Calibrating the Mean-Reversion Parameter in the Hull-White Model Using Neural Networks

Georgios Moysiadis[1]([✉]), Ioannis Anagnostou[1,2], and Drona Kandhai[1,2]

[1] Quantitative Analytics, ING Bank,
Foppingadreef 7, 1102BD Amsterdam, Netherlands
`georgios.moysiadis@ing.com`
[2] Computational Science Lab, University of Amsterdam,
Science Park 904, 1098XH Amsterdam, Netherlands

Abstract. Interest rate models are widely used for simulations of interest rate movements and pricing of interest rate derivatives. We focus on the Hull-White model, for which we develop a technique for calibrating the speed of mean reversion. We examine the theoretical time-dependent version of mean reversion function and propose a neural network approach to perform the calibration based solely on historical interest rate data. The experiments indicate the suitability of depth-wise convolution and provide evidence for the advantages of neural network approach over existing methodologies. The proposed models produce mean reversion comparable to rolling-window linear regression's results, allowing for greater flexibility while being less sensitive to market turbulence.

Keywords: Neural networks · Time-dependent mean-reversion · Calibration · Interest rate models · Hull-White model

1 Introduction

Stochastic models for the evolution of interest rates are a key component of financial risk management and the pricing of interest rate derivatives. The value of these products is enormous, with the outstanding notional amount currently being in the order of hundreds of trillions of US dollars [1]. One of the most widely used interest rate models is the Hull-White model, introduced in [2].

As with every computational model, the performance of the Hull-White model is significantly affected by its parameters and an improper calibration may lead to predictive inconsistencies. Specifically, the speed of mean-reversion can influence notably its results; a small value may produce more trending simulation paths, while a larger value can result in steady evolution of the interest

---

Disclaimer: The opinions expressed in this work are solely those of the authors and do not represent in anyway those of their current and past employers.

rate. If the calibrated value does not reflect the actual market conditions, this may result in unrealistic estimates of risk exposure.

Neural networks have been proposed for the calibration of the speed of mean-reversion and volatility [3,4], as they are able to learn more complicated structures and associations in comparison to linear models. The existence of such structures is apparent, considering the relations described by the term structure of interest rate. Using neural networks to estimate variables for explicit computational models enables the experimentation with more complex and larger datasets, which can ultimately improve their performance.

The inability of simple models to exploit complex associations is perceived as an opportunity to expand the current body of research. We contribute a novel calibration method, with neural models that handle multiple historical input points of several parallel time series. Given the wide use of interest rate models and the data limitations, a future extension of neural models to exploit common features from several datasets is deemed crucial.

The rest of this paper is organized as follows. In Sect. 2 we summarize the basic features of the underlying models and methodologies that will be used. Additionally we outline several approaches related to mean-reversion calibration. In Sect. 3 we study the Hull-White model. In Sect. 4 we discuss the proposed neural network approach and the evaluation. Finally, in Sect. 5 we present our results and in Sect. 6 we summarize our findings.

2   Background and Related Work

2.1   Hull-White Interest Rate Model

The model we consider, describes interest rate movements driven by only a single source of risk, one source of uncertainty, hence it is one-factor model. This translates in mathematical terms having only one factor driven by a stochastic process. Apart from the stochastic term, the models are defined under the assumption that the future interest rate is a function of the current rates and that their movement is mean reverting. The first model to introduce the mean reverting behaviour of interest rate was proposed by Vasicek [5]. The Hull-White model is considered its extension and its SDE reads:

$$dr(t) = (\theta(t) - \alpha r(t))dt + \sigma dW(t) \tag{1}$$

where θ stands for the long-term mean, α the mean reversion, σ the volatility parameter and W the stochastic factor, a Wiener process. Calibrating the model refers to the process of determining the parameters α and σ based on historical data. $\theta(t)$ is generally selected so that the model fits the initial term structure using the instantaneous forward rate. However, its calculation involves both the volatility and mean-reversion, increasing the complexity when both are time-dependent functions.

Studying the Eq. (1) we observe the direct dependence on previous instances of interest rate which become even more obvious if we consider that θ indirectly

relies on the term structure of interest rate as well. Clearly this model incorporates both temporal patterns, expressed as temporal dependencies and the market's current expectations, while the mean reversion term suggests a cyclical behaviour, also observed in many other financial indicators.

The concept of mean reversion suggests that the interest rates cannot increase indefinitely, like stocks, but they tend to move around a mean [6], as they are limited by economic and political factors. There is more than one definition of mean reversion, varying by model and scope. Mean reversion can be defined by historical floors and peaks or by the autocorrelation of the average return of an asset [7]. In the Vasicek model family, it is defined against the long term mean value towards which the rate is moving with a certain speed.

2.2   Neural Networks

Neural networks are devised as a computational equivalent of the human brain; Every neuron is a computation node that applies a non-linear function f (e.g. sigmoid), which applies a mathematical operation on the input and the connecting weights. These nodes, like in the human brain, are interconnected. They form layers that move information forward from one layer of neurons to the next. The neurons of each layer are not connected, allowing communication only with previous and succeeding layers.

Neural networks have evolved both technically and theoretically. The latest theories that are studied propose that real life data forms lower-dimensional manifolds in its embedding space [8]. For example, a set of images of a hand-written letters that include size, rotation and other transformations can be mapped to a lower dimensional space. In this manifold, the distinct variations of the same letter are topologically close. In this context, neural networks are able to learn such manifolds from the given high dimensional datasets.

This property is largely exploited in computer vision, where transformation invariance is essential for most applications. Specific modules that are formed in this weight-learning schema, such as convolutional layers, have scaled up the performance of neural networks in problems such as tracking or image recognition. It is not unusual that handcrafted models are constantly updated or replaced by neural networks.

These qualities are also applied in financial time-series for modelling and prediction tasks [9]. Models with convolutional architectures extract temporal patterns as shapes, while models with recurrent architectures learn correlations through time points. Recurrent neural networks, extend the feed-forward paradigm allowing information to move both ways. The output of a layer is connected back to its input with the appropriate trainable weights, while feeding the results to the next layer as well.

2.3   Related Work

Currently, practitioners are calibrating the Hull-White model with a variety of methods: generic global optimizers [10], Jamshidian decomposition and swap

market model approximation [11], but arguably the most popular is linear regression [12]. The simplicity of the model, together with the flexibility that can be achieved by tweaking the length of the input window, make linear regression able to approximate both the market reality and one's expectations. The terms of the Hull-White model are reproduced by a linear model, given $t > s$:

$$r(t) = \hat{\alpha}r(s) + \hat{\theta} + \epsilon(t) \tag{2}$$

where $r(t)$ and $r(s)$ are the interest for the respective time-points t, s, $\hat{\theta}, \hat{\alpha}$ the historically implied long term-mean and speed of mean-reversion, and ϵ the measured error.

The trained variables are translated back to the Hull-White parameters:

$$\hat{\alpha} = e^{-\alpha(t-s)} \Rightarrow \alpha = \frac{-\ln\hat{\alpha}}{t-s}$$

$$\hat{\theta} = \frac{\theta}{\alpha}\left(1 - e^{-\alpha(t-s)}\right) \Rightarrow \theta = \frac{\alpha\hat{\theta}}{\left(1 - e^{-\alpha(t-s)}\right)} \tag{3}$$

$$sd_\epsilon = \sigma\sqrt{\frac{1 - e^{-2\alpha(t-s)}}{2\alpha}} \Rightarrow \sigma = sd_\epsilon\sqrt{\frac{-2\ln\hat{\alpha}}{(1-\hat{\alpha}^2)(t-s)}}$$

Although machine learning and neural networks have been widely used in finance for stock prediction [13], volatility modelling [14], currency exchange rate movement [15] and many more, only a few attempts have been made to address the calibration problem, and specifically, the estimation of theoretically consistent speed of mean-reversion.

Neural networks are utilized to calibrate the time-dependent mean-reversion of the Ornstein-Uhlenbeck process used for temperature derivatives in [16]. The neural network, as the approximator γ, is trained to predict the temperature of the next day. In this way, it is incorporating the dynamics of the process without explicit parameterization. Then by calculating the derivative with respect to the input of the network, they yield the value of mean-reversion. Starting from a simplified discretized version of Ornstein-Uhlenbeck process for $dt = 1$, using the paper's notation:

$$\hat{T}(t+1) = \alpha\hat{T}(t) + e(t)$$
$$\hat{T}(t+1) = \gamma(\hat{T}(t)) + e(t)$$

where $\hat{T}$ denotes the pre-processed temperature data at time t, α the simplified form of speed of mean reversion, σ the volatility and $e(t)$ the differential of the stochastic term[1]. Then computing the derivative, we calculate the time-dependent mean reversion value:

$$\alpha(t) = \frac{d\hat{T}(t+1)}{d\hat{T}(t)} = d\gamma/d\hat{T} \tag{4}$$

---

[1] Find the full expressions and derivation in [16].

3   Mean Reversion in the Hull-White Model

We have seen that mean reversion calculation can be approached in different ways depending on the underlying model. In the Vasicek model, the mean reversion parameter is assumed to be constant for certain period of time. This assumption is lifted in more generic versions, accepting it as a function of time. Under Hull-White, when calibrating with linear regression, α is re-calculated periodically on historical data as the day-to-day changes are not significant for sufficiently long historical data. This approximation results in values usually within the interval (0.01–0.1) [11]. It is common among practitioners to set alpha by hand, based on their experience and current view of the market.

Consider the generic Hull-White formula:

$$dr(t) = (\theta(t) - \alpha r(t))dt + \sigma dW(t) \tag{5}$$

Applying Ito's lemma for some $s < t$ yields:

$$r(t) = r(s)e^{-\alpha(t-s)} + \frac{\theta}{\alpha}\left(1 - e^{-\alpha(t-s)}\right) + \sigma e^{-\alpha(t-s)} \int_s^t e^{\alpha(t-u)}dW(u) \tag{6}$$

The discretized form of Eq. (5) for $dt = \delta t$ is:

$$r(t + \delta t) \approx \theta(t)\delta t + r(t)(1 - \alpha\delta t) + \sigma\epsilon(t) \tag{7}$$

where $\epsilon(t)$ denotes a value sampled from a Gaussian distribution with mean 0 and variance δt. To avoid notation abuse, this discretized approximation will be written as an equation in the following sections. Hull defined the expression for θ with constant α and σ as:

$$\theta(t) = F_t(0, t) + \alpha f(0, t) + \frac{\sigma^2}{\alpha}\left(1 - e^{-\alpha t)}\right) \tag{8}$$

where $F_t(0, t)$ is the derivative with respect to time t of $f(0, t)$, which denotes the instantaneous forward rate at maturity t as seen at time zero. Similar to [6], in our dataset the last term of the expression is fairly small and can be ignored.

4   Methodology

Our main approach for mean reversion calculation, is based on the assumption that historical rates can explain the future movement incorporating the sense of long-term or period mean value. The evolution of the yield curve is exploited in order to learn latent temporal patterns. This is achieved by training a generic function approximator that learns to predict the next-day interest rate. Starting from Hull-White model (Eq. (1)), by discretizing similar to (7) yields:

$$\begin{aligned} dr(t) &= (\theta(t) - \alpha r(t))dt + \sigma dW(t) \\ r(t + \delta t) &= \theta(t)\delta t + r(t)(1 - \alpha\delta t) + \sigma\epsilon(t) \end{aligned} \tag{9}$$

Let $\delta t = 1$ and $\kappa = (1 - \alpha)$

$$r(t + 1) = \theta(t) + \kappa r(t) + \sigma \epsilon(t) \qquad (10)$$

We train a neural network with input r as the function approximator to learn the generalized version of (10) expressed as:

$$r(t + 1) = \gamma(r(t)) + e(t) \qquad (11)$$

where $e(t)$ denotes the measured error. Then by calculating the derivative with respect to the input of the neural network, we can compute the values of the time function $\kappa(t)$ as:

$$\kappa(t) = \frac{dr(t + 1)}{dr(t)} = \frac{d\gamma}{dr} \qquad (12)$$

Neural networks are constructed to be fully differentiable functions and their derivation is a well defined procedure, which can be found in [17].

4.1   Can Theta Be Replaced?

Notice that in Eq. (11), θ is not an input of γ, under the assumption that since the long term mean should be based on the current and historical interest rates, only them are needed as input to the network, which will infer the implied relations. In terms of [16], by not including θ, the equivalent preprocessing of the data is omitted. However, under the Hull-White model long term mean relies on the forward rate which describes more complex relation between $r(t+1)$ and $r(t)$ that is not modeled explicitly in the network. Removing this information from the neural model makes the approximation to have a looser connection to the assumptions of the Hull-White model; in principle, it breaks the mean reverting character. To address that, we will describe two different approaches along with Eqs. (10) and (11). The first incorporates θ, calculated on top of linear regression-calibrated alpha, resulting to:

$$\hat{r}(t + 1) = r(t + 1) - \theta^{LR}(t) \qquad (13)$$

relying on a different model, such linear regression, introduces limitations and specific assumptions. This approach is not explored in the scope of this work.

As previously mentioned, the prevailing feature in the calculation of a long term mean is the forward rate. By keeping this information, the theoretical dependence of the movement of interest rate to the market expectations is partially embodied in the training data. This is realized by subtracting the first term of θ, F_t , to offer an approximation to long term mean value:

$$\hat{r}(t + 1) = r(t + 1) - F_t \qquad (14)$$

Using Eqs. (13) and (14) the neural network is used to learn the evolution of $\hat{r}$:

$$\hat{r}(t + 1) = \gamma(r(t)) + e(t) \qquad (15)$$

4.2   Mapping

In Eq. (3) mappings from linear regression parameters to Hull-White were provided. Similarly, the calculated values from neural networks need to be transformed to be usable within the interest rate model. From Eq. (12) let $\frac{dr(t+1)}{dr(t)} = \frac{d\gamma(r)}{dr} = Y(t)$. Beginning with the simple discrete form of the Hull-White model (Eq. (7)) as expressed in Eq. (12) we get:

$$
\begin{aligned}
Y(t) &= \kappa(t) \\
Y(t) &= 1 - \alpha(t) \\
\alpha(t) &= 1 - Y(t)
\end{aligned}
\tag{16}
$$

The mapping in Eq. (16), is appropriate for networks accepting input only the rate of the previous time-step $\delta t = 1$. To be consistent with the assumption that multiple historical input points can battle high mean-reversion volatility, the mapping used should be suitable for functions regardless the distance of the sampled points. We move back to constant α and Y for simplicity, and starting with the general continuous expression of Hull-White model, we insert the simple form of Hull-White model in our derivation:

$$
\begin{aligned}
Y &= \frac{d\left(r(t)e^{-\alpha\delta t} + \frac{\theta}{\alpha}\left(1 - e^{-\alpha\delta t}\right) + \sigma e^{-\alpha(t-s)}\int_s^t e^{\alpha(t-u)}dW(u)\right)}{dr(t)} \\
Y &= e^{-\alpha\delta t} \\
\alpha &= \frac{-ln(Y)}{\delta t}
\end{aligned}
\tag{17}
$$

The expression for α is the same as the one used for linear regression. The output of the networks define a partial solution for each historical time-step that is provided, i.e. the partial derivative with respect to each input, that is mapped to the time-dependent alpha function. In that sense, the output of the neural network is treated as the result of the procedure $Y(t)$:

$$
\alpha(t) = \frac{-ln(Y(t))}{\delta t}
\tag{18}
$$

4.3   Evaluation and Datasets

Three interest rate yield curve datasets were used for the experiments EUR (Fig. 2), USD (Fig. 3) and GBP (Fig. 1); The EUR dataset is comprised by 3223 data points from July 19, 2005 until November 22, 2017 with 22 maturities ranging from 1–12, 18 months, plus 2–11 years, while USD dataset starts from January 14 2002 until June 6 2011, 2446 data points with 16 maturities 1–10 and 12, 15, 20, 25, 30 and 40 years. Our datasets are fairly limited as the recorded values are daily, and in order to include all the existing rate regimes, we restricted the available maturities to 22 and 16. Our GBP data consists of 891 time-points
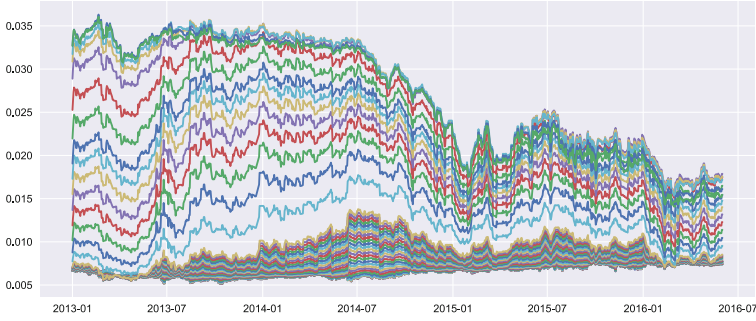
Fig. 1. Bootstrapped GBP Libor rates per maturity

from January 2, 2013 until June 1, 2016 with 44 maturities ranging from 0, 1, 7, 14 days, 1 to 24 months, 2 to 10 years and 12, 15, 20, 25, 30, 40, 50 years. The curve has been bootstrapped on top of OIS and only FRA and swap rates have been used.
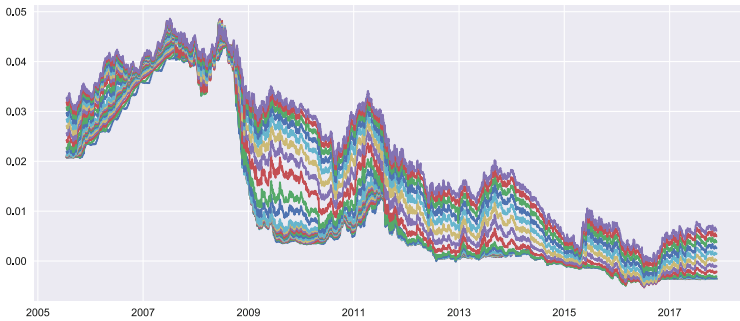

Fig. 2. European swap rate per maturity

In order to approximate the time-dependent speed of mean-reversion similarly to the linear regression method, we calculate the Pearson correlation $\in [-1, 1]$ in a rolling window manner. This way we avoid negative alpha values which cannot be used in Hull-White model. Being limited by the extent of the datasets, the window size varies from 300, 400 to 500 time points.

5   Results

Our experiments were conducted with convolution and recurrent (LSTM) based neural networks for input length of 5, 15 and 30 data points. Throughout our tests, CNNs achieved better test-set accuracy than LSTMs in all datasets, Table 1, which is not the main interest of this work, but affects the calibration procedure overall and can play a role in deciding which is the most efficient network.
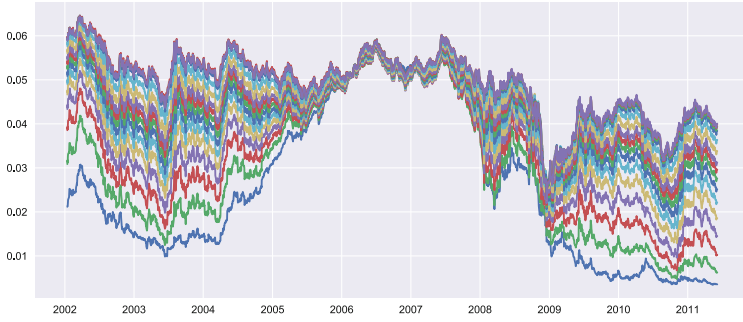
Fig. 3. USD Libor per maturity

Table 1. Average prediction error

CNN	5	15	30	LSTM	5	15	30
EUR	1.61E−04	1.08E−04	8.18E−05		3.53E−04	3.06E−04	2.96E−04
USD	1.02E−04	5.07E−05	4.92E−05		2.83E−04	2.40E−04	2.35E−04
GBP	2.10E−05	2.01E−05	2.01E−05		2.17E−05	2.24E−05	2.20E−05

The CNN approach (Fig. 5) seem to be in relative accordance with LR trends, following some of the patterns but not always agree on the levels of mean reversion. While there are mismatches and lags, we can spot several similarities in the movement. The evolution of alpha from CNN-5 suggests that a shorter history length results to less flexible model, at least for CNNs. Here we have three cases, with 5, 15 and 30 input length. On the lower end this behaviour is quite obvious but moving to more historical data points, the changes become stronger, resulting to higher and lower levels. In our tests, we have observed significantly different behaviours of CNNs based on the kernel_size/history_size
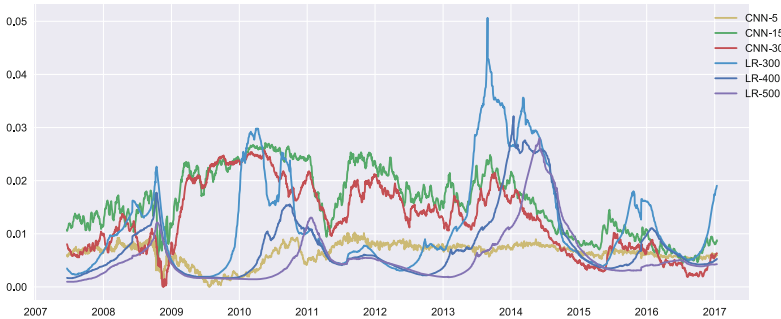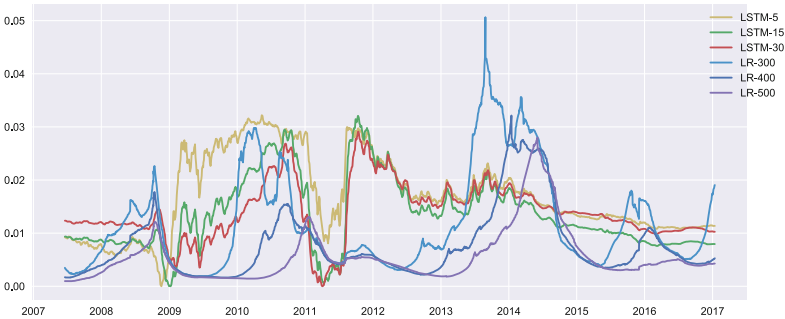

Fig. 4. EUR Mean reversion calibrated by CNN trained with Euro data

ratio affecting the smoothness and the overall performance. The networks used here, have ratios of 2/5, 1/3 and 1/3 for CNN 5, 15 and 30 respectively.
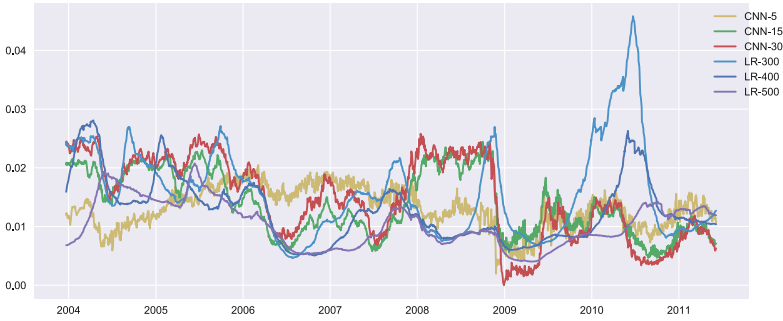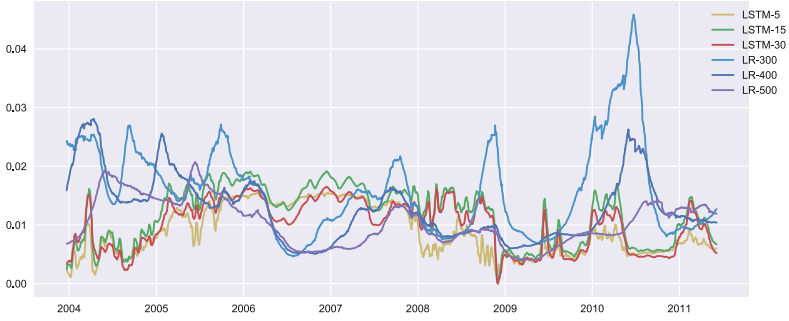

Fig. 5. USD Mean reversion calibrated by CNN trained with USD data

On the LSTM side, there is not a similar smoothness adjustment tool, which enables us to study the effect of history length easier. In Fig. 6 we see that shorter and longer history affect the outcome, but the relative movement is in close sync. In comparison to the CNN networks the levels are generally the same. We observe parallel evolution that is consistent in major changes, the two significant jumps are followed by all three networks. The first major difference from the evolution of LR (2012), seems as if the movement of LSTM is an exaggerated jump of the respective LR and CNN slopes parts. In the second (2014), we can see that the peak of the NNs coincide with the peak observed in LR-300 only, similar to CNN-30 but does not follow its level. For the USD dataset both CNN (Fig. 4) and LSTM (Fig. 7) preserve the same characteristics, while CNN-5 yields results closer to LSTMs, mostly because of the first part of the dataset (2004–2006), where lower alpha than LR and CNN (15 & 30) is reported. For the rest, LSTMs are closer to LR-500 but follow the same trends with CNNs.


Fig. 6. EUR Mean reversion calibrated by LSTM trained with Euro data

Fig. 7. USD Mean reversion calibrated by LSTM trained with USD data

In Sect. 4.1 we have discussed the role of θ in Hull-White and the significance of forward rate for its calculation. In the previous tests, the forward rate was not present in any form in the data. Following Eq. (14), we encode forward rate information by introducing the prime term of θ. The CNN-30 network is trained on this data producing the results in Fig. 8. The levels of alpha are closer to LR-500 and generally undergo smoother transitions than the simply trained CNN.
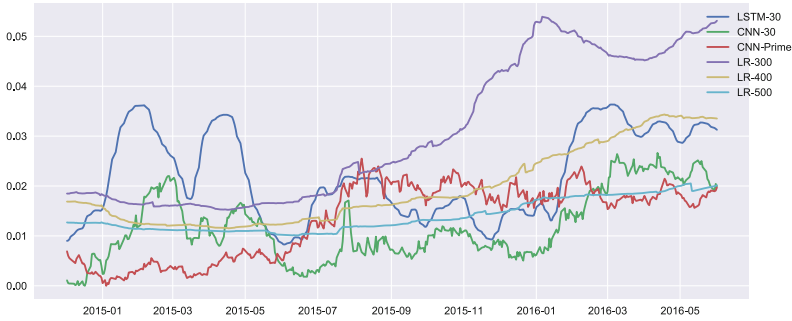
Fig. 8. CNN, LSTM, Prime calibrating GBP-Libor

Depth-wise convolution seems indeed to be suitable for parallel time-series. We observed a steady change in predictive accuracy from shorter history CNNs to longer, among the three, the network with 30-step history depth achieved the best results. However, this comparison is not simple since CNNs require certain hyper-parameter tuning, which exceeds layer count or history depth, but concerns the characteristics of the convolution layer itself; separable module, kernel size and pooling. While we focused on kernel size effects, our early tests did not favour max-pooling nor separable convolution, which led us to not further explore these options.

CNNs seem to be affected by input length and kernel size with regard to alpha calculation. 5-step networks with 2/5 kernel size ratio yield steadier results with

small day-to-day changes, whereas 15 and 30 step networks generally seem to be more flexible. Particularly, CNN-15 evolution is closer to CNN-30 movement, undergoing slightly less abrupt changes. CNN-30 is the best performing network, in terms of accuracy, but also yields alpha values comparable to linear regression's more consistently among all datasets. LSTMs did not follow exactly the same pattern in terms of predictive accuracy, all three networks, regardless the history depth, resulted in approximately the same levels.

This can be observed in alpha calculations as well, where all three networks are mutually consistent and tend to follow the same slope more closely. However, the outcome in the GBP trained networks suggests that this is the case only when they are sufficiently trained.

Using the forward prime in the training phase yields smoother results, but with level mismatches with CNN-30 and LSTM-30. Observe that, overall, all networks follow the increasing trend of linear regressions, but CNN-Prime is closer and undergoes smoother changes, while big jumps are absent (Fig. 8). This is an indication that forward rate, and market expectations as an extension, can be indeed used to extract knowledge from the market and confirms Hull's indication to use the forward rate.

Generally, the behaviour of the networks is consistent in all datasets. We see that the reported mean reversion moves upward when IR experiences fast changes. The sensitivity to these changes varies depending on the network and especially the number of maturities observed. Studying these results, we recognize the main advantage that the inclusion of forward prime offers; it lessens the sensitivity of the network to partial curve changes, augmenting the importance of the curve as a whole. Even if the evaluation of these results cannot be exact, the CNN networks seem to have greater potential and proved more suitable for prediction and mean reversion calculation. However, in real life conditions the use of a smoothing factor and the inclusion of prime are deemed necessary. LSTMs may require more data for training, but produce consistent results regardless of the number of historical data points supplied. These findings suggest that history depth can play greater role in the performance of CNN networks than LSTM.

6   Conclusion

We presented a method to calibrate the speed of the mean reversion in the Hull-White model using neural networks, based only on historical interest rate data. Our results demonstrate the suitability of depth-wise convolution and provide evidence for the advantages of neural network approach over existing methodologies. The obtained mean reversion is comparable to rolling-window linear regression's results, allowing for greater flexibility while being less sensitive to market turbulence.

In the future, we would like investigate the comparison with more advanced econometric methods than linear regression, e.g. Kalman filters [18]. Another interesting direction is the understanding of learning with neural networks.

Sound model governance and model risk management frameworks are essential for financial institutions in order to comply with regulatory requirements. This creates a constrained environment for the growth of black box methods such as neural networks. Nevertheless, the latest trends in machine learning reveal that the research in the direction of opening the black box is growing [19].

References

1. BIS: Over-the-counter derivatives statistics. https://www.bis.org/statistics/derstats.htm. Accessed 05 Feb 2018
2. Hull, J., White, A.: Pricing interest-rate-derivative securities. Rev. Financ. Stud. **3**(4), 573–592 (1990)
3. Suarez, E.D., Aminian, F., Aminian, M.: The use of neural networks for modeling nonlinear mean reversion: measuring efficiency and integration in ADR markets. IEEE (2012)
4. Zapranis, A., Alexandridis, A.: Weather derivatives pricing: modeling the seasonal residual variance of an Ornstein-Uhlenbeck temperature process with neural networks. Neurocomputing **73**, 37–48 (2009)
5. Vasicek, O.: An equilibrium characterization of the term structure. J. Financ. Econ. **5**(2), 177–188 (1977)
6. Hull, J.: Options, Futures, and Other Derivatives. Pearson/Prentice Hall, Upper Saddle River (2006)
7. Exley, J., Mehta, S., Smith, A.: Mean reversion. In: Finance and Investment Conference, pp. 1–31. Citeseer (2004)
8. Narayanan, H., Mitter, S.: Sample complexity of testing the manifold hypothesis. In: Advances in Neural Information Processing Systems, vol. 23, Curran Associates Inc., Red Hook (2010)
9. Wei, L.-Y., Cheng, C.-H.: A hybrid recurrent neural networks model based on synthesis features to forecast the Taiwan stock market. Int. J. Innov. Comput. Inf. Control **8**(8), 5559–5571 (2012)
10. Hernandez, A.: Model calibration with neural networks. Risk.net, July 2016
11. Gurrieri, S., Nakabayashi, M., Wong, T.: Calibration methods of Hull-White model, November 2009. https://doi.org/10.2139/ssrn.1514192
12. Sepp, A.: Numerical implementation of Hull-White interest rate model: Hull-white tree vs finite differences. Technical report, Working Paper, Faculty of Mathematics and Computer Science, Institute of Mathematical Statistics, University of Tartu (2002)
13. Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., Iosifidis, A.: Forecasting stock prices from the limit order book using convolutional neural networks. IEEE (2017)
14. Luo, R., Zhang, W., Xu, X., Wang, J.: A neural stochastic volatility model. arXiv preprint arXiv:1712.00504 (2017)

15. Galeshchuk, S., Mukherjee, S.: Deep networks for predicting direction of change in foreign exchange rates. Intell. Syst. Acc. Financ. Manag. **24**(4), 100–110 (2017)
16. Zapranis, A., Alexandridis, A.: Modelling the temperature time-dependent speed of mean reversion in the context of weather derivatives pricing. Appl. Math. Financ. **15**(4), 355–386 (2008)
17. LeCun, Y., Touresky, D., Hinton, G., Sejnowski, T.: A theoretical framework for back-propagation. In: Proceedings of the 1988 Connectionist Models Summer School (1988)
18. El Kolei, S., Patras, F.: Analysis, detection and correction of misspecified discrete time state space models. J. Comput. Appl. Math. **333**, 200–214 (2018)
19. Shwartz-Ziv, R., Tishby, N.: Opening the black box of deep neural networks via information. arXiv preprint arXiv:1703.00810 (2017)