- Time Forecasting Series -
Projecting French birth

DEGIOANNI – KLIKEL - SIMIER

DATASET DESCRIPTION

Source: INSEE

Description: Number of births in France

Start date: January 1946

End date: December 2018

Frequency: Monthly

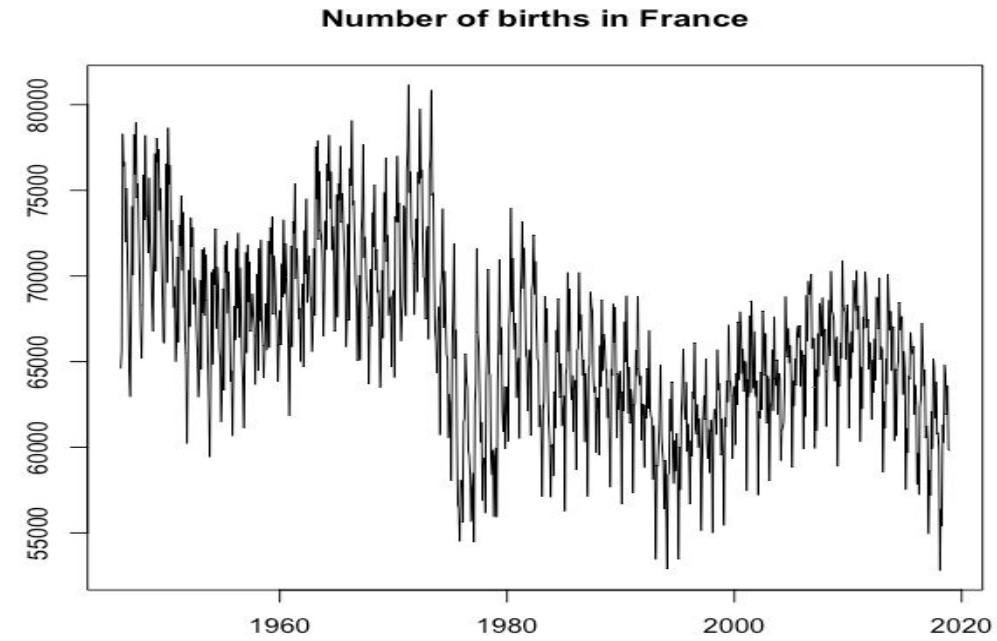
Number of observation: 876

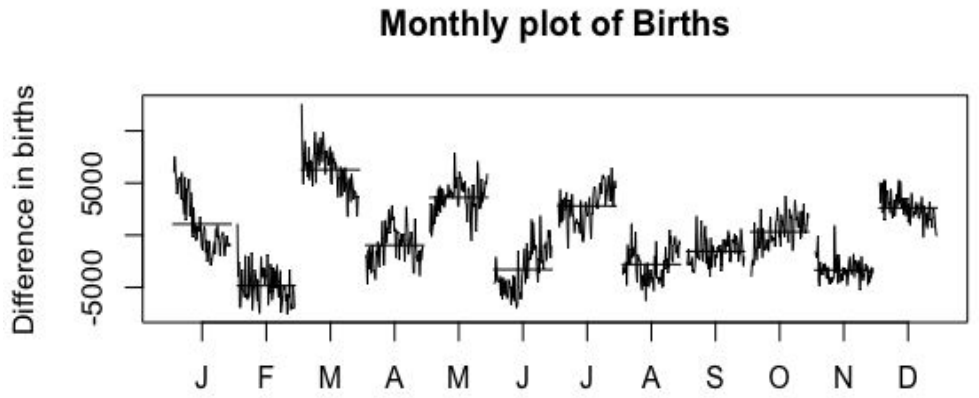
Source : https://bit.ly/2E8m7Q3

Overview dataset

period	Births
1946-01	64599
1946-02	65702
1946-03	78294
1946-04	76400
1946-05	76636
1946-06	71970
1946-07	75104
1946-08	71297
1946-09	68450
1946-10	64468
1946-11	62949
1946-12	68035
1947-01	74079
1947-02	70065
1947-03	78247
1947-04	75911
1947-05	78953
1947-06	74571
1947-07	75384
1947-08	72383

Time Series Overview - Birth in France

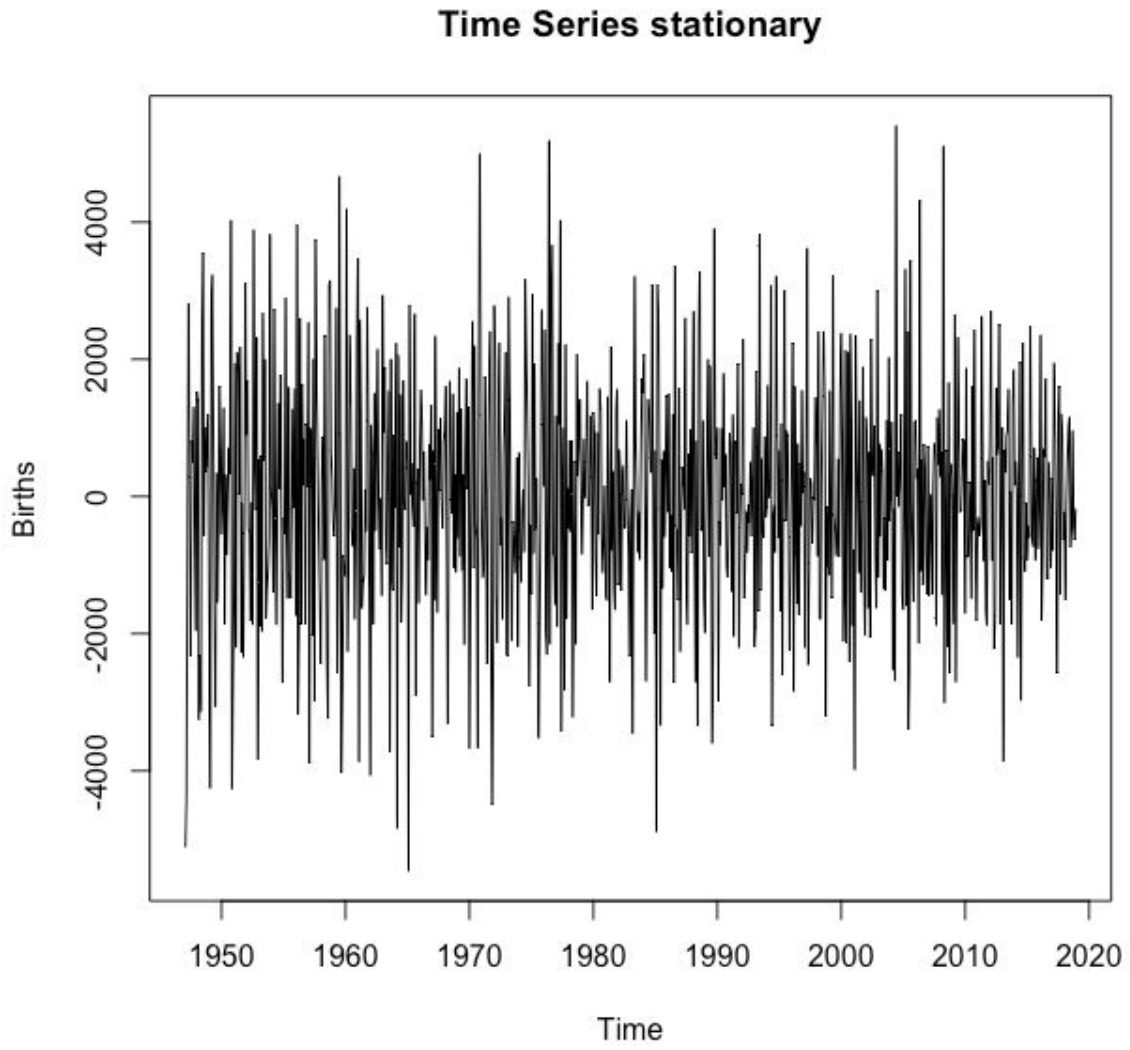




We see a downward trend with some bumps in the number of births. It is difficult to see here if there is seasonality but if we look at the monthplot, the seasonality is confirmed.

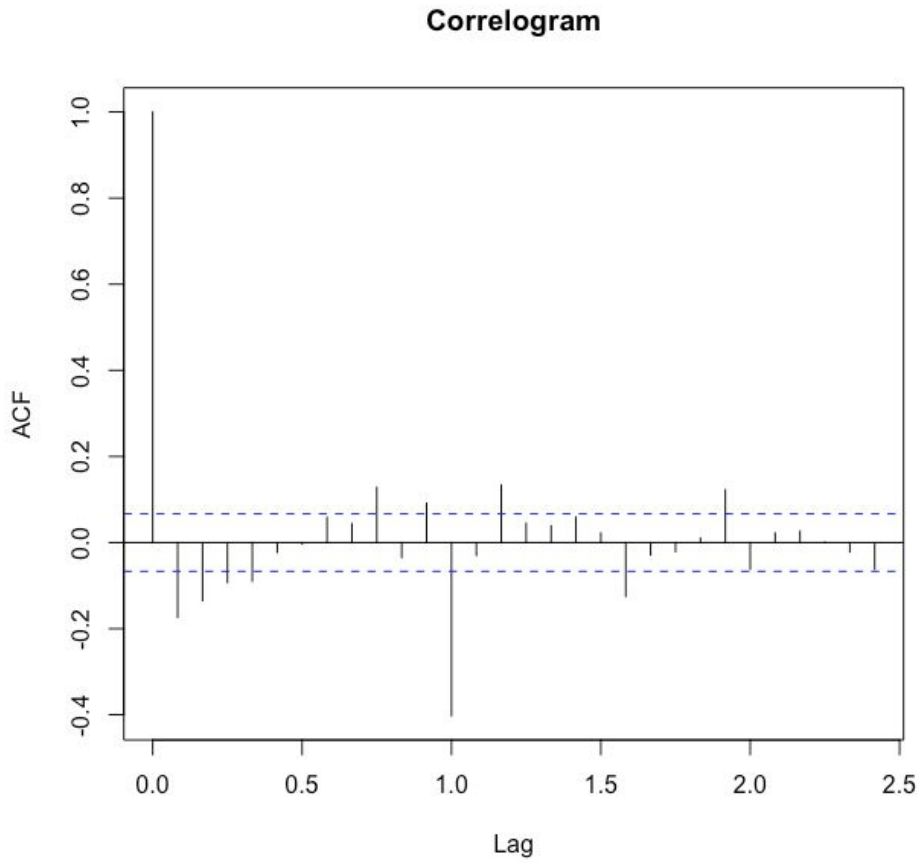
Time Series Overview - Birth in France

After removing the trend and the seasonality, we obtain the following time series which is **stationary**:

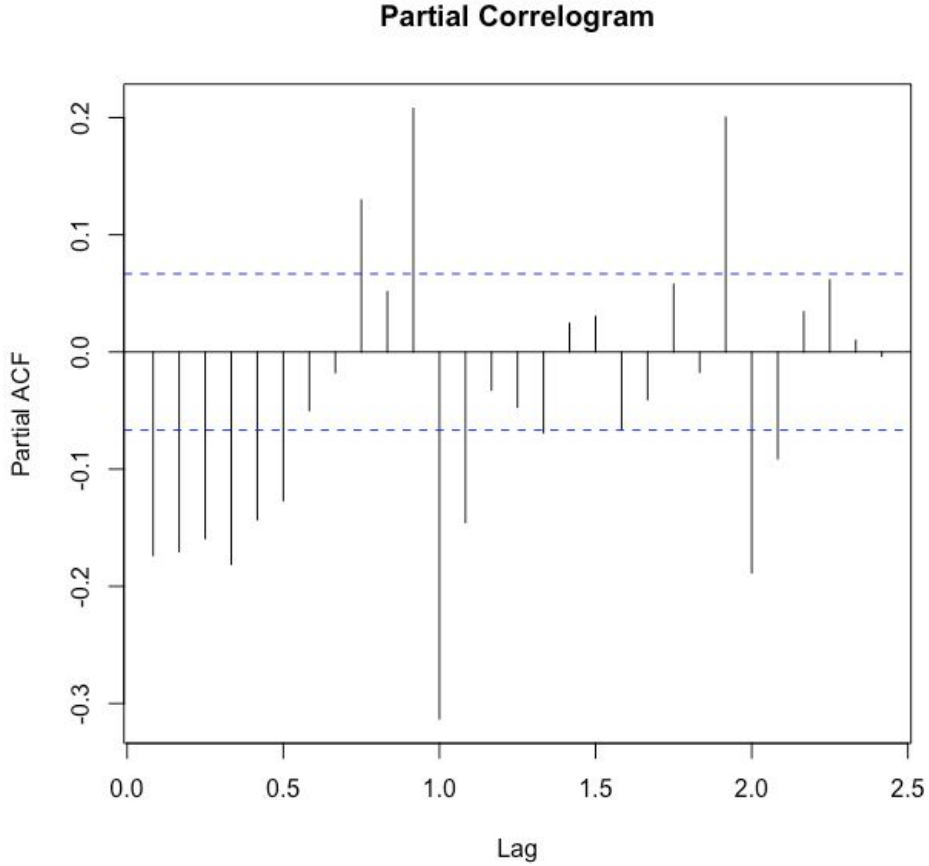


Not a white noise as the Box-Ljung test indicates a p-value < 0.05, we strongly reject the null hypothesis (H0: ρ0 = … = ρkmax = 0, ⇒ a white noise)

Correlogram - Birth in France



The Correlogram indicates that a **Moving average MA(4)** repeated once or twice could be appropriate.



The Partial Correlogram indicates that a **Autoregressive AR(6)** repeated once or twice could be appropriate.

Models Comparison - Birth in France

TEST	Higher order terms significant	AIC	BIC	Correlogram	Box test	MAE
AR(6) repeated twice	Yes	14947.58	14990.42	Small correlations	OK	1066.724
MA(4) repeated once	Not all	14852.59	14881.15	Small correlations	OK	995.6268
SARIMA(2,2)(1,1)	Not all	14842.52	14875.84	Small correlations	OK	995.0669

At first sight, if we decide to weigh more the BIC criteria as it penalizes even more for the complexity of the model, the **SARIMA(2,2)(1,1) model** seems to be the best.

NB: With the argument ‘order’ of the function ‘arima’, we have remove the trend and with the argument ‘seasonal’ we have removed the seasonality for the 3 models.

Out-Sample Comparison - Birth in France

TEST	MAE	MAPE
MA(4) repeated once	3623.213	8.31%
SARIMA (2,2)(1,1)	3628.549	8.34%

We compute a **Diebold Mariano** test to assess the significance of the difference in MAE and MAPE between our 2 models.

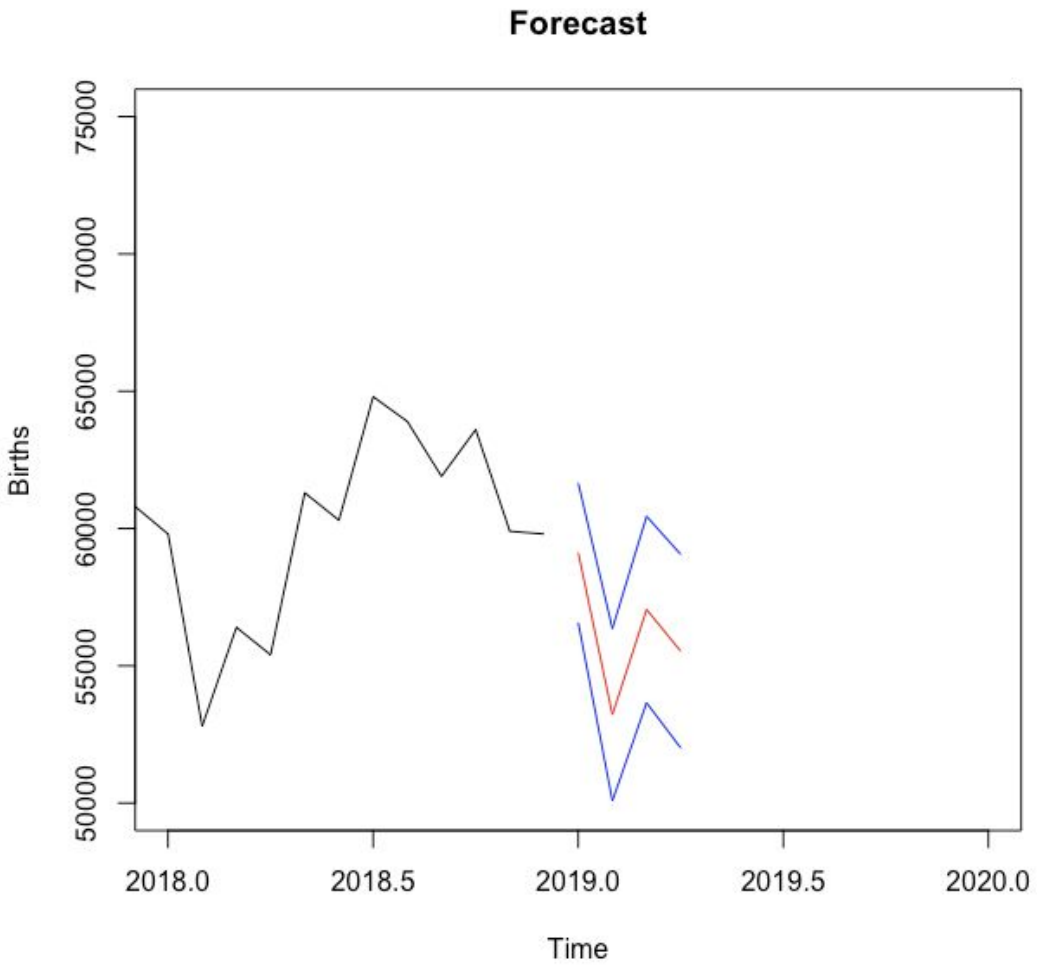
```
        Diebold-Mariano Test

data:  error1.herror2.h
DM = 0.99314, Forecast horizon = 4, Loss function power = 1, p-value = 0.3212
alternative hypothesis: two.sided
```

We conclude that the forecast performance of the two models, measured by the MAE, is **not significantly different**.

Thus, we stick to the BIC criteria and **choose the model SARIMA(2,2)(1,1)**. The model is quite good as we can see with an error of 8.3% or 3628 births.

Forecasting with best model - Birth in France



Prediction			
Month	**Lower**	**Prediction**	**Upper**
Jan 2019	56 578	59 117	61 655
Feb 2019	50 079	53 218	56 356
Mar 2019	53 653	57 047	60 442
Apr 2019	52 004	55 529	59 054

Conclusion

During modelling, we discovered that different models can be used to forecast birth in France. We decided to choose an **Arma(2,2) model** which was the most accurate model, according to the criteria chosen .

Our model emphasizes for the coming months a global downward trend in French births. Note that in March, we oversee a potential up in birth.

Trying to predict the number of birth in France isn’t an easy thing and other parameters must be taken into account in parallel to statistical models to seize all the **complexity of demographic forecast**.