

High dimensional data visualisation for systematic FX trading

Evaluating the effect of training period length on the performance of machine learning algorithms

Does more data mean better models? Maybe not for neural network systematic FX trading strategies. In this paper, we explain a qualitative approach to estimating the optimal look-back window size to create training sets for deep neural network-based FX strategies.

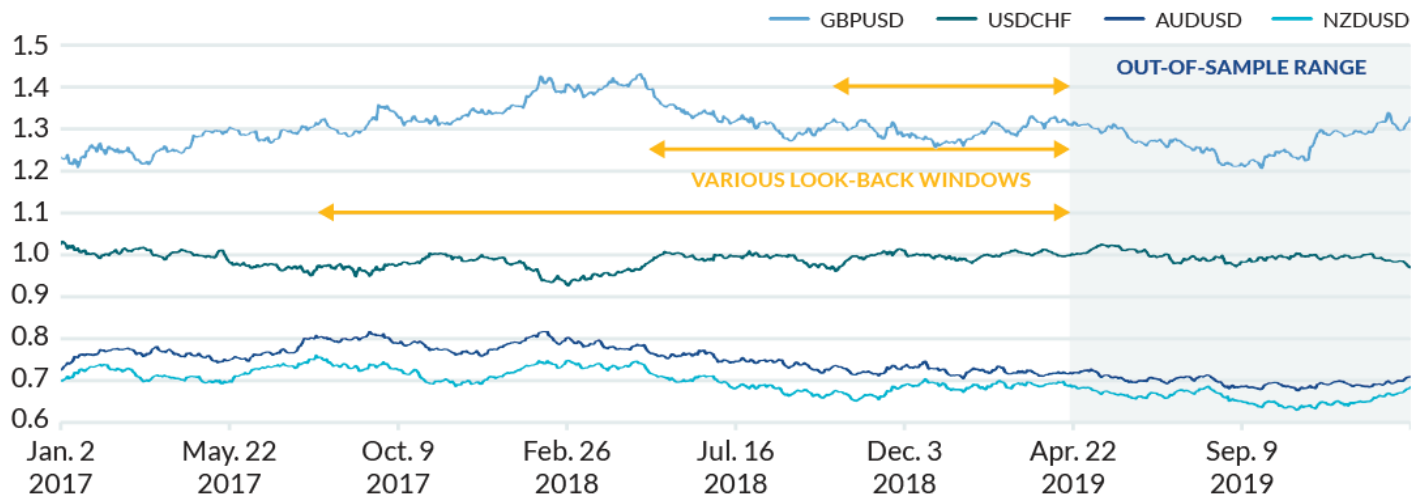
Explaining the problem: class over-representation vs. lack of data dilemma

Machine learning-based systematic FX trading algorithms utilise training sets to learn optimal model parameters. These training sets are usually constructed from various time series, sometimes with different granularities, and the main training set is usually split into several subsets to perform training, validation, model selection, and (simulated) out-of-sample evaluation.

A common approach to extracting training samples from raw sequential data is to allocate a (temporal) look-back window (LBW) length. To capture the latest market trends, these LBWs are usually selected to be close in time to the out-of-sample period. Different LBW sizes generate different training sets with their own specific statistical features. In Figure 1 we show daily spot FX rates for four (example) currency pairs. While the gray area shows the out-of-sample (test) range, various LBWs (shown as yellow arrows) can be used to construct different training sets.

The captured samples within the selected LBW range are then pre-processed and prepared for the next steps. Stationarity analysis may be necessary if the statistics of the signal change over time. Another step could be normalisation: the input signal to a machine learning pipeline is usually normalised, which can be an extremely important part of its learning performance.

FIGURE 1: LOOK-BACK WINDOWS (LBWs)



Source: Mesirow

Mesirow Currency Management

High dimensional data visualisation for systematic FX trading

Evaluating the effect of training period length on the performance of machine learning algorithms

Does more data mean better models? Maybe not for neural network systematic FX trading strategies. In this paper, we explain a qualitative approach to estimating the optimal look-back window size to create training sets for deep neural network based FX strategies.

Explaining the problem: class over-representation vs. lack of data diversity

Machine learning based systematic FX trading algorithms often training sets to learn optimal model parameters. These training sets are usually constructed from various time series, sometimes with different granularities, and the most training set is usually split into several subsets to perform training, validation, model selection, and operational out-of-sample evaluation. A common approach to estimating training samples from new requested data is to allocate a temporal look-back window (LBW) length. To capture the latest market trends, these LBWs are usually selected to be close to time to the out-of-sample period. Different LBW sizes generate different training sets with their own specific statistical features. In Figure 1 we show daily spot FX rates for four important currency pairs. While the grey area shows the out-of-sample look-back range, various LBWs (shown as yellow arrows) can be used to construct different training sets.

FIGURE 1: LOOK-BACK WINDOWS (LBW)

Jan 2 2017 May 22 2017 Oct 9 2017 Feb 26 2018 Jul 16 2018 Dec 3 2018 Apr 22 2019 Sep 9 2019

Mehmet Ozdemir
Vice President,
Senior Research
Scientist

Richard Turner
Managing Director,
Head of Research

[Read more](#)

The information contained herein is intended for institutional clients, Qualified Eligible Persons, Eligible Contract Participants and Wholesale Clients, or the equivalent classification in the recipient's jurisdiction, and is for informational purposes only. Nothing contained herein constitutes an offer to sell an interest in any Mesirow Financial investment vehicle. It should not be assumed that any trading strategy incorporated herein will be profitable or will equal past performance. Please see the disclaimer at the end of the materials for important additional information.