https://github.com/marl2en/predictFuture

Births in Sweden

Part I: Analyse of monthly data from January 2019 to November 2022 with an attempt to forecast next 12 months.

Part II: Analyse of yearly data from 1749 to 2021.

1. Get datasets

Datasets can be dowmloaded from scb.se (Statistiska Centralbyrå – Statistic Sweden)

CSV files were edited to represent clear data columns with header.

Load data with:

```
births_month = pd.read_csv(save_path + 'births_month3.csv',delimiter=',')
```

2. Show data and visualize

Show header of dataset:

```
>>> births_month.head()
   Year     Month  Population at the end of period  Population growth1  Live births  Still births  Deaths  Population surplus  Immigrants  Emigrants  Surplus of immigrants        Date
0  2019   January                         10242296               12111         9620            41    8372                1248       11858       4293                   7565  2019-01-01
1  2019  February                         10250006                7710         8821            30    7487                1334        9720       3344                   6376  2019-02-01
2  2019     March                         10258037                8031         9811            36    7791                2020        9528       3517                   6011  2019-03-01
3  2019     April                         10265728                7691        10064            29    7373                2691        8249       3249                   5000  2019-04-01
4  2019       May                         10274848                9120        10278            34    7147                3131        9380       3391                   5989  2019-05-01
```

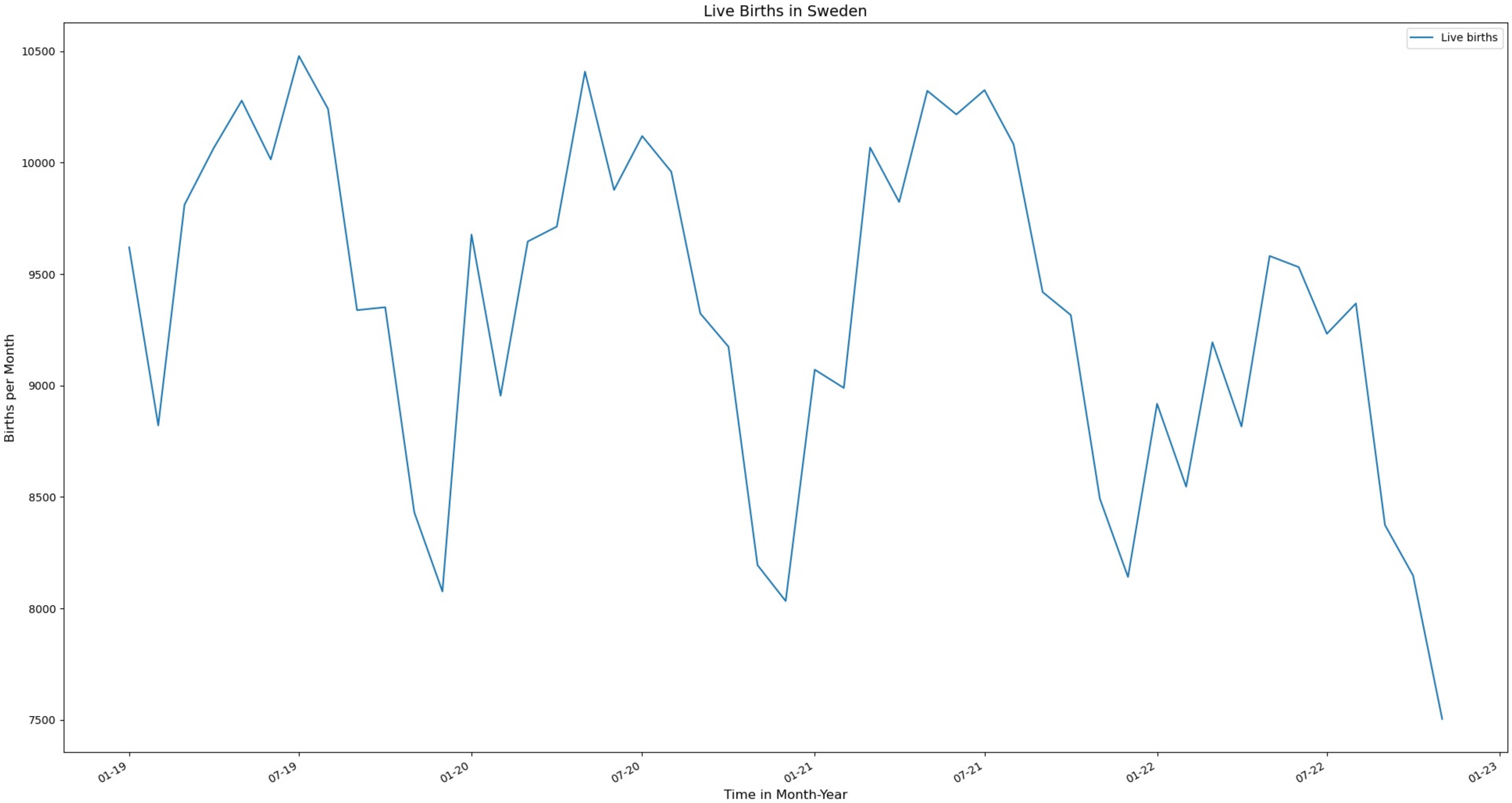
and tail

```
>>> births_month.tail()
    Year      Month  Population at the end of period  Population growth1  Live births  Still births  Deaths  Population surplus  Immigrants  Emigrants  Surplus of immigrants        Date
42  2022       July                         10499430                6547         9232            34    7366                1866        8743       4062                   4681  2022-07-01
43  2022     August                         10509402                9972         9368            26    7566                1802       14586       6416                   8170  2022-08-01
44  2022  September                         10514895                5493         8374            24    7381                 993       10289       5789                   4500  2022-09-01
45  2022    October                         10518309                3414         8147            30    7592                 555        7355       4496                   2859  2022-10-01
46  2022   November                         10520558                2249         7504            28    7597                 -93        6129       3787                   2342  2022-11-01
```

Now lets visualize birth per month in Sweden from January 2019 to November 2022.

```
showDTplot(births_month,col2show=['Live births'],offset=0,title='Live Births in Sweden',ylabel='Births per Month')
```

https://github.com/marl2en/predictFuture



X-axis is a timeline in Month-Year. Every month has a data point.
Y-axis shows total number of Live births per months.
There are four montains, one for every year with less births at winter and more at summer. The last mountain (2022) is smaller.
Conslusion: time serie shows a trend and seasonality.

3. Analyze the data and make predictions

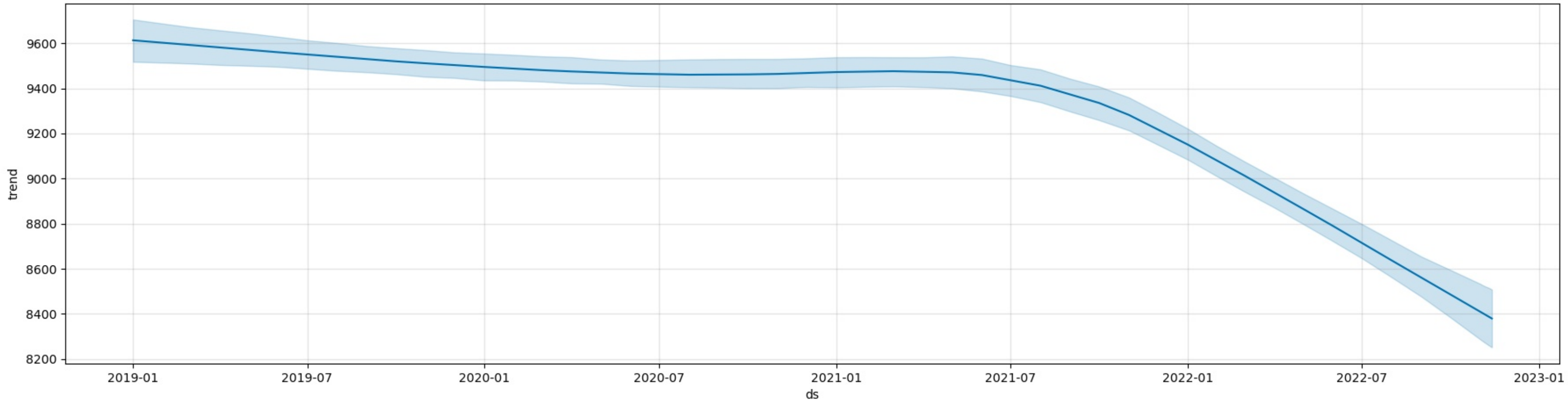
The problem boils down to analyze an uni-variate time series and forecast the future into a chosen time horizon (12 months).
At first we want to decompose data into trend and cycle.
This could easily be done with Hodrick-Prescott (HP) filter inculded in statsmodel python library (statsmodels.org), but let's chose another approach.

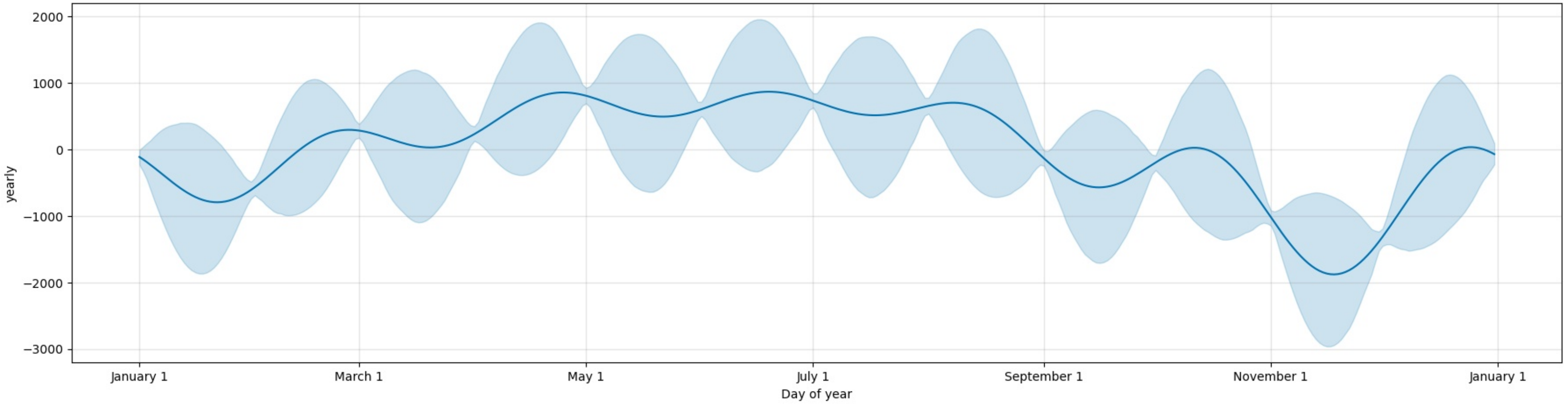
The Prophet is a python library provided by the Facebook development team to make predictions of time series.

Just install using pip, import, create a model, fit model to data and predict the future, validate model and save results.

```
from prophet import Prophet
from prophet.diagnostics import cross_validation
from prophet.diagnostics import performance_metrics
from prophet.plot import plot_cross_validation_metric

horizon = 12
df = births_month[['Date','Live births']].copy()
df.columns = ['ds','y']
df['ds'] = pd.to_datetime(df['ds'])
df['y']=df['y'].astype(float)
m = Prophet(changepoint_prior_scale=0.03,
daily_seasonality=False,mcmc_samples=100,interval_width=0.75,changepoint_range=0.95,yearly_seasonality=6,weekly_seasonality=False,seasonality_prior_scale= 0.1)
m.fit(df)
future = m.make_future_dataframe(periods = horizon) # , freq = 1
forecast = m.predict(future)

fig1 = m.plot(forecast);plt.show()
fig2 = m.plot_components(forecast);plt.show()

# validate model
df_cv = cross_validation(m, initial='730 days', period='180 days', horizon = '365 days')
df_p = performance_metrics(df_cv)
fig = plot_cross_validation_metric(df_cv, metric = 'mape');plt.show()

#save results
forecast.to_csv(save_path+'births_month_forecast.csv', index=False,header=True)
```

https://github.com/marl2en/predictFuture

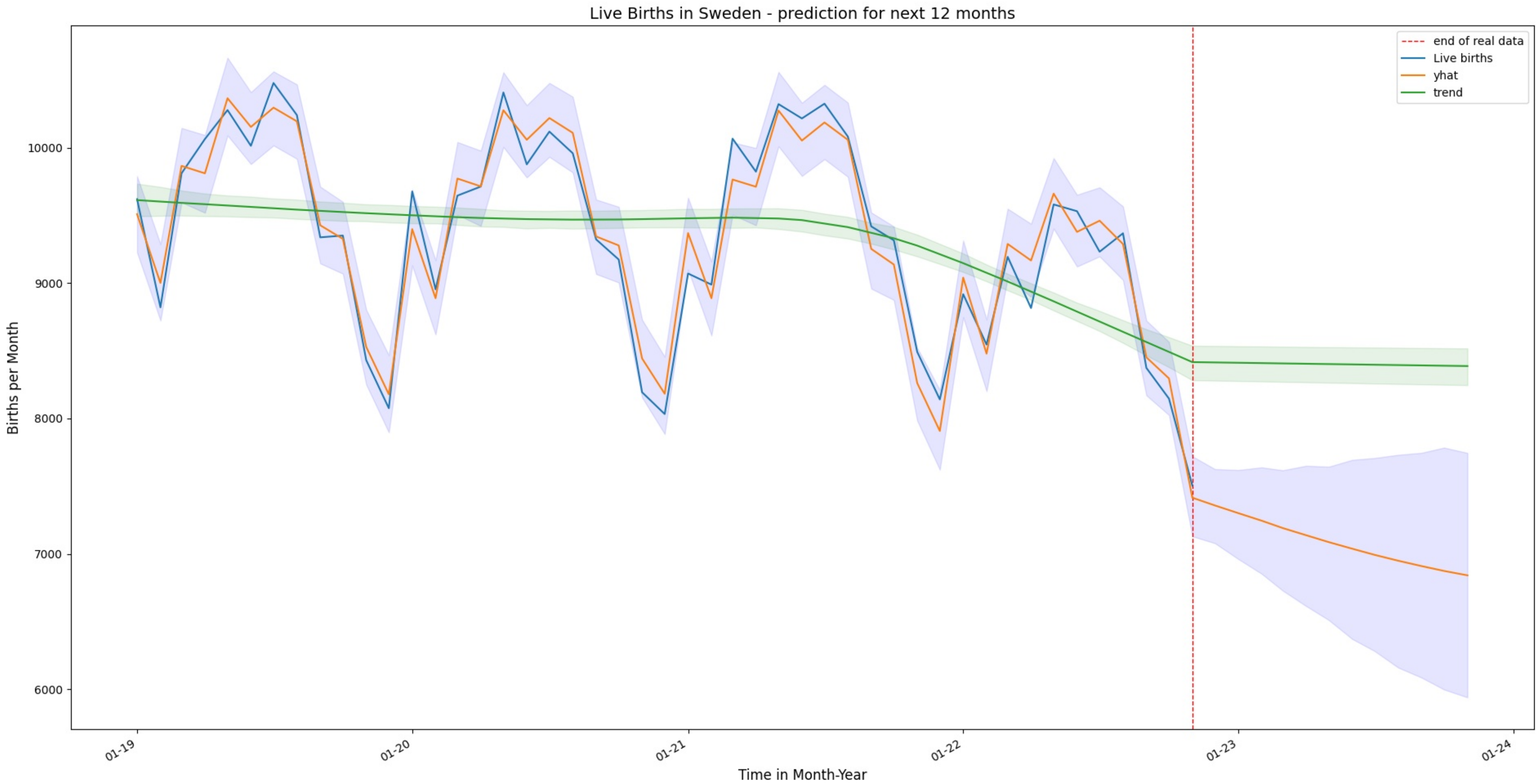




The upper graph shows the trend declining slowly from 9600 births per month to about 9500 during 2019/01-2021/7.
July 2021 marks a change point in trend which is now falling steeper to 8400 births per month. Light blue marks the models confidence interval.

The lower graph visualize variation of births per month (more births at summer, less at winter).
Conclusion: Summer 2021 is a change point of trend in births to a more negative outlook.

https://github.com/marl2en/predictFuture



This image shows the forecast of the prophet model for the next 12 months.

Left of the dashed red line shows real data (dark blue), fitted values of the model (orange) with confidence interval (light blue), trend with interval in green.
Forecast is shown right of dashed line.
Predicted birth's mean is in orange and confidence interval is in light blue. Predicted trend looks very much optimistic.
If you don't like the prophecy, don't blame the prophet.
Disadvantage: The model don't get seasonality right (no maximum in summer 2023)

https://github.com/marl2en/predictFuture

The prophet model does a great job in estimating trend of fitted data, but for yearly cycle it performs badly. Maybe it can be tuned better.
To estimate seasonal component of the data we use SARIMAX instead.
Cycle = real data − trend

```
forecast['cycle'] = forecast['Live births'] - forecast['trend']
```

SARIMAX is a seasonal AutoRegressiv Integrated Moving Average regressor model of the statsmodel library.

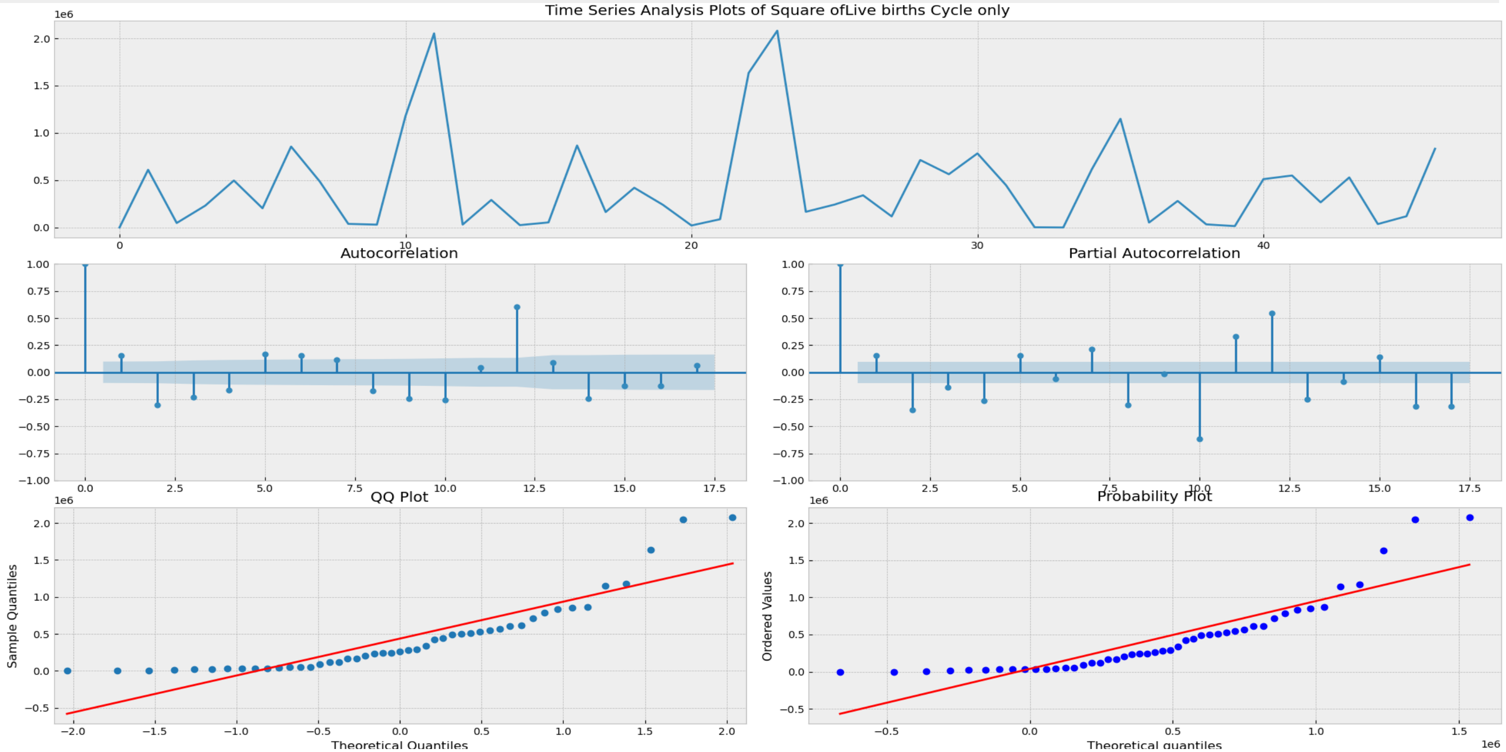
Load libraries

```
import statsmodels.api as sm
import statsmodels.tsa.api as smt
from scipy import stats
from statsmodels.tsa.api import SARIMAX
```

Estimate order of the model by:

```
tsplot(y=X, lags=None, figsize=(18, 16), style='bmh',target='Live births Cycle only',show=True,takeSquare=True)
```

Like anyone would gess, there is autocorrelation at lag 12 months. For the sarimax model we choose order=(2,12), see graph:



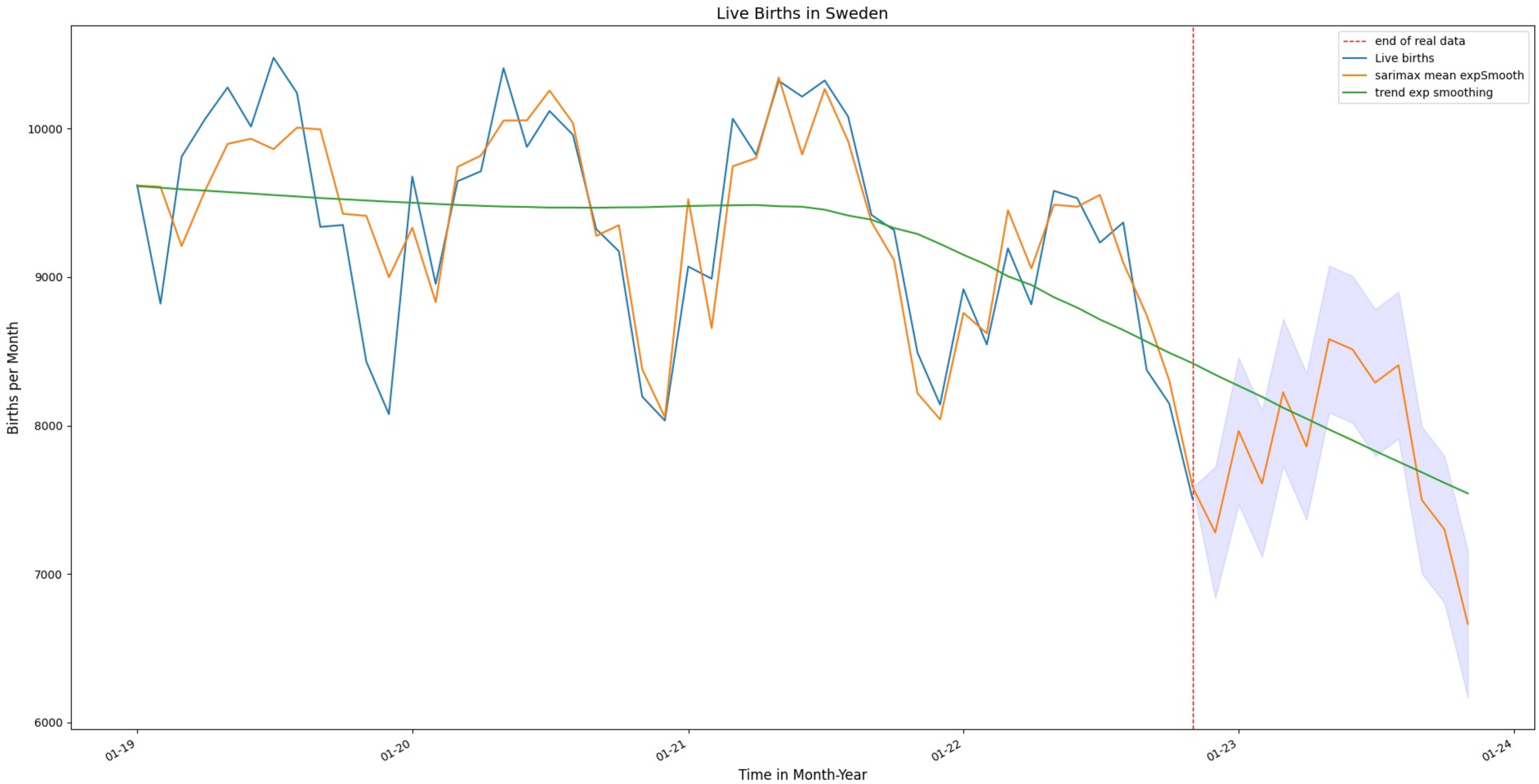
https://github.com/marl2en/predictFuture

Create model, fit and print summary.

```
sarimax_mod = SARIMAX(X, order=((2,12), 0, 1), trend="ct")
sarimax_res = sarimax_mod.fit()
print(sarimax_res.summary())
                               SARIMAX Results
==============================================================================
Dep. Variable:                        y   No. Observations:                   47
Model:            SARIMAX([2, 12], 0, 1)   Log Likelihood                -332.109
Date:                  Mon, 30 Jan 2023   AIC                            676.217
Time:                          13:10:20   BIC                            687.318
Sample:                               0   HQIC                           680.395
                                   - 47
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.9866     18.104      0.054      0.957     -34.497      36.471
drift         -0.8220      2.067     -0.398      0.691      -4.874       3.230
ar.L2          0.0648      0.042      1.541      0.123      -0.018       0.147
ar.L12         0.8999      0.034     26.330      0.000       0.833       0.967
ma.L1          0.5001      0.139      3.609      0.000       0.229       0.772
sigma2      5.075e+04   1.14e+04      4.463      0.000    2.85e+04     7.3e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.15   Jarque-Bera (JB):                 1.71
Prob(Q):                              0.70   Prob(JB):                         0.43
Heteroskedasticity (H):               0.79   Skew:                            -0.21
Prob(H) (two-sided):                  0.64   Kurtosis:                         2.17
===================================================================================
```

Now we’ve got 2 scenarios.
1. Trend reverse to previously levels according to prophets forecast. (image sarimaxPrediction_with_prophet_trend.png)
2. Trend continues with unchanged slope. See image below

https://github.com/marl2en/predictFuture



The green line is the trend. Trend forecast with statsmodels.tsa.holtwinters ExponentialSmoothing.
The blue line is real data.
Orange: fitted values of model and forecast.
Area with light blue: confidence interval

https://github.com/marl2en/predictFuture

Predicted values with 95% confidence interval

```
mask = (forecast.index >= dt.date(2022,12,1)) & (forecast.index <= dt.date(2023,11,30))
forecastvalues = forecast[["Date","sarimax lower expSmooth","sarimax mean expSmooth","sarimax upper expSmooth"]].loc[mask].values

Year-Month      lower   mean    upper
'2022-12',      6836,   7277,   7719,
'2023-01',      7469,   7963,   8456,
'2023-02',      7113,   7607,   8102,
'2023-03',      7728,   8223,   8718,
'2023-04',      7363,   7857,   8352,
'2023-05',      8087,   8581,   9076,
'2023-06',      8018,   8512,   9007,
'2023-07',      7793,   8288,   8783,
'2023-08',      7911,   8406,   8901,
'2023-09',      7002,   7497,   7992,
'2023-10',      6804,   7299,   7793,
'2023-11',      6167,   6662,   7157
```

Part II.

The real births 2022 from january to november are:

```
mask = (births_month.index >= dt.date(2022,1,1)) & (births_month.index <= dt.date(2022,12,31))
births22 = int(births_month['Live births'].loc[mask].values.sum() )
births22 # jan-nov
97211
```

Let’s add forecast for december’s upper prediction:
97211+7719= 104930 as the total number of births for 2022.

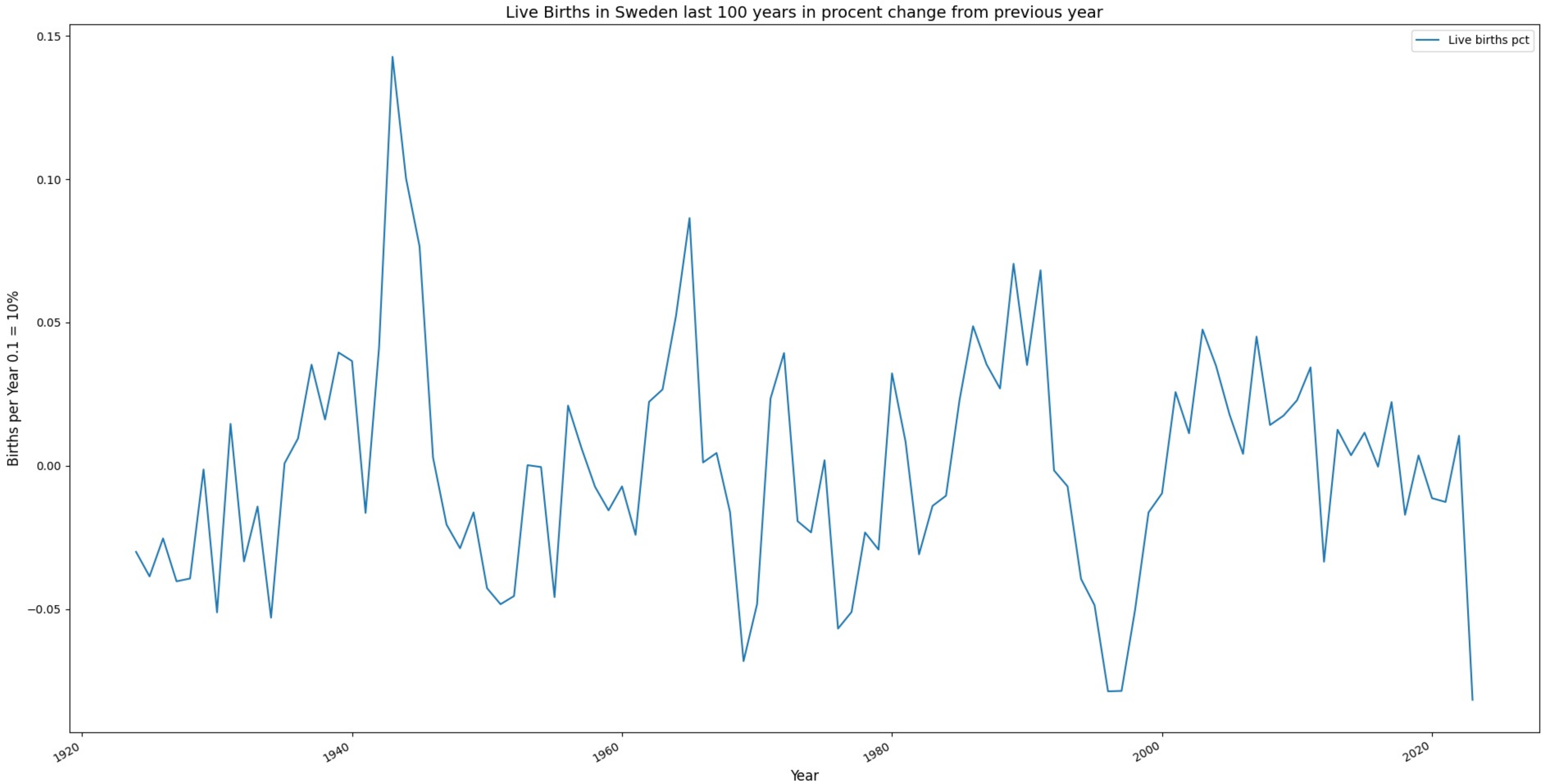
```
births = pd.read_csv(save_path + 'be0101_tab9utv1749-2021-1a.csv',delimiter=',')
birthsPerYear = births['Live births'].values.tolist() + [births22]
years = births['year'].values.tolist() + [2022]
```

Calculate change in procent from one year to the next:

```
birth_proc = births1749_2022['Live births'].pct_change().fillna(0.)  #.dropna()
births1749_2022['Live births pct'] = birth_proc.values
```

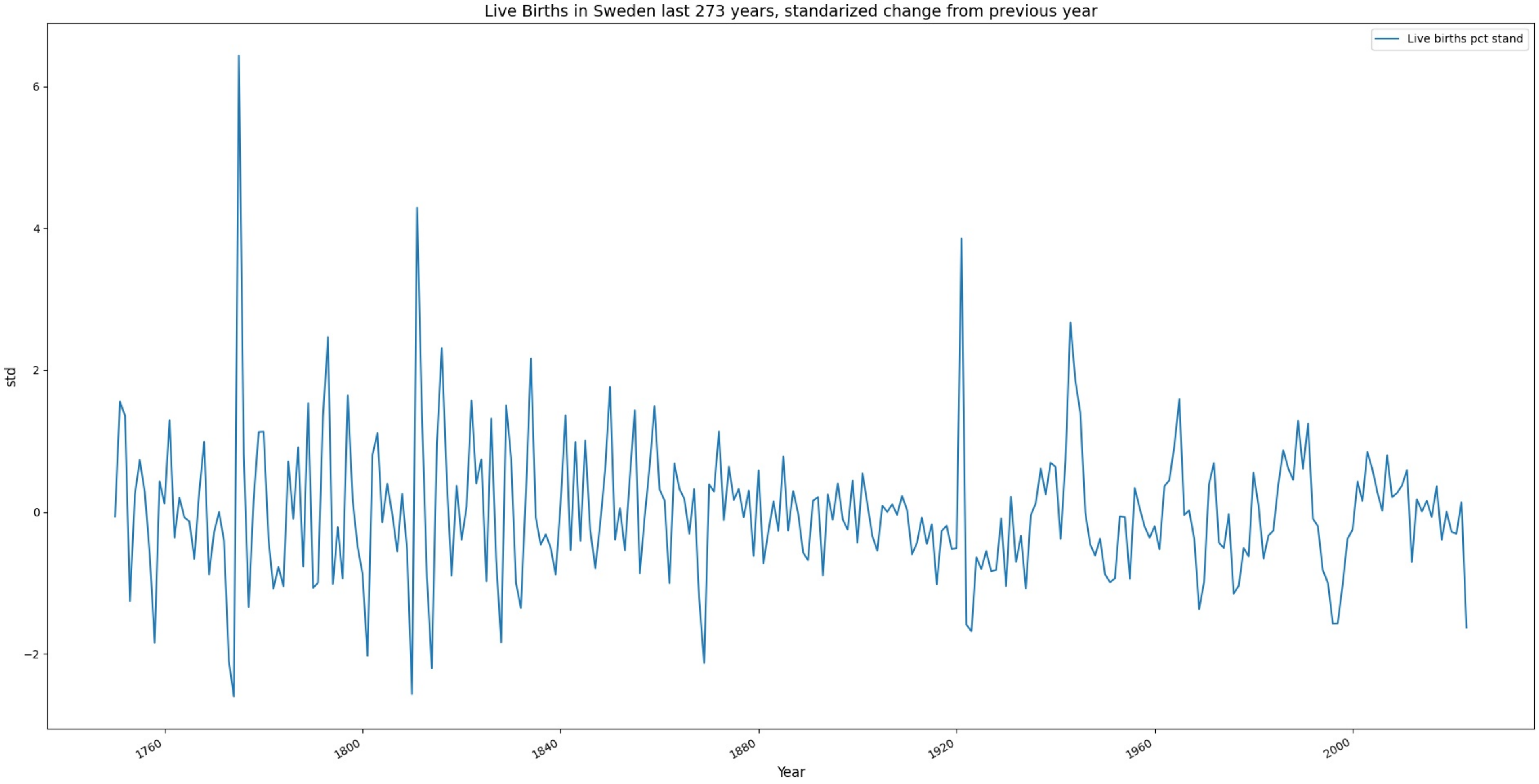
https://github.com/marl2en/predictFuture

The change from 2021 to 2022 is estimated to: -8.168 %
A plot of the last 100 years is shown below.




```
births1749_2022.iloc[-1]
year                        2022
Live births               104930
Date                  2022-12-31
Live births pct         -0.08168
Live births pct stand  -1.629892
Name: 2022-12-31, dtype: object
```

https://github.com/marl2en/predictFuture



A standardized version of the last 273 years.

Final Thoughts

For this study the total number of births were used and not the birth rate.

During january 2019 to november 2022 the population increased by 278262

```
births_month['Population at the end of period'].iloc[-1]-births_month['Population at the end of period'].iloc[0]
278262
```

There was a surplus in immigration. See image immigrationSurplus.png

In summer 2021 occurred an undenialable change of trend.

For how long this trends will last is uncertain.

Hypothesis about the cause/causes:

A: Uncertainty because of the pandemic especially autumn 2020 has led to postponed motherhood.

Argument against A: 2022 was quite normal with regard to the pandemic. The trend should recover.

B: Uncertainty because of high living cost, inflation.

Against B: Inflation was not an issue in 2020 and not so much in 2021 either. No reports of low birth rate from developing countries I know of.

C: An experimential vaccin approved under emergency rules only without long time safety data and not tested for possible impact on fertility or cancer.

Against C: Without a control group there is no clear evidence to prove hypothesis C wrong.