Semester project: Forecasting on CHF with ML methods

Valentin Parisot

July 18, 2022

1 Main idea

Forecasting a currency is a process of making predictions about the future exchange rate against others currencies based on trend analysis and past and present data. So essentially data is collected, studied and the analysis is done to forecast future scenarios that are likely to occur. We are interested in CHF forecasting against major currencies, we will focus on CHF/EUR and CHF/USD. To achieve the goal of forecasting we can use classic methods as conventional rolling or recursive OLS regression, but our goal is to achieve forecasting with machine learning methods (ML) we will then discuss of two methods stemming from the field of sequential learning. Those methods are general forecasting methods that were not specifically designed for exchange rate forecasting, we can use them to solve others various problems such as air quality and electricity consumption forecasting; see Cesa-Bianchi and Lugosi, (2006); Mauricette et al., (2009); Stoltz, (2010).

- The first method is the exponentially weighted average strategy with discount factors (EWA). This method was introduced in the early 90s by Vovk(1990) and Littlestone and Warmuth(1994). An exponentially weighted average is a measure used to model or describe a time series in finance this is useful for technical analysis and volatility modelling. In machine learning an exponential moving average is a type of moving average which applies more weight to the most recent data points, in other words this method is like giving more importance to the last experience or memories than to older one. In addition, a moving average in ML is a statistical tool to determine the direction of a trend.
- The second method is the sequential ridge regression with discount factors (SR). This method of ridge regression was introduced by Hoerl and Kennard(1970) in a stochastic setting. SR method is an evolution in the ML field of the basic OLS regressions, an OLS regression is the method used to find the simple linear regression of a set of data, the problem with the OLS regression is that it tends to overfit past data. The goal is to prevent this, then we add a regularization term to the squared error to help control the range of components $\beta_{j,t}$. In general ridge regression solves the problem of overfitting, as just regular squared regression fails to recognize the less important features and uses all of them, leading to overfitting as discussed above and this is one of the main problem of OLS regression. Ridge regression adds a slight bias/regularization term, to fit the model according to the true values of the data. This bias or regularization term will be λ for us, which is an hyperparameter.

Both methods are used with discount factors, in modelling it represents a decimal number multiplied by a cash flow value to discount it back to its present value. For us the discount factor will always be represented by γ which is a hyperparameter. Also we will use $\beta_{j,t}$ as our slope coefficients/weights based on information available up to time t and j is just representing the fundamental considered at time t. The fundamentals used here stem from the simple exchange rate models of the 1970s and Taylor-rule based models. Every methods and formula that we will discuss are based on the general forcasting equation for exchange rate for a currency pair wich follow the formula below (cf Figure 1) with the following parameters:

s_t is the logartihm of the exchange rate at time t.

α_t are the intercept up to time t.

$\beta_{j,t}$ are the slope coefficients/weights up to time t.

$f_{j,t}$ are the fundamentals at time t.

$$\widehat{s}_{t+1} - s_t = \alpha_t + \sum_{j=1}^{N} \beta_{j,t} f_{j,t} \,,$$

Figure 1: general forecasting equation

Note that the time unit is months.

Our goal is then to find $\beta_{j,t}$ with ML methods.

2 Methods

2.1 EWA:

Above we presented the idea of EWA now we will focus on the method and how it works, this forecasting methods involves different parameters, a sequence of η_t positive numbers which are the learning rates, γ which is the discount factor and $\kappa \geq 0$ which is the discount power, they are all hyperparameters. Then we need to picks weights for $\beta_{j,t}$ the rules is given by figure 2 below.

$$\beta_{j,t} = \frac{1}{Z_t} \exp\left(-\eta_t \sum_{\tau=1}^{t} \left(1 + \frac{\gamma}{(t+1-\tau)^\kappa} \right) \left(s_\tau - s_{\tau-1} - f_{j,\tau-1} \right)^2 \right),$$

where Z_t is a normalization factor:

$$Z_t = 2 \exp\left(-\eta_t \sum_{\tau=1}^{t} \left(1 + \frac{\gamma}{(t+1-\tau)^\kappa} \right) \left(s_\tau - s_{\tau-1} \right)^2 \right) + \sum_{j=1}^{N} \exp\left(-\eta_t \sum_{\tau=1}^{t} \left(1 + \frac{\gamma}{(t+1-\tau)^\kappa} \right) \left(s_\tau - s_{\tau-1} - f_{j,\tau-1} \right)^2 \right)$$

Figure 2: rules to determine $\beta_{j,t}$ in EWA

The weights obtained create a sub-convex weighted vector then all $\beta_{j,t}$ are non-negative and sum up to something smaller or equal to 1. This method has been studied and developed by principally Cesa-Bianchi et al. (1997), Cesa-Bianchi (1999), Auer et al. (2002).

2.2 SR:

Above we presented the idea of SR now we will focus on the methods and how it works, the SR method is very close to OLS regression, the OLS regression is just a special case when λ is equal to 0. This method involves 2 parameters γ which is the discount factor and $\lambda \geq 0$ which is the regularization. Then we need to picks weights for $\beta_{j,t}$ the rules is given by figure 3 below. What follows relies on recent new analyses of ridge regression in the machine learning community, see the papers by Vovk (2001) and Azoury and Warmuth (2001), as well as the survey in the book by Cesa-Bianchi and Lugosi (2006).

$$(\beta_{1,t},\, \ldots,\, \beta_{N,t}) \\ = \mathop{\arg\min}_{\beta_1,\ldots,\beta_N \in \mathbb{R}} \left\{ \lambda \sum_{j=1}^{N} \beta_j^2 + \sum_{\tau=1}^{t} \left(1 + \frac{\gamma}{(t+1-\tau)^{\kappa}}\right) \left(s_\tau - s_{\tau-1} - \sum_{j=1}^{N} \beta_j f_{j,\tau-1} \right)^2 \right\}$$

Figure 3: Figure 3: rules to determine $\beta_{j,t}$ in SR

3 Data and hyperparameters

ML methods should have superior performance over long time periods hence our data should be over the longest time series possible.

There are different samples :

• Based on end-of-month exchange rate (EOM) data from the International Monetary Fund (IMF) from March 1973 to December 2014. An end-of-month exchange rate is the exchange rate ruling on the final working day of a given month.

• Based on Federal Reserve Economic Data (FRED) averaged exchange rate data (from the St. Louis Federal Reserve), thia is a database maintained by the Research division of the Federal Reserve Bank of St. Louis that has more than 816,000 economic time series from various sources. They cover in particular exchange rates, interest rates, monetary aggregates,... The time series are compiled by the Federal Reserve and many are collected from government agencies such as the U.S. Census and the Bureau of Labor Statistics

• Based on end-of-month exchange rate (EOM) real time data from Organisation for Economic Co-operation and Development (OECD) from February 1999 to April 2017, those are for robustness check. The main difference with the first sample is that here we have real time data, real-time is information that is delivered immediately after collection. There is no delay in the timeliness of the information provided

Also, we need to pick the best hyperparameters for the ML methods discussed above, to achieve that we use a sequential grid search and thus only report the results obtained for a single instance of the method.The idea is to use the optimisation strategy, the classic one that we always in Ml to find those parameters.

References

https://halshs.archives-ouvertes.fr/halshs-01003914v6/document
https://data.mendeley.com/datasets/yxystdn2hz/1

Amat, C., Michalski, T., Stoltz, G., 2018a. Data set associated with the article “Fun- damentals and exchange rate forecastability with simple machine learning methods”. Posted on Mendeley.com.
URL http://dx.doi.org/10.17632/yxystdn2hz.1

Amat, C., Michalski, T., Stoltz, G., 2018b. Fundamentals and exchange rate forecasta- bility with simple machine learning methods. Supplementary material.
URL https://halshs.archives-ouvertes.fr/halshs-01003914