Deep Learning for Predictions in Emerging Currency Markets

Svitlana Galeshchuk[1,2] and Sumitra Mukherjee[3]
[1]Department of Accounting and Audit, Ternopil National Economic University, Ternopil, Ukraine
[2]Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes, Grenoble, France
[3]College of Engineering and Computing, Nova Southeastern University, Fort Lauderdale, U.S.A.

Keywords: Neural Networks, Deep Learning, Convolution Networks, Exchange Rate Prediction, Emerging Markets.

Abstract: Accurate prediction of exchange rates is critical for devising robust monetary policies. Machine learning methods such as shallow neural networks have higher predictive accuracy than time series models when trained on input features carefully crafted by domain knowledge experts. This suggests that deep neural networks, with their ability to learn abstract features from raw data, may provide improved predictive accuracy with raw exchange rates as inputs. The preponderance of research focuses on developed currency markets. The paucity of research in emerging currency markets, and the crucial role that stable currencies play in such economies, motivates us to investigate the effectiveness of deep networks for exchange rate prediction in emerging markets. Literature suggests that the Efficient Market Hypothesis, which posits that asset prices reflect all relevant information, may not hold in such markets because of extraneous factors such as political instability and governmental interventions. This motivates our hypothesis that inclusion of carefully chosen macroeconomic factors as input features may improve the predictive accuracy of deep networks in emerging currency markets. This position paper proposes novel input features based on currency clusters and presents our method for investigating the hypothesis using exchange rates from developed as well as emerging currency markets.

1 INTRODUCTION

Transactions worth billions of dollars a day take place in the foreign exchange market, making it one of the largest financial markets in the world (Report on global foreign exchange market activity in 2013). Exchange rates are expressed in terms of a base-quote currency pair that represents the number of units of quote currency that may be exchanged for each unit of the base currency. Accurate prediction of forex rate rates is critical for formulating robust monetary policies and developing effective trading and hedging strategies in the foreign exchange market (Lukas and Taylor, 2007)

Econometric models are not effective for exchange rate predictions when the forecast horizon is less than a year (Meese and Rogoff, 1983). Time series models are poor at predicting the direction of change in rates. Shallow artificial neural networks and support vector machines perform marginally better when using carefully crafted input features; significant efforts by domain experts may be needed to obtain such features from raw input data.

The recent success of deep neural networks in a variety of domains may be partially attributable to their ability to learn abstract features from raw data (LeCun et al., 2015). This suggests that deep networks may be effective in predicting foreign exchange rates based on raw time series data.

Our first objective is to investigate whether deep neural networks are significantly better at foreign exchange rate prediction than time series models and shallow networks when raw exchange rate data are used as input features. Our preliminary results using exchange rates between the US dollar and three major currencies in mature markets–Euro, British Pound, and Japanese Yen–suggest that indeed deep convolution networks perform better than extant methods.

The preponderance of research in foreign exchange prediction focuses on established markets. In response to the paucity of research in emerging currency markets, and in recognition of the fact that stable currency markets play a crucial role in determining the well-being of such economies, our second objective is to adapt deep network models for predicting exchange rates in emerging markets.

Galeshchuk S. and Mukherjee S.
Deep Learning for Predictions in Emerging Currency Markets.
DOI: 10.5220/0006250506810686
In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, pages 681-686
ISBN: 978-989-758-220-2

As representative emerging markets we consider countries in the Eastern Partnership (EaP). The Eastern Partnership is an initiative of the European Union that aims to foster improved economic relationship with the post-Soviet states of Armenia, Azerbaijan, Belarus, Georgia, Moldova, and Ukraine. Improved macroeconomic conditions in the EaP countries is a pre-condition for their economic integration with European Union. Research suggests that currency market stability is one of the most important indicators of sustainable development and growth in these economies and that accurate prediction of exchange rate is critical to the formulation of robust monetary policies. This lends further impetus to our study of developing improved models for exchange rate prediction in emerging markets.

Literature suggests that the Efficient Market Hypothesis, which posits that asset prices reflect all relevant information, may not hold in emerging markets because of extraneous factors such as political instability and governmental interventions. This motivates our hypothesis that inclusion of carefully chosen macroeconomic factors as input features may improve the predictive accuracy of deep networks in emerging currency markets. An ancillary goal of this study is to develop a novel set of input features that are obtained by forming clusters of currency markets based on distance metrics derived from correlation measures.

The roadmap for the remainder of this position paper is as follows: Section 2 formally defines the exchange rate prediction problem. Section 3 briefly discusses the related literature. Section 4 describes our proposed methodology. Section 5 concludes with some observations.

2 THE PREDICTION PROBLEM

We use a standard formulation of the exchange rate prediction problem where our goal is to predict the direction of change: Let y_t and y_{t+k} denote the values of an exchange rate between a pair of currencies in periods t and $t+k$, respectively, for some $k > 0$. Define the direction of change $z_k(t) = 1$ if the rate increases in k periods, *i.e.* if $y_{t+k} - y_t > 0$; otherwise, $z_k(t) = 0$. Our objective is to learn a function $f_k\colon \mathbb{R}^p \to \{0,1\}$ such that $f_k(y_t, y_{t-1}, \dots, y_{t-p+1}) = z_k(t)$. We train models to predict the direction of change. Let $\hat{z}_k(t) = \hat{f}_k(y_t, y_{t-1}, \dots, y_{t-p+1})$ be the predicted direction of change k periods forward, where $\hat{f}_k$ is a function learnt by a model. A k period forward prediction model model is evaluated by its classification accuracy on out-of-sample observations, where classification accuracy is defined as the percentage of test cases for which the predicted direction of change $\hat{z}_k(t)$ equals the true direction of change $z_k(t)$.

3 RELATED WORK

Exchange rate prediction methods may be categorized into econometric methods, time series models, and machine learning techniques. We re-view these approaches briefly and then discuss deep neural networks.

3.1 Econometric Models

Econometric models predict exchange rate based on economic factors. The Mundell-Fleming model (1962), Dornbusch's (1976) asset-market approach to exchange-rates, and New Keynesian models are examples of such models and a good survey of such models can be found in Engel (2013). These models are widely used by central bankers around the world. However, research indicates that these models are not effective when the prediction horizon is less than a year (Neely and Sarno, 2002).

Meese and Rogoff (1983) demonstrated that such models fail to outperform a random walk in out of sample predictions and their findings are still widely accepted.

3.2 Time Series Models

An excellent survey of time series forecasting models can be found in Box et al. (2015). Autoregressive Integrated Moving Average (ARIMA) models and Exponential Smoothing (ETS) models are the most commonly used time series models for foreign exchange rate prediction. ARIMA models can deal with non-stationary data by differencing transformations and subsume autoregressive models and moving average models as special cases. ETS models are non-stationary and can capture trends and seasonality. Time series models may provide satisfactory point estimates for exchange rates, but the direction of change implied by these estimates are often poor indicators of the true direction.

3.3 Artificial Neural Networks

Artificial neural network (ANN) with a single hidden layer often outperform time series models in providing point estimates for exchange rates as demonstrated in Dunis (2015) Thinyane and Millin (2011), Nag (2002), and Galeshchuk (2016). However, the direction of change implied by these point estimates are often unacceptably inaccurate. This renders these method less useful as a basis for formulating monetary policies. This further motivates us to investigate the ability of deep networks to predict the direction of change in forex rates.

3.4 Deep Neural Networks

Deep learning techniques originally introduced by Ivakhnenko (1971) and then Hinton (2002, 2006) has been successfully applied in a variety of domains including face detection (Osadchy et al., 2013), speech recognition (Sukittanon et al., 2004), object recognition (Schmidhuber, 2005), document categorization (Hinton and Salakhutdinov, 2006), and natural language processing (Lee et al., 2009). Deep learning networks have also been used for time series predictions (Busseti et al., 2012; Langkvist et al., 2014) and for financial predictions (Ribeiro and Noel, 2011; Chao et al. 2011; Yeh et al., 2014; Lai et al.). Restricted Boltzmann machines and auto-encoders machines have been used for dimensionality reduction and unsupervised pre-training. Applications are discussed in Larochelle et al. (2009), Masci et al. (2011), and Vincent et al. (2007).

Deep convolution networks (DN) are attractive for high dimensional prediction and classification problems (LeCun et al 2015). DNs are suitable for exchange rate prediction for two main reasons: First, high level features abstracted by the network may serve as noise filters and dimensionality reduction techniques may help abstract input features. Secondly, the temporally-local correlation between consecutive observations may be exploited to reduce the number of parameters to be estimated in the network by connecting only a small number of adjacent inputs to each unit in a hidden layer.

Our work is motivated by results from experiments to compare the accuracy of deep networks with baseline models (ARIMA, ETS, and ANN) to predict the direction of changes of exchange rates for EUR/USD, GBP/USD, and USD/JPY (Galeshchuk and Mukherjee, 2017). Results demonstrate that trained deep networks achieve better out-of-sample prediction accuracy than baseline methods.

Units in a DN receive inputs from small contiguous receptive fields that collectively cover the entire set of input features. This allows units to act as local filters and to exploit local correlation between contiguous inputs. Units share weights and bias parameters to create a feature map and this not only results in a significant reduction in the number of parameters to be estimated but also facilitates detection of features irrespectively of their actual position in the input field. The reduction in the number of parameters may be very significant as the number layers in the network and the number of units in each layer increases.

Recurrent neural networks are an effective class of neural network designed to handle sequence dependence. Stacked Long Short-Term Memory (LSTM) is a type of recurrent neural network used in deep learning which makes effective use of model parameters, converges quickly, and outperforms deep feed forward neural networks. That is why, it is often used for time-series predictions. Being adapted for dimensionality reduction and unsupervised pre-training tasks, LSTMs have been successfully used for unsupervised extraction of abstract input features for prediction problems. The approach has also proved effective in financial predictions.

4 METHODOLOGY

In this section we describe the data sets to be used in this study, discuss additional features to be used for prediction in emerging markets, present baseline models including shallow neural networks, and describe our deep convolution networks.

4.1 Data Sets

For developed currency markets, we use the daily closing rates between three currency pairs: Euro and US Dollar (EUR/USD), British Pound and US Dollar (GBP/USD), and US Dollar and Japanese Yen (USD/JPY) to train and test our models. The rates may be downloaded from: http://www.global-view.com/forex-trading-tools/forex-history/. Data for the years 2000 to 2015 are considered. For emerging currency markets we use the exchange rates of EaP countries to US Dollar: AZN/USD, AMD/USD, BYR/USD, MDL/USD, UAH/USD, GEL/USD. For each data set we train models for daily, monthly, and quarterly predictions.

4.2 Input Macroeconomic Features

In order to provide better exchange-rates prediction on the macroeconomic level, researchers develop monetary models of exchange rates based on fundamental economic data. We will include the indicators of real sector (GDP growth, unemployment, wages), current and capital account (current account balance, openness as ratio of total import and export to GDP), public and private foreign debt, capital flows, and ratio of international reserves to 3 months import, international variables (interest rates and price ratios). Some additional factors that may need to be considered include: money growth, fiscal growth, and a measure for the degree of political instability and market liberalization.

Improved exchange rate prediction models are particularly challenging to develop in volatile emerging markets with political instability as is the case in EaP economies. The EU is the main economic partners of EaP states. Financial markets of EaP countries and Russia are still highly coupled through trade and political relationships in post-soviet period. The high co-volatility of these markets requires us to identify distinct patterns of linkages among European, EaP, and Russian markets. Furthermore, contagious effect of crises is observed widely as local currency deterioration worsens macroeconomic indicators in trading partners.

The core currencies in EU-EaP-Russia area will be modelled as a network. The correlation between these exchange rates will be computed for a selected time horizon. We will use a 3 month horizon since international trading the payments are made up to 90 days. Then, each correlation coefficient in the correlation matrix of the N markets will be mapped to a metric distance between pairs of indices to form an N×N distance matrix with values ranging between 0 and 1. This distance matrix will be used to construct a minimal spanning tree (MST) in a fully connected graph where the vertices represent the currencies and the arc lengths inversely proportion to the strength of the correlations between the currencies. Clusters will be formed by removing the longest edges of the MST. Strongly correlated currencies are connected by short links and belong to the same cluster; unrelated currencies connected by longer links belong to different clusters. This will provide insights regarding the pattern of currency crises spread in the EaP economies and permit us to investigate synchronization among the currency markets in the EaP area.

4.3 Baseline Models

We use a random walk model, two time series models (ARIMA and ETS), and a single layered neural network as baseline models. The time series models provide point estimates $\hat{y}_{t+k}$ for the rates. We predict output class $\hat{z}_k(t) = 1$ if $\hat{y}_{t+k} > y_t$, and 0 otherwise. The predicted direction of change $\hat{z}_k(t)$ is compared with the actual direction of change $z_k(t)$. Results for ARIMA and ETS are obtained using the *auto.arima* model and the *ets* model from the R library *forecast* with default parameters (Hyndman and Khandakar 2008).

A neural network model with a single hidden layer will also be used in our study as a baseline model. The units have sigmoid transfer functions and use gradient descent and backpropagation for training. The model is trained on vectors with p features $(y_t, y_{t-1}, \dots, y_{t-p+1})$ as inputs and y_{t+k} as output to predict a point estimate $\hat{y}_{t+k}$ for the k period forward rate. As in the case of the time series models, we predict the output class $\hat{z}_k(t) = 1$ if $\hat{y}_{t+k} > y_t$, and 0 otherwise to compare the actual and predicted directions of change. Results are obtained using the R package *nnet*. Models parameters are tuned through cross-validation by performing a grid search over the parameter ranges using the *tune* function from the R package *e1071*. For details of these packages, see https://cran.r-project.org/web/packages/nnet/nnet.pdf and: https://cran.r-project.org/web/packages/e1071/e1071.pdf).

4.4 Deep Convolution Network

The deep convolution network has l layers of hidden units separating the input layer from the output unit. We use b_j^i to denote the internal bias of the j^{th} unit in the i^{th} layer and W_{jk}^i to represent the weight of the connection to that unit from the k^{th} unit in the $(i-1)^{\text{th}}$ layer. For an input vector $\boldsymbol{x}$, the output of j^{th} unit in the i^{th} layer is computed as $h_j^i(\boldsymbol{x}) = ReLU(a_j^i)$, where $a_j^i = b_j^i + \sum_k W_{jk}^i h_k^{i-1}(\boldsymbol{x})$, and $ReLU(a) = \max(0, a)$ is the rectified linear unit function. The output uses a *softmax* transfer function. Adam optimizer (Kingma et al 2015) is used to minimize a cross-entropy loss function. The open source library *TensorFlow* is used to create the DN models (https://www.tensorflow.org/).

4.5 Stacked Long Short-term Memory

We intend to use Stacked Long Short-Term Memory (LSTM) deep network with mechanisms for exchange-rate prediction in this experiment. LSTM network is a type of recurrent neural network used in deep learning because very large architectures can be successfully trained.

The output value of recurrent neural network (Galeshchuk, 2014) can be formulated as:

$$y = F_3(\sum_{t=1}^{n} v_t p_t - b_t),$$

$$h_t = F_2(\sum_{i=1}^{m} W_{ij}\, x_i + \sum_{k=1}^{n} v'_{kj}\, p_k(t-1) + W3\, y(t-1) - b_{2j})$$

where $F_{3,}F_2$ are logistic activation functions, n is the number of neurons in the hidden layer, v_t is the weight coefficient from j-neuron of the hidden layer to the output neuron, p_t is the output value of j-neuron of the hidden layer, b_3 is the threshold of the output neuron, m is the number of neurons in the input layer, $w_{ij}$are the weight coefficients from the i-input neuron to j-neuron of the hidden layer, x_i are the input values, b_{2f} are the thresholds of the neurons of the hidden layer, v'_k is the synapse from k context neuron of the hidden layer to the j-neuron of the same (hidden) layer, $p_k(t-1)$ is the output value of k context neuron of hidden layer in the previous moment of time $t-1$, $w3$ is the synapse from context output neuron to the j-neuron of the hidden layer, $y(t-1)$is the value of context output neuron in the previous moment of time $t-1$

For the version of LSTM used, F is implemented by the following composite function (see Graves at al., 2013):

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$$
$$C_t = f_t C_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$$
$$h_t = o_t \tanh(c_t)$$

where σ is the logistic sigmoid function, and i, f, o, c are respectively the input gate, forget gate, output gate and cell activation vectors, all of which are the same size as the hidden vector h.

5 CONCLUSIONS

This position paper outlines our approach for developing improved models for exchange rate prediction using deep neural networks. The ability of deep networks to learn abstract features from raw data motivates this approach. Preliminary results confirm that our deep network produces significantly higher predictive accuracy than the baseline models for developed currency markets. We now plan to adapt this model for exchange rate prediction in emerging currency markets by including macroeconomic factors as input features. A novel set of input features based on currency clusters may help improve predictive accuracy of such models. This study will be among the first to integrate information about market liberalization and political stability with macroeconomic indicators and time-series data on exchange rate and transaction volume. Inclusion of these factors as predictors should improve predictive accuracy for exchange rate, especially in emerging markets.

REFERENCES

Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. 2015. Time Series Analysis: Forecasting and Control, 5th Edition, Wiley.

Busseti E., Osband I., Wong S. 2012. Deep Learning for Time Series Modeling. *CS 229 Final Project Report*.

Chao J., Shen F., Zhao J. 2011. Forecasting Exchange Rate with Deep Belief Networks. *Proceedings of International Joint Conference on Neural Networks*, San Jose, California, USA.

Dornbusch R. 1976. Exchange Rate Expectations and Monetary Policy. *Journal of International Economics* 6 (3): 231–244.

Dunis C. L., Laws J., Sermpinis G. 2011. Higher order and recurrent neural architectures for trading the EUR/USD exchange rate. *Quantitative Finance* 11(4): 615-629.

Engel. C. 2013. Exchange rates and interest parity. National Bureau of Economic Research: 77.

Fleming J. M. 1962. Domestic financial policies under fixed and floating exchange rates. *IMF Staff Papers* 9: 369–379.

Galeshchuk S. 2016. Neural networks performance in exchange rate prediction. *Neurocomputing* 172: 446-452.

Galeshchuk, S., 2014. Neural-based method of measuring exchange-rate impact on international companies' revenue. *In Distributed Computing and Artificial Intelligence, 11th International Conference. Springer International Publishing:* 529-536.

Galeshchuk, S., Mukherjee S., 2016 Deep Networks for Predicting Direction of Change in Foreign Exchange Rates. *Intelligent Systems in Accounting, Finance and Management: early view papers.*

Graves, A., Mohamed, A.R. and Hinton, G., 2013, May. Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645-6649). IEEE.

Hinton G. E. 2002. Training products of experts by minimizing contrastive divergence. *Neural Comput.* 14: 1771–1800.

Hinton G. E., Osindero S., The Y. 2006. A fast learning algorithm for deep belief nets. *Neural Computations.* 18: 1527–1554.

Hinton G. E., Salakhutdinov R. 2006.Reducing the dimensionality of data with neural networks. *Science*. 313 (5786): 504–507.

Hyndman R. J., Khandakar Y. 2008. Automatic time series forecasting: the forecast package for R,. *Journal of Statistical Software* 26 (3): 1-22, 2008 DOI: http://ideas.repec.org/a/jss/jstsof/27i03.html.

Kingma, D. P., Ba, J. L. (2015). Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, 1–13.

Lai A., Li M. K., Pong F.W. Forecasting Trade Direction and Size of Future Contracts Using Deep Belief Network. Stanford University.

Langkvist M., Karlsson L., A. Loutfi. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters* 42: 11–24.

Larochelle H., Bengio Y., Louradour, P. Lamblin. 2009. Exploring strategies for training deep neural networks. *The Journal of Machine Learning Research* 10: 1-40.

LeCun Y., Bengio Y., Hinton G. 2015. Deep Learning. *Nature* **521**: 436–444.

LeCun Y., Bottou L., Bengio Y., Haffner P. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE*. 86(11); 2278–2324.

Lee H., Largman Y., Pham P., Ng A. 2009. Unsupervised feature learning for audio classification using convolutional deep belief networks. *Advances in Neural Information Processing Systems* 22.

Lukas M., Taylor M. 2007. The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis. *Journal of Economic Literature* 45 (4): 936–972.

Masci J., Meier U., Ciresan D., Schmidhuber J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. *Lecture Notes in Computer Science* 6791: 52-59.

Meese R., Rogoff K. 1983. The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification? *NBER Chapters, in Exchange Rates and International Macroeconomics*: pp. 67–112.

Mundell R. A. 1963. Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economic and Political Science* 29 (4): 475–485.

Nag A. 2002. Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting* 21(7), pp. 501- 511, 2002.

Neely C., Sarno L. 2002. How well do monetary fundamentals forecast exchange rates? *Federal Reserve Bank of St. Louis Working Paper Series:* 2002-2007.

Osadchy M., LeCun Y., Miller M. 2013. Synergistic face detection and pose estimation with energybased models. *Journal of Machine Learning Research* **8**: 1197–1215.

Report on global foreign exchange market activity in 2013. April 2013. *Triennial Central Bank Survey*. Basel, Switzerland: Bank for International Settlements. http://www.bis.org/publ/rpfx13fx.pdf.

Ribeiro B., Noel L. Deep Belief Networks for Financial Prediction. *Proceedings of ICONIP 2011*, Part III, LNCS 7064; 766–773.

Schmidhuber J. 2005. Deep Learning in Neural Networks: An Overview. *Neural Networks*: 85-117.

Simard P. Y., Steinkraus D., Platt J. C.,. 2003. Best Sukittanon S., Surendran A.C., Platt J. C., Burges C. J. 2004. Convolutional networks for speech detection. Interspeech: 1077–1080.

Thinyane H., Millin J. 2011. An investigation into the use of intelligent systems for currency trading. *Computational Economics* 37(4): 363-374.

Vincent P., Larochelle H., Bengio Y., Manzagol P. Extracting and Composing Robust Features with Denoising Autoencoders,” Proceedings of the 25th International Conference on Machine Learning (ICML 08); 1096-1103.

Wagner N., Michalewicz Z., Khouja M., McGregor R. R. 2007. Time Series Forecasting for Dynamic Environments: The DyFor Genetic Program Model. *Trans. Evol. Comp.* 11 (4): 433-452.

Xiao R. 2014. Deepnet: deep learning toolkit in R. R package version 0.2. http://CRAN.R-project.org/package=deepnet.

Yeh S-H., Wang C.J., Tsai M.F. 2014. Corporate Default Prediction via Deep Learning. ISF.