Aurouet Lucas and Meynier Thibaud M1 Ekap

# Final Report

## Analysis and forecasting of housing credit production in France

**ABSTRACT**

In this paper, we tried to forecast housing credit production in France for the year 2020. We use monthly data from January 2011 to December 2019. At first, we used JDemetra + to remove any outliers and seasonal components from our dataset. With clean data we found that a very few if not none of the models we tried could outperform the naive forecasting method. This result also holds with non-seasonally adjusted data. Last, we found that adding exogenous variables such as google trend data, CAC 40 or economic activity indicators could be used to improve forecasting accuracy, however the precision gains are not sufficient to cope for the addition of exogenous parameters and their estimation.

University of Nantes

Academic year 2019-2020

# Table of contents

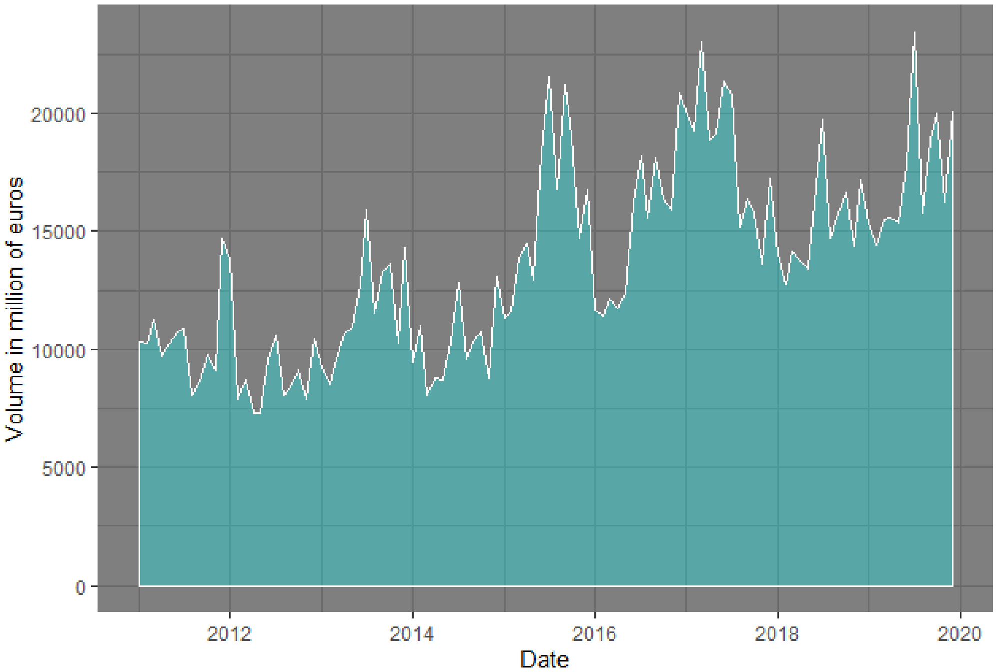
# I. SEASONAL SERIES AND CYCLIC ANALYSIS

## I.1. Preliminary analysis

In this section, we first are going to done a brief review of our data set. In a second time we will deeply analyze our series, i.e we will check if there are any outliers which could impact our future regressions done by ARIMA, and our forecasts. Then, in a third time, we will verify if our series is seasonal or not. If so, we are going to do a seasonal adjustment to remove the seasonality and then, we will proceed to forecasting. Two methods will be used here for the seasonal adjustment of our series: X13 and Tramo Seats. A comparison will be made between both.

*Descriptive statistics*

Figure 1. Housing credit production in France since 2011




Our data set comes from the Banque de France, which does a monthly follow up on the housing credit production, since 2011. The data has been download from [here](). Table 1 summaries main indicators about the housing credit production distribution. The values are in million of euros.

Table 1. Summary of housing credit production series

| Skewness | Kurtosis | Normality test p-value[1] | Mean | Median | Std error |
|---|---|---|---|---|---|
| 0,377 | -0,766 | 0,004 | 13703,524 | 13619,916 | 4031,469 |

According to table 1, our distribution is not normal (p-value under 0.05). The mean is not far from the median which mean that the distribution of value is quite symmetric. The Skewness value means that the series is is spread out on the right (positive value), and the negative kurtosis told us that our distribution is wider than a standard Normal. Finally, according to the Standard error value, our distribution varies around the mean about +/- 4031.47.

*Seasonal detection and scheme decomposure*

The statistics test have been computed through the JDemetra+ software . According to these different statistical tests, the housing credit production series is seasonal (Figure 2). The series passes all the statistical tests, and notably the auto-correlations at seasonal lags. As we can see on Figure 1. There are peaks in July and December. That could be explained by the fact that just after July, in France, summer holidays begin, and most part of the people are going on vacation during this month. After this break, the production of credit is rising again (in general) and reaches another peak in December.

Figure 2. Summary of seasonality test

Summary

*Data have been differenced and corrected for mean*

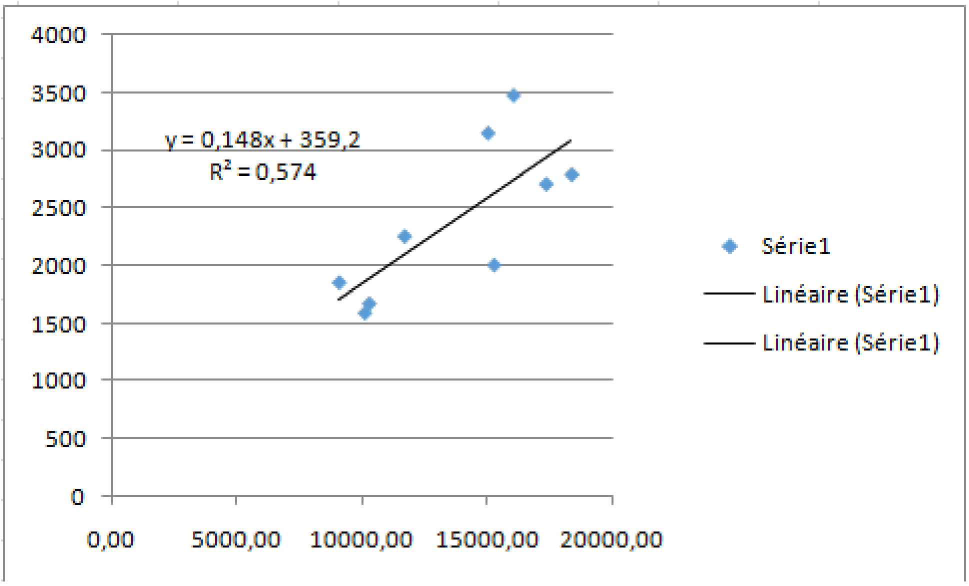
| Test | Seasonality |
|---|---|
| 1. Auto-correlations at seasonal lags | YES |
| 2. Friedman (non parametric) | YES |
| 3. Kruskall-Wallis (non parametric) | YES |
| 4. Spectral peaks | YES |
| 5. Periodogram | YES |
| 6. Seasonal dummies | YES |
| 6bis. Seasonal dummies (AMI) | YES |

Now, we are going to detect potential outliers on our series, using an ARIMA regression. According to Figure 2, we detect 2 outliers. One in January 2012, and the second one in February 2014.

[1] p-value of Shapiro-Wilk normality test

Moreover, if we look at figure 1, we can see that the scheme decomposure is more multiplicative than additive. We can observe that the peaks are higher and higher from the beginning of the series in 2011; and from 2017, they seem to be decreasing, which mean that the scheme is not additive (which assumes the peaks are similar on each period). But to confirm this, we made the Buys Ballot test with a simple linear regression, where our dependent variable is the standard deviation from each year. Our regressor is the mean for each year.  The test states that if the coefficient of the regressor of the linear model is different from 0, then, the scheme decomposure is multiplicative, if not, it is additive. According to this test and our data, scheme decomposure is multiplicative (see figure 3). The coefficient of the regressor is different from 0 and statistically significant.

Figure 3. Buys Ballot's test on housing credit production series




result from linear regression for Buys Ballot's test

| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| (Intercept) | 359.24880 | 679.00675 | 0.529 | 0.613 | |
| mean_buys | 0.14815 | 0.04822 | 3.072 | 0.018 | * |

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 468.8 on 7 degrees of freedom

Multiple R-squared:  0.5742,     Adjusted R-squared:  0.5134

F-statistic: 9.439 on 1 and 7 DF,  p-value: 0.01801

*Outliers detection*

Figure 4. Outliers detected in housing credit production series

| | Period | Value | StdErr | TStat |
|---|---|---|---|---|
| AO | 1-2012 | 4327,8491 | 927,8345 | 4,6645 |
| AO | 2-2014 | 3248,1011 | 772,8465 | 4,2028 |

We can note that these 2 outliers are AO typed (see Figure 4), i.e they are additive. In other words, they are affecting only one observation at a given moment in the series. These outliers will be treated with the use of Tramo Seats and X13 methods to produce seasonal adjustment on our series. These points seems to be more related to a stochastic occurrence than an economical justification. But we can note that at the middle on 2015, the credit production is going down and reaches 11 500 millions of euros at the start of 2016. This downward slope can be explained by the recession which occured in the eurozone in 2014-2015. The European central bank decided, to counter this recession and the deflation risk, to set on a quantitative easing policy. The aim of these policies was to lower interest rates and stimulate the credit demand.

### I.2.Seasonal adjustment

Figure 5. Housing credit production CVS series with Tramo Seats and trend with Tramo Seats

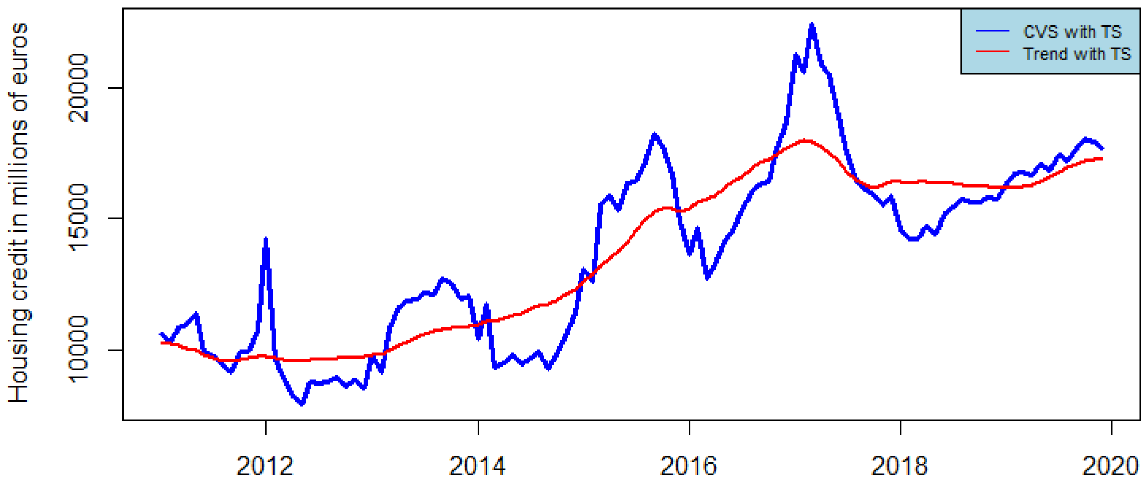


Figure 6. Housing credit production CVS series with X13 and trend with X13

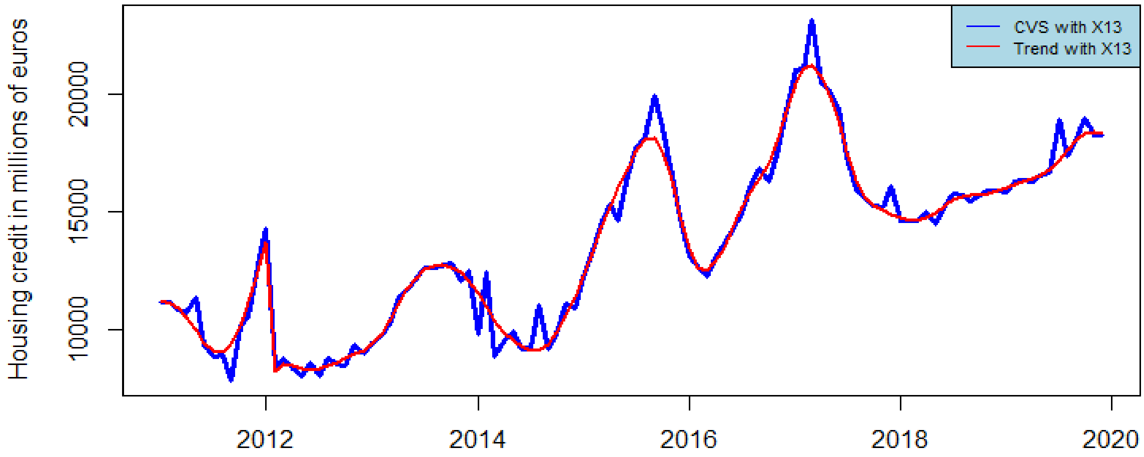



Figure 5 and 6 show us the seasonal adjusted series of housing credit production with TS , and X13. We can observe that TS resulted in a stronger smoothing than X13 for the trend of the series. This result drives us to say that we lose more information with TS which might be an issue when we dating of the business cycles of our series. Nevertheless, we will do the dating of business cycles for both trend series, and we will compare the results, but we think that the dating cycles with X13 trend will be more interesting.  In a second time, we will do the dating of growth cycle. To do the datating of growth cycle, we gonna do the first differentiation for each series to have the variation between 2 periods, as opposed to the actual values of the series.

The table 2 summaries both methods employed to make the seasonal adjustment of our series. We can see that both methods respect the normality of residuals. They both detect one LS type outliers in February 2012. That is different from the previous analysis where we found 2 outliers, one in 2012, but in January not in February.

We can note that the ARIMA modelization between both methods is different. TS uses a simpler form than X13. TS has just 2 parameters against 3 for X13. TS has estimated 2 moving average parameters (MA), based on the mean of the series, when X13 has estimated 3 auto-regressives parameters which take in count the past of the series. Moreover, TS only did one seasonal differentiation while X13 also added a differentiation on the classical part.

On the AIC, AICC and BIC criterion, both models are close but TS has the best BIC value (-4.53 vs 14.59). Moreover, the SD error of the TS ARIMA model is close to 0 when the one of

X13 ARIMA model is really high (1275). Table 3 shows the coefficient estimated by both methods. According to this table, we can note that for TS ARIMA model, the influence of the past is quite high, the coefficient is equal to -0.91 for the classical part. For the seasonal side, the past have less weight, but still has significant impact, with a coefficient of 0.7.

Table 2. Summaries of X13 and TS seasonal adjustment method

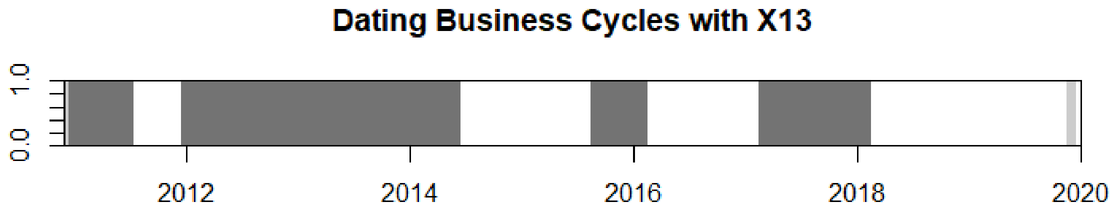
| | X13 | Tramo Seats |
|---|---|---|
| ARIMA model | [(0,1,2)(0,1,1)] | [(1,0,0)(1,1,0)] |
| outliers | 1 detected: 2-2012, LS | 1 detected: 2-2012, LS |
| AIC , AICC, BIC | AIC:1649.62, AICC:1650.91, BIC:14.59 | AIC:1662.83, AICC:1663,78, BIC:-4.53 |
| Log likelihood | -817.81 | 87.88 |
| Normality of residuals | 0.63 | 0.4339 |
| SD error | 1275.92 | 0.09 |

Table 3. Coefficients estimated by both methods

| X13 | | | | TS | | | |
|---|---|---|---|---|---|---|---|
| | Coefficients | T-Stat | P[\|T\| > t] | | Coefficients | T-Stat | P[\|T\| > t] |
| Theta(1) | -0,2289 | -2,54 | 0,0128 | Phi(1) | -0,9154 | -22,71 | 0,0000 |
| Theta(2) | 0,5533 | 6,16 | 0,0000 | BPhi(1) | 0,7011 | 10,34 | 0,0000 |
| BTheta(1) | -0,6478 | -6,86 | 0,0000 | | | | |

### I.3 Cyclic analysis

Figure 7. Dating business cycles with X13 and TS



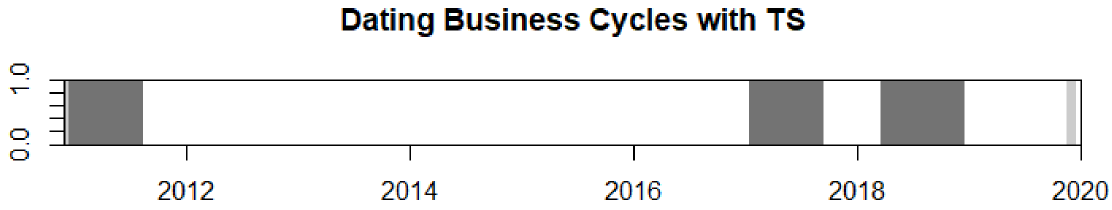




Table 4. Dating business cycles with X13

| Phase | Start | End | Duration in month | Level at start | Level at end | Amplitude |
|---|---|---|---|---|---|---|
| R | - | 2011M8 | - | - | 9068 | - |
| E | 2011M8 | 2012M1 | 5 | 9068 | 13730 | 4661.5 |
| R | 2012M1 | 2014M7 | 30 | 13730 | 9122 | 4607.1 |
| E | 2014M7 | 2015M9 | 14 | 9122 | 18185 | 9062.2 |
| R | 2015M9 | 2016M3 | 6 | 18185 | 12535 | 5649.6 |
| E | 2016M3 | 2017M3 | 12 | 12535 | 21218 | 8683.4 |
| R | 2017M3 | 2018M3 | 12 | 21218 | 14646 | 6572.4 |
| E | 2018M3 | - | - | 14646 | - | - |

Note: E= expansion, R=recession

Table 5. Dating business cycles with TS

| Phase | Start | End | Duration in month | Level at start | Level at end | Amplitude |
|---|---|---|---|---|---|---|
| R | - | 2011M9 | - | - | 9563 | - |
| E | 2011M9 | 2017M2 | 65 | 9563 | 17957 | 8394.1 |
| R | 2017M2 | 2017M10 | 8 | 17957 | 16194 | 1762.7 |
| E | 2017M10 | 2018M4 | 6 | 16194 | 16423 | 228.4 |
| R | 2018M4 | 2019M1 | 9 | 16423 | 16146 | 277.1 |
| E | 2019M1 | - | - | 16146 | - | - |

Figure 8. First differentiation of CVS by TS and smoothing with HP filter

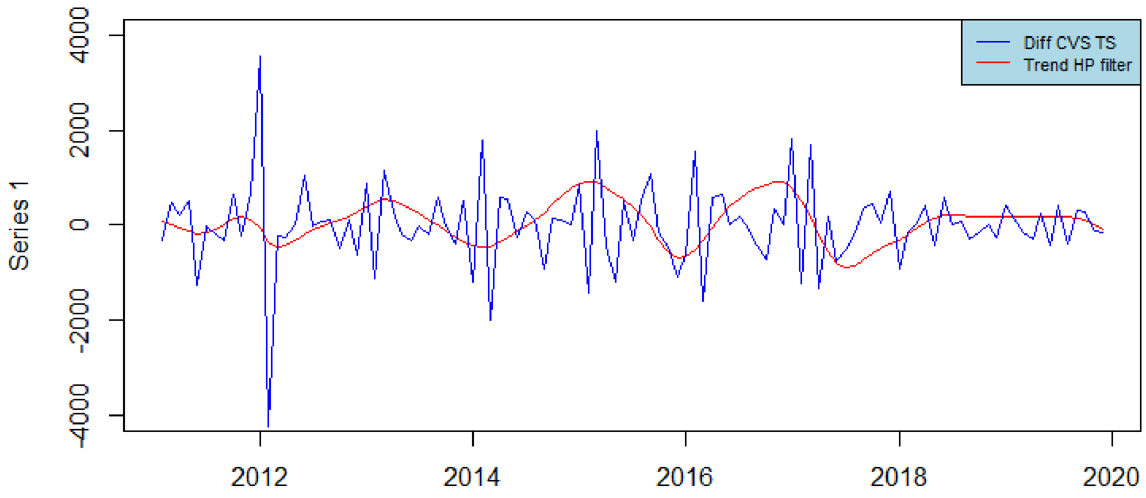



Figure 9. First differentiation of CVS by X13 and smoothing with HP filter

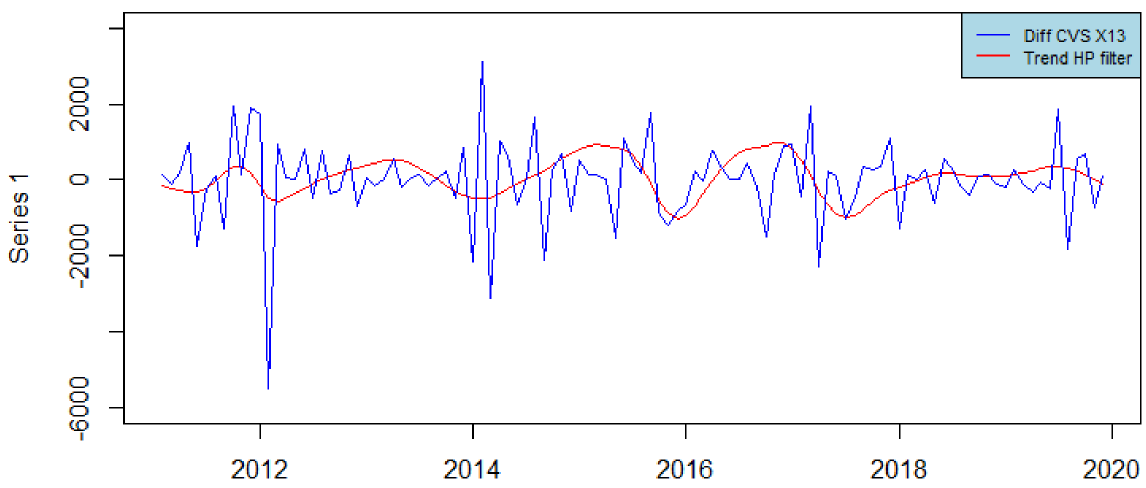



Figure 10. Dating acceleration cycles with X13 and TS

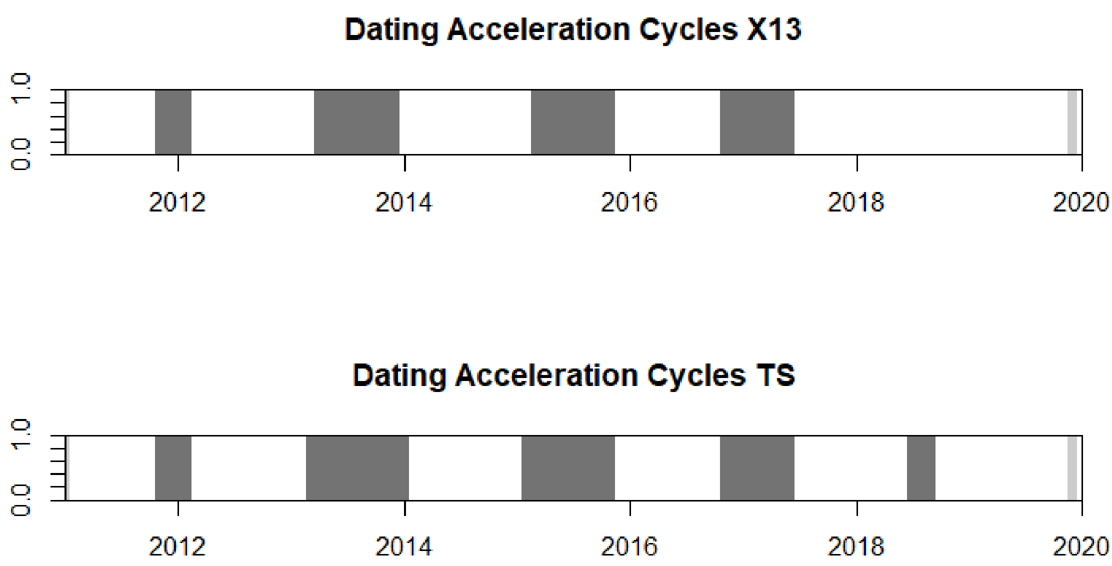


Table 6. Dating acceleration cycles with X13

| Phase | Start | End | Duration in month |
|---|---|---|---|
| E | - | 2011M11 | - |
| R | 2011M11 | 2012M3 | 4 |
| E | 2012M3 | 2013M4 | 13 |
| R | 2013M4 | 2014M1 | 9 |
| E | 2014M1 | 2015M3 | 14 |
| R | 2015M3 | 2015M12 | 9 |
| E | 2015M12 | 2016M11 | 11 |
| R | 2016M11 | 2017M7 | 8 |
| E | 2017M7 | 2018M7 | 12 |
| R | 2018M7 | - | - |

According to table 4, with the trend obtained with X13, we can tell that our seasonal series contains 8 business cycles (4 expansion cycles and 4 recession cycles). In our case, expansion means that the production of housing credit in growing during all the phase of the cycle. Moreover, we can note that the production of housing credit in France; from March 2018 to December 2019 (last date of the series) is in a expansion phase. To have a more detailed analysis concerning this expansion phase, we have to look at the acceleration cycles, summarized in table 6.

The theory suggests that a peak in an expansion phase, is always preceded by a peak in the acceleration cycle. But, if we have a peak in the acceleration cycle during this phase of expansion, it is means that the expansion phase is potentially in a deceleration phase, i.e simply that the growth of the housing credit production will eventually fall down to 0, and then, it will enter a recession phase. Figure 9 and 10 represent the first differentiation of both CVS computed by X13 and TS, as well as the trend obtained with an Holdrick-Prescott smoothing filter. As we can see, at the end of the acceleration cycles, the trend is going down.

According to table 6, we can note that we are in this case. Actually, the expansion phase started in march 2018, and the phase of deceleration of the series occurred in July 2018 so 4 months later. Therefore, we can say that the expansion phase of the housing credit production in France, will not continue in 2020.

Why do we think that the expansion phase of housing credit production will not continues in 2020?

In one hand, if we assume that the expansion of the housing credit production stops in December 2019, the duration of this expansion phase will be around 21 months. Hence it would be the largest expansion phase of our series since 2011. Maybe this phase could end after December 2019, but we think this phase will not continue beyond February 2020, simply because currently, France is facing an economic and sanitary crisis. Therefore, the expansion phase is unlikely to continue in 2020. Another argument that w can add is the following. Because we do not have more data, we cannot date the end of this expansion phase, maybe the peak has already been reached. This idea is plausible because we usually observe 12 months durations for each cycles in our series.

On the other hand, our analysis is supported with the dating of business cycles and acceleration cycles with trend done with TS. If we look at table 5. This expansion phase is dated at January 2019, although the trend done with TS is quite different of the one done with X13, both methods arrive to the same conclusion: the housing credit production is in an expansion phase. Moreover, there is a peak in the acceleration cycle, at the same period as dated with TS (view table 7). This is comforting the idea that the expansion phase will soon encounter a peak, and enter in recession.

Concerning the characteristics of our business cycles, with X13, our cycles have a duration of 5 months at least, but in most of the cases, the duration is around a year (12 months). There is just one recession phase which has a duration of 30 months (2 years and a half). Moreover, we can note that the amplitudes are large (around 6000 million per cycle), Indeed, the amplitude of each cycle represent between 1/3 and a 1/2 of the level of the series at the start of the cycles, which implies a large cycle amplitude. This number drive us to say that business cycles are correctly dated, in opposition with the dating with TS, because the characteristics are different. There are 2 phases with an amplitude superior to 1500 million, and the last 2 phases have a low amplitude (around 250 million). Moreover, with the TS method, we only have one expansion phase between the end of 2011 and the start en 2017, while with X13 dating, we have a recession phase about 30 months. These results are completely different (see figure 8). Hence, we will retain the dating done with X13 because it seems more precise and interesting to sum up the business cycles.

Table 7. Dating acceleration cycles with TS

| Phase | Start | End | Duration in month |
|---|---|---|---|
| E | - | 2011M11 | - |
| R | 2011M11 | 2012M3 | 4 |
| E | 2012M3 | 2013M3 | 12 |
| R | 2013M3 | 2014M2 | 11 |
| E | 2014M2 | 2015M2 | 12 |
| R | 2015M2 | 2015M12 | 10 |
| E | 2015M12 | 2016M11 | 11 |
| R | 2016M11 | 2017M7 | 8 |
| E | 2017M7 | 2018M7 | 12 |
| R | 2018M7 | 2018M10 | 3 |
| E | 2018M10 | 2019M1 | 3 |
| R | 2019M1 | - | - |

## II. SEASONAL SERIES AND FORECASTING

We then proceed to forecasting. We will use a particular method which requires two different datasets, one used for estimation and in-sample adjustment quality control. The other one is the testing set, used for forecasting quality assessment. The first dataset contains most of the values, these values are used to estimate different models. Once we settle on a final, or a few final models, we forecast 12 values based on the model we chose. These 12 forecasts will then be compared to actual values kept in the testing set. This will give us an idea of the prediction quality of the different models. The forecast are produced in a rolling fashion, meaning we produce 12 1-step-ahead forecast and we add the observed value back to the original dataset at each step.

Here, the testing set will contains values for the year 2019 (the series is monthly, hence we have 12 values for 2019). The training set will contains every other observation, from 2011 to 2018, for a total of 96 values. We use these values to estimate and adjust 7 different models, before producing 12 rolling forecasts. We will try to fit a SARIMA(0,1,1)*(0,1,1) s = 12, a SARIMA(p,d,q)*(P,D,Q) using the auto.arima function in R. The other forecasts will be computed through other types of models and exponential smoothing; such as TBATS (**T**rigonometric seasonality, **B**ox-Cox transformation, **A**RMA errors, **T**rend and **S**easonal components.), ETS (Exponential Smoothing State Space Model), STL (Seasonnal and Trend decomposition using Local regression) and NNETAR (neural netwotk fitted non-linear autoregression), and finally; Holt-Winters filtering.

The first SARIMA model, (0,1,1)*(0,1,1) is quite close to the SARIMA model fitted by JDemetra + using the X-13 method. Indeed, the X-13 method gave us a (0,1,2)*(0,1,1). However the (0,1,1)*(0,1,1) has a slightly smaller Log-Likelihood value, indicating a better good-of-fit. The AIC of this model also is smaller (1480.24 against 1649.62 in the (0,1,2)*(0,1,1) model). The form of these two models is quite similar with one differenciation for the classical part and one for the seasonnal part and using only MA terms. The model chosen by the Tramo-Seats method on the other hand is different. First, this model ((1,0,0)*(1,1,0)) has no differentiation term for the seasonnal part and is only composed of AR parameters as opposed to the 2 other SARIMA models wee looked at before. It also has a larger AIC (1662.83), it seems this model has the worst goodness-of-fit between these 3 models. We also fitted a SARIMA (2,1,1)*(2,1,2) on our data.  This model has a log-likelihood value of 735.36 and an AIC of 1486.71. So far, this model is the best SARIMA model in-sample. But we will have to see if we are not over-fitting. Good in-sample accuracy measures do not necessarily mean good prediction as the model might not be stable over time.

The auto.arima() function gave us a ARIMA(2,1,2) model even though our data displays a strong seasonal persistence as we can notice on the PACF bellow (see figure 11). We used the BIC as the optimization criterion in order to obtain a more stable model. However, even by using the AIC the model computed by R has no seasonal component. We can compare this model to the SARIMA (0,1,1)*(0,1,1) s = 12 that was also fitted to our data. Both SARIMA models are summarized in table 8.

Figure 11. ACF of housing credit production series

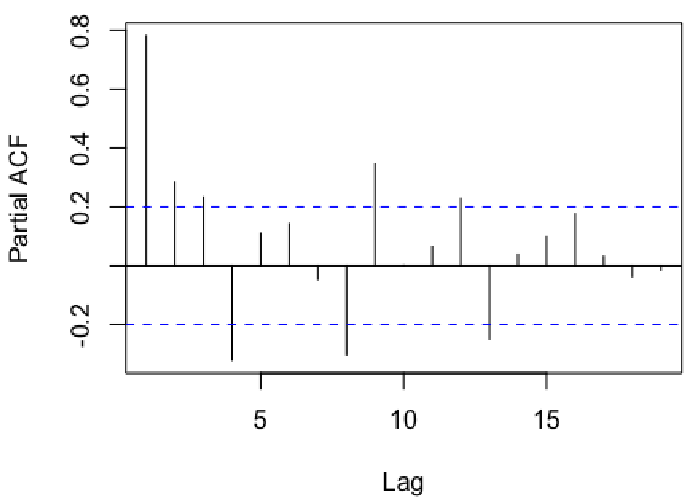


Table 8. Summary of ARIMA Models

```
arima(x = Data[1:96, 2], order = c(0, 1, 1), seasonal = list(order = c(0, 1,1), period = 12))

Coefficients:
         ma1      sma1
      -0.1203  -1.0000
s.e.   0.0930   0.3256

sigma^2 estimated as 2246514:  log likelihood = -737.12,  aic = 1480.24

Training set error measures:
ME       RMSE   MAE     MPE      MAPE   MASE              ACF1
-26.1401 1393.666 970.2492 -0.8079295 7.391025 0.481619 -0.002610507
```

```
ARIMA(2,1,2)

Coefficients:
        ar1       ar2      ma1      ma2
      -0.9651  -0.9714  0.7542  0.8624
s.e.   0.0394   0.0327  0.1061  0.0813

sigma^2 estimated as 4072069:  log likelihood=-857.04
AIC=1724.08   AICc=1724.75   BIC=1736.85

Training set error measures:
ME       RMSE   MAE     MPE      MAPE   MASE              ACF1
69.73232 1964.684 1507.397 -0.7367568 12.08164 0.748252 -0.08298078
```

The ARIMA (2,1,2) chosen by R is apparently worse in terms of adjustment quality. The AICc has a smaller value in the (0,1,1)*(0,1,1) model, same goes for the Log-Likelihood (in absolute value). The different error measures also seem to indicate that the model chosen by auto.arima() is less precise in-sample. Because we want a SARIMA model we cannot use the the auto.arima() function which drops the seasonal part. We try a few different models and we settle for a SARIMA (2,1,1)*(2,1,2) s = 12, summarize in table 9.

Table 9. Summary of SARIMA (2,1,1)*(2,1,2) s = 12

```
arima(x = Data[1:96, 2], order = c(2, 1, 1), seasonal = list(order = c(2, 1,2), period = 12))
Coefficients:
         ar1     ar2      ma1     sar1    sar2      sma1      sma2
      0.1227  0.1950  -0.2808  0.0015  0.1563  -0.9481  -0.0515
s.e.  0.3136  0.1152   0.3032  0.5033  0.1564   0.5368   0.4714
sigma^2 estimated as 2244577:  log likelihood = -735.36,  aic = 1486.71
```

Training set error measures:

| ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|---|---|---|---|---|---|---|
| -40.26682 | 1393.065 | 987.2185 | -0.7752446 | 7.571835 | 0.4900424 | 0.002740171 |

This new model is slightly better than the SARIMA (0,1,1)*(0,1,1). We have a smaller AIC and Log-Likelihood as well as smaller RMSE.

We will also use a naive prediction as a benchmark to assess whether or not a model is better at predicting in the sense that it gives more information than a naive prediction.

The table 10 shows forecasts using the 7 different methods, as well as the actual values from the testing set.

Table 10. Forecasting out-of-sample of all models

| | SARIMA | Auto_SARIMA | TBATS | STL | ETS | NNETAR | HW | Naive | Observed |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 14253.71 | 14714.22 | 15860.36 | 14619.30 | 15863.76 | 15850.00 | 16394.76 | 15855.01 | 15855.01 |
| 2 | 15423.57 | 15427.50 | 16249.55 | 15571.33 | 16230.73 | 16234.41 | 18220.62 | 16277.26 | 16277.26 |
| 3 | 14910.10 | 13482.19 | 16372.40 | 14676.82 | 16364.13 | 16385.53 | 14287.53 | 16380.83 | 16380.83 |
| 4 | 15379.41 | 15603.43 | 16325.10 | 15152.72 | 16326.28 | 16394.71 | 13665.15 | 16321.48 | 16321.48 |
| 5 | 18355.01 | 17465.92 | 16564.25 | 18298.98 | 16551.75 | 16488.00 | 16956.51 | 16580.71 | 16580.71 |
| 6 | 19880.19 | 19762.43 | 16684.74 | 19710.20 | 16676.37 | 16622.40 | 18995.17 | 16693.10 | 16693.10 |
| 7 | 18928.37 | 19195.42 | 18780.62 | 18092.16 | 18651.71 | 18878.75 | 18565.81 | 18913.23 | 18913.23 |
| 8 | 17947.82 | 18778.71 | 17505.68 | 18006.32 | 17515.79 | 17420.12 | 17344.29 | 17387.77 | 17387.77 |
| 9 | 18325.12 | 18989.40 | 18136.14 | 18066.93 | 18119.34 | 18167.37 | 18922.62 | 18192.54 | 18192.54 |
| 10 | 17377.26 | 16117.92 | 18948.17 | 17104.52 | 18908.45 | 18318.67 | 16920.74 | 19018.39 | 19018.39 |
| 11 | 19944.39 | 19418.04 | 18362.60 | 20303.70 | 18386.73 | 18414.05 | 21149.56 | 18309.06 | 18309.06 |
| 12 | 17477.52 | 17204.99 | 18379.20 | 17360.62 | 18322.74 | 18399.92 | 17688.10 | 18315.30 | 18315.30 |

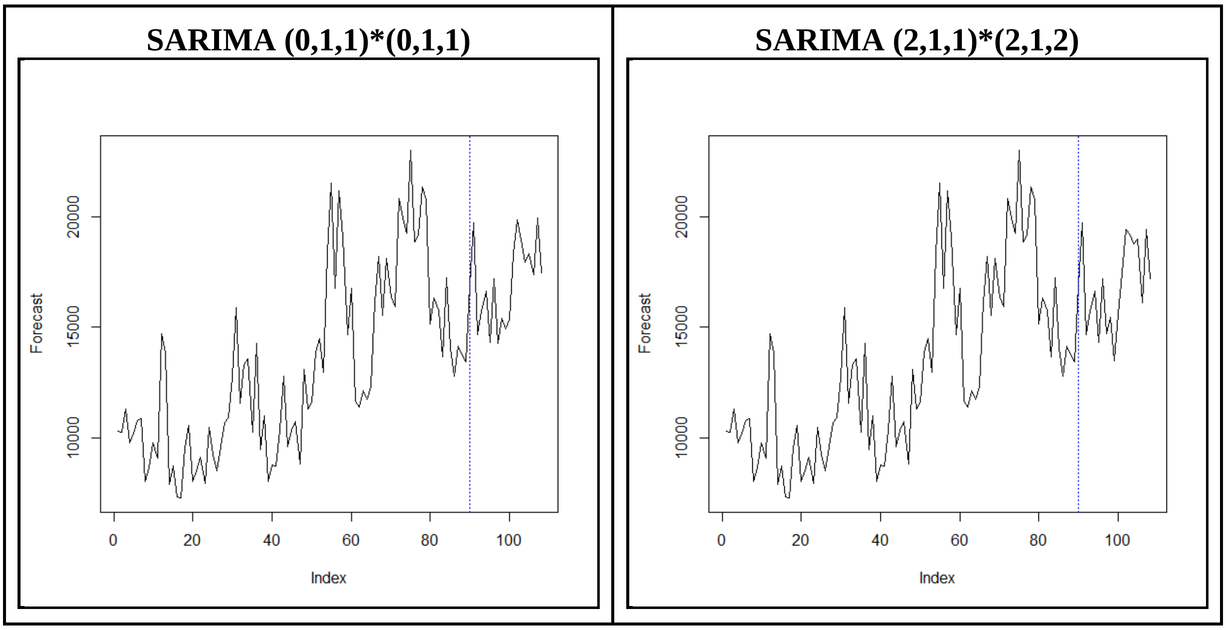
We can now compare the different forecasting qualities. We computed the Mean Squared Error, Mean Error and Mean Absolute Error for each model. The naive prediction has one the best forecasting quality so far. It has the smallest MSE, MAE and ME. Unfortunately, none of the models outperform the benchmark naive prediction. Holt-Winters seems to give the worst forecasts (see table 11).

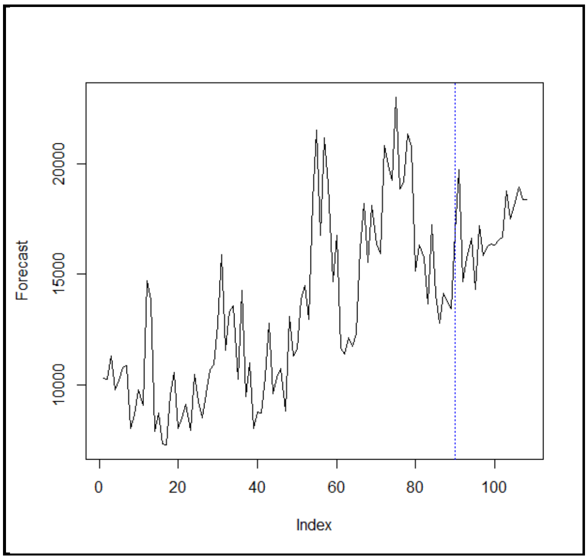
Table 11. Forecasting error measures

| | SARIMA (0,1,1)*(0,1,1) | SARIMA (2,1,1)*(2,1,2) | TBATS | ETS | STL | NNETAR | HW | Naive | ARIMA X13 | ARIMA Seats |
|---|---|---|---|---|---|---|---|---|---|---|
| **MSE** | 5672676 | 6056474 | 3431118 | 3561438 | 6738730 | 3553987 | 7461421 | 3288662 | 7910814 | 3139217 |
| **ME** | 5.456249 | 202.6709 | 8.261204 | 29.1822 | 108.694 | 47.438 | -70.2421 | 1.939346 | -82.35 | 320.18 |
| **MAE** | 1995.597 | 2059.901 | 1524.402 | 1545.19 | 2155.42 | 1580.209 | 2297.83 | 1496.773 | 2402.022 | 1399.695 |

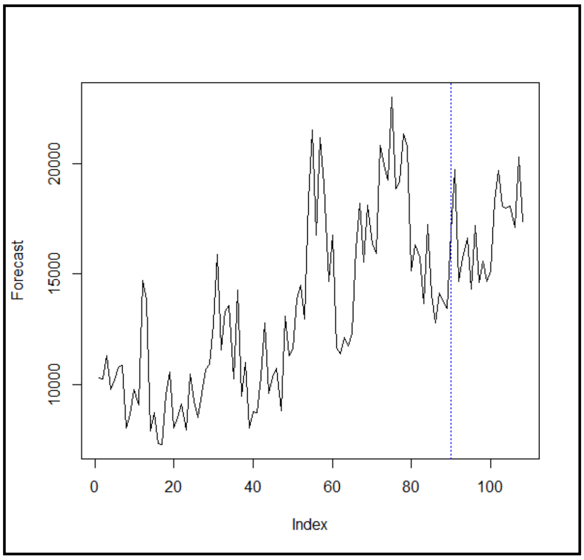
Figure 12 shows the different results from each forecasting method:

Figure 12. Housing credit production and forecasting out-sample done by each model



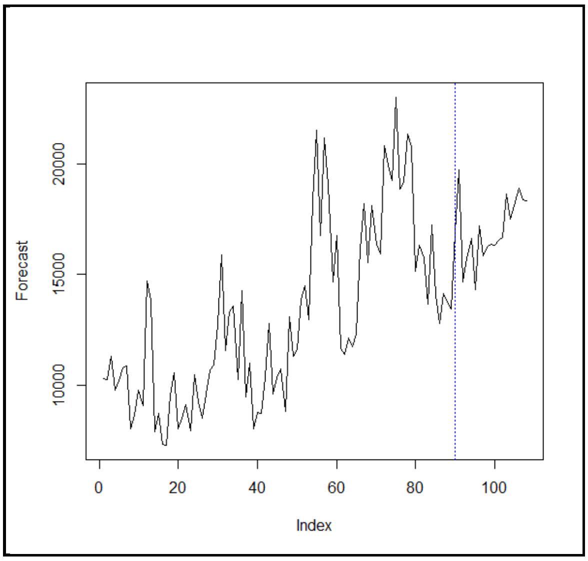



**TBATS**

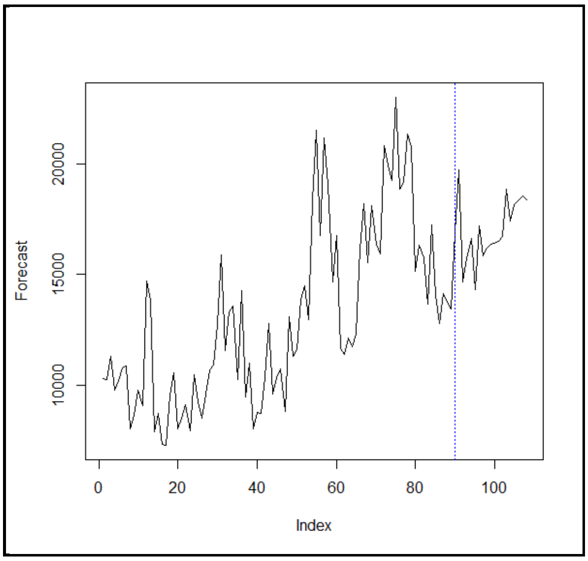


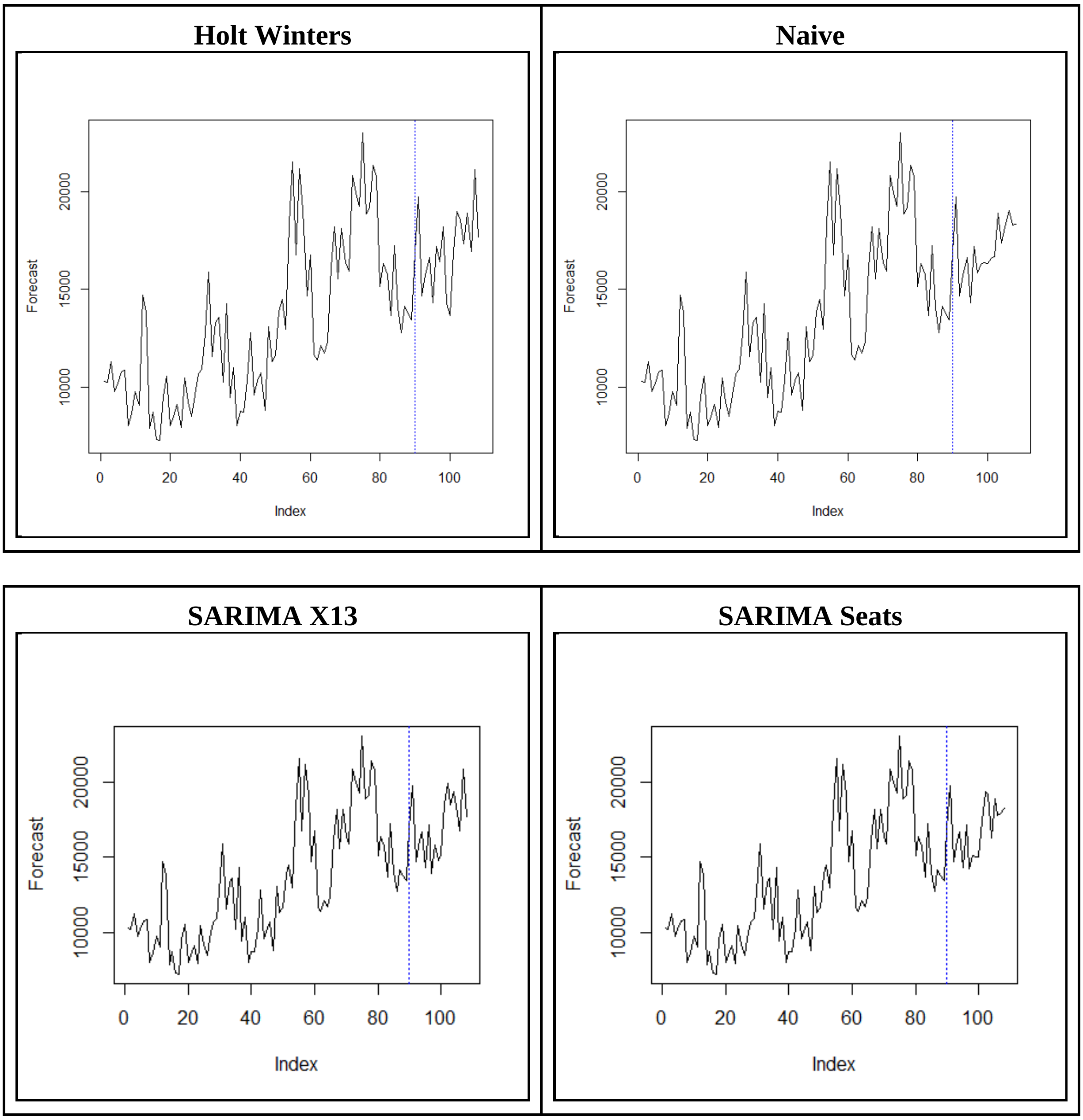

**STL**

**ETS**




**NNETAR**






We will now use the Diebold & Mariano test to determine if the predictions are statistically different depending on the model considered.

multivariate Diebold-Mariano test

data:  Data[97:108, 2] and Forecasts

statistic = 0.2011, lag length = 5, p-value = 3.93e-06

alternative hypothesis: Equal predictive accuracy does not hold.

The null hypothesis assumes all models have the same predictive power, the alternative hypothesis assumes they do not. As we can see we can reject the null hypothesis with a 1% risk. The models' predictive accuracy differ. We can run individual tests on each model. Here the null hypothesis is the same as before: the two models have the same predictive accuracy. The alternative is that model 2 is less accurate than model 1. All p-value of the DM test are summarize in table 12.

Table 12. Dieblod Mariano test on out-sample forecasting

| MODEL 2 | MODEL 1: Naive | SARIMA (0,1,1)*(0,1,1) | SARIMA (2,1,1)*(2,1,2) | TBATS | STL | ETS | NNETAR | HW |
|---|---|---|---|---|---|---|---|---|
| **NAIVE** | | | | | | | | |
| **SARIMA (0,1,1)*(0,1,1)** | 0.0217 | | | | | | | |
| **SARIMA (2,1,1)*(2,1,2)** | 0.0328 | 0.3593 | | | | | | |
| **TBATS** | 0.1058 | 0.971 | 0.9588 | | | | | |
| **STL** | 0.0097 | 0.079 | 0.311 | 0.11 | | | | |
| **ETS** | 0.1017 | 0.9613 | 0.948 | 0.0101 | 0.9893 | | | |
| **NNETAR** | 0.1026 | 0.9766 | 0.9676 | 0.3374 | 0.991 | 0.597 | | |
| **HW** | 0.0179 | 0.1428 | 0.2409 | 0.0207 | 0.3232 | 0.024 | 0.0181 | |
| **SARIMA X13** | 0.008 | 0.015 | 0.09 | 0.009 | 0.1028 | 0.0102 | 0.007 | |
| **SARIMA Seats** | 0.60 | 0.99 | 0.98 | 0.67 | 0.9918 | 0.7216 | 0.7277 | |

We notice that the Naive prediction is statistically better from all the other models' prediction. It is statically better than the two SARIMA models and also better than the Holt-Winters Filtering and the STL method at a 5% risk level. We also notice the SARIMA models performing worse than almost every other models (p-values higher than 0.1). Naive prediction had the smallest out-of-sample error measures and also produces statistically different forecasts than the other models. We can conclude it is the best model to forecast this series.

## III. NON SEASONAL SERIES AND FORECASTING

### III.1. Linear forecasting

The goal is to produce one-step-ahead rolling forecasts using different methods as we did in the second part. We will use an AR(p), an AR(1) and a SARIMA(p, d, q)*(P, D, Q). Both AR(p) and SARIMA orders will be given by the auto.arima function from R, and also by the visual analysis of the ACF & PACF functions. The first 96 observations are used for modelling and the remaining 12 will be used for forecasting.

First, we need to check for stationarity, as the AR(.) models need to be fitted on stationary data in order to produce reliable forecasts. We apply the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. Both tests are designed to detect the presence of a unit root in the autocorrelations in our data. For ADF, we test H0: presence of a unit root against Ha: no unit root. In other words; H0: the data isn't stationary against Ha: the data is stationary. KPSS works the other way around with H0: the series is stationary against Ha: the data is not stationary.

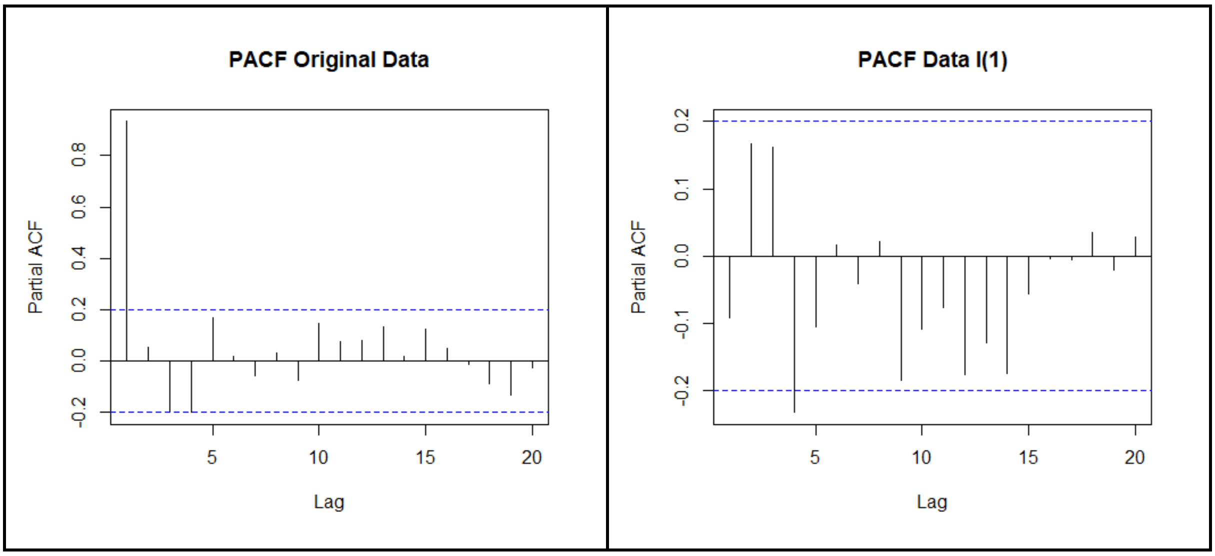
Table 13. ADF and KPSS test on housing credit production CVS series

| | ADF p-value | KPSS p-value | Result |
|---|---|---|---|
| Original Data | 0.07258 | < 0.01 | stationary at a 10% confidence interval |
| Differentiated Data I(1) | < 0.01 | > 0.1 | data is stationary |

According to table 13, the results from ADF on the original series suggests stationarity with a 10% risk but KPSS rejects the null hypothesis of stationarity. Because the p-value from ADF is quite close to 5% we decide to differentiate the serie once. Once the data has been integrated, both tests confirm the stationary hypothesis. The decision of keeping either the original data or the differentiated data is quite subjective. In the next part, we will use the original data to fit an ARX, so we decide to trust the ADF test and to use the original data.

However we will try to fit models on both dataset (original and differentiated) to see which performs better in-sample. We decide to fit the AR(p) model using auto.arima() function, and also using the PACF, to determine the required order (see figure 13).

Figure 13. PCAF on CVS and differentiated series




The PACF tells us that the series displays a strong persistence. The first autocorrelation is highly statistically significant, as well as the third and fourth lagged values. The PACF of the first order difference suggests that the data behave as a completely random process, hence impossible to model. Therefore we will start using the original dataset. The pattern of the PACF indicates a AR(1) model might be appropriate.

We first decide to fit an AR(1) model on our original dataset (non-seasonally adjusted), before trying to fit an AR(p) model using auto.arima().

arima(x = Data[1:96, 2], order = c(1, 0, 0))

Coefficients:

| | ar1 | intercept |
|---|---|---|
| | 0.9320 | 13318.68 |
| s.e. | 0.0337 | 1670.36 |

sigma^2 estimated as 1592230:  log likelihood = -822.7,  aic = 1651.41

As we can see the autoregressive coefficient is 0.93, close to 1, another proof that we are facing a serie with a lot of time persistence, maybe not stationary. Further, we will compare the adjustment errors to those from other models (SARIMA and Holt-Winters exponential smoothing) along with their residuals.

We decide to test out a AR(p) using the auto.arima() function. The optimization algorithm selects a model based on one of the information criteria (AIC, AICc, BIC). We will use both BIC and AIC, as BIC tends to penalize the model based on the number of parameters used (the AIC does the same but the BIC formula gives more weight to the number of estimated parameters). BIC will give a more parsimonious model than AIC most of the time. Because we are trying to forecast future values, having a parsimonious model is essential. Indeed, the more parsimonious the model, the more stable it is.

Once again, we are not perfectly certain which dataset is the most appropriate (either the original or the integrated one). To answer that question, we use the auto.arima() function on the original data in order to find an ARIMA(p, d, 0) model. The idea is to check whether or not R decides to differentiate the data if we just feed the optimization algorithm with raw data.

ARIMA(0,1,0)

sigma^2 estimated as 1653195:  log likelihood=-814.91

AIC=1631.83   AICc=1631.87   BIC=1634.38

It seems that our serie cannot be modelled by an AR(p) process. The auto.arima() function cannot find a suitable model no matter what information criterion we use. The model is a (0, 1, 0) that contains no autoregressive parameters. Our serie may be following an MA(q) process, we will test this hypothesis as we try to fit an ARIMA(p, d, q) model. However, an interesting point is that R decided to differentiate the serie once, this is an indication that our data is probably not trend-stationary.

Because we won’t be able to find an appropriate AR(p) model, we will try to fit an ARIMA(p, d ,q). We use the auto.arima() function once again. We face the same problem, even if we allow MA parameters, the final model is a null model automatically fitted on the first order difference. The issue is that the original data is not trend-stationary, but the first order difference follows a white noise process. Therefore finding a linear time series model is difficult nay impossible.

One option is to force R to apply an ARIMA model on the raw data. If we let the auto.arima() function choose the parameters, the algorithm decides to differentiate the serie once and then finds an ARIMA(0,1,0) again. That is why we force R to choose d = 0 (use the original dataset) to find a suitable ARIMA model.

ARIMA(1,0,3) with non-zero mean

Coefficients:

| | ar1 | ma1 | ma2 | ma3 | mean |
|---|---|---|---|---|---|
| | 0.8658 | 0.0227 | 0.3645 | 0.2025 | 13315.933 |
| s.e. | 0.0622 | 0.1273 | 0.1118 | 0.1231 | 1340.007 |

sigma^2 estimated as 1493279:  log likelihood=-817.3

AIC=1646.61   AICc=1647.55   BIC=1661.99

By forcing R to apply a ARIMA on the raw data, we find another model. This new model is a (1, 0, 3). The autoregressive coefficient is close to 1 (0.87) as we first observed in the AR(1) model. The model has three moving average parameters. But we can also notice that this model as a larger Log-Likelihood than the null model, which seems to indicate that this model has very little predictive power (see table 14).

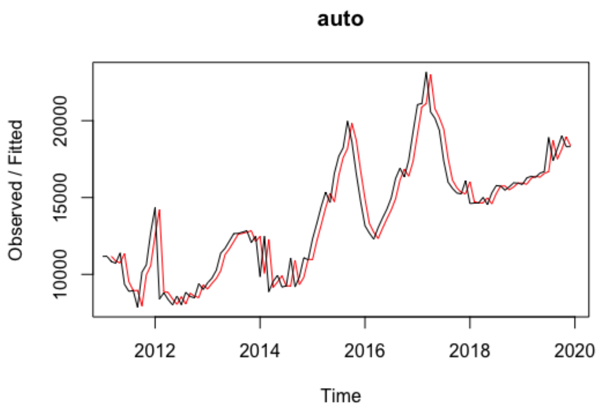
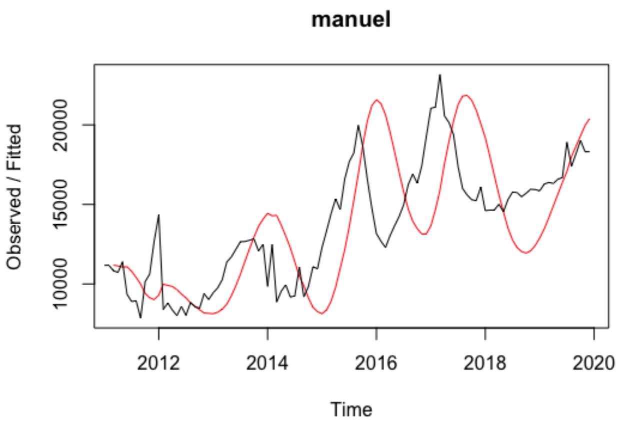
Table 14. Comparison of ARIMA criterion

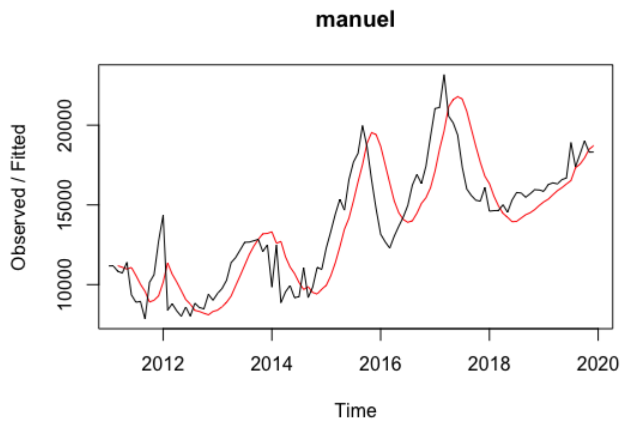
| | **AIC** | **BIC** | **Log-Likelihood** | **In-Sample RMSE** | **In-Sample ME** |
|---|---|---|---|---|---|
| **AR(1)** | 1651.41 | NA | -822.7 | 1261.836 | 34.45184 |
| **ARIMA(1, 0, 3)** | 1646.61 | 1661.99 | -817.3 | 1189.749 | 21.31336 |

We tried different models some with fixed number of parameters, some fitted based on the AIC or BIC criteria. We were not sure whether or not our data was stationary, so we decided to use the optimization algorithm to determine if the data needs to be differentiated. We found that R used I(1) data everytime, indicating that the data wasn't stationary. Our data is non-stationary (at 5% risk level). Meanwhile, the first order difference seems to follow a random pattern which cannot be modelled. To find an appropriate model we had to accept the stationarity hypothesis at a 10% risk level and fit an ARIMA model on the data. We will now compare the residuals and the goodness-of-fit of these 2 differents parametric models. We excluded the null models. So far, we only tried parametrics methods, we will now be using a non-parametric forecasting technique (Holt-Winters double exponential smoothing, shorten as HW). Afterward, we will compare the forecast quality of these different models.

We first apply an automatic non-seasonal HW filter (gamma = FALSE). Automatic filtering parameters are chosen according to the RMSE. This can produce over-fitting, depending on the results we might have to adjust the parameters manually. The figure 14 shows how the filtered series changes depending on the parameters' value.

Figure 14. Holt Winter double exponential smoothing with different alpha and beta parameters

| **Alpha** | **Beta** | **Output** |
|---|---|---|
| **0.9176327** | **0** |  |
| **0.1** | **0.5** |  |

| | | |
|---|---|---|
| **0.25** | **0.15** |  |



As we thought the automatic filtering (alpha = 0.97, beta = 0) may be over-fitting the data, giving a lot of importance to recent observations. The filtered series overlap the original one almost perfectly. Even though the filter fits the in-sample data very well (hence, very small RMSE), these parameters may induce large forecast errors. On the other hand, with an alpha of 0.1 and a beta of 0.5, the filter is a lot smoother. Beta = 0.5 amplifies the variations in the series because it gives a lot of weight to past trends. We decide to settle for alpha = 0.25 and beta = 0.15. With these parameters, we fit the data quite well but the filter is also smooth enough to be of use when trying to forecast out-of-sample values.

Now, we are going to produce 12 one-step-ahead forecast in a rolling fashion. We will use the AR(1) and ARIMA(1,0,3) as well as two of the HW filters stated above. Some of the models we tried out are left out because of their weak in sample adjustment measures. Table 15 shows the out-of-sample RMSE, ME, and MAE from the 4 forecasting models we used.

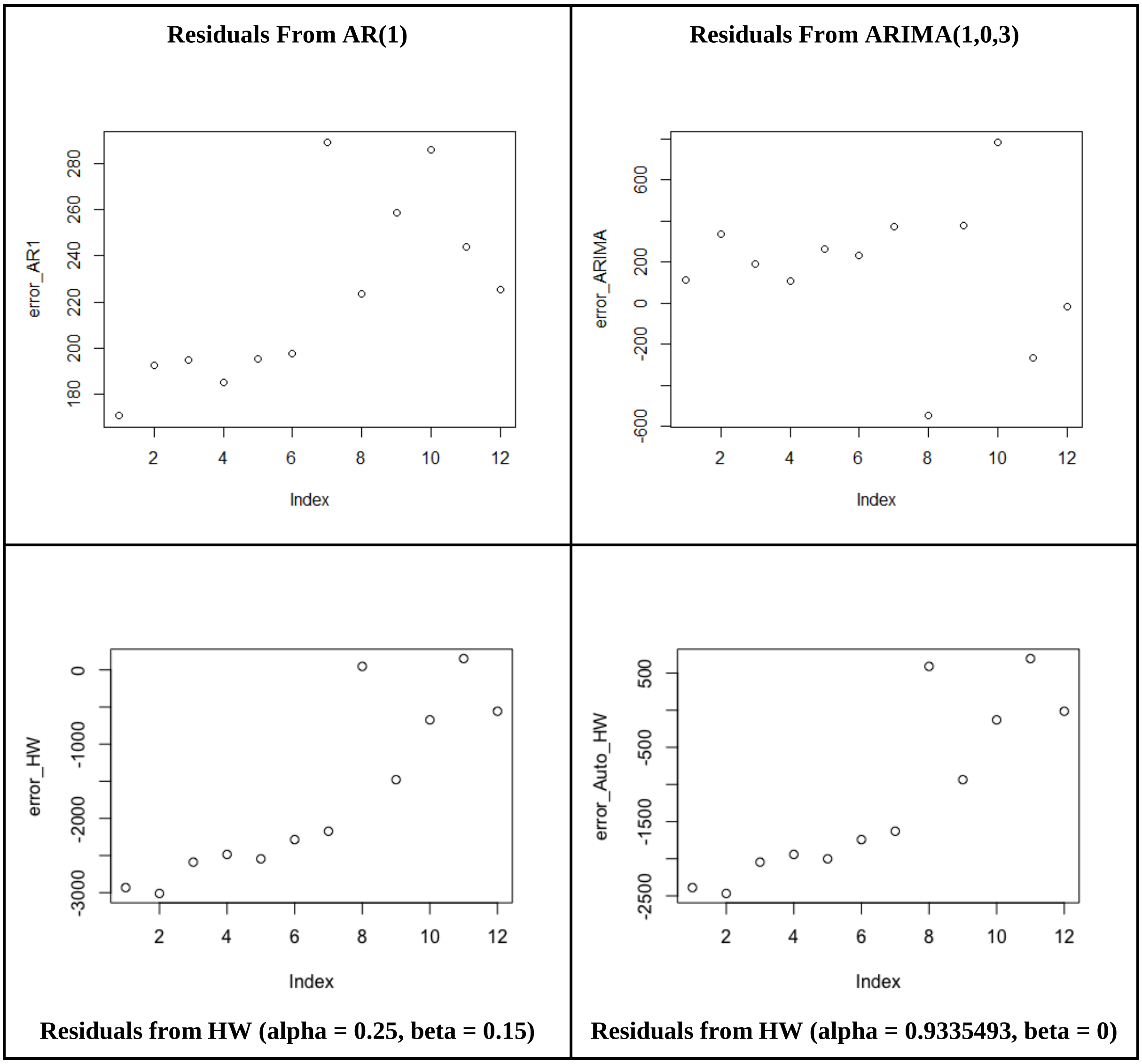
Table 15. Out-of-sample error measures from 4 forecasting models

| | **ME** | **RMSE** | **MAE** |
|---|---|---|---|
| **AR(1)** | 221.8686 | 225.084 | 221.8686 |
| **ARIMA(1, 0, 3)** | 162.0267 | 360.9254 | 300.9103 |
| **HW (alpha = 0.25, beta = 0.15)** | -1512.15 | 1862.69 | 1545.463 |
| **HW (alpha = 0.9335493, beta = 0)** | -969.2475 | 1456.865 | 1183.528 |

We can notice that the ARIMA(1) has the best forecasting accuracy. This is not surprising in the sense that we forced R to use d = 0 to fit an ARIMA model, we know that the data is at least not strongly stationary (ADF test p-value = 0.07764). The main issue with non-stationarity is the model's instability when fitted to out-of-sample data. ARIMA models make the hypothesis that the data is stationary and they do not take into account the trend or seasonality. Even if a model fits well on in-sample non-stationary data, it is usually the case that the model loses its predictive power when fitted out-of-sample. However here, because the data still is stationary at 10%, this might not be an issue for short-term forecasting.

We will then plot the residuals from the differents models summarize in figure 15.

Figure 15. Residuals from ARIMA and HW models



The residuals from the two HW filters have the same pattern, they seem to increase exponentially. However the ARIMA (1, 0, 3) residuals are less concentrated around 0 than the AR(1) model. This is rather logical as the ARIMA (1, 0, 3) model has the largest RMSE and MAE. It has a smaller ME error then the AR(1) model because, as we can see on the residuals plot, the residuals of the ARIMA model are spread between -600 and 600 against 180 and 280 for the AR(1) residuals. This is why we need to use either squarred or absolute errors values when comparing forecasting quality.

We can test for Normality using Jarques-Bera test. The null hypothesis is that the data is Normally distributed and the alternative is that the data follows any other distribution.

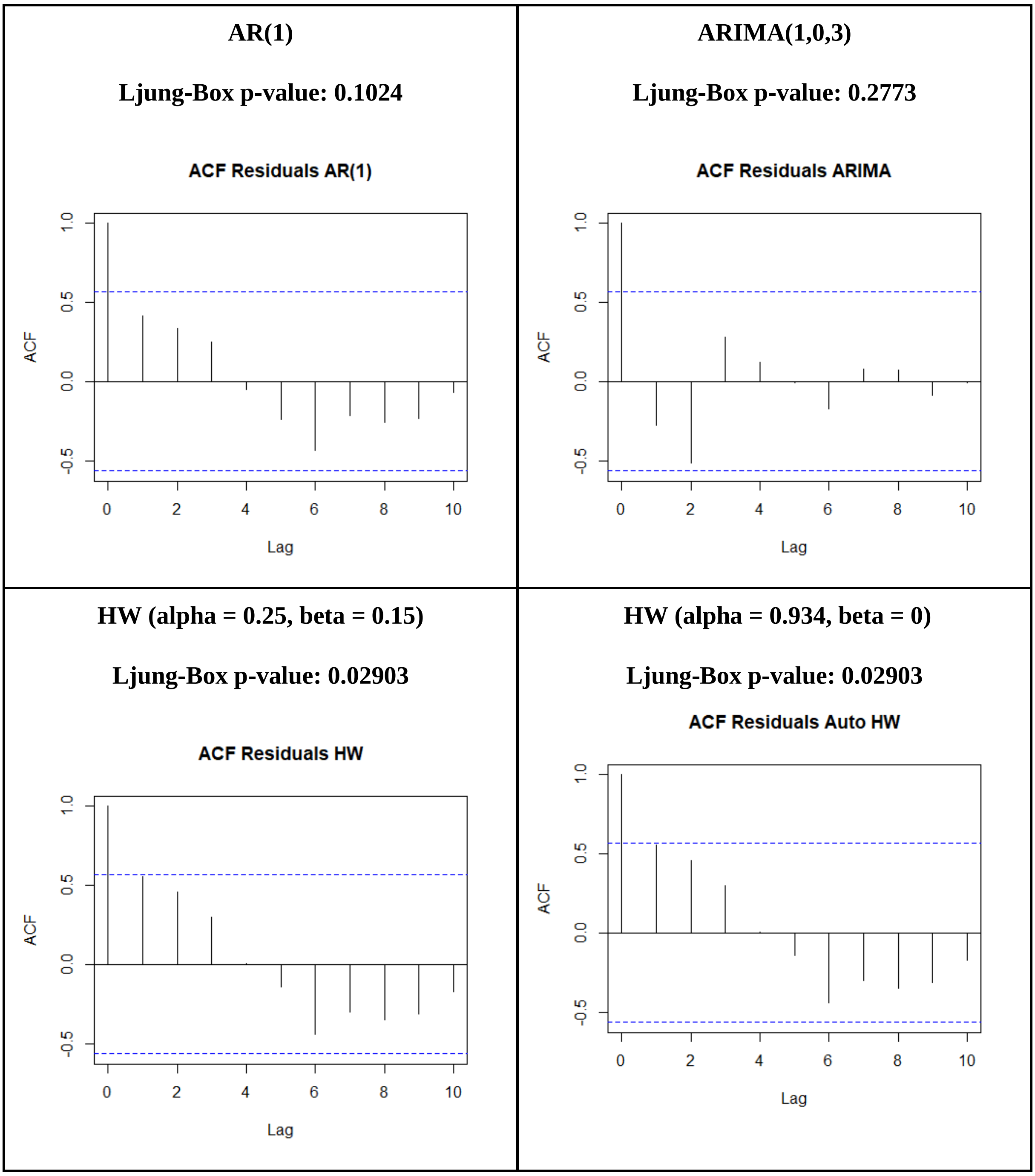
Table 16. Jarque-Bera normality test on ARIMA and HW residuals

| **Model** | **AR(1)** | **ARIMA(1,0,3)** | **HW (alpha = 0.25, beta = 0.15)** | **HW (alpha = 0.934, beta = 0)** |
|---|---|---|---|---|
| **Jarque-Bera p-value** | 0.5741 | 0.7817 | 0.5324 | 0.5324 |

We find that the residuals from the 4 models are Normally distributed which indicates our models function properly and are not spurious (see table 16). We will also check for significant autocorrelations using the ACF and the Ljung-Box test on the residuals summarize in figure 16 .

The ARIMA model has the largest p-value (0.2773) indicating it's residuals displays no significant autocorrelation at lag 1. We also notice that it's ACF is quite erratic, which is consistent with the large p-value. For the other models, we see that the two Holt-Winters filterings residuals show significant autocorrelations at lag 1 while their ACF have a exponentially decreasing pattern. We notice the same elements for the AR(1) model but milder, the ACF pattern is similar but less distinct and the Ljung-Box p-value is slightly larger (0.1024) indicating that we cannot reject the null hypothesis of not significant autocorrelation but we are quite close from rejecting it at a 10% risk level. These informations, especially the remaining time persistence is an indication that our models are not working as well as expected. However they are not surprising taking into account the fact the auto.arima() function led us to use null models. Maybe a linear model nor an exponential smoothing are sufficiently complex to explain our data.

Figure 16. ACF and Ljung-Box test on residuals of ARIMA and HW models



### III.2. Forecasting with regressor

In this section, we will try to forecast the seasonally adjusted housing credit production, with other economic variables. We will test a google trend variable representing the number of research on 'crédit immobilier' in France. The interest to use such a variable is because the data is easily accessible allowing quick forecasting.

*Literature review*

Concerning the economics variables, there is a debate about the effects of economic activity on the credit consumption. On one side, some economists, based on theoretical background, suggest that a strong economic activity would have a positive effect on the expected profit and income for households and firms (non financial). In other words, a strong economic activity enables the agents to face an higher indebtment level, and then, they will be able to finance more consumption and investment through credit (Kashyap, Stein and Wilcox, (1993)).The same reasoning is applicable on the supply side. Expectations of robust economic activity implies higher productivity and demand leading to more profitable projects, and a higher net expected present value, thus the credit demand will be higher.

In contrast, other theoretical arguments contradict previous mechanisms. The argument is the following: an increase in productivity (and not expected productivity), results in an increase of the output and profits (Bernanke and Gertler (1995) and Friedman and Kuttner (1993)).

Following the reasoning, in period of expansion, corporation might prefer financing with internal resources and reduce their demand of credit or external financing. The same way, households, might prefer to reduce their level of debt due to higher income. In recession period, households and firm may increase their demand for credit to smooth the impact of declining income and profits.

Hence, it is difficult to anticipate the impact of any economic activity indicator on the housing credit production. Nevertheless, because we are studying the supply side of credit market here, we think that in period of economic expansion, the production of credit will follow, because it is less risky for banks to grant loans in expansion period than in recession.

To measure the economic activity, we will retain 2 indicators. The choice of these indicators is restricted due to monthly frequency of our data. Hence the growth of GDP is impossible to use as a regressor. Thus we take an indicator of business climate in retail sell and car repairs, built on a monthly survey published by the INSEE. The second indicator of the economic activity is the CAC40 index, which has been used by Levieuge (2017) to forecast bank loans to firms. It has been shown that this indicator is one of the most pertinent indicators of bank loans. It is explained with balance-sheet effects. The stock price of a firm represents the future cash flow that it could generate. According to Levieuge, the stock price accurately represents the quality of the balance sheet of the borrower. In other words, it more or less shows the capacity to pay back. The higher the price, the higher the solvability, according to this reasoning, if the index of CAC 40 goes up, the loans volume has to goes up as well.

We think that this indicator could be useful to explain the housing credit production too. Assuming that some households have saving indexed on financial markets, if the price index is raising, then it could generate a wealth effect that can push households to borrow money to finance housing investments.

In addition, to explain the variation of the housing credit production, we will use long term interest rates as used Calza, Gartner and J. Sousa (2001). Because we are studying the loans market, theoretically, the relation between interest rates and credit volume is positive, i.e the banks are able to produce more loans  because the expected return are higher.

However, nowadays, this reasoning could be reviewed. Indeed, high level of interest rates means, higher expected return on loans for the banks, but on the demand side, it means higher cost of credit. Hence, it is difficult to tell the sign of the relation between the credit production and interest rates. For long term interest rate, we take French treasury bills with a 10 years maturity.

*Descriptive statistics*

As we can see according on figure 17, only the google trend variable (immo) seems to contain outliers. Therefore, we use tsoutliers function on all variables, and 3 variables have been adjusted from this outliers; immo, commerce and long term interest rate (see appendix 2). Moreover, we do the seasonal adjustment of these series (immo and long term interest rate) with seasonal packages. The following summary statistics are done on cleaned datas.

Figure 17. Box plot of variables

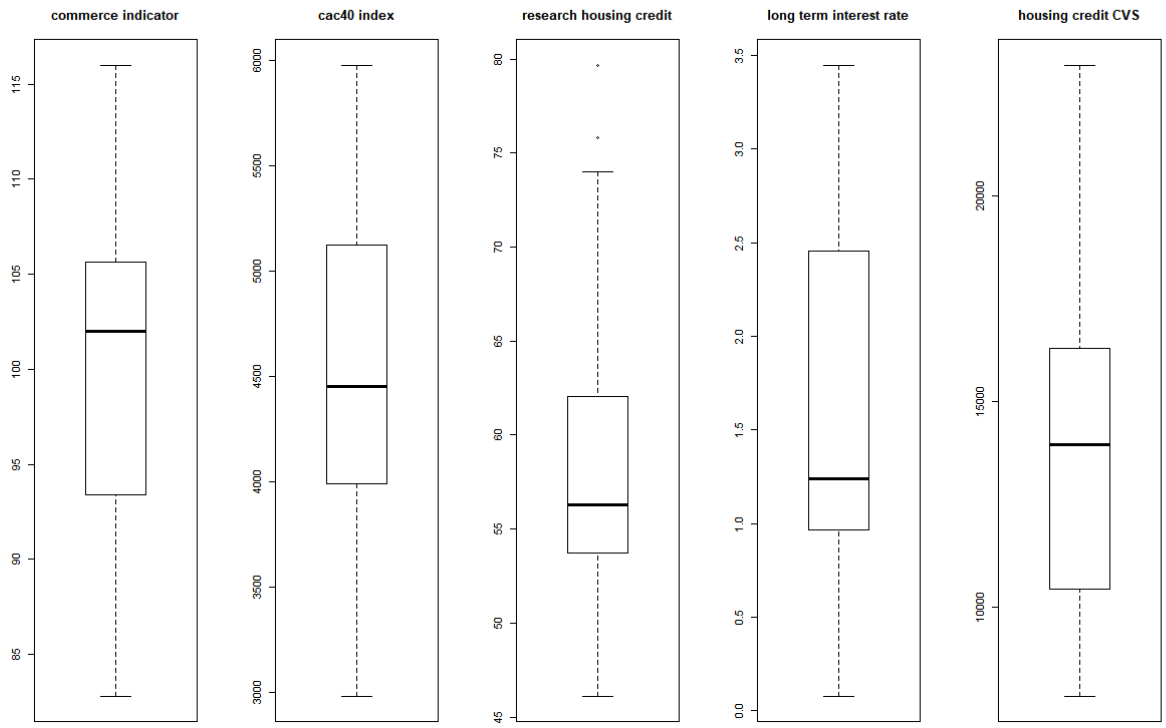



Table 17. Descriptive statistics on regressor and regressed

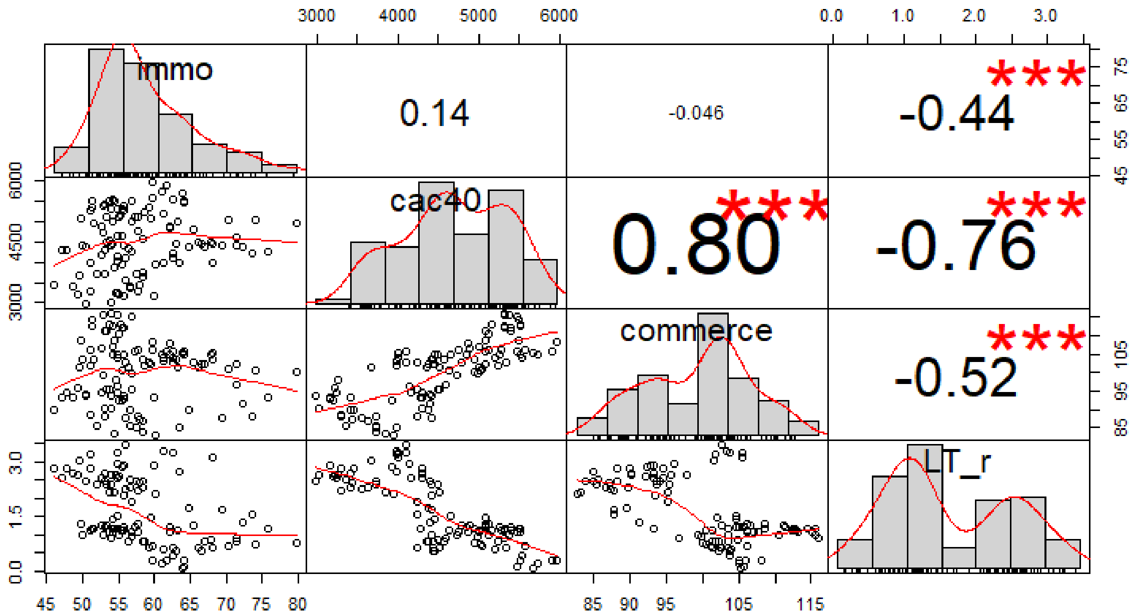
| | sd | median | kurtosis | skewness | normality | mean | stationarity |
|---|---|---|---|---|---|---|---|
| cred | 3666,51 | 13950,715 | -0,831 | 0,217 | 0,005 | 13699,206 | 0,0437 |
| immo | 6,673 | 56,271 | 0,539 | 0,915 | 0 | 58,059 | 0,0640 |
| cac40 | 754,054 | 4451,775 | -0,89 | -0,205 | 0,011 | 4513,559 | 0,0535 |
| commerce | 8,504 | 102 | -0,823 | -0,183 | 0,007 | 100,057 | 0,0896 |
| LT_r | 0,896 | 1,238 | -1,191 | 0,305 | 0 | 1,619 | 0,4028 |

According to table 17, the Skewness value means that the series is spread out on the right (positive value) for the credit series, immo and the long term interest rate. For the commerce indicator and the cac40, the values are spread out on the left. The negative kurtosis told us that the distribution is wide for all variables. Moreover, we can note that none of the variables follow a Normal distribution.

Finally, we did the ADF test on each variable to check if they were following the stationarity hypothesis. As we can see, 4 of 5 variables are following stationarity assumption at a 10% risk level.

According to figure 18, our regressor are quite correlated especially commerce,cac40, long term interest rate. Immo seems to be not correlated to much with economic activity indicator, only with long term interest rate. In this way, there is potentially a problem of multicollinearity.

Figure 18. Chart correlation of regressor



*Modelization*

Table 18. Models with regressor using ARX

| | AR(1) | ARX1 | ARX_immo | ARX_cac40 | ARX_commerce | ARX_rate[2] |
|---|---|---|---|---|---|---|
| Constant | 793.82# | 0.00048** | 0.00029*** | -278.46 | -736 | 960.93* |
| AR | 0.946*** | 0.9*** | 0.94*** | 0.87*** | 0.91*** | 0.94*** |
| Immo | - | 0.64*** | 64*** | - | - | - |
| Cac40 | - | 0.077 | - | 0.46# | - | - |
| Commerce | - | 0.21 | - | - | 19.76 | - |
| LT_r | - | - | - | - | - | 416.17 |
| R² | 0.887 | 0.902 | 0.90 | 0.89 | 0.888 | 0.886 |
| SE of regression | 1241.84 | 1172.2 | 1171.37 | 1227.97 | 1241.87 | 1252.24 |
| Log-lik | -913.13 | -905.45 | -906.36 | -911.43 | -912.63 | -904.97 |

Table 18 summaries the ARX models with explanatory variables and a simple AR(1). The coefficient of AR parameter tells us that the series is close to be non stationary because this coefficient is close to the unit, but it is acceptable. Besides, this means the past of the series has a big weight on the present values of the series, and we can see it with the high $R^2$ equal to 0.887. Afterward, we tested an ARX with 3 explanatory variables, stationary at a 10% risk level.

Only the Google trend variable "Immo" seems to be statistically significant. The $R^2$ of this model has increased a bit in comparison to the AR(1) model. The fact that Commerce and Cac40 aren't significant can be related to the fact that these variables are correlated between each other(cf. Fig chart correlation). Hence, to see their potential explanatory effect on our regressed series, we decided to estimate 3 models with only one regressor in each model.

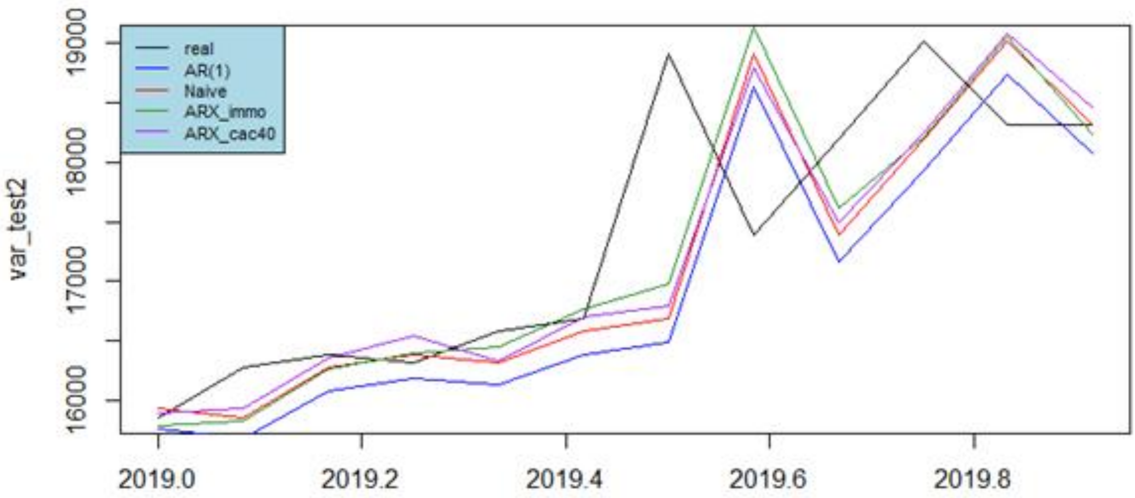
We can see that ARX_immo and ARX_cac40 have an higher $R^2$ than AR(1) and, the regressors are statistically significant at 1% for immo, and 10% for cac40 (we can note that the p-value is close to 0.05). Only ARX_commerce seems to have no difference with AR(1) model. The $R^2$ is relatively close and the standard error of both models is the same. Hence, 2 models will be retained: ARX_immo and ARX_cac40.

We will compare these 2 models with the AR(1) model and naïve prediction to see if adding explanatory variables is meaningful. Thus, to determinate the best model, we will produce out-of-sample forecasts, with a training data set (96 observations) and a testing data set (12 observations). We will compare forecasting accuracy according to 4 criterion which are mean

[2] The model here is done with differentiated series of long term interest rate of France because the series was not stationary. Hence there is one observation difference between this model and the others one.

square error (MSE), mean absolute error (MAE), median square error (MDSE) and median absolute error (MDAE).

Figure 19. One step forecasts for 2019 housing credit production in France, 4 models and real values




As we can see on figure 19, all models have a similar curvature. We observe that AR(1) seems to be the poorest model to produce forecast of our series because its curve is far from the black one which represents the real values. Both ARX models (green and purple curves) are the closest from the real values, that will likely be the 2 best models in term of forecasting according to errors indicator that we calculate afterward. Surprisingly, the naïve forecasting is not far from the 2 ARX models, and seems to be better than AR(1). Nonetheless, all these forecasts are close from each other, and we need to differentiate them. Hence we will calculate MSE, MAE, MDSE, MDAE, and Diebold & Mariano (DM) test on forecasts for each model. in addition, we will do the DM test on fitted values for each model too (see table 19, 20,21) .

Table 19. Comparative errors indicator on one step forecasts (h = 12) out-of-sample.

| | MSE | MAE | MDSE | MDAE |
|---|---|---|---|---|
| AR(1) | 887837.8 | 693.8511 | 191922.8 | 438.0334 |
| ARX_immo | 718920 | 568.7249 | 110790.3 | 292.3841 |
| ARX_cac40 | 700702.3 | 567.4158 | 92608.76 | 300.444 |
| naive | 780626.2 | 593.975 | 122747.6 | 340.74 |

Table 20. Comparison between each model with Diebold Mariano test out-of-sample

| | Naive | AR(1) | ARX_immo | ARX_Cac40 |
|---|---|---|---|---|
| Naive | - | 0.95 | 0.31 | 0.054 |
| AR(1) | | - | 0.75 | 0.90 |
| ARX_immo | | | - | 0.55 |
| ARX_cac40 | | | | - |

According to table 20, the DM test on out-sample forecasts tells us that the ARX_Cac40 seems to be the best model between all four models. However, the test just tells us that this model is better than ARX_immo only. Hence, it his tough to select one model. That is confirming the intuition that we have got when we saw the graph with each forecast compared to the real values. All models were close. Meanwhile, taking in account the error indicator calculated in table X, ARX_cac40 might be the best model in term of prediction because it has 3 of the lowest values of these indicator. Afterward, ARX_immo has the lowest MDAE of all models; that positioning it as the 2nd best model. To refine the selection of the best model, we do the DM test on our 4 models, this time in sample. Table X2 shows the p-value of the DM test.

Table 21. Comparison between each model with Diebold Mariano test in-sample

| | Naive | AR(1) | ARX_immo | ARX_Cac40 |
|---|---|---|---|---|
| Naive | | | | |
| AR(1) | 0.15 | | | |
| ARX_immo | 0.008 | 0.16 | | 0.054 |
| ARX_cac40 | 0.07 | 0.07 | | |

According to table 21, ARX models seems perform better in sample than AR(1) and naïve model. We can notice that ARX_immo is better than naïve but does not perform better than AR(1). ARX_cac40, is the best model according to DM test. Neither ARX_immo, AR(1) or naive perform better than ARX_cac40 at 10% of risk. Hence, ARX_cac40 is retained as the best model in term of forecasting accuracy and will be retained as the final model to forecast seasonally adjusted housing credit production for the first quarter of 2020. Nevertheless, because the results are statistically significant only at a 10% risk level, forecasts will also be computed with the 3 other models.

Multivariate Diebold Mariano test on all models out-sample

data:  var_test and Rpred3

statistic = -20,94, lag length = 7, p-value < 2.2e-16

alternative hypothesis: Equal predictive accuracy does not hold.

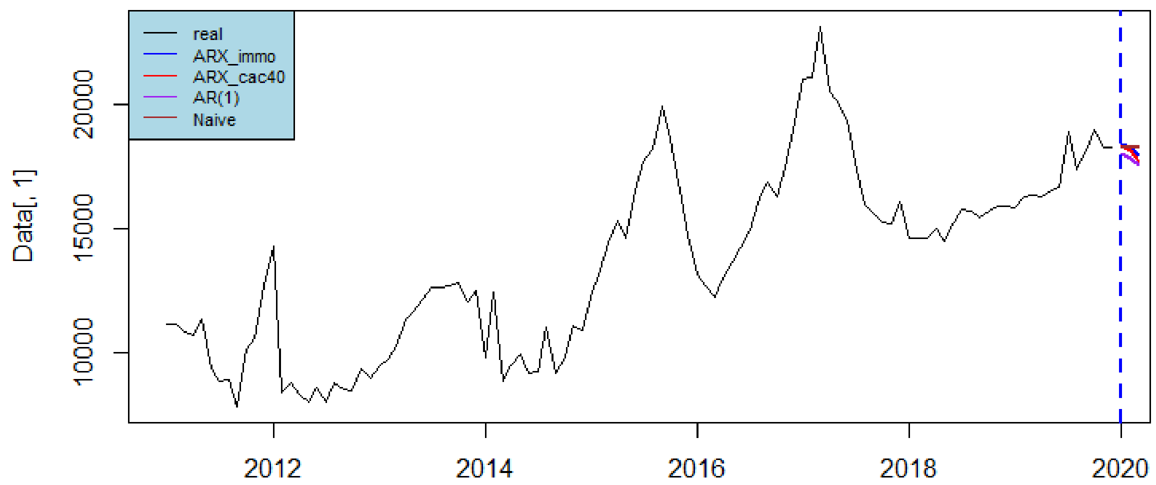
This test states that the one step forecasts are statistically different between the different models because the p-value is lower than 0.05. Hence, it seems that the ARX_cac40 might be better than naïve forecasting according to multivariate DM test done on out-sample forecast, summarized in table 21. We will now produce forecast for first quarter of 2020 for our series.

*Forecasting for the first quarter of 2020*

Figure 20 shows the forecasts for the first quarter of 2020 with our four previous models. Except naïve prediction, all other 3 models predict that the housing credit production will decrease. A surprising result is that AR(1) is producing the lowest values, in comparison of ARX_cac40 and ARX_immo; knowing that the cac40 index actually encounters a massive decrease, as all other stock index all around the world, and that the interest for housing credit in France measured by google trend is also decreasing. Hence, AR(1) forecast might be closer to real values.

IMF declared in march, at the beginning of the coronavirus pandemic that world was in recession. Thus, we expected that these explanatory variables will produce forecast with lower values than AR(1). This could be explain essentially by the low weight of each variable in the ARX model (the variation of $R^2$ by adding an explanatory variable is less than 0.01).

Figure 20. Forecasting for first quarter of 2020 with 4 models



Therefore, we build a OLS model with our 3 variables immo, cac40 and the retail commerce indicator. One model has been retained to produce forecast. This model is summarized in the table 22. We took the same variables than those we put in ARX model, hence, they are stationary at a 10% risk level, seasonally adjusted and they are also cleaned of potential outliers.

We can see than log(cac40) and log(commerce) are statistically significant. The adjusted $R^2$ of this regression is 0.65 which mean that the model quality in term of adjustment is quite good. Note that this model has a VIF (appendix 4) which is under 5 for each variables, hence, there is no multicollinearity between these 2 variables, while we saw that they were correlated. Moreover, normality of residuals hypothesis is respected, as the hypothesis of homoskedasticity of error (appendix 4). The interest to build an OLS with our variables here is to see if they have a real predictive power, and it seems to be the case. Nevertheless, this OLS model is by far the poorest model in term of prediction out-of-sample as table 22 and figure 21 shows. Besides, this result could be explained by the fact that this model as no autoregressive parameter which are there to predict the series with its past values. Thus, it is normal to have this result.

Table 22. OLS model with 2 regressor and error indicator

Coefficients:

| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| (Intercept) | -2.1058 | 0.8665 | -2.430 | 0.0168 | * |
| log(cac40) | 1.0044 | 0.1372 | 7.319 | 5.25e-11 | *** |
| log(commerce) | 0.6860 | 0.2774 | 2.473 | 0.0150 | * |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

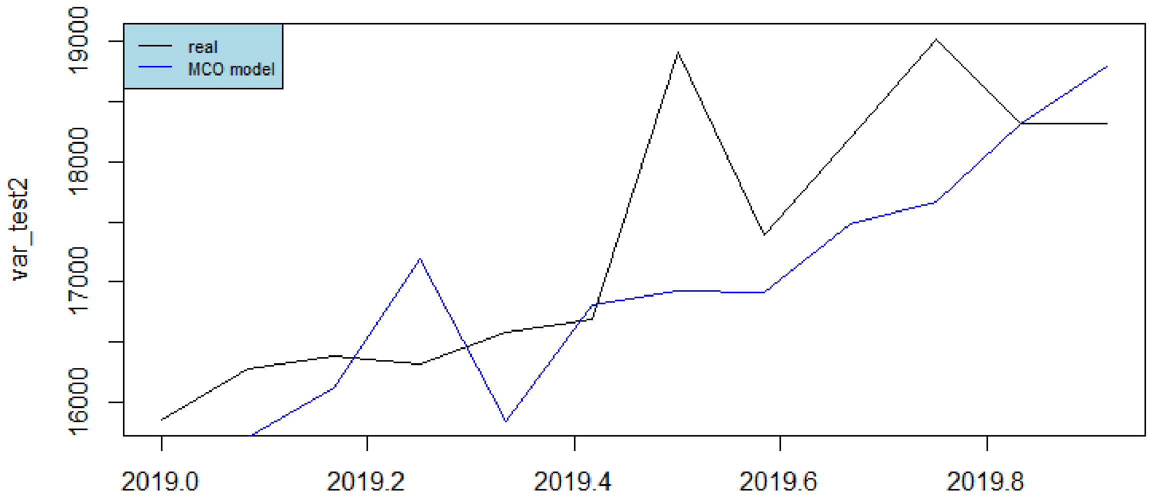
Residual standard error: 0.1635 on 105 degrees of freedom

Multiple R-squared: 0.6541, Adjusted R-squared: 0.6476

F-statistic: 99.3 on 2 and 105 DF, p-value: < 2.2e-16

| MSE | MAE | MDSE | MDAE |
|---|---|---|---|
| 805979.7 | 722.7873 | 412713.8 | 639.0601 |

Figure 21. Forecasting out-of-sample with OLS model




The forecast for the first quarter of 2020 by all models are included in table 23. We can note that the forecast made by OLS give us lower values than forecast made with previous models. That way, we think these forecast for the first quarter of 2020 are more realistic according to the economic context that we know actually even though the models performance are the poorest out-sample. The figure 22 shows the forecasting done by the OLS model.

Figure 22. CVS series and OLS model with forecasting for first quarter of 2020

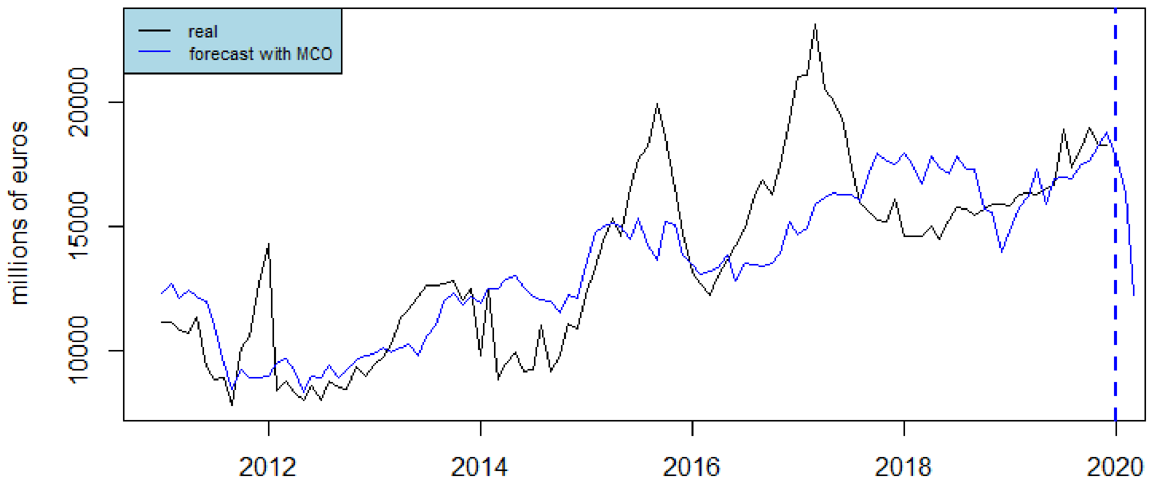


Table 23. Forecasting for the first quarter of 2020

| | immo | cac40 | AR | Naive | MCO |
|---|---|---|---|---|---|
| January-20 | 18402,88 | 18382,78 | 18060,3228 | 18315,3 | 17753.56 |
| February-20 | 18292,58 | 18153,58 | 17818,8731 | 18315,3 | 16346.92 |
| March-20 | 17961,71 | 17723,56 | 17593,6714 | 18315,3 | 12291.55 |

## IV. CONCLUSION

The aim of this work was to produce forecast on the housing credit production in France. We removed the seasonality from the series with TS and X13 methods. X13 was more relevant for the analysis of the business cycles. An interesting result that emerges from the study of the cycles is that the series gonna probably enter in recession in 2020 because the acceleration cycle is in a recession phase (TS cycle tell us that there is a growth cycle starting in January 2019, meanwhile, the acceleration cycle has also encountered a peak at the same period). This result pushes us to conclude that the growth of the series will slow down, or even be negative soon (if it not already the case but with the lack of data, we cannot date the peak of the expansion phase). We justified this by the fact that the duration of this expansion phase is the longest of all expansion phases that the series experiencied since 2011 (according to X13 dating cycle). Hence it is likely that we are close to the end of this phase. In addition, the actual economic recession will probably impact significantly the production of the housing credit in the future.

In the second part, we tried to produce forecast on the original series (not seasonal adjusted). The main result of this part is that the naïve prediction and SARIMA seats seems to be the best to predict the housing credit production, in comparison with SARIMA models, HW double exponential smoothing, TBATS, ETS, STL and NNETAR models. The only model giving at least as accurate predictions is the SARIMA(1,0,0)*(1,0,0). Their MSE, ME and MAE are the smallest between all the model we tried. However the Diebold & Mariano tests told us the none of these two models statistically outperforms the other one.

Finally, in the last section, we also produce forecast of the housing credit production, but this time, by using seasonally adjusted series. In addition, we added explanatory variables to improve the prediction accuracy. We compare these results with those from a naïve prediction, and a AR(1) model. The main result is that adding explanatory variables help to produce better forecasting, but these exogenous variables are not drastically helpful in term of accuracy gains, this is mostly due to the fact that the AR(1) parameter explain most of the variance of the series, hence, adding new variable only has a marginal impact. That way, to see the explanatory power of our regressor, we decided to build an OLS model. The model retained is by far the poorest in term of accuracy forecasting out-sample. However, his forecast for the first quarter of 2020 seems to be closer to futures values, taking into account the economic background that we are actually going through with coronavirus pandemic. In this way, using OLS rather than time series model could be interesting.  Notwithstanding the models developed here are quite simple. It would be interesting to see if other form of modelization, using VAR for example, could produce better results than times series models or OLS.

## V. Bibliography

Kashyap,A., J. Stein and D.Wilcox (1993) Monetary policy and credit conditions: evidence from the composition of external finance, American Economic Review, 83, pp. 8-98.

Bernanke, B. and M. Gertler (1995) Inside the black box: the credit channel of monetary policy transmission, NBER working paper No. 5146.

Friedman, B. and K. Kuttner (1993) Economic activity and the short-term credit markets: an analysis of prices and quantities, Brookings Papers on Economic Activity, 2, pp. 193-283.

G. Levieuge (2017) Explaining and forecasting bank loans. Good times and crisis, Applied Economics, Vol. 49, Issue 8 Pages 823-843.

A. Calza, G. Gartner and J. Sousa (2001) Modeling the demand for loans to the private sector in the euro area, working paper No. 55.

## VI. Data sources

Housing credit production

Google trend immo: here

Cac40 index

Retail sell indicator survey

Long term interest rate

# VIII. Appendix

## VIII.1. Appendix of section II

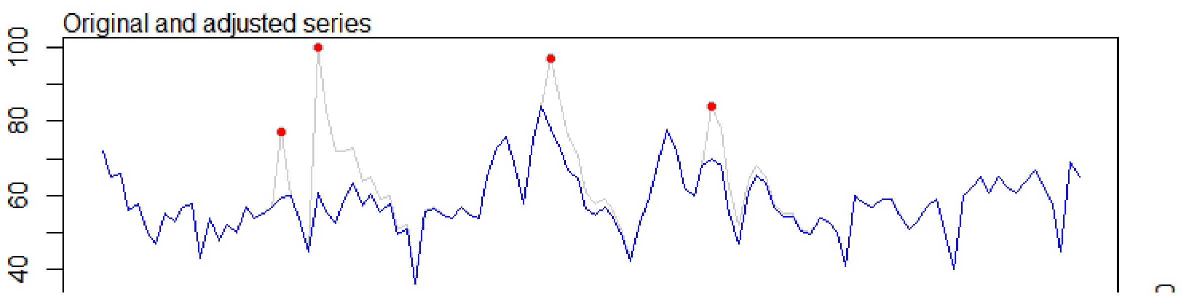
Appendix 1. Algorithm to produce one step ahead forecast for h=12

```
i - 1

for (i in 1 : 12){

  Data_stl - ts(Data[1  96 + i, 2], start = c(2011, 01), frequency = 12)

  Rpred[i, 1] - forecast(arima(Data[1  96 + i, 2], order = c(0, 1, 1),
seasonal = list(order = c(0, 1, 1), period = 12)), h = 1)[['mean']]

  Rpred[i, 2] - forecast(arima(Data[1  96 + i, 2], order = c(2, 1, 2),
seasonal = list(order = c(2, 1, 2), period = 12)), h = 1)[['mean']]

  Rpred[i, 3] - forecast(tbats(Data[1  96 + i, 1]), h = 1)[['mean']]

  Rpred[i, 4] - forecast(stlf(Data_stl), h = 1)[['mean']]

  Rpred[i, 5] - forecast(ets(Data[1  96 + i, 1]), h = 1)[['mean']]

  Rpred[i, 6] - forecast(nnetar(Data[1  96 + i, 1]), h = 1)[['mean']]

  Rpred[i, 7] - forecast(HoltWinters(Data_stl, alpha = NULL, beta = 0.2,
gamma = TRUE, seasonal = 'mul'), h = 1)[['mean']]

  Rpred[i, 8] - snaive(Data[1  96 + i, 1], h = 1)[['mean']]

  i = i + 1

}
```

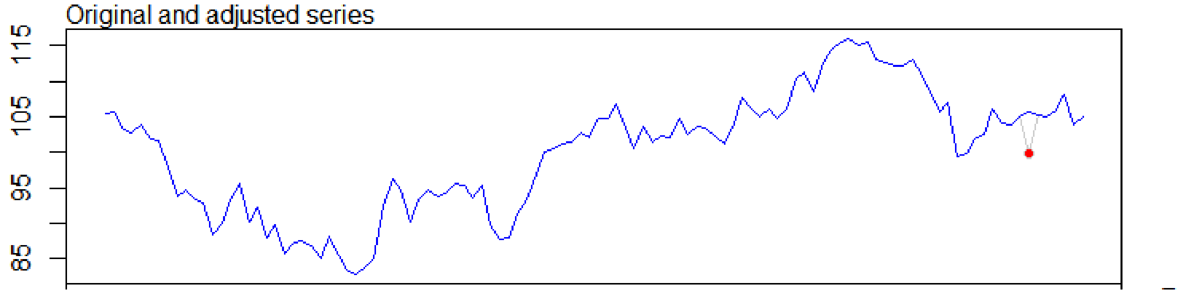
## VIII.2. Appendix of section III

Appendix 2. Outliers adjustment

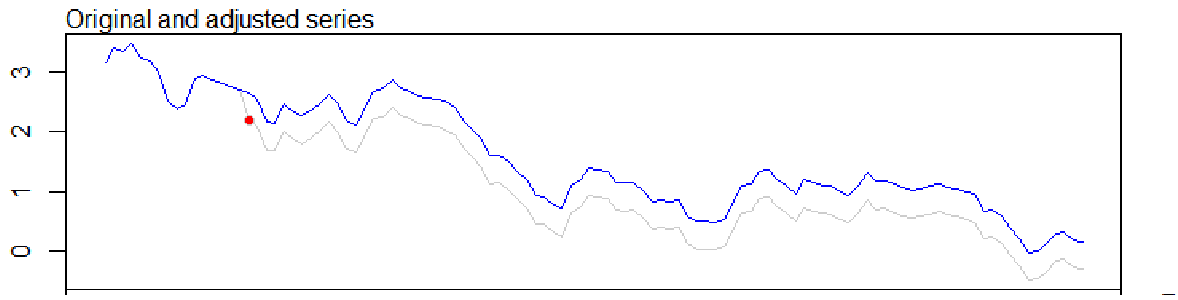
Outliers from google trend series



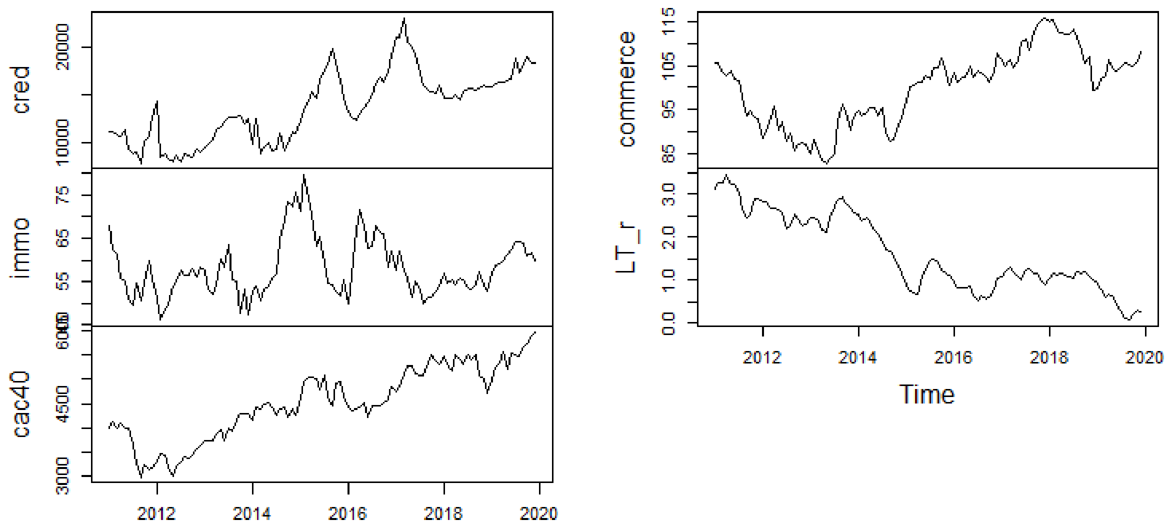
outliers from retail sell indicator




outliers from long term interest rate of France




Appendix 3. Plot of all variables adjusted from seasonality and outliers



## Appendix 4. Statistical test on OLS model (VIF, normality of residuals and homoskedasticity of error)

*Vif*

vif(model_mco)

| log(cac40) | log(commerce) |
|---|---|
| 2.286985 | 2.286985 |

*Normality of residuals*

Shapiro-Wilk normality test

data:  residuals(model_mco)

W = 0.97698, p-value = 0.0577

*Homoskedasticity of error*

studentized Breusch-Pagan test

data:  model_mco

BP = 0.86784, df = 2, p-value = 0.648