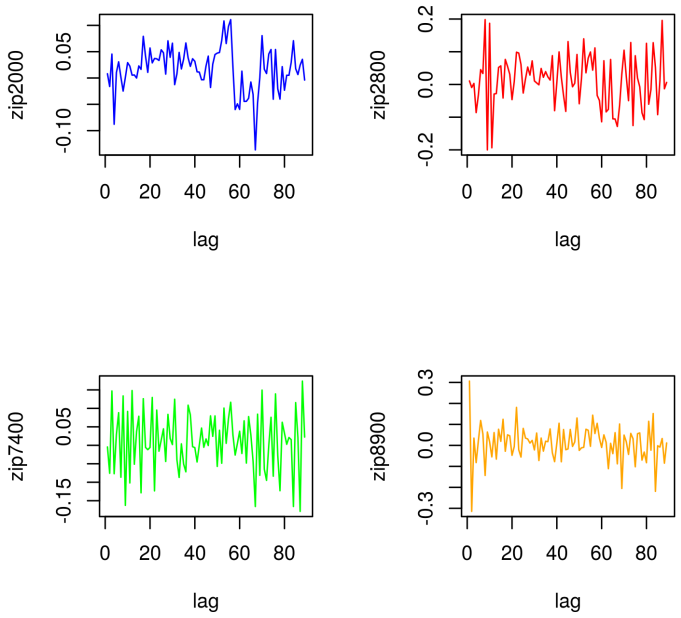

Figure 4: First difference and logarithmic transformation of prices.

We can see how in the Figure 4, the prices variance are more compact.

2 Question 2: ACF, PACF, CCF

In the following plots the ACF and PACF are presented for non-transformed and transformed Zips, cpi, discount and real30yr.

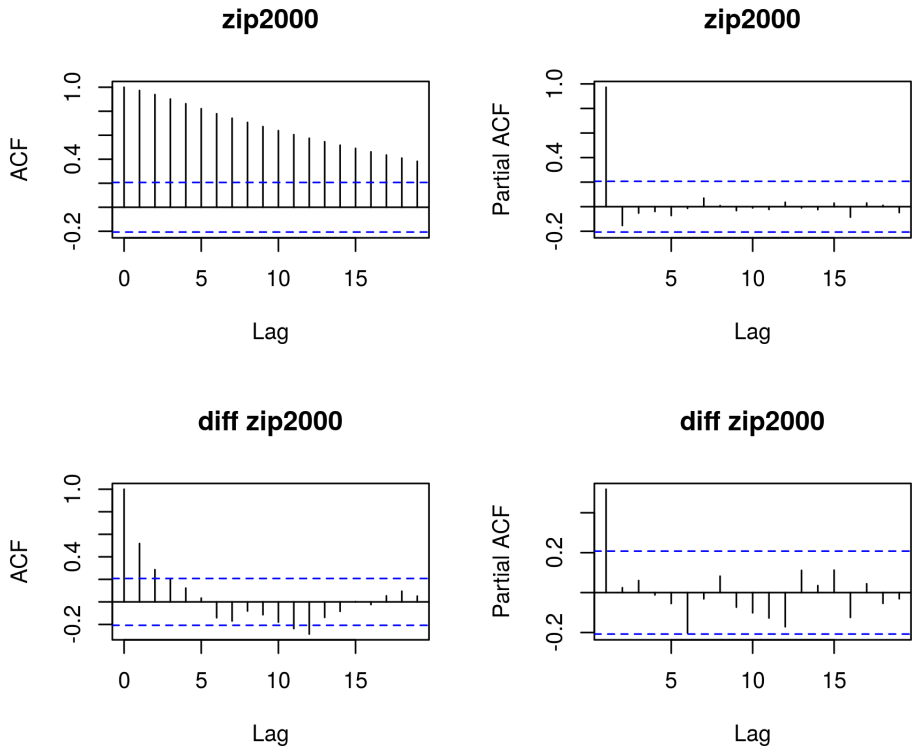

Figure 5: ACF and PACF zip 2000.

Zip2000 series ACF and PACF are slightly under-differentiated. The sharp cut at PACF, and the ACF suggest a 2 or 3 AR term.

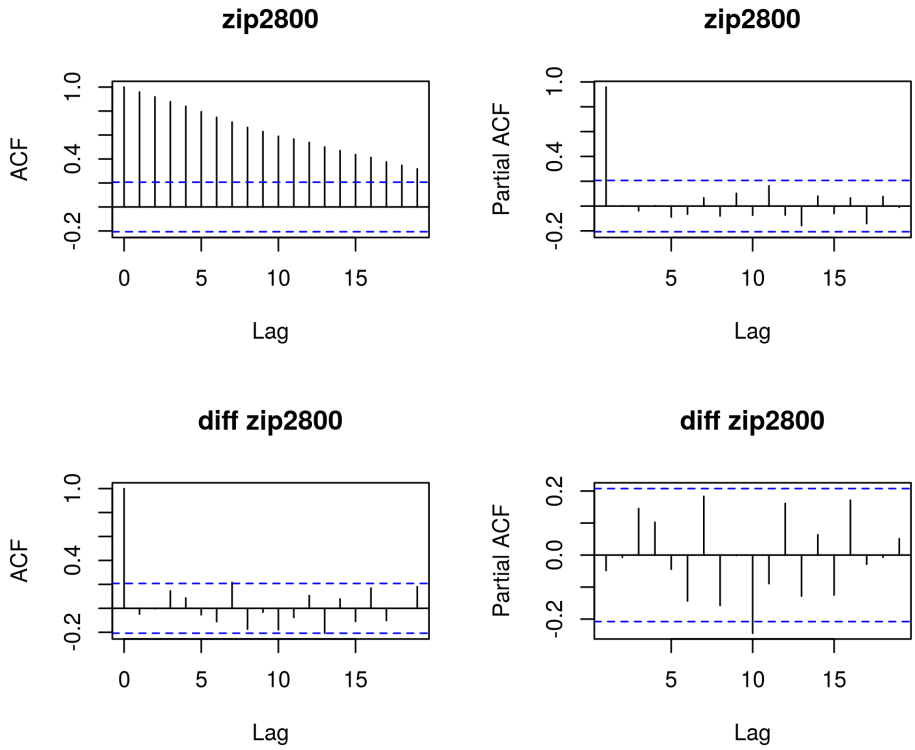

Figure 6: ACF and PACF zip 2800.

The zip2000 series ACF and PACF do not suggest any further update besides the differentiation.

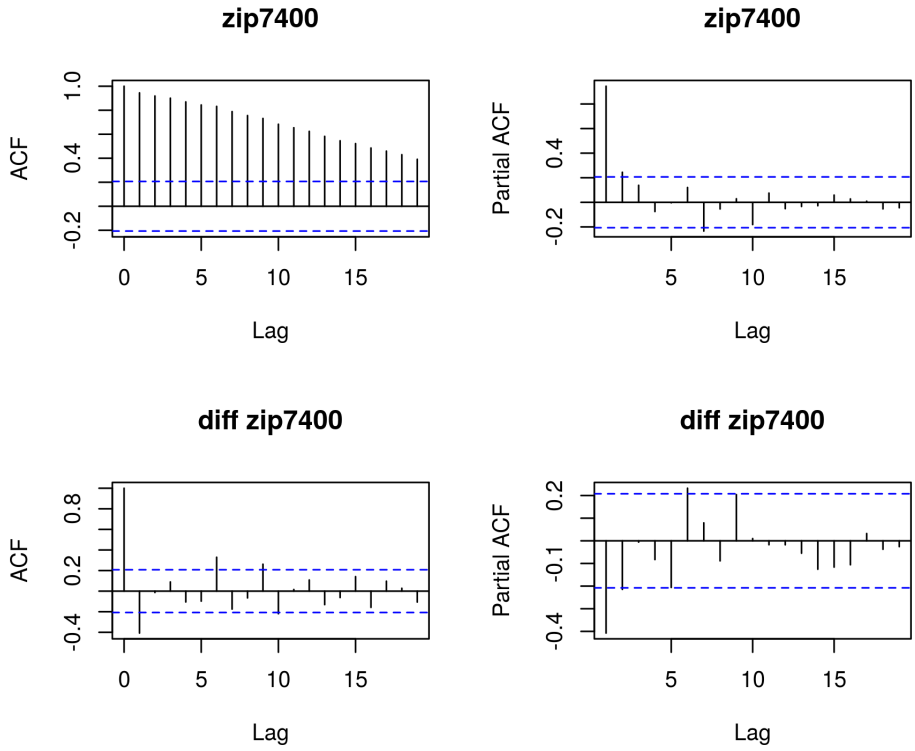

Figure 7: ACF and PACF zip 7400.

Zip7400 series is slightly over-differentiated. This suggest a MA(1) term.

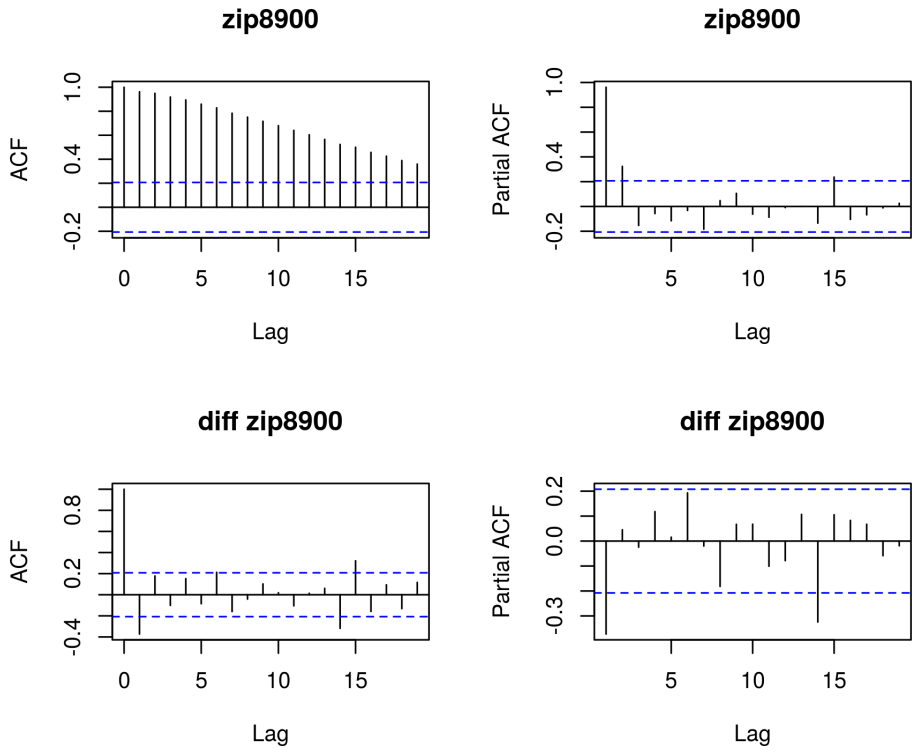

Figure 8: ACF and PACF zip 8900.

Zip8900 series is also slightly over-differentiated. This suggest, again, a MA(1) term.

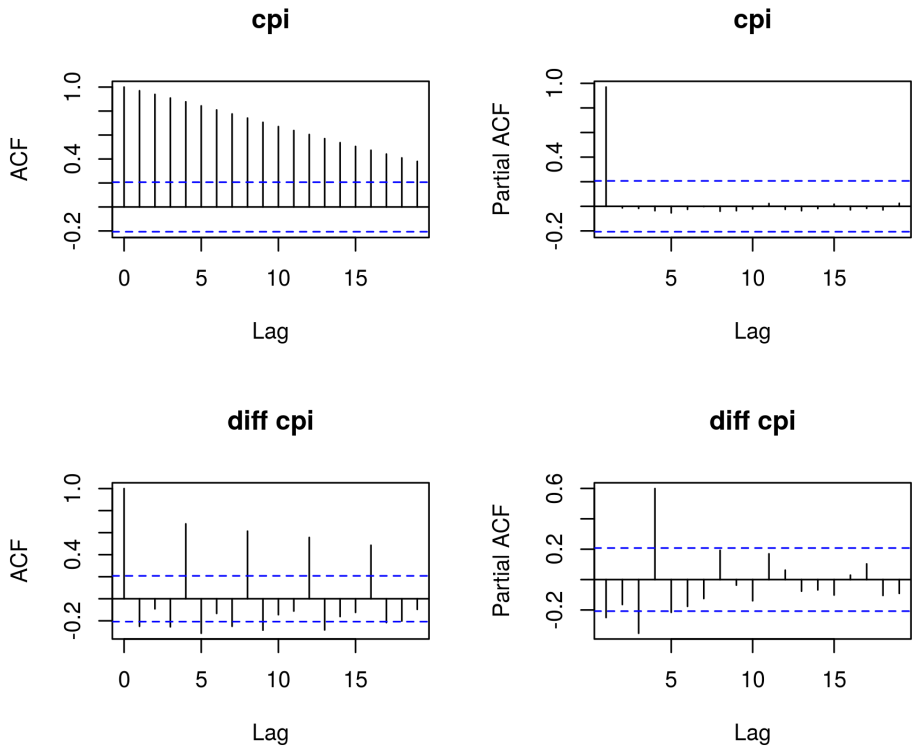

Figure 9: ACF and PACF cpi.

CPI series show a clear seasonal pattern with a period of 4.

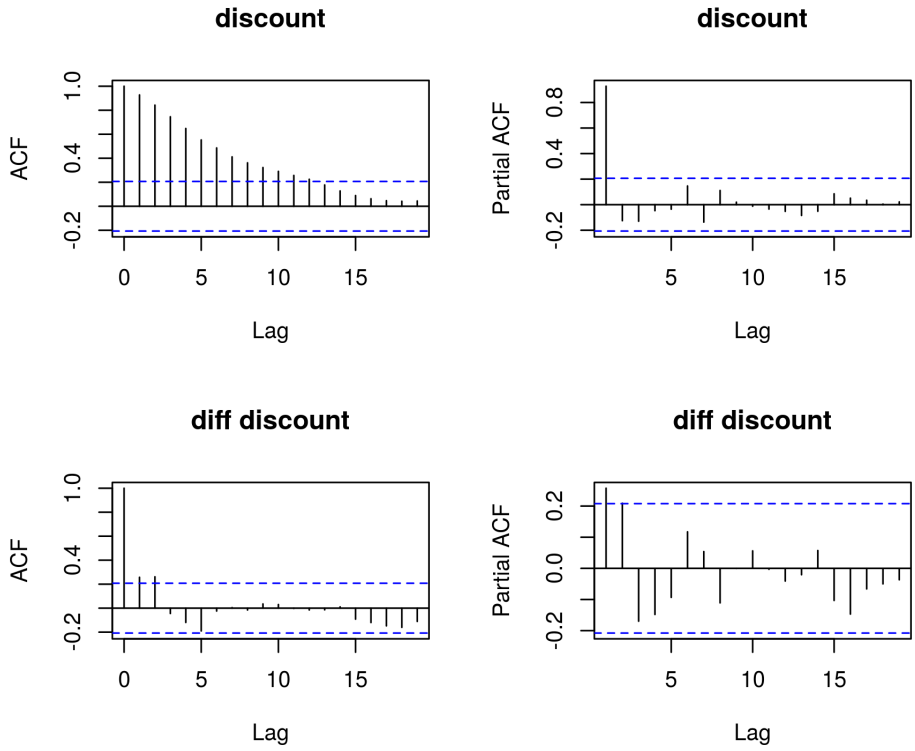

Figure 10: ACF and PACF discount.

Discount series suggests a AR(1) term.

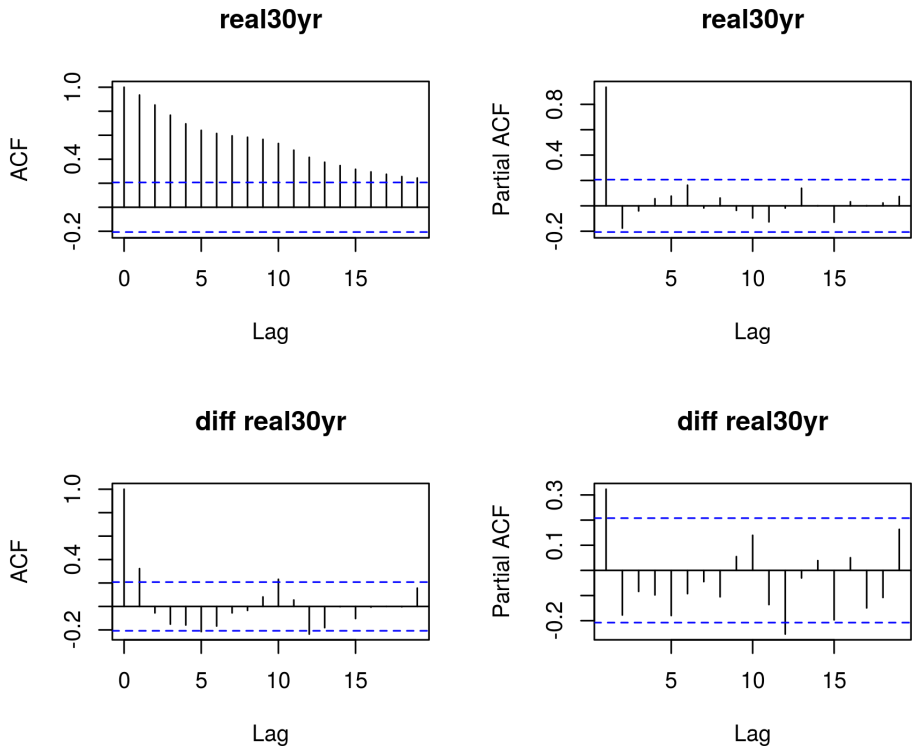

Figure 11: ACF and PACF zip real30yr.

real30yr series also suggests a AR(1) term.

Here some CCF plots are shown. We can not substract much information from them since the correlations are propagated along the lags. Pre-whitening is necessary, which will be made after constructing the ARIMA models. The rest of the CCF plots are at appendix A.

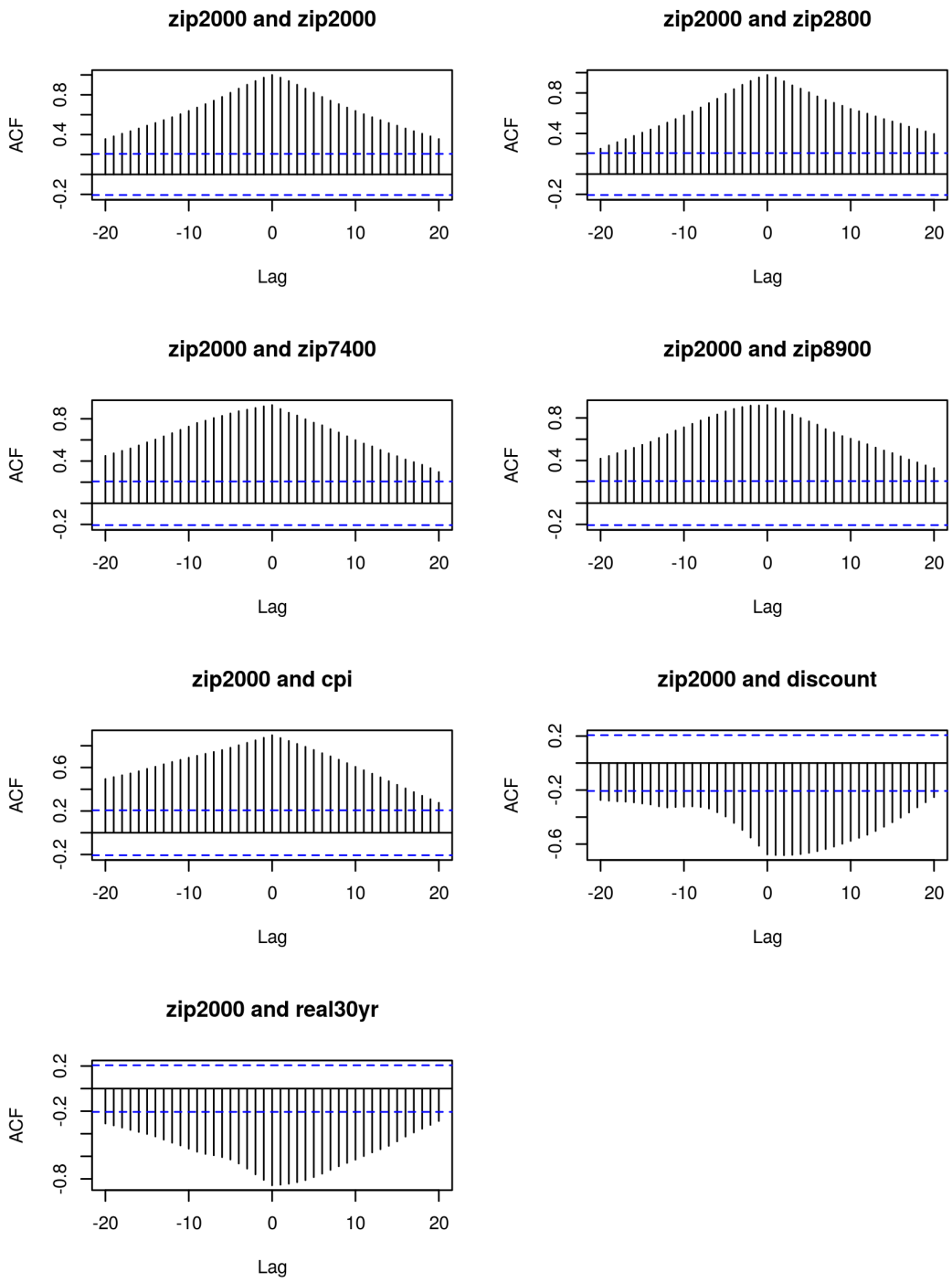

Figure 12: CCF

3 Question 3: Individual ARIMA models

The following ARIMA models have being fit for each of the series, following the criteria mentioned in the question 2.

- **Zip2000:** ARMA(1,1,0)
- **Zip2800:** ARMA(0,1,0)
- **Zip7400:** ARMA(0,1,1)
- **Zip8900:** ARMA(0,1,1)
- **cpi:** ARMA(0,1,0)x(0,1,1)
- **discount:** ARMA(1,1,0)
- **real30yr:** ARMA(1,1,0)

We aimed to make simple models, rather than more accurate but very complex models. As stated in the assignment, we will go through model building of some of the series. Here we show cpi ARIMA building. If we look at the first non-seasonal difference ACF and PACF plots in Figure 13, we can see how clearly a seasonal difference of period 4 is needed.

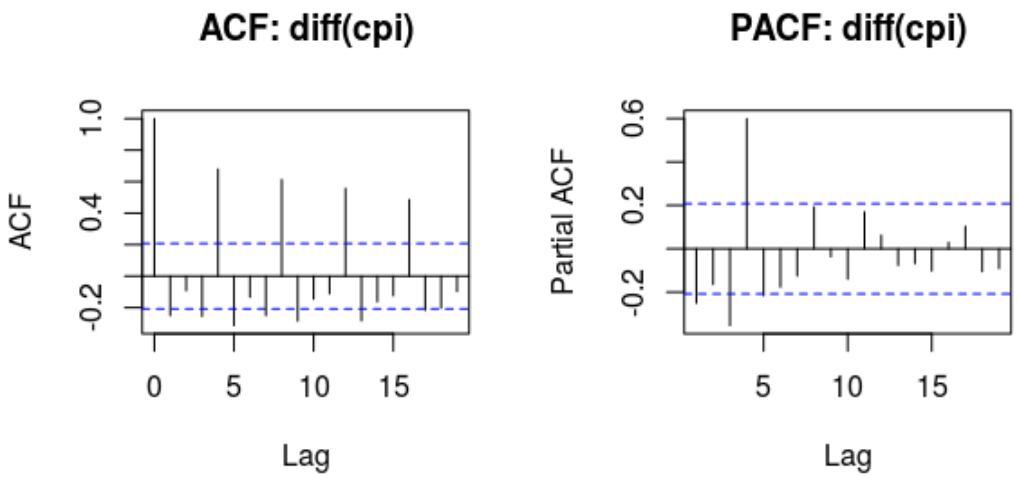

Figure 13: ACF and PACF from diff(cpi)

A ARIMA(0,1,0)x(0,1,0), i.e and ARIMA model twice difference, one seasonal and one non-seasonal. Now if we look at the remaining residuals in Figure 14, a seasonal AR term is suggested.

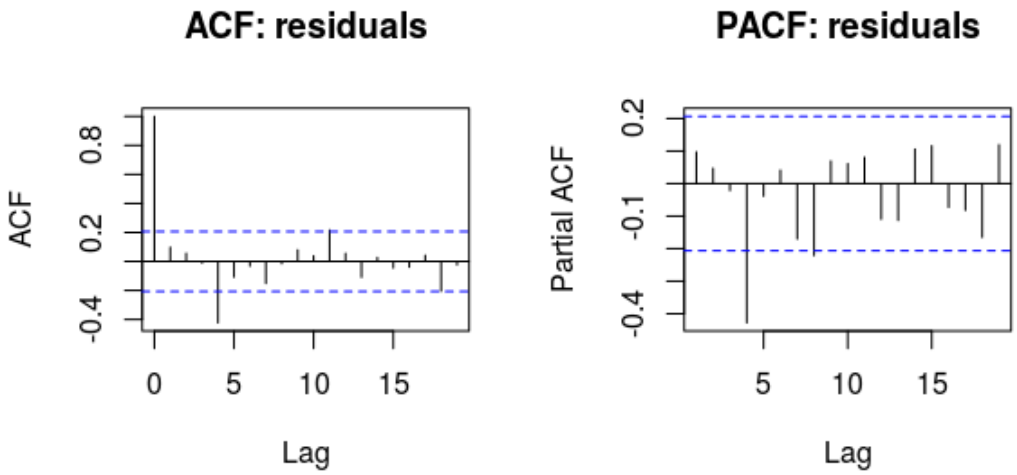

Figure 14: ACF and PACF from ARMA(0,1,0)x(0,1,0) residuals

If we look at the new residuals we have in Figure 15, we do not find any relevant auto-correlation left.

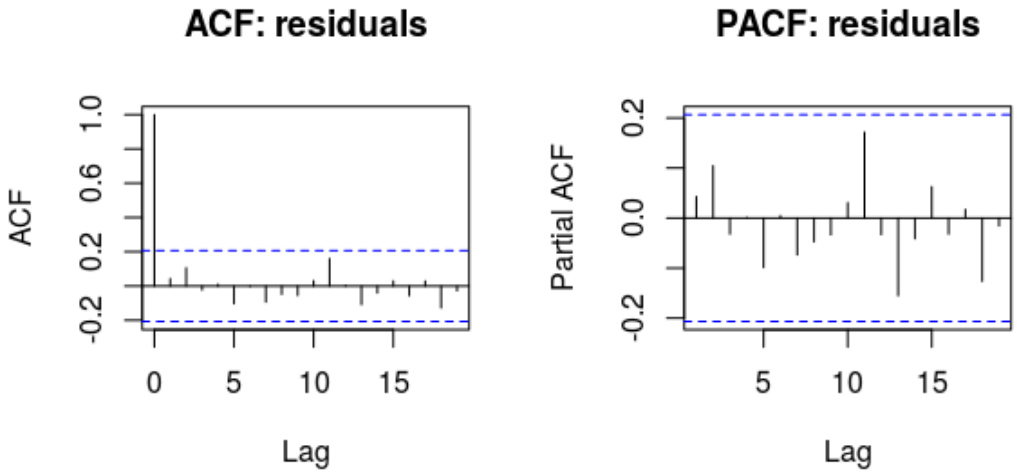

Figure 15: ACF and PACF from ARMA(0,1,0)x(0,1,1) residuals

The QQplot from Figure 18 look like a good approximation of white noise. Also the **sign test** give us 42 negative residuals against 47 positive, which is also a good result.

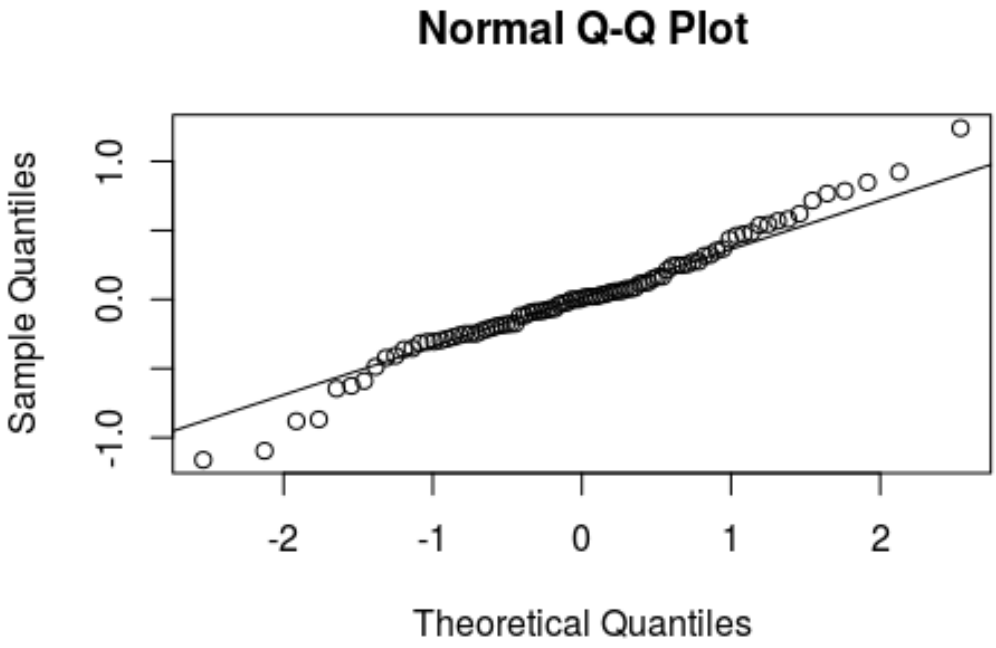

Figure 16: ACF and PACF from ARMA(0,1,0)x(0,1,1) residuals

In the following the CCF between zip2000 and the rest of series are presented. Now that we have created and ARIMA model, we are able to prewhite the series.

After the process of prewhitening we can see how zip2000 is correlated with previous or future lags of the rest of the series. However, this information will not be used for the multivariate model selection.

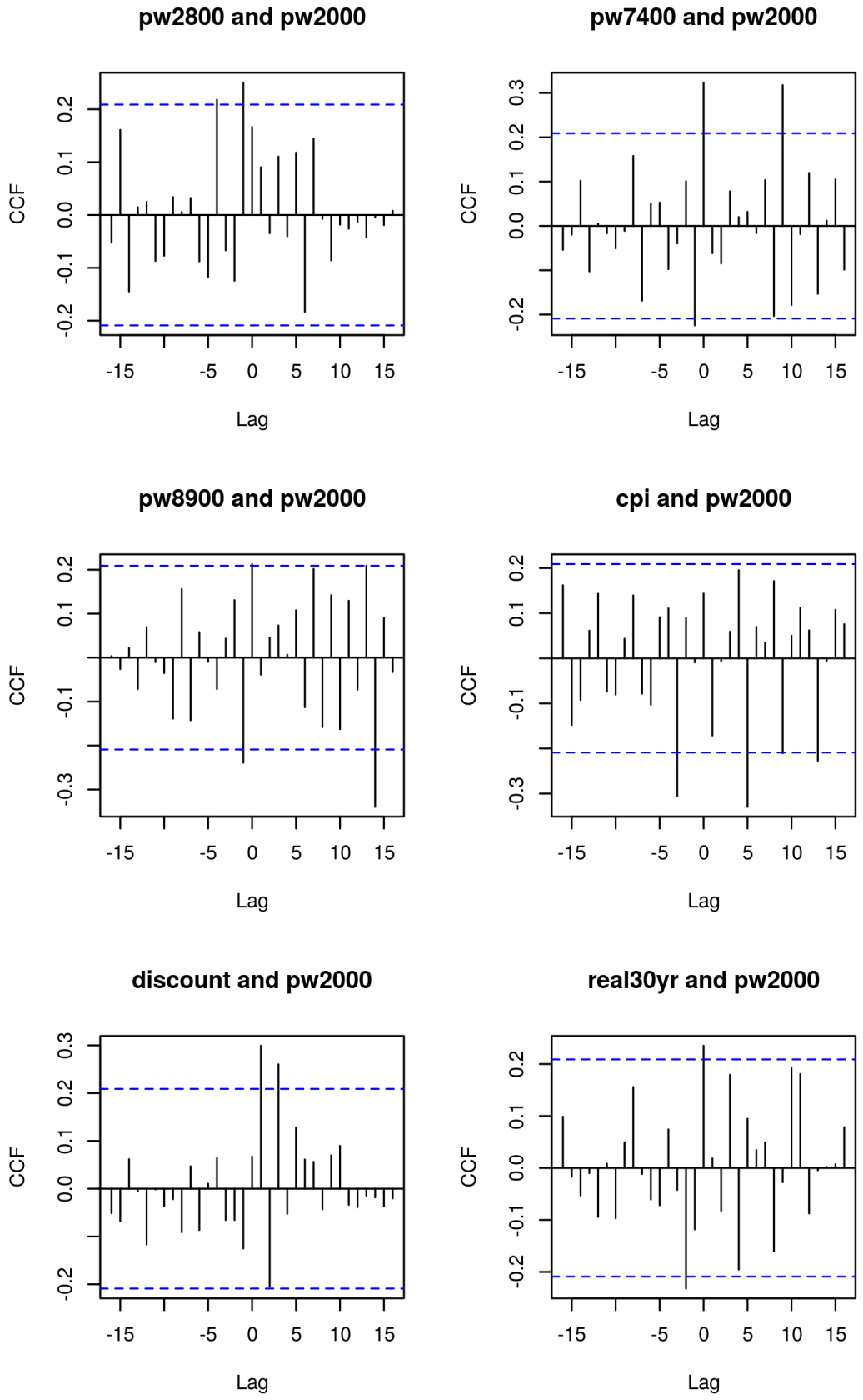

Figure 17: CCF after prewhitening

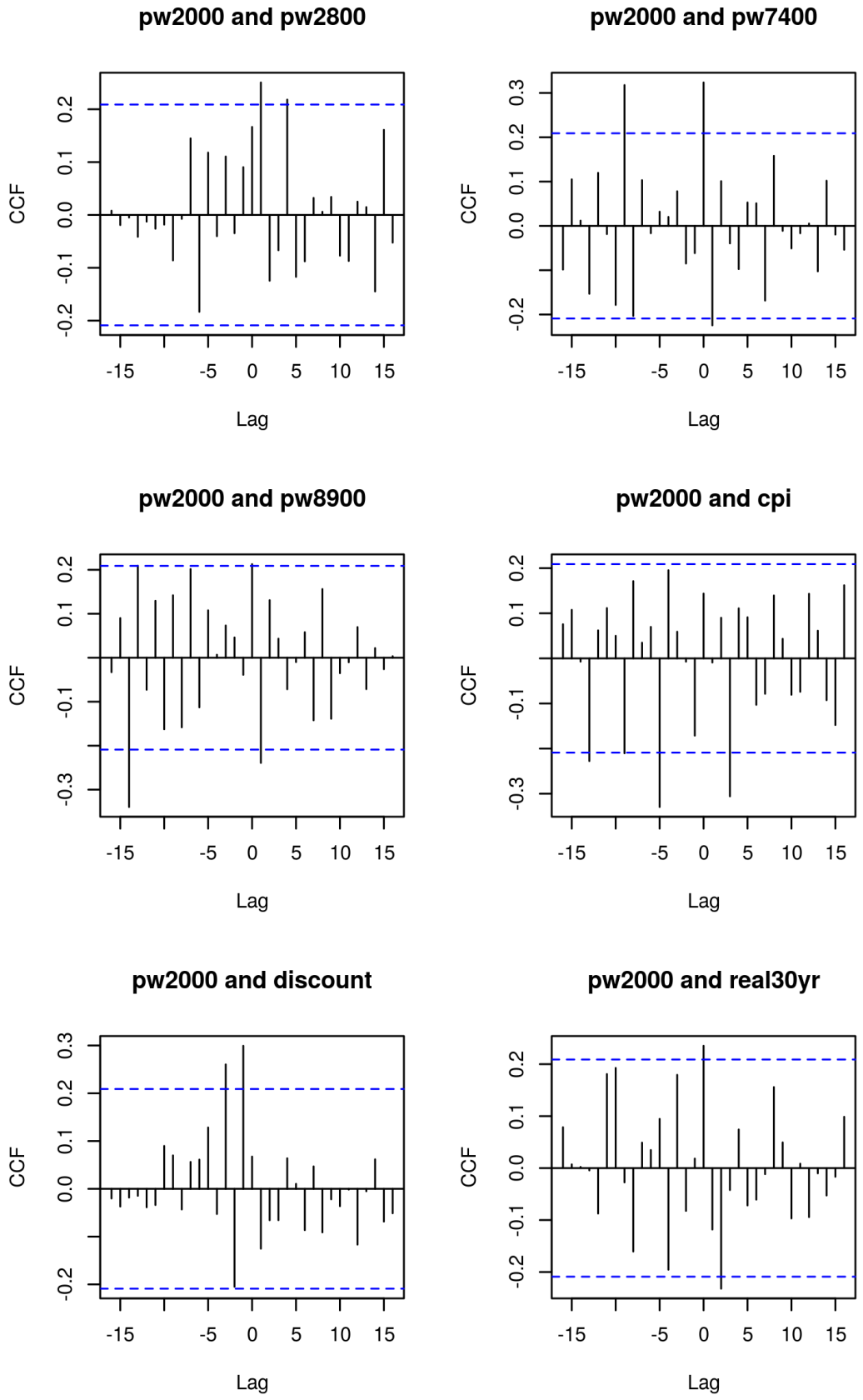

Figure 18: ACCF after prewhitening

4 Question 4: Multivariate model selection

Since all the univariate models implemented in previous section are order one or less, the first model we will try is going to be a AR(1) and MA(1) MARIMA model. The variables cpi, real30yr and discount are going to be treated as regression variables.

With this implementation the residuals covariance matrix converge in around 30 iterations.

Table 1: AR(1) terms

	y1	y2	y3	y4	y5	y6	y7
1	-0.317	-0.261	-0.198	0.581	32.975	147.993	32.167
2	-0.331	-0.227	-0.193	0.439	-203.268	456.248	35.824
3	-0.131	-0.050	0.090	0.551	12.369	305.573	0.749
4	-0.130	-0.128	0.257	0.047	93.721	-110.643	-27.435
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: MA(1) terms

	y1	y2	y3	y4	y5	y6	y7
1	0.158	-0.099	-0.462	0.452	0.000	0.000	0.000
2	0.685	-0.690	-0.155	0.372	0.000	0.000	0.000
3	-0.033	0.072	-0.348	0.223	0.000	0.000	0.000
4	-0.023	-0.130	0.482	-0.562	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The multiple correlation is shown in the table.

Table 3: Correlation residuals

	u1	u2	u3	u4	u5	u6	u7
1	0.42	0.3	0.37	0.31	0	0	0

In the Figure 19 are shown the predictions for the training data in blue, and the real values with circles.

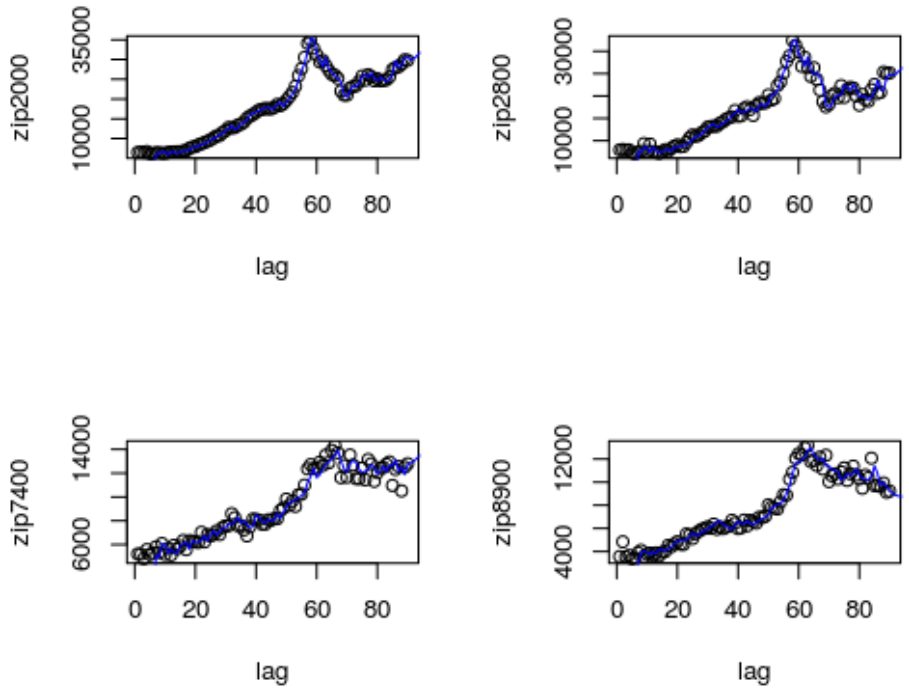

Figure 19: First difference of cpi, discount and real30yr.

Now if we look at the f values of the AR and MA coefficients, we can see how not all the coefficients are relevant.

Table 4: AR fvalues

	x1.y1.Lag.1	x2.y2.Lag.1	x3.y3.Lag.1	x4.y4.Lag.1	x5.y5.Lag.1	x6.y6.Lag.1	x7.y7.Lag.1
y1	2.42	4.01	0.65	4.78	0.03	0.82	0.33
y2	0.95	1.09	0.22	0.98	0.45	2.8	0.15
y3	0.59	0.21	0.19	6.18	0.01	5.01	0
y4	0.76	1.82	2.04	0.06	0.5	0.86	0.45
y5	0	0	0	0	0	0	0
y6	0	0	0	0	0	0	0
y7	0	0	0	0	0	0	0

Table 5: MA fvalues

	x1.y1.Lag.1	x2.y2.Lag.1	x3.y3.Lag.1	x4.y4.Lag.1	x5.y5.Lag.1	x6.y6.Lag.1	x7.y7.Lag.1
y1	0.41	0.41	2.5	2.03	0	0	0
y2	2.74	7.22	0.1	0.5	0	0	0
y3	0.03	0.31	2.04	0.71	0	0	0
y4	0.02	1.34	5.12	5.92	0	0	0
y5	0	0	0	0	0	0	0
y6	0	0	0	0	0	0	0
y7	0	0	0	0	0	0	0

Table 6: AR terms

	x1.y1.Lag.1	x2.y2.Lag.1	x3.y3.Lag.1	x4.y4.Lag.1	x5.y5.Lag.1	x6.y6.Lag.1	x7.y7.Lag.1
y1	0	0	2.5	0	0	0	0
y2	16.54	12.4	0.63	1.15	0	0	0
y3	0	0	13.02	1.21	0	0	0
y4	0	0	8.5	25.65	0	0	0
y5	0	0	0	0	0	0	0
y6	0	0	0	0	0	0	0
y7	0	0	0	0	0	0	0

Table 7: MA terms

	x1.y1.Lag.1	x2.y2.Lag.1	x3.y3.Lag.1	x4.y4.Lag.1	x5.y5.Lag.1	x6.y6.Lag.1	x7.y7.Lag.1
y1	19.09	7.86	0	3.39	0	0	0
y2	0	3.46	1.15	0.59	0	0	0
y3	0	8.68	0	7.9	0	6.08	0
y4	3.96	0	4.04	0	0	0	0
y5	0	0	0	0	0	0	0
y6	0	0	0	0	0	0	0
y7	0	0	0	0	0	0	0

Table 8: Correlation

	u1	u2	u3	u4	u5	u6	u7
1	0.39	0.28	0.36	0.28	0	0	0

As we can see the new model have a similar correlations, with less coefficients estimated.

5 Question 5: Predictions

In the graphics we can see the forecasting for 4 steps ahead(red) and the real values(green). While most of the predictions look reasonable, we should notice that the predictions for the zip 2000 are in the boundary of the upper limit 95 percent confidence interval. Also the prediction, instead of following the ascending trend enter in a kind of stationary behave.

Some more models should be implemented to seek better predictions. One idea is including cpi, discount and real30yr as non-regression variables.

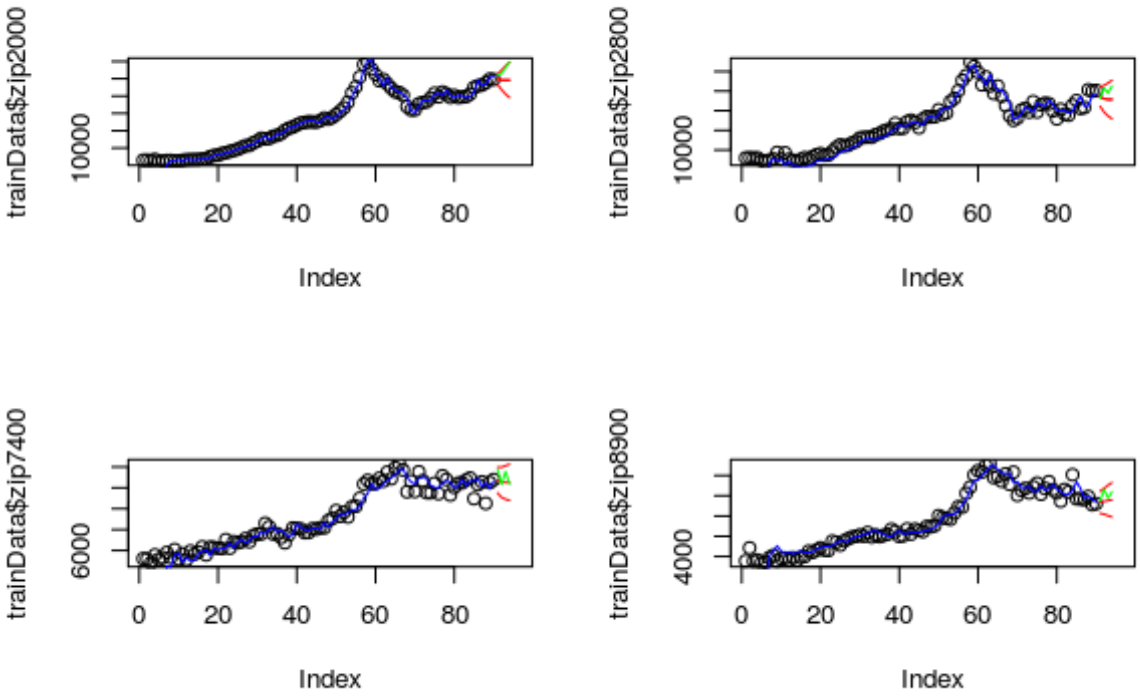

Figure 20: CCF

6 Appendix A: CCF plots from question 3

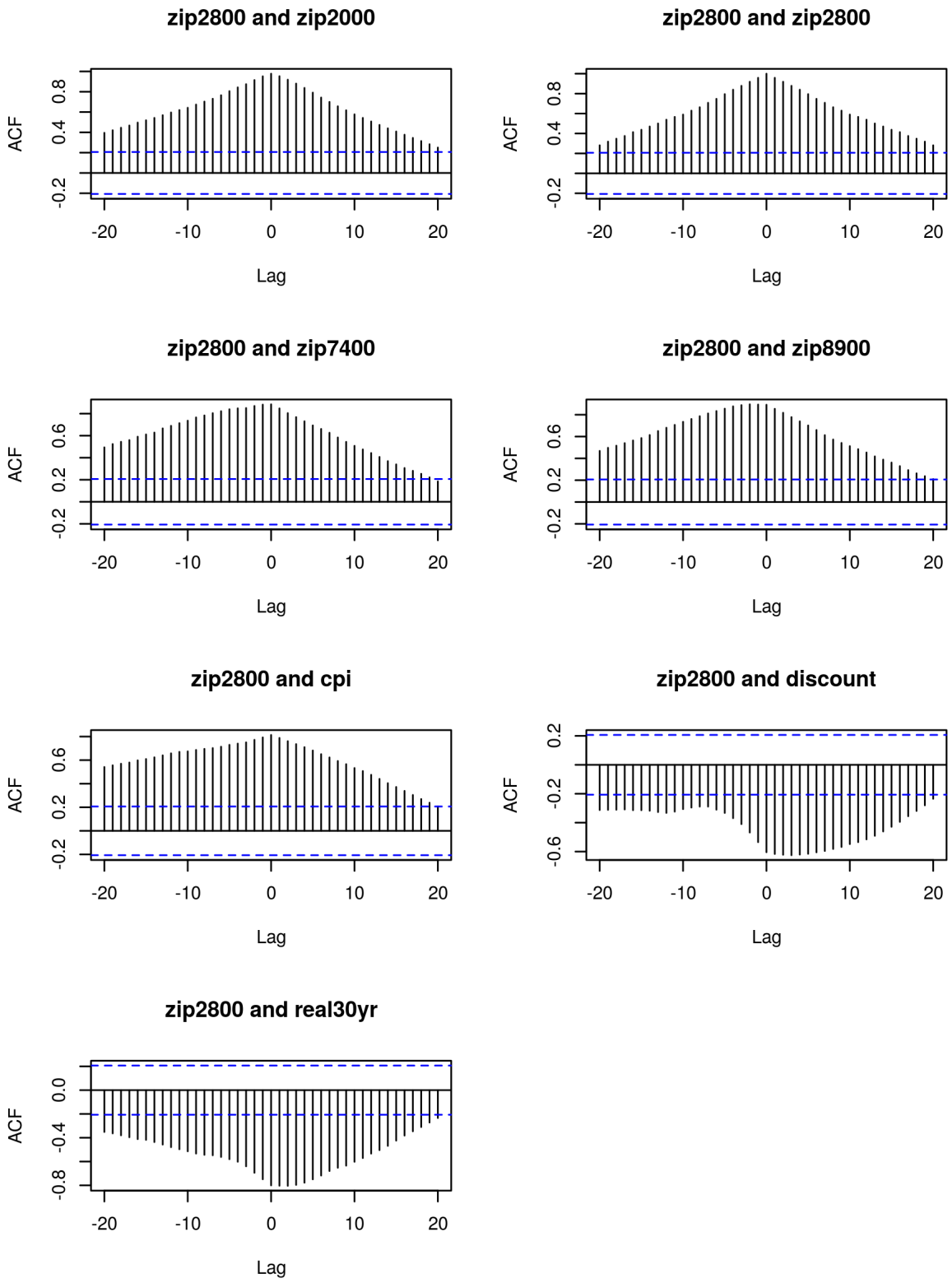

Figure 21: CCF

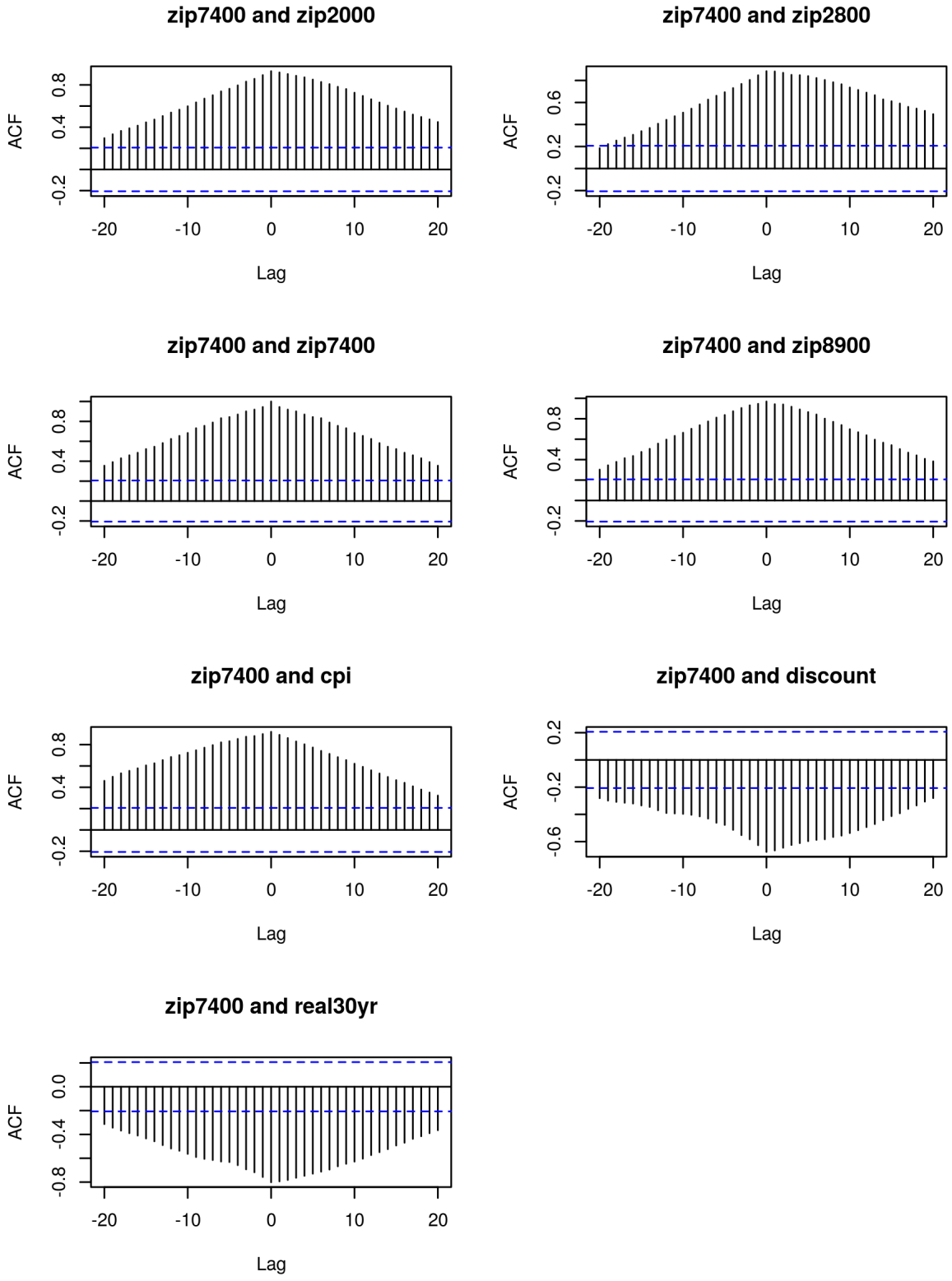

Figure 22: CCF

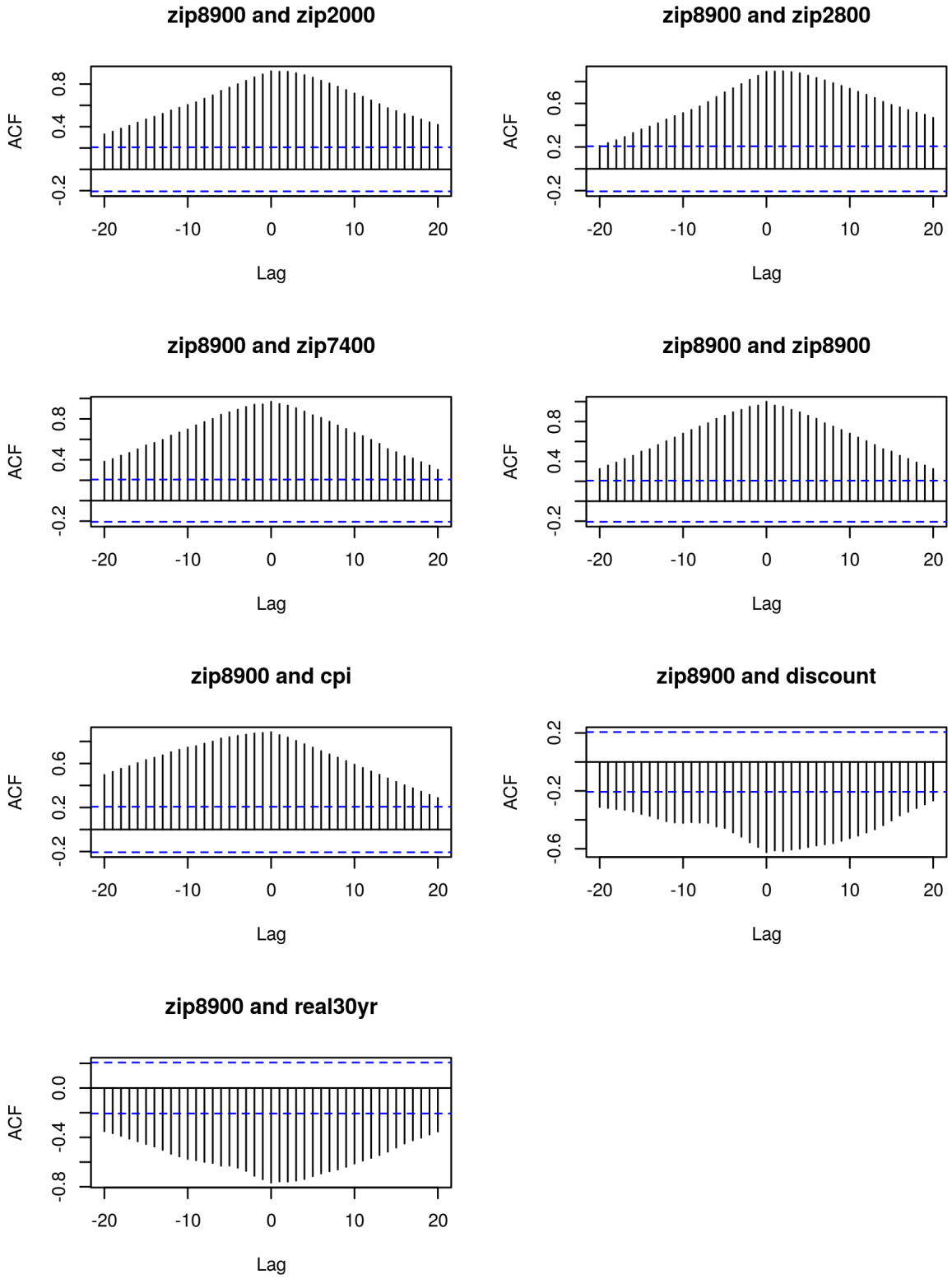

Figure 23: CCF

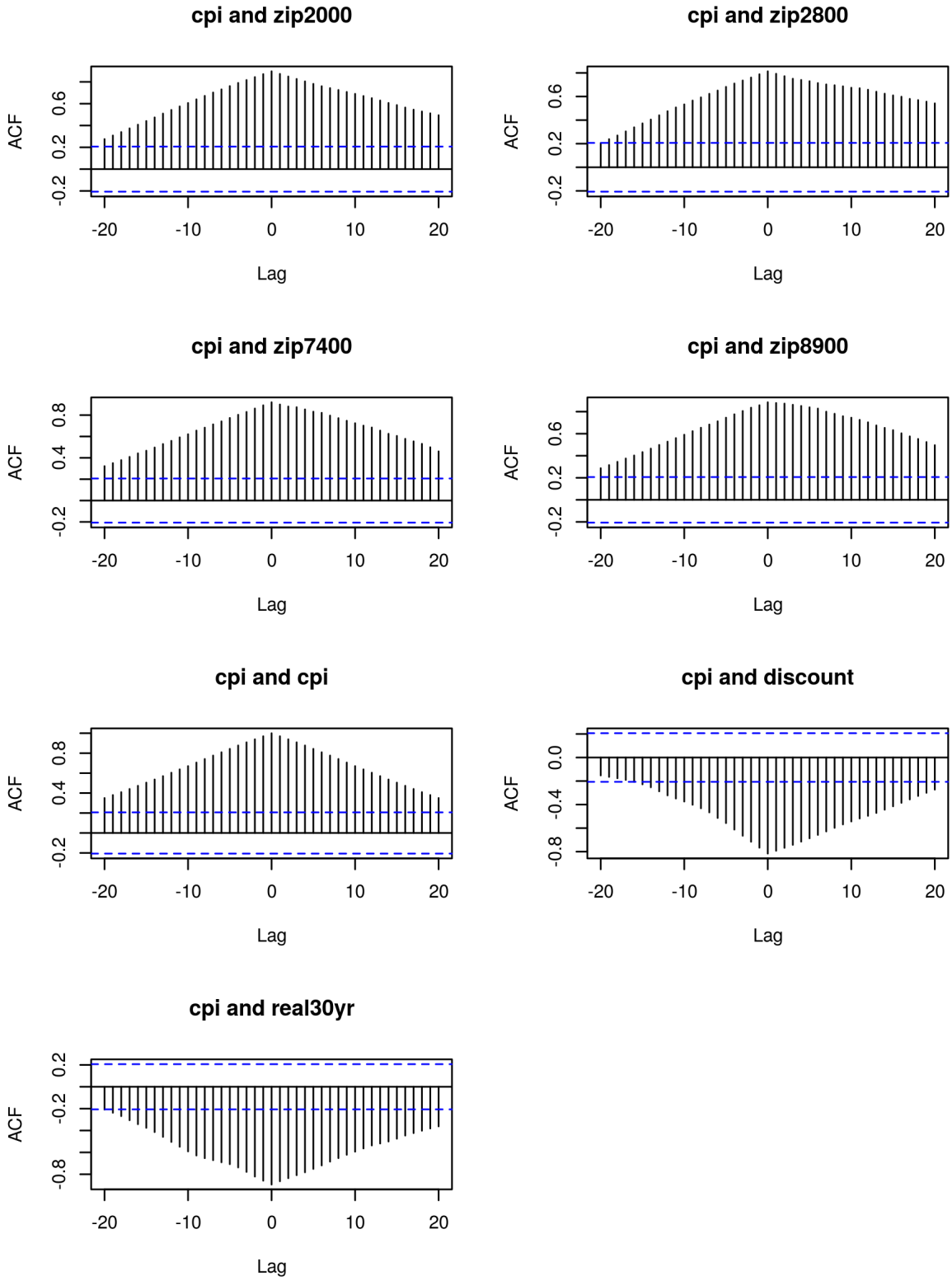

Figure 24: CCF

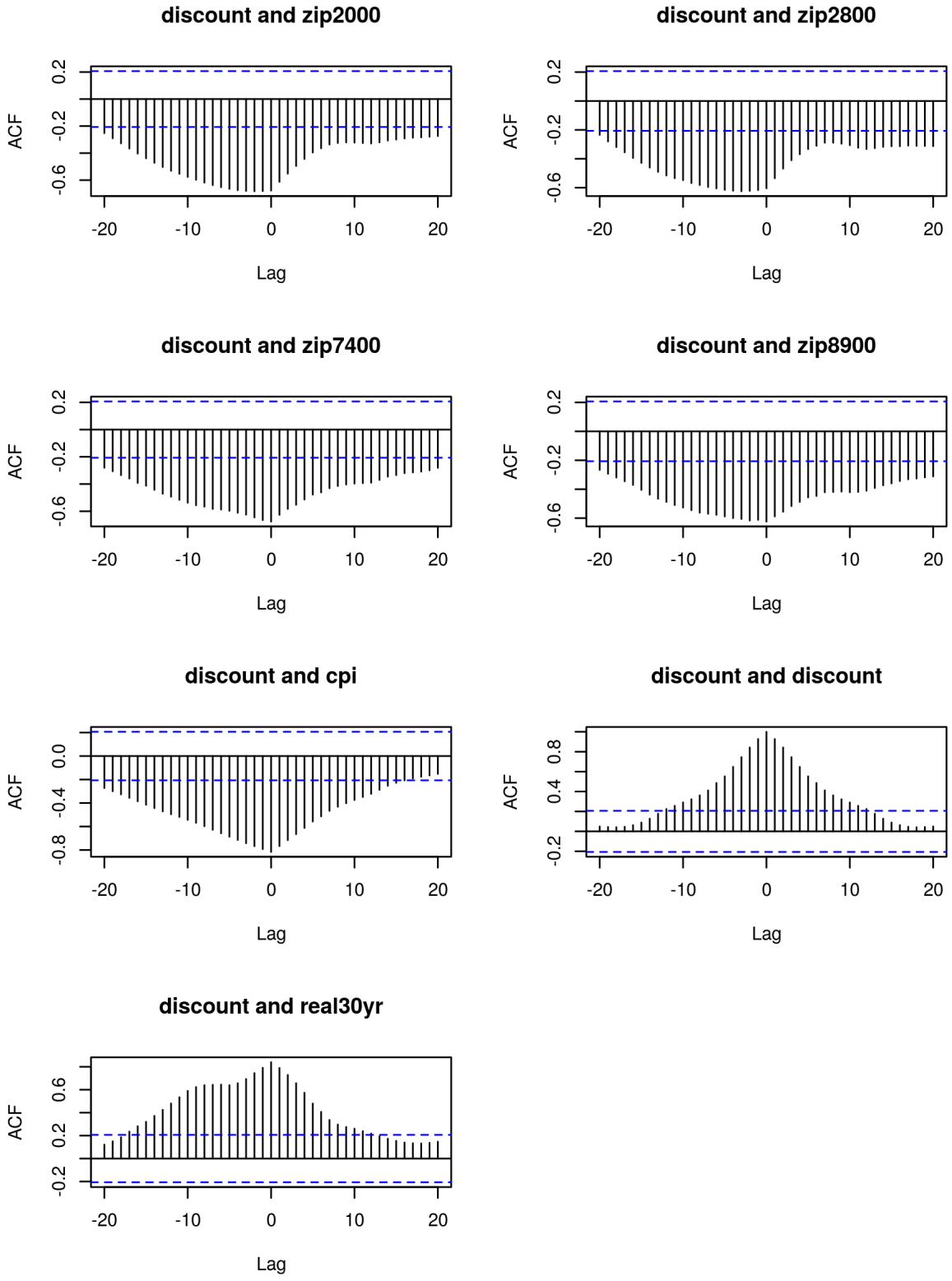

Figure 25: CCF

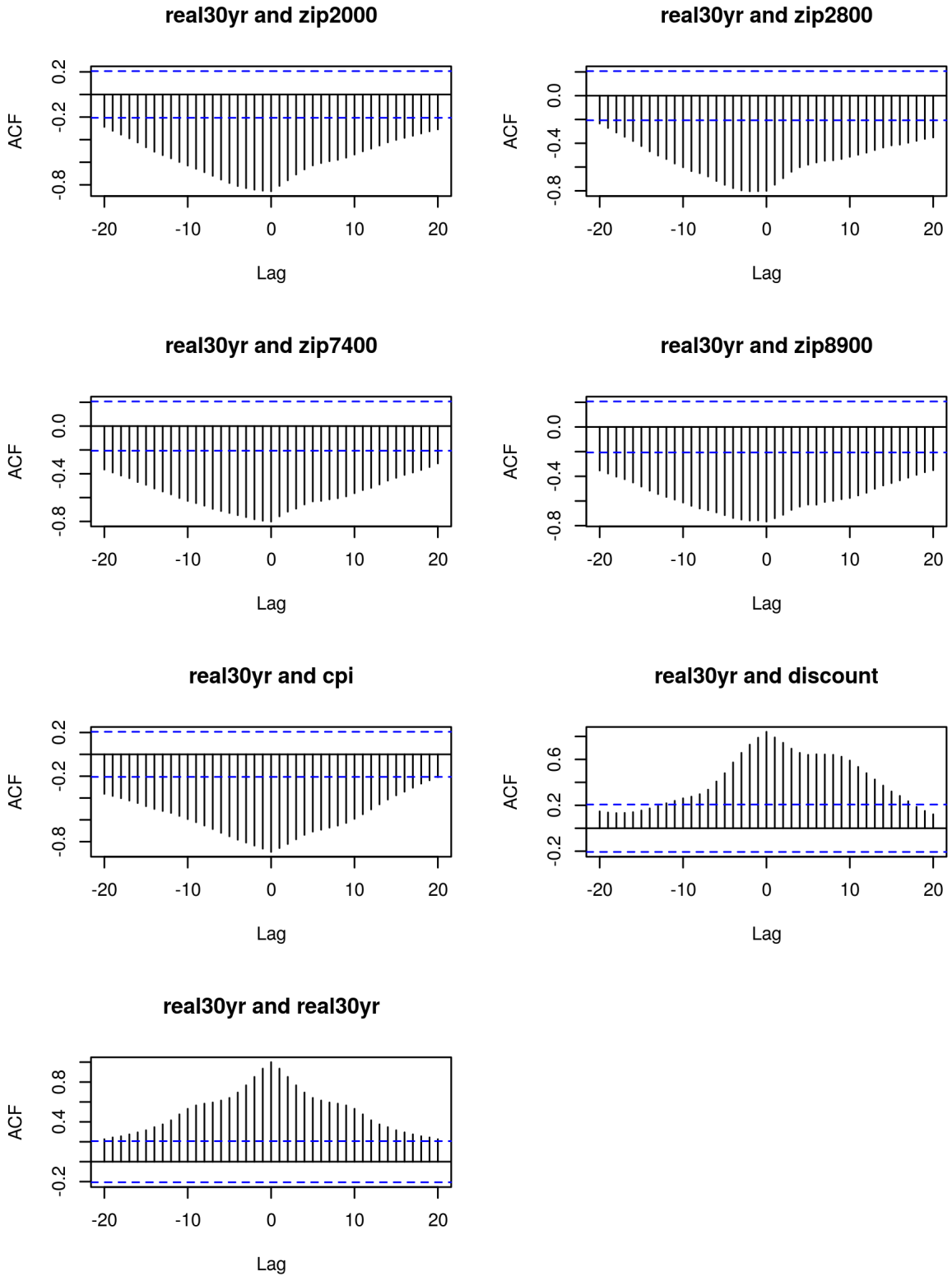

Figure 26: CCF