Multi-Scale Predictability for Emerging Foreign Exchange Markets

Flerskalig predikterbarhet på tillväxtmarknadens valutamarknader

Anton Ask Åström & Olle Dahlén

Supervisor: Gazi Salah Uddin
ABSTRACT

Expanding the work done by Bekiros and Marcellino (2013) to emerging markets will give a better understanding of the prediction power of artificial neural networks with a wavelet design. The financial perspective discussed will be from an investor or traders view with possible strategies and policies. Investors who seek a better understanding of future volatility or a trader that wishes to improve the prediction of returns can learn from our conclusions. Continuously improving prediction models is a difficult but rewarding task, especially considering the effort and resources already invested in the area. The investments in emerging markets have been on a decline the past years but still has the possibility to bloom into a more attractive choice.
ACKNOWLEDGEMENTS

Special thanks to Gazi Salah Uddin for his guidelines and aid on the topic. He has with his expertise in the area helped us to get started and polish our work. Thanks to Robert Forschheimer and Martin Singull who instructed us on technical matters which had a high complexity.
CONTENTS

1 INTRODUCTION 6
  1.1 Aim 7
  1.2 Research Questions 7
  1.3 Delimitation 7
  1.4 Background 8
  1.5 Machine Learning - A quick introduction 8

2 EMERGING MARKETS 10
  2.1 Basic concept 10
  2.2 Central Bank Autonomy 11
  2.3 Exchange Rate Arrangement 11
  2.4 Choice of Countries to Investigate 13

3 LITERATURE REVIEW AND THEORY 15
  3.1 Literature Review 15
  3.2 Theory 19
    3.2.1 Artificial Neural Networks 19
    3.2.2 GARCH and Random Walk 20
    3.2.3 Shannon entropy 21

4 METHOD 22
  4.1 Artificial Neural Networks 22
  4.2 Wavelet Design 23
  4.3 Benchmark Results 23
  4.4 Overfitting 24
  4.5 Wavelet multiscale analysis 24
  4.6 Prediction Performance Measurements 25
    4.6.1 DM test 26
    4.6.2 Methodological contribution 26

5 DATA 27
  5.1 Summary of Data 28

6 RESULTS 34
  6.1 Discussion 37
  6.2 Conclusion 40
  6.3 Further research 40

A DATA 42

B ANN AND WAVELETS SIMPLIFIED EXPLANATION 46
  B.1 ANN, Artificial Neural Network 46
  B.2 Wavelets 47
  B.3 Wavelets and the ANN 47
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Basic ANN model</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Entropy of Different Input</td>
<td>21</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Abstraction of ANN as a mathematical function</td>
<td>22</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Sections of the data</td>
<td>27</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Raw series, log series and histogram of the log</td>
<td>29</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Raw series, log series and histogram of the log</td>
<td>29</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Raw series, log series and histogram of the log</td>
<td>30</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Raw series, log series and histogram of the log</td>
<td>30</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Decomposition of the WON Volatility</td>
<td>32</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Decomposition of the WON Returns</td>
<td>33</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Decomposition of the EUR Returns</td>
<td>42</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Decomposition of the EUR Volatility</td>
<td>43</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Decomposition of the PESO Returns</td>
<td>43</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Decomposition of the PESO Volatility</td>
<td>44</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Decomposition of the TAIWAND Returns</td>
<td>44</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Decomposition of the TAIWAND Volatility</td>
<td>45</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1  Central Bank Autonomy  11
Table 2  Exchange Rate Arrangements according to IMF  13
Table 3  Economic data  14
Table 4  Literature table  18
Table 5  Time Horizons  28
Table 6  Descriptive Statistics  31
Table 7  Shannon Entropy  36
Table 8  Results for Returns  38
Table 9  Results Volatility  39
INTRODUCTION

Emerging markets have long been interesting for investors, much due to their potential economic growth in comparison to developed markets. Though, as the very definition of an emerging market states, these markets are not as mature as a developed market and therefore might seem less safe than the more established developed market, where one can assume more predictable and hence stable markets.

The main goal of this paper is to investigate if emerging markets really are that unpredictable, where the focus will be on these markets foreign exchange markets. There are many investigations similar to this done before, the absolutely best one in the authors opinion is Bekiros and Marcellino [2013] where an investigation similar to the one made in this paper is done. The difference is that their main focus is developed foreign exchange market.

Forecasting, whether on an emerging market or not, plays a big role in today’s economics. No one can know exactly what will happen tomorrow, but with models like the one used in this paper, markets might be able to act more predictable and therefore help the stability of the overall economy. As, in times of crises, markets act unpredictably and investors easily overreact to newly published news Lim et al. [2008], one objective of forecasting market movements is to better understand and prevent future economic shocks, such as the financial crisis in 2008 or the dot-com bubble in 2000. Forecasting also has useful applications of correcting market discrepancies and leading to market efficient prices. The foreign exchange market of a certain country is important to consider in every investment decision made concerning that country. An unpredictable and unstable foreign exchange market can make a good investment turn into a bad one. Considering investments into emerging markets has been on decline for several years [IMF 2016b], an investigation that could help understanding the nature of these countries’ foreign exchange markets might be helpful to once again increase investments into emerging markets.

As for what methods to use when forecasting, a number of them are presented in the literature review and theory section of this paper. This thesis aims to replicate the prediction model of another paper,
namely Bekiros and Marcellino’s “The multiscale causal dynamics of
foreign exchange markets” and investigate if it also improves forecast-
ing on emerging foreign exchange markets compared to benchmark
models. The model used in Bekiros and Marcellino [2013], and many
other of the papers referred to in the literature review section is based
on Artificial Neural Networks, which itself is a family of machine
learning models.

In today’s information society machine learning algorithms has a
bigger impact on daily life than most of us can imagine. Examples
of features that are based on machine learning are Facebook’s face
recognition [Becker and Ortiz 2008], Apple’s Siri and an application
we might be using daily in the future, the so called autonomous cars,
or smart cars [Oliver and Pentland 2000].

ANN will be explained in more detail further on in this paper, but
as a brief introduction it is a model that is inspired by the human
brain and the way the neurons and neural networks function. This
neural network is used to estimate non-linear functions that depend
on large number of inputs.

The currencies that we use in our models are: South Korean Won
(WON), Mexican Peso (PESO), Taiwanese New Taiwan Dollar (TAI-
WAND), all of which are valued relatively to the US Dollar (USD).
For comparison purposes we will also apply the model on the Euro
(EUR). Explanations to why we chose these currencies will follow fur-
ther on in the paper.

1.1 AIM

The aim of this thesis is to examine the performance of the new
methodology by Bekiros and Marcellino [2013] on emerging markets
and discuss financial benefits of a better prediction power.

1.2 RESEARCH QUESTIONS

How can we replicate the ANN forecasting model from Bekiros and
Marcellino [2013]? Does the method perform different on emerging
markets? To whom and how can a better prediction be of importance?

1.3 DELIMITATION

This thesis will only investigate the ANN part of Bekiros and Mar-
cellino [2013]. There will be no general analysis regarding other time
series data than emerging foreign exchange markets. The research
will not investigate the hybrid model with General Autoregressive Conditional Heteroskedasticity, which will be described in the theory section 3.2.2 and ANN. In the mentioned paper by Bekiros and Marcellino, they investigate causality dependencies cross related to other currencies which this paper will not handle.

1.4 BACKGROUND

The two authors of this paper are currently enrolled in two bachelor programs, one in computer science and one in economics, so both are interested in writing this thesis where they could combine our two main areas of interest into one subject. The area of study in this paper was given to us by our supervisor, since the areas concerning the prediction is quite new to us, he helped us getting started by giving us reading materials and instructions.

Described in a very simplified way, the Bekiros and Marcellino article investigates a new way of using ANN as a prediction model for currencies, using a decomposed time-series of currency data as input, where the time-series is decomposed into different perspectives: short-term, medium-term and long-term, before being used as inputs to the ANN. The level of decomposition is chosen using Shannon entropy, which will be explained in more depth further on in the paper. This simplifies into providing the ANN with improved tools to create predictions of the inputs. The ANN is fed with information about how short-term traders impact the market, as well as how long-term traders impact the market, and it will be investigated if this method can predict the future values of the currencies more accurately.

1.5 MACHINE LEARNING - A QUICK INTRODUCTION

Machine learning is an area of computer science that in 1959 was defined by Arthur Samuel as "A field of study that gives computers the ability to learn without being explicitly programmed" [Samuel 1959]. There are a lot of different machine learning algorithms, for instance: decision tree learning, clustering, bayesian networks, Support vector machines and, what will be used in this paper, artificial neural networks. All of these have in common that they, to different extents, can learn from new data and that they build their model from the inputs and do not follow explicit imperative programming instructions.

A commonly used definition of a machine learning algorithm is: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E" and was originally stated by Tom M. Mitchell [Mitchell 1997, p. 2].
Machine learning can be categorized into supervised, unsupervised and reinforcement learning. In supervised learning the algorithm gets the example inputs and the desired outputs and the objective is to learn the relationship between them [Kohavi and Provost 1998]. Unsupervised learning is the kind of learning where no additional information is given to the algorithm, the algorithm shall itself find patterns and structure of the input data [Kohavi and Provost 1998]. In reinforcement learning the algorithm is used to interact dynamically, such as it would do with, for instance, self-driving cars [Oliver and Pentland 2000], and it is not specifically told whether it is close to the goal or not [Sutton and Barto 1998, p. 3-5].
EMERGING MARKETS

Investigating predictability of foreign exchange markets in emerging countries brings a couple of questions needed to be answered about these countries foreign exchange. Is the central bank autonomous or is it influenced by the political landscape in the country? How is the exchange rate of the investigated currency arranged? In other words, is it fixed or floating? However, first of all, what is an emerging market? All of these questions will be answered in this section.

2.1 BASIC CONCEPT

Emerging markets are countries that has some characteristics similar to that of a developed market, but still doesn’t meet all of the criteria to be called a developed market. An emerging market is still considered more economically developed than a frontier market, which is the category of countries that are considered to be the least economically developed when categorizing countries into these three groups [Barra, 2010]. There are several institutes and organizations worldwide that list countries according to this categorization, e.g MSCI, IMF, FTSE, S&P and The Economist. As for this paper we will use the classification provided by MSCI. Here, three emerging markets are investigated: South Korea, Taiwan and Mexico, all of which are categorized as emerging markets according to MSCI [MSCI, 2016].

When categorizing countries into these three groups, MSCI is using a so called classification framework [MSCI, 2016]. MSCI is using a model where a country needs to reach different levels of three different criteria to be placed in a certain group. The criteria are: economic development, size and liquidity as well as market accessibility [MSCI, 2016].

Emerging markets suffered greatly from crises during the 80s and 90s: the Latin American debt crisis in the 80s and the Asian financial crisis in the 90s. In the last decade, up until 2010, emerging markets had an overall strong growth rate and approximately 60 percent of emerging markets had a higher growth rate than in the previous decades [Cubeddu et al., 2014]. From the last decade until today, growth in emerging markets has slowed down and so has also investments into these countries [ECB, 2015]. [IMF, 2016b] reports that both a decline in capital inflows and an increase in capital outflows
Table 1: Central Bank Autonomy

<table>
<thead>
<tr>
<th>Country</th>
<th>Political Autonomy</th>
<th>Economic Autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECB</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>South Korea</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Mexico</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Taiwan</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Note: The number in the table stands for the number of criteria met according to the model presented in Arnone et al. [2007].

Note: Data for Taiwan was excluded from the report referred to, according to the IMF similar data has previously been excluded from similar reports “by request of the Chinese authorities” [IMF [2016a]].

have contributed to a slowdown in net inflows for emerging markets, much due to a narrowing gap in growth prospectives between emerging and developed markets [IMF [2016b]]. According to Cubeddu et al. [2014] the main conditions that drove the growth in the emerging markets during the last decade, in particular vigorous global trade, high commodity prices and easy financing conditions, are forecasted not to continue in the 2010:s and forward.

2.2 CENTRAL BANK AUTONOMY

A definition of central bank autonomy is presented in Arnone et al. [2007] where the autonomy is divided into political and economical autonomy. Political autonomy is defined as the “ability of the central bank to select the objectives of monetary policy”, based on eight criteria. Economic autonomy is measured on a basis of seven different criteria and is defined to be the “ability of the central bank to select its instruments” [Arnone et al. [2007]. Presented in table 1 you may find how many of these criteria each central bank meet.

2.3 EXCHANGE RATE ARRANGEMENT

The importance of whether an emerging market has a flexible exchange rate or not is briefly discussed in Cubeddu et al. [2014] where it is shown that for the median emerging market with a fixed exchange rate, higher US long-term interest rates have a statistically significant negative impact on growth, whereas the impact is not statistically significant on markets with floating exchange rate and could therefor be both positive or negative. In Reinhart and Reinhart [2001] findings show that flexible exchange rates can help soften the effects of external financial shocks. On the other hand, in Calvo and Mishkin [2003] it is argued that less attention should be focused on whether the emerging markets exchange rate is floating or not, and more focus should be on institu-
tional reforms, in order to create economic success in these countries. The main base of a successful prediction with a model like the one used in this paper though, is that the exchange rate arrangement is floating, as non pegged currencies fluctuate more and therefore are fundamentally harder to predict and interesting to investigate with these new methods.

There are many different exchange rate arrangements around the world, with different arrangements for almost every country. Whereas the most actively traded currencies fluctuate against each other, many currencies are pegged to single currencies, such as the US dollar or the Euro [Eun et al., 2012]. IMF currently classifies the different exchange rate arrangements into 10 separate regimes as shown in table 2 [IMF, 2014]. As for this paper’s sake, the currencies investigated are classified as follows: The Mexican peso has a free floating arrangement with an inflation targeting framework as their monetary policy and so has the South Korean won, aside from that they are classified as having a floating arrangement [IMF, 2014]. The Euro also has a free floating arrangement, but are classified as “Other” when it comes to the monetary policy framework [IMF, 2014]. As for Taiwan it is stated on the official website of the Central Bank of the Republic of China (Taiwan) that “The NT dollar exchange rate is determined by market forces. However, when seasonal or irregular factors disrupt the market, the CBC will step in to maintain an orderly foreign exchange market.” Central Bank of the Republic of China [2016] and might be classified as “Other managed arrangements” or some kind of soft peg according to IMF’s classification in table 2 [Eun et al., 2012].
Table 2: Exchange Rate Arrangements according to IMF

<table>
<thead>
<tr>
<th>Type</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard pegs</td>
<td>Exchange arrangement with no separate legal tender</td>
</tr>
<tr>
<td></td>
<td>Currency board arrangement</td>
</tr>
<tr>
<td>Soft pegs</td>
<td>Conventional peg</td>
</tr>
<tr>
<td></td>
<td>Pegged exchange rate within horizontal bands</td>
</tr>
<tr>
<td></td>
<td>Stabilized arrangement</td>
</tr>
<tr>
<td></td>
<td>Crawling peg</td>
</tr>
<tr>
<td></td>
<td>Crawl-like arrangement</td>
</tr>
<tr>
<td>Floating</td>
<td>Floating</td>
</tr>
<tr>
<td>regimes</td>
<td>Free floating</td>
</tr>
<tr>
<td>Residual</td>
<td>Other managed arrangement</td>
</tr>
</tbody>
</table>

Note: Exchange rate arrangements taken from [IMF 2016b].

2.4 CHOICE OF COUNTRIES TO INVESTIGATE

When choosing countries to investigate with the model presented in this paper, there were several criteria to take into consideration. Firstly, as briefly discussed in section 2.3, in the predictability aspect of a model like this one, one rather chose a floating exchange rate than one that is pegged. Otherwise the model would mostly try to predict what level the central bank wants the rate to stay on rather than what rate the market thinks is the correct rate at the moment. Secondly, it is important that the market is reasonably large. Meaning that if the market is too small and large market participants can influence the market, it is of no interest for this model. In table [3] the GDP, FDI, market capitalization etc. of the chosen countries are presented. Lastly, and perhaps the most essential aspect to take into consideration, the country has to be an emerging market. There were several countries considered that fell out of the selection when they where either considered to be a frontier or a developed market, such as Singapore or Argentina [Barra 2010].
Table 3: Economic data

<table>
<thead>
<tr>
<th>Series Name</th>
<th>Country/Year</th>
<th>2000</th>
<th>2005</th>
<th>2010</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP at market prices (current US$, million)</td>
<td>Mexico</td>
<td>683 648</td>
<td>866 346</td>
<td>1 049 925</td>
<td>1 294 690</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
<td>561 633</td>
<td>898 137</td>
<td>1 094 499</td>
<td>1 410 383</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>296 724</td>
<td>361 595</td>
<td>446 137</td>
<td>502 039</td>
</tr>
<tr>
<td>Market capitalization of listed domestic companies (% of GDP)</td>
<td>Mexico</td>
<td>18.31%</td>
<td>27.60%</td>
<td>43.27%</td>
<td>37.09%</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
<td>30.49%</td>
<td>79.94%</td>
<td>99.76%</td>
<td>85.99%</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>85.36%</td>
<td>133.69%</td>
<td>165.04%</td>
<td>165.63%</td>
</tr>
<tr>
<td>Market capitalization of listed domestic companies (current US$, million)</td>
<td>Mexico</td>
<td>125 204</td>
<td>239 128</td>
<td>454 345</td>
<td>480 245</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
<td>171 262</td>
<td>718 011</td>
<td>1 091 911</td>
<td>1 212 759</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>253 297</td>
<td>483 430</td>
<td>736 207</td>
<td>831 539</td>
</tr>
<tr>
<td>Merchandise trade (% of GDP)</td>
<td>Mexico</td>
<td>50.59%</td>
<td>51.07%</td>
<td>57.96%</td>
<td>62.49%</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
<td>59.25%</td>
<td>60.75%</td>
<td>81.46%</td>
<td>77.86%</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>98.51%</td>
<td>106.41%</td>
<td>120.10%</td>
<td>119.68%</td>
</tr>
<tr>
<td>Net foreign assets (current LCU, million)</td>
<td>Mexico</td>
<td>205 729</td>
<td>707 711</td>
<td>1 333 940</td>
<td>2 835 684</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
<td>126 061 890</td>
<td>187 629 775</td>
<td>254 496 479</td>
<td>284 302 460</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>N/A</td>
<td>10 613 957</td>
<td>14 590 580</td>
<td>19 184 751</td>
</tr>
<tr>
<td>Primary income on FDI, payments (current US$, thousand)</td>
<td>Mexico</td>
<td>6 076 294</td>
<td>8 486 329</td>
<td>9 682 339</td>
<td>16 581 483</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
<td>15 265 315</td>
<td>11 565 541</td>
<td>13 071 461</td>
<td>19 003 003</td>
</tr>
<tr>
<td></td>
<td>Taiwan</td>
<td>7 608 000</td>
<td>4 228 000</td>
<td>3 812 000</td>
<td>5 770 000</td>
</tr>
</tbody>
</table>

Note: Data obtained from Datastream [International 2016] & The World Bank [2016].
LITERATURE REVIEW AND THEORY

Here the existing works of prediction and theories of ANN will be presented in a simplified way and the specific technical details will be excluded due to the scope of this thesis. The advanced theories behind predicting financial time series with machine learning algorithms such as ANN will be explained to match the level of bachelor students in finance. For a simplified explanation see appendix B. A summary of the literature included exists in table 4.

3.1 LITERATURE REVIEW

METHODS The hybrid model using General Auto-regressive Conditional Heteroskedasticity, GARCH(1,1), and ANN for forecasting have been used in Bildirici and Ersin [2009], Kristjanpoller et al. [2014], Monfared and Enke [2014], Khashei and Bijari [2010], Roh [2007], Kristjanpoller and Minutolo [2015], Chaabane [2014] with different performances. The prediction was improved, using the same method, in Khandelwal et al. [2015], Adhikari and Agrawal [2014], Chaabane [2014], Bildirici and Ersin [2009], Roh [2007] and in some cases the previous research is not clear [Monfared and Enke, 2014]. The possibility to examine an input signal through decomposing it into a linear and a non-linear part exists. Khandelwal et al. [2015] decomposes the signal and uses Auto-regressive integrated moving average to model the linear characteristics of the input signal and ANN for the non-linear part. All of the above mentioned articles use ANN as a general tool to predict different time series. This gives an indication that ANN can be used in different fields or at least in different financial time series. This also supports the idea being able to use ANN as an universal estimator. The use of the absolute value of the returns as volatility is used in Bekiros and Marcellino [2013], but can be argued as a simplified way of measuring the volatility. The volatility could be analyzed with both negative and positive movements instead of just movements as the absolute measure describes. This may be a current delimitation but not our aim to analyze.

DATA The literature presented uses different data sets, even non-financial series such as in Khandelwal et al. [2015], Khashei and Bijari
But in most cases stock indexes has been analyzed and this type of decomposition and benchmark test will preferably be tested on volatile series that contains linear and non-linear properties. Two of the articles analyzed emerging markets, Kristjanpoller et al. [2014], Carvalhal and Ribeiro [2008], but they did not use the foreign market exchange series instead they used stock indexes. Most of the data used in the articles have been time series with many data points. A financial time series such as a foreign exchange are well documented and can be analyzed in multiple time frames due to its high resolution. Dividing a data set into training, validation and testing one need many data points to produce a high performance of the ANN network. The articles using stock indexes do not face this issue regarding the lack of data points but is still an important aspect to understand.

MODELS  The common modeling by the literature in the literature table is combining and making models which use certain characteristics and the combination creates a better prediction power. Khanddelwal et al. [2015], Kristjanpoller and Minutolo [2015], Chaâbane 2014, Wang [2009], Bildirici and Ersin [2009] use different models such as Discrete Wavelet Transform and hybrids that beats the prediction power of a single model. This selection of models to combine does not seem to be certain but an arbitrary selection. The knowledge of a data set’s characteristics seems to play an important role when selecting a model. The usage of classical models such as Black and Scholes in Wang [2009], Tseng et al. [2008] seems to increase the performance of the prediction as well. The Autoregressive conditional heteroskedasticity model have been extended and tweaked into a large group of subcategories. The main focus in most articles such as Roh [2007], Tseng et al. [2008] focus on combining these different ARCH families with ANN. This spotlight on tweaking the linear part (ARCH) of the model contributes to better performance, Roh [2007], but one might consider tweaking the non-linear part, ANN, also. Possible tweaking of ANN are using an information criterion such as Aikake, Bayesian or Schwarz information criterion. Bekiros and Marcellino [2013] uses the Schwarz information criterion to determine the amount of input lags for the ANN, but also the Ljung-Box statistics.

DECOMPOSITION  Only one paper uses the method with Shift Invariant Discrete Wavelet Transform and ANN to compare against the RW model, that paper is the one from Bekiros and Marcellino [2013]. This makes it interesting to further investigate the theory of SIDWT as a mean of aiding the ANN in predicting the next time series value. Another decomposition that Babu and Reddy [2014]
uses is low volatile and high volatile. This differs from linear and non linear but as in Bekiros and Marcellino [2013] the volatility series (absolute returns) with a higher standard deviation compared to the returns, the decomposition model performs better. Decomposing aims to aid the model with better descriptive variables of the signal.

Another possible approach would be to predict not the signal itself but the decomposition levels and after predicting reconstruct the signal from the predicted decomposition levels. This means one would predict for eight different series if the decomposition level was chosen to be eight. No paper in our study uses this approach but mentioning the idea is important for further research. This work will focus on predicting the series and not the decomposition levels.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Variables used</th>
<th>Method</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Khandelwal, R. Adhikari, G. Verma</td>
<td>2015</td>
<td>Lynx, FX, Temperature, Mining</td>
<td>DWT, linear/nonlinear</td>
<td>ARIMA linear, ANN nonlinear, DWT</td>
<td>The three together better than one each</td>
</tr>
<tr>
<td>C. Narendra Babu, B. Eswara Reddy</td>
<td>2015</td>
<td>NSE India stock index</td>
<td>High/low volatile data, no ANN</td>
<td>Decomposition, ARIMA low volatile, GARCH high volatile</td>
<td>Better than ANN, ARIMA, GARCH on their own</td>
</tr>
<tr>
<td>W. Kristjanpoller, Marcel C. Minutolo</td>
<td>2015</td>
<td>Gold-price volatility</td>
<td>Only Data</td>
<td>ANN-GARCH hybrid</td>
<td>Compared to GARCH alone, beats it</td>
</tr>
<tr>
<td>R. Adhikari, R. K. Agrawal</td>
<td>2014</td>
<td>SP500, IBM stock, GBP/USD</td>
<td>Data and feedback input</td>
<td>Hybrid, BIC, RW for linear, ANN for non linear</td>
<td>Improved performance over individual methods</td>
</tr>
<tr>
<td>W. Kristjanpoller et Al</td>
<td>2014</td>
<td>Emerging markets, Brazil, Chile, Mexico</td>
<td>Volatility, GARCH</td>
<td>ANN-GARCH hybrid</td>
<td>ANN wins sometimes, results are robust</td>
</tr>
<tr>
<td>S. Almasi Montaneda, D. Enkeb</td>
<td>2014</td>
<td>Nasdaq stocks</td>
<td>Only Data</td>
<td>ANN-GJR(1,1) Hybrid</td>
<td>ANN better in crisis, Hybrid good for volatility</td>
</tr>
<tr>
<td>S. Bekiros, M. Marcellino</td>
<td>2013</td>
<td>Foreign Exchange markets</td>
<td>DWT</td>
<td>Wavelet, Shannon Entropy</td>
<td>Wavelet ANN better Volatility</td>
</tr>
<tr>
<td>N. Chabane</td>
<td>2013</td>
<td>Electricity prices</td>
<td>Linear/nonlinear data</td>
<td>Hybrid ARIMA-ANN, linear/nonlinear</td>
<td>The hybrid beats all one each</td>
</tr>
<tr>
<td>M. Bildirici, zgr. Ersin</td>
<td>2013</td>
<td>oil prices and volatility</td>
<td>Data, volatility</td>
<td>MLP based NN, 1st of GARCH</td>
<td>Better with ANN than without</td>
</tr>
<tr>
<td>M. Khoshe, M. Bijari</td>
<td>2010</td>
<td>Sunspot, Lynx, GBP/USD</td>
<td>Only Data</td>
<td>ANN, ARMA, Hybrid</td>
<td>Good, no comparison to RW</td>
</tr>
<tr>
<td>Yi-Hsien Wang</td>
<td>2009</td>
<td>Taiwan stock index options volatility</td>
<td>BS model variables in ANN</td>
<td>Volatility with Grey-GJR-GARCH, BS inputs to ANN</td>
<td>Grey-GJR-GARCH beats other methods, no RW</td>
</tr>
<tr>
<td>M. Bildirici, zgr. Ersin</td>
<td>2009</td>
<td>Istanbul Stock Exchange</td>
<td>Only Data</td>
<td>ANN-GARCH (families) hybrid</td>
<td>Hybrid beats GARCH (families)</td>
</tr>
<tr>
<td>C. Tseng, S. Cheng, Y. Wang, J Peng</td>
<td>2008</td>
<td>Volatility of Taiwan stock index options</td>
<td>BS with ANN, Volatility EGARCH, EGARCH/Grey-EGARCH for volatility, then ANN</td>
<td>Grey-EGARCH makes ANN better than without</td>
<td></td>
</tr>
<tr>
<td>Tae Hyup Roh</td>
<td>2007</td>
<td>Volatility Korea stock market</td>
<td>Extracted using GARCH</td>
<td>NN-GARCH/EWMA/EGARCH</td>
<td>NN-EGARCH best, hybrid better than alone</td>
</tr>
<tr>
<td>R. Glen Donaldson, M. Karmatza</td>
<td>1997</td>
<td>International stock indexes</td>
<td>GARCH volatility and Data</td>
<td>GARCH into ANN for volatility</td>
<td>ANN over GJR-GARCH,</td>
</tr>
</tbody>
</table>
Figure 1: Basic ANN model

Input layer | Hidden layer | Output layer
---|---|---
Input 1 | | Output
Input 2 | | |
Input 3 | | |
Input 4 | | |
Input 5 | | |

Note: Example neural network with 5 inputs and 3 neurons in one hidden layer. In our case input 1 would be the previous day's value of the approximation, input 2 the day before that and so forth.

3.2 THEORY

For the scope of this financial thesis the methods will not be described in depth but enough to conceptually understand them. The main concepts are ANN, random walk and wavelet decomposition.

3.2.1 Artificial Neural Networks

The history behind ANN stretches all the way back to 1958 when Frank Rosenblatt created the model to replicate the brain's neurons for visualizing data and to recognize objects [Rosenblatt, 1958]. In 1969 Marvin and Seymour Papert discovered that problems such as processing power of the current computers limited the performance of the model and the popularity of ANN stagnated [Minsky and Papert, 1969]. In the last decades the processing power of computers has improved and ANN has been developed and improved to create intelligent data analytics in image processing [Ciresan et al., 2012], time series analysis [Khashei and Bijari, 2010] and in our case currency exchange [Bekiros and Marcellino, 2013]. The model in this thesis will follow the ideas from Bekiros and Marcellino with a decomposition...
of the input signal and an ANN to produce the output [Bekiros and Marcellino, 2013].

3.2.2 GARCH and Random Walk

Generalized Autoregressive Conditional Heteroskedasticity will need some basic explanation due to the involvement of the presented literature in this work. In the field of econometrics, GARCH can be used to characterize and model time series data. When testing a new model, GARCH or Random Walk are the benchmark to test against. The model assumes that the variance of the error can be determined from a function of the sizes from previous time periods error. The theory behind this was founded by Bollerslev [1986] and has been used in econometrics to compute the volatility of an economic time series.

The mathematical expression of the GARCH(p,q) model:

\[ y_t = x_t + \epsilon_t \]
\[ \epsilon_t \sim \mathcal{N}(0, \sigma^2_t) \]
\[ \sigma^2_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \sigma^2_{t-i} + \sum_{j=1}^{p} \beta_j \epsilon^2_{t-j} \] (1)

In this model we have two parameters, q and p. Therefore it is common to refer to the GARCH model as GARCH(p,q) where p is the order of \( \sigma^2 \) terms and q is the order of \( \lambda^2 \) ARCH terms. This handles the past values of the model.

When considering a financial time series as a stochastic process, the random walk model (RW) can be suitable to apply. The RW model is also set as a benchmark in the article by Carvalhal and Ribeiro [2008]. Due to its non complicated approach, the RW model is simple and robust. The next time period can be expressed as the previous time period plus an error, see equation 2. The model has proven effective in analyzing financial time series data [Kilian and Taylor, 2003].

The mathematical expression of the RW model:

\[ y_t = y_{t-1} + \epsilon_t \] (2)

This simple stochastic mathematical model can be applied to several different problems and can be used as a general model much like the ANN. This makes ANN a competitor to this model. ARIMA(p,q,d) is a model for predicting a linear series with parameters p = orders of autoregressive terms, q = order of the moving average and d = the degree of differencing. The random walk is equal to the ARIMA(0,1,0) as described in equation 2 but both terms will be used in the thesis.
## Shannon entropy

The optimum decomposition level will be chosen by finding the lowest possible entropy based on the amount of decomposition levels of the signal $y$. If $c_j$ represents the details in the j-th level approximation coefficient of $y$ in $j = 1, \ldots, J$. With properties such as the entropy $E$ is additive $E(0)$ and $E(y) = \sum_j E(c_j)$.

Shannon entropy is defined as:

$$E_{\text{shannon}}(c_j) = -c_j^2 \log(c_j^2)$$

(3)

![Figure 2: Entropy of Different Input](image)

Note: A lower entropy, left, will be easier to guess than a higher entropy, right.

The entropy measures how much the input signal fluctuates. Meaning that if a decomposition level has a lower entropy, predicting the next value of the time series will be easier. A conclusion to be drawn can be that the lower entropy the easier to guess the next value. The entropy can be explained using a histogram, see figure 2. The input with the most similar value, the left one, has a lower absolute entropy making it easier to predict that the next value should be around zero.
METHOD

The code will be implemented using MATLAB’s Neural Network and Wavelet toolboxes. The results will be presented by calculating the error of the model and comparing it to the RW benchmark model. When predicting future prices the first day will be predicted through the neural network and then fed as an input to the system and so on to create a vector of values that represents daily future values. This shift will contribute with less bias in the model. The prediction error will most likely increase in each of these iterations and the exact amount of future days have not yet been decided due to the uncertainty of the increase in error.

4.1 ARTIFICIAL NEURAL NETWORKS

The work in this thesis will, as stated above, use an optimal decomposition level according to the Shannon entropy. This decomposition will be fed as an input signal to an ANN network using feed-forward and back-propagation as a training algorithm. This will, as argued from theories above, hopefully predict time series better which have a high amount of noise and volatility. The reason for using ANN will be to capture non linear components in the time series.

Figure 3: Abstraction of ANN as a mathematical function

Note: Visualizing an ANN as a function depending on inputs.
See figure [1] on page [16] that shows an ANN network with its inputs, one hidden layer, and the connections to the output. The input variables when applying the model to a foreign exchange market is the different time series values, input 1 is today’s value, input 2 is yesterday’s value and so on. The hidden layer and connections creates the non-linear mathematical function which outputs the predicted values. The predicted output will then be a vector corresponding to next days value, the day after that and so on. Figure [3] shows an abstract model of Σ that symbols the non-linear function.

The results from an ANN model relies some on network design and on randomization of the initial weights. By testing several designs and multiple randomization initiations the design parameters of the network will be determined. How many inputs (days before) should the network handle. How many hidden layers, in order to extract more advanced features of the data, shall the network consist of. Also the choice of how many neurons per layer must be determined. Another design aspect will be to decide to use a recurrent network or not. A recurrent network can memorize the events from previous time periods which make it suitable for time series analysis.

Our work use five neurons and one hidden layer. We re-train the model for each step ahead. It has been tested with two to ten neurons and two to five hidden layers but no conclusive improvement could be determined. A larger network demands is more compute intensive.

4.2 Wavelet Design

This work follows the previous work of Bekiros and Marcellino [2013] with a decomposition of the input signal. The selection of decomposition will, as stated above, be determined using the lowest possible entropy. The reason for using wavelets in this work is to analyze the time series in different time horizons. Also aiding the ANN which is not in using simple operations such as adding, subtracting etc.

4.3 Benchmark Results

To examine the performance of our model a comparison against the random walk is done. Comparison against previous results from Bekiros and Marcellino [2013] is also analyzed to see if the performance of our work shows similar results or not. The model will be tested one to five days ahead for prediction error.
Due to the fact that ANN can be considered as a general function approximation to a certain level of error, the possibility to fit the function to the data can result in overfitting. This means that the ANN can have a low error during the training set but when predicting new values that are not from a past pattern the error will be high. This can be controlled by using cross validation when training the ANN \cite{Tetko et al., 1995}. When training, the error of predicting a future value will be determined from the cross validation set and if the error decreases during the training the performance increases. Stopping the training will only be done when the error of prediction increases when predicting the values of the cross validation, meaning that the ANN overfits the data and the out of sample error increases. This process will increase the stability and robustness of the ANN.

4.5 WAVELET MULTISCALE ANALYSIS

When analysing a time series, such as a currency, previous work has included statistical features such as GARCH and different families of it \cite{Bildirici and Ersin, 2009}. To improve the prediction further, another method decomposes the signal into different wavelets through a process called SIDWT. The decomposition of the signal has economical interpretations, where for example the first order of decomposition will capture short term trades from, for instance, day-traders and the last level of decomposition will be from a longer interval such as months or years. Another reason for using decomposition is to handle noise in a volatile and changing market. When a market have drastic changes such as in financial crisis the high amount of trade volume and activity creates disturbances in the market making it harder to make predictions \cite{Lim et al., 2008}. The aim when using different strategies are to find how different models can detect certain characteristics of a time series.

A special case is when the volatility of the input is high, such as in a crisis in the financial markets, then the model with a wavelet design seems to be the most effective \cite{Monfared and Enke, 2014}. The work done in this paper will use SIDWT for the decomposition. Improving predictions in financial crisis can be of great value and result in less economic loss due to the fact that the market price will more easily be determined. This paper will not consider the causality part of \cite{Bekiros and Marcellino, 2013}, therefore the shift-invariant property of SIDWT are not fully motivated. The analysis of different time horizons across currencies will not be examined and then the possibility to shift the signal in time might be unnecessary.
To explain the decomposition in easier terms, one selects a decomposition level that encapsulates as much as possible information of the signal and to make it easier for the ANN to analyze different shocks but still maintaining a high level of information. It will encapsulate the information by selecting a low entropy interpreted as losing a low amount of information and aid the ANN to handle noise and different fluctuations caused in different time horizons.

There are different wavelets which can be applied but following [Bekiros and Marcellino 2013] this work will use Daubechies wavelets of length eight (db8). They are commonly used on different data types. The will not be further explained due to the complexity.

When applying wavelet decomposition the time series must be of length $N = 2^j$. This is not the general case and there exists different methods to extend or overcome this issue. Our work uses a periodic extension of the signal. The signal is assumed to be periodic and can be extended with the last values until it reaches the desired length. The scope of our work does not cover developing any other method.

### 4.6 Prediction Performance Measurements

In our paper we will use Mean Forecast Square Error and Mean Average Forecast Error to examine out of sample forecast and evaluate the models performances. The error will be determined for each day for MSFE and MAFE.

The mathematical expression for MSFE:

$$MSFE = \frac{1}{N} \sum_{i=1}^{N} (Y_i - F_i)^2$$

$Y_i = \text{Known value}$

$F_i = \text{Forecasted value}$

(4)

The mathematical expression for MAFE:

$$MAFE = \frac{1}{N} \sum_{i=1}^{N} |Y_i - F_i|$$

$Y_i = \text{Known value}$

$F_i = \text{Forecasted value}$

(5)
4.6.1 DM test

The DM test will be applied to test the models predicting power. It uses the difference between the squared errors for its test variable. In our case the DM test variable will be:

\[ d = e_{\text{ANN}}^2 - e_{\text{RW}}^2 \]

\[ H_0 : E(d) = 0 \]

\[ H_1 : E(d) \neq 0 \]  \hspace{1cm} (6)

\[ |DM| = \frac{\overline{d} - \mu}{\sqrt{\frac{2\pi f(0)}{4}}} \sim N(0,1) \]

The method for using this test is to check the absolute value of the DM test variable. If it is significantly large compared to a critical value then one can check if the test variable is negative, first model has a stronger predictive power, or positive, second model has a stronger predictive power.

Using only the mean of the forecasted errors does not give enough information about the performance and therefore the DM-test is applied. Due to some haphazardly existing outliers the mean could be lower for one model. The consistency can be measured with the DM-test and therefore it reveals more information about the performance of the prediction.

As a trader one could want the information about if the series will gain or decrease the next time period. And so, instead of measuring the value of the series there could be a signal for buying or selling a financial product. This paper does not include any information about this, but it could be interesting for traders.

4.6.2 Methodological contribution

The difference between the method presented in this paper and Bekiros and Marcellino [2013] is that we choose the decomposition level individually per time series. The reason for this is to handle differences in characteristics for each time series. The series handled in Bekiros and Marcellino [2013] has similar decomposition levels but this might not be the case with emerging markets.
DATA

The data is selected on the criterion’s described in chapter 2. The selection is the emerging markets presented earlier. Markets with higher volatility may be more easily predicted due to certain characteristics such as trends [Carvalhal and Ribeiro, 2008] but they are essential for trying the idea of decomposition for a signal with a high amount of noise. The data will also cover financial crisis such as the recession in 2008 and the dot-com bubble in 2000. These financial crises can be important to examine due to the aim of handling financial data with noise and volatility.

The data will be daily closing price data of currencies valued relatively to the USD\(^1\). It will be separated into training, cross validation and testing data for the ANN. Our model will use 70% of the data sample for training, 15% for cross validation and 15% for testing the accuracy of the model. The cross validation will be used to examine how the model will perform out of sample when training. This will prevent the model from being overfitted. See figure 4 for a visual explanation of the data separation. The returns from the foreign exchange is defined as \( r_t = \log(x_t) - \log(x_{t-1}) \) where \( x_t \) represents the closing price on day \( t \).

\(^1\) United States Dollar

Figure 4: Sections of the data

Note: 70% for training, 15% cross validation, 15% test
Table 5: Time Horizons

<table>
<thead>
<tr>
<th>Decomposition Level</th>
<th>Time Horizon (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 - 4</td>
</tr>
<tr>
<td>2</td>
<td>4 - 8</td>
</tr>
<tr>
<td>3</td>
<td>8 - 16</td>
</tr>
<tr>
<td>4</td>
<td>16 - 32</td>
</tr>
<tr>
<td>5</td>
<td>32 - 64</td>
</tr>
<tr>
<td>6</td>
<td>64 - 128</td>
</tr>
<tr>
<td>7</td>
<td>128 - 256</td>
</tr>
<tr>
<td>8</td>
<td>256 - 512</td>
</tr>
</tbody>
</table>

Note: Bekiros and Marcellino [2013] uses maximum decomposition level eight due to less economic interest in higher levels. The economic interpretation of higher levels quickly reaches large timescales.

The performance of the model will be evaluated on different time horizons. Due to the decomposition of the input signal different decomposition levels will represent different time horizons. See table 5 for translations of the different decomposition levels to time horizons in days. The performance will also be benchmarked against the RW model and against a developed market to investigate differences in performance. The forecast error will be evaluated with the MSFE and MAFE. The lower error the better. See appendix A for data from each currency and the decomposition levels.

5.1 Summary of Data

The data is taken from International [2016]. The raw data is daily values of the foreign currency valued relatively to the US dollar. The period is from 1996-06-03 to 2014-02-21, 4603 samples. This makes the training period of the model between 1996-06-03 and 2013-08-05, 4403 samples. The two models are then tested on the 200 days from 2013-08-05 to 2014-02-21. This is a long period and many data points which could make both models biased on past data, but as we are consistent and shifts the time series each time we predict 5 steps ahead in both models, they have the same biased effect.

The currencies are, as stated before, South Korean Won, Mexican Peso, Taiwan Dollar and Euro. The Euro is used as a control variable against Bekiros and Marcellino [2013] to check our model and compare emerging markets against developed markets.
Figure 5: Raw series, log series and histogram of the log

![Graph of EUR]

Figure 6: Raw series, log series and histogram of the log

![Graph of WON]

The graphs of the actual market price of the currency, log graph and the histogram of the log series will be shown in figure 5 to figure 8. In the histogram both WON and PESO looks normally distributed and their skewness in table 6 are the lowest.
Figure 7: Raw series, log series and histogram of the log

Figure 8: Raw series, log series and histogram of the log
Table 6: Descriptive Statistics

<table>
<thead>
<tr>
<th>Name</th>
<th>EUR</th>
<th>PESO</th>
<th>WON</th>
<th>TAIWAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.18</td>
<td>2.37</td>
<td>7.02</td>
<td>3.46</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.15</td>
<td>0.16</td>
<td>0.13</td>
<td>0.06</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.61</td>
<td>-0.27</td>
<td>-0.04</td>
<td>-0.56</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.53</td>
<td>2.37</td>
<td>3.32</td>
<td>2.48</td>
</tr>
<tr>
<td>JB Statistics</td>
<td>326</td>
<td>132</td>
<td>21</td>
<td>292</td>
</tr>
<tr>
<td>JB Pval</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARCH Statistics</td>
<td>4576</td>
<td>4586</td>
<td>4575</td>
<td>4585</td>
</tr>
<tr>
<td>ARCH Pval</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q Statistics</td>
<td>45638</td>
<td>45369</td>
<td>44505</td>
<td>45237</td>
</tr>
<tr>
<td>Q Pval</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The Statistics are calculated on the log value of the raw input signal

Table 6: Emerging markets have a remarkably higher mean than the developed market EUR. The Jaque-Bera statistics is lower for the WON compared to the others but still all of the currencies p-value for the JB-test is significant low and the hypothesis of a normal distribution can be rejected. The Q-test are significant in all of the log series meaning that the null hypothesis of an independent distribution of the data points can be rejected. The lag for the Q-test was 10 days. The results from the ARCH-test is significant for all series and the null hypothesis that the squared residuals are not auto-correlated. One of many possibilities to explain this could be that the series is dependent on past values and are not distributed from one single distribution. This could add to the reason for using an ANN that inputs many past days and can construct a complex structure of weights to find this irregular behavior.
Here follows the raw signal and the decomposition levels of WON. The decompositions of the other currencies can be found in appendix A. We do not analyze the dependencies of the decomposition levels as Bekiros and Marcellino [2013] do so a visual representation of all currencies will not contribute to the thesis.
Figure 10: Decomposition of the WON Returns
RESULTS

Previously in this paper the forecasting model has been extensively explained and relevant theories has been presented. What follows in this section are the results of the model. Does it perform as good on emerging markets as it does on the developed markets presented in Bekiros and Marcellino [2013]? Does the model beat the RW at all, and in that case, in what situations? Lastly, the importance of the result from the DM-test is clarified.

In table 7 on page 36 the results from calculating the entropy and selecting the decomposition level in bold. The selected decomposition level never exceeds eight due to economic interpretations of the time as seen in table 5 on page 28. For the return series the maximum level eight is chosen for all series, but for the volatility it varies from currency to currency. We differ from Bekiros and Marcellino [2013] in selecting the decomposition level separately. Bekiros and Marcellino [2013] uses an average level for both the returns and volatility based on the results but our work use the lowest value for each specific currency. This makes a difference especially for the TAIWAND which in the volatility also has decomposition level eighth.

The results from our 200 days rolling prediction can be seen in table 8 on page 38 and in table 9 on page 39. The DM test statistics shows the comparison of the models. If ANN has a better predicting power the DM-statistics will be negative and if the RW predicts better it will be positive.

RETURNS In most of the cases for the returns the RW beats the ANN. It is significantly better for EUR for day 1, 3 and 5. Also significantly better for TAIWAND on day 1, 2 and for the WON on day 3, 4 and 5. But in some cases the results are inconclusive and neither the models are better than each other. For example day 2 and day 3 for PESO have a high p-value indicating no difference between the models. Also day 3 and 4 for TAIWAND, as well as day 2 for EUR, gives a high p-value. Even though the ARIMA beats the ANN, the MSFE are not much lower. The MAFE and MSFE have a low difference in all currencies and prediction steps. This finding is interesting because of the different structures and characteristics of the models.
The ANN gives more promising results in the volatility series. It is significantly better for all the days in the EUR market. Day 3, 4 and 5 for WON also have low p-values. For PESO it varies, the first day ANN is better but day 3 to 5 the ARIMA has better results, though not significant on the 10% level. One interesting observation which Bekiros and Marcellino [2013] made in the volatility series was that the ANN model predicted better the more steps ahead you predict. This behavior can be seen in TAIWAND and WON. The results are not as consistent as in Bekiros and Marcellino [2013], their work show a significantly better predicting power in all developed markets for the volatility. The reason for emerging markets to differ from developed markets are interesting but hard to determine.

One reason for using the DM-test is that one can not draw conclusions only based on the average value due to the fact that there might exist outliers which contributes to the average. This might lead to false conclusions, when the DM-test gives a more robust and descriptive results of the consistency of the models. For example, day 2 on volatility for the WON MSFE are lower for the ANN, but the DM-test does not give a significant result in favor of the ANN. When there are a big difference in MSFE the DM-test does not seem to provide contradictory results though. In most return results the MSFE are lower for ARIMA and the DM-test show a high value with significant result in some cases.
Table 7: Shannon Entropy

<table>
<thead>
<tr>
<th>Decomposition level</th>
<th>PESO Return</th>
<th>PESO Volatility</th>
<th>EUR Return</th>
<th>EUR Volatility</th>
<th>TAIWAND Return</th>
<th>TAIWAND Volatility</th>
<th>WON Return</th>
<th>WON Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.7245</td>
<td>2.1539</td>
<td>4.2724</td>
<td>1.6507</td>
<td>4.6383</td>
<td>2.7983</td>
<td>5.2458</td>
<td>2.75</td>
</tr>
<tr>
<td>3</td>
<td>2.1625</td>
<td>1.2655</td>
<td>2.312</td>
<td>0.7677</td>
<td>2.5844</td>
<td>1.9131</td>
<td>4.0912</td>
<td>1.393</td>
</tr>
<tr>
<td>4</td>
<td>1.0127</td>
<td>0.8568</td>
<td>1.1922</td>
<td>0.4165</td>
<td>1.477</td>
<td>1.3651</td>
<td>1.223</td>
<td><strong>0.8695</strong></td>
</tr>
<tr>
<td>5</td>
<td>0.5195</td>
<td>0.6122</td>
<td>0.6336</td>
<td>0.208</td>
<td>0.8255</td>
<td>0.7969</td>
<td>0.7158</td>
<td>1.2997</td>
</tr>
<tr>
<td>6</td>
<td>0.3249</td>
<td><strong>0.5756</strong></td>
<td>0.4182</td>
<td><strong>0.12</strong></td>
<td>0.5054</td>
<td>0.6349</td>
<td>0.53</td>
<td>2.0057</td>
</tr>
<tr>
<td>7</td>
<td>0.1748</td>
<td>0.7163</td>
<td>0.1979</td>
<td>0.1594</td>
<td>0.3951</td>
<td>0.6111</td>
<td>0.376</td>
<td>2.3562</td>
</tr>
<tr>
<td>8</td>
<td><strong>0.096</strong></td>
<td>0.566</td>
<td><strong>0.1246</strong></td>
<td><strong>0.1598</strong></td>
<td><strong>0.2394</strong></td>
<td><strong>0.5194</strong></td>
<td><strong>0.1936</strong></td>
<td>2.2847</td>
</tr>
<tr>
<td>9</td>
<td>0.05</td>
<td>0.711</td>
<td>0.0462</td>
<td>0.2279</td>
<td>0.1301</td>
<td>0.8166</td>
<td>0.1117</td>
<td>2.2539</td>
</tr>
<tr>
<td>10</td>
<td>0.0216</td>
<td>0.3183</td>
<td>0.0282</td>
<td>0.1713</td>
<td>0.058</td>
<td>0.5982</td>
<td>0.0502</td>
<td>1.5077</td>
</tr>
</tbody>
</table>

Note: The shannon entropy values for each decomposition level. The chosen level is when the decomposition values increases, interpreted as harder for the model to guess. No decomposition level beyond level eight is chosen. Bold marks the lowest value.
6.1 Discussion

Until now, the hard facts has been introduced. From here on the authors will discuss what conclusions can be drawn from those facts and what research there is still to be made.

Recreation of the method by Bekiros and Marcellino [2013]. As we got the same results in the EUR series as Bekiros and Marcellino [2013], ANN significantly better in volatility but not in returns we conclude that we did. We managed to reproduce similar results following the same steps in decomposing the signal and training the network. Making this reproduction have been one aim for this thesis. Even though only a reproduction of a method can be of less academic interest, the reproduction of the method was not an easy task. The complexity of the model and method are high and we discussed many problems outside of the main scope of the thesis only to reproduce the method. Theories about wavelets, ANN, ARIMA and the DM-test where new and needed research.

Differences between developed and emerging markets?
The return series does not show a big difference. But the ANN predicts the developed market, EUR, significantly better for the volatility but not in all cases for the emerging markets. This can be interpreted as the model works better for developed markets. This might be due to a high average on the log series which indicates a linear trend for the emerging markets which maybe the ARIMA model can predict better. This conclusion might vary due to other reasons and cannot be strictly determined.

How can investors use these results? As Bekiros and Marcellino [2013] concludes, the new method using SIDWT and ANN predicts the volatility series more accurate and the volatility of an investment can more easily be determined. An investor interested in reducing risk, the knowledge of future volatility can improve the confidence and aid in decision processes. They might be interested in using the model used in this work to predict future currency prices which might affect the value of their investment.

How can traders use these results? Traders, whether it is long-term traders or so called day-traders, could probably use a model much like this one. The model made and the results concluded are based on the errors of the model compared to the actual rate that day, which means that in this paper there are no buy-sell signal analysis made. Though, the model could easily be converted into an algorithm that tries to predict whether the exchange rate is going to go up or down the next time period, which could be the next day,
Table 8: Results for Returns

<table>
<thead>
<tr>
<th>Day</th>
<th>Returns</th>
<th>ANN</th>
<th>ARIMA</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAFE</td>
<td>MAFE</td>
<td>MSFE</td>
</tr>
<tr>
<td></td>
<td>EUR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.26E-03</td>
<td>1.93E-05</td>
<td>3.23E-03</td>
<td>1.88E-05</td>
</tr>
<tr>
<td>2</td>
<td>3.24E-03</td>
<td>1.91E-05</td>
<td>3.23E-03</td>
<td>1.88E-05</td>
</tr>
<tr>
<td>3</td>
<td>3.29E-03</td>
<td>1.93E-05</td>
<td>3.25E-03</td>
<td>1.89E-05</td>
</tr>
<tr>
<td>4</td>
<td>3.22E-03</td>
<td>1.87E-05</td>
<td>3.20E-03</td>
<td>1.85E-05</td>
</tr>
<tr>
<td>5</td>
<td>3.26E-03</td>
<td>1.95E-05</td>
<td>3.20E-03</td>
<td>1.84E-05</td>
</tr>
<tr>
<td></td>
<td>PESO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5.06E-03</td>
<td>4.42E-05</td>
<td>5.06E-03</td>
<td>4.34E-05</td>
</tr>
<tr>
<td>2</td>
<td>5.08E-03</td>
<td>4.37E-05</td>
<td>5.06E-03</td>
<td>4.35E-05</td>
</tr>
<tr>
<td>3</td>
<td>5.05E-03</td>
<td>4.36E-05</td>
<td>5.05E-03</td>
<td>4.34E-05</td>
</tr>
<tr>
<td>4</td>
<td>5.15E-03</td>
<td>4.47E-05</td>
<td>5.07E-03</td>
<td>4.36E-05</td>
</tr>
<tr>
<td>5</td>
<td>5.14E-03</td>
<td>4.46E-05</td>
<td>5.07E-03</td>
<td>4.36E-05</td>
</tr>
<tr>
<td></td>
<td>TAIWAND</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.22E-03</td>
<td>3.02E-06</td>
<td>1.19E-03</td>
<td>2.85E-06</td>
</tr>
<tr>
<td>2</td>
<td>1.18E-03</td>
<td>2.84E-06</td>
<td>1.17E-03</td>
<td>2.74E-06</td>
</tr>
<tr>
<td>3</td>
<td>1.17E-03</td>
<td>2.70E-06</td>
<td>1.17E-03</td>
<td>2.75E-06</td>
</tr>
<tr>
<td>4</td>
<td>1.19E-03</td>
<td>2.75E-06</td>
<td>1.16E-03</td>
<td>2.73E-06</td>
</tr>
<tr>
<td>5</td>
<td>1.17E-03</td>
<td>2.77E-06</td>
<td>1.15E-03</td>
<td>2.70E-06</td>
</tr>
<tr>
<td></td>
<td>WON</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.12E-03</td>
<td>1.62E-05</td>
<td>2.97E-03</td>
<td>1.55E-05</td>
</tr>
<tr>
<td>2</td>
<td>3.06E-03</td>
<td>1.60E-05</td>
<td>2.94E-03</td>
<td>1.51E-05</td>
</tr>
<tr>
<td>3</td>
<td>3.06E-03</td>
<td>1.68E-05</td>
<td>2.89E-03</td>
<td>1.49E-05</td>
</tr>
<tr>
<td>4</td>
<td>3.05E-03</td>
<td>1.62E-05</td>
<td>2.89E-03</td>
<td>1.49E-05</td>
</tr>
<tr>
<td>5</td>
<td>3.08E-03</td>
<td>1.63E-05</td>
<td>2.91E-03</td>
<td>1.49E-05</td>
</tr>
</tbody>
</table>

Note: The Pval is calculated on the absolute value of the DM-statistics
Table 9: Results Volatility

<table>
<thead>
<tr>
<th>Volatility</th>
<th>ANN</th>
<th>ARIMA</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAFE</td>
<td>MSFE</td>
<td>MAFE</td>
</tr>
<tr>
<td>Day</td>
<td>EUR</td>
<td>PESO</td>
<td>TAIWAND</td>
</tr>
<tr>
<td>1</td>
<td>2.52E-03</td>
<td>9.44E-06</td>
<td>2.72E-03</td>
</tr>
<tr>
<td>2</td>
<td>2.54E-03</td>
<td>9.40E-06</td>
<td>2.72E-03</td>
</tr>
<tr>
<td>3</td>
<td>2.53E-03</td>
<td>9.63E-06</td>
<td>2.71E-03</td>
</tr>
<tr>
<td>4</td>
<td>2.58E-03</td>
<td>9.73E-06</td>
<td>2.71E-03</td>
</tr>
<tr>
<td>5</td>
<td>2.61E-03</td>
<td>9.77E-06</td>
<td>2.71E-03</td>
</tr>
</tbody>
</table>

Note: The Pval is calculated on the absolute value of the DM-statistics
week or minute. As stated above, the error of the volatility series are in many of the cases significantly better than that of the RW, meaning that if trying to predict the volatility is of use for you as a trader, the model presented here could be of great use. The volatility is e.g. used for price determination in the famous Black-Scholes model \cite{Jensen1972}.

**Risk and Pricing** Many decisions can be aided with the information acquired from predictions. From an economical point of view, risk defined as volatility could beneficially use this model to predict future volatility. As risk often is defined as volatility and based on our results and Bekiros and Marcellino \cite{Bekiros2013} concludes that this wavelet ANN based model performs well with the volatility, risk prediction could, from an economic perspective, be improved with our model. Volatility have other applications as well, such as derivative pricing. Using the volatility as future pricing with the Black-Scholes could be one application. The log return series could still be assumed to be a random walk but the volatility could be predicted with the wavelet ANN model presented.

### 6.2 Conclusion

The performance of the model on emerging markets are similar to developed markets and the same results can be expected. It is not for certain that the model predicts better in the return series but in the volatility the ANN with Wavelet inputs predicts better in most cases. This also rises the interesting question if traders can trust the information given from the classical random walk method when predicting returns. Investors will benefit in predicting the volatility with the ANN model due to the lower errors with this model. They can also trust the robustness of this model to a certain extent due to the fact that our study and Bekiros and Marcellino \cite{Bekiros2013} have concluded this.

**Model Selection** As classical models continues to give robust results the effort of searching for improvements continues. The field of advanced models, such as machine learning algorithms, definitely adds an interesting field to investigate and evaluate. As this area of study is relatively new, the selection of model seems ambiguous even though ANN has a proven track record as mentioned in the presented theory.

### 6.3 Further Research

**Prediction steps** As Bekiros and Marcellino \cite{Bekiros2013} mentions, predicting further ahead could be of interest for additional research.
The trend on developed markets that the ANN model predicts better for longer prediction steps can be of interest.

**SELECTING PARAMETERS AND MODEL** The best way of determine which model to use and the parameters needed for the chosen model can also further be developed. The Schwartz criteria for ANN together with the Lung-Box test to determine input parameters are solid but determining which model works best on which data characteristics can be further researched.

**DM-TEST** As the aim is to minimize the errors of the model and compare it to another model the DM-test could also be used when determining the parameters of the model. One could focus in understanding the input data and selecting the parameters based on input data analysis. But this might lead to over-optimizing and degenerate the robustness of the model. It might indicate in which direction to go, a linear model, a non linear model but setting parameters based on past values might lead to over-fitting. One solution, which is compute intensive, might be to first get an indication on where to start with the parameters such as lags, neurons or hidden layers with Ljung-Box test. Then change the parameters within a reasonable interval determining with the DM-test which configuration of the ANN that gives the best prediction power on out of sample data. This will probably generate a more robust model which the parameters has actually been set to optimize the performance of the prediction.
Figure 11: Decomposition of the EUR Returns
Figure 12: Decomposition of the EUR Volatility

Figure 13: Decomposition of the PESO Returns
Figure 14: Decomposition of the PESO Volatility

Figure 15: Decomposition of the TAIWAND Returns
Figure 16: Decomposition of the TAIWAND Volatility
ANN AND WAVELETS SIMPLIFIED EXPLANATION

Since the topic chosen for this Bachelor thesis is quite a technically advanced one, this appendix will try to explain the two main technical concepts, namely ANN and Wavelets, in a more simplified manner. The concepts will be explained to match the level of bachelor students of finance and will have a financially applicable approach as opposed to the mathematical and technical one taken in the other parts of the paper, hence the somewhat informal explanations in the eyes of someone familiarized with the topic.

B.1 ANN, ARTIFICIAL NEURAL NETWORK

For an introduction to the history of the model and its’ super-category, machine learning, see sections 3.2.3 and 1.5. The ANN is basically a mathematical function trying to fit itself to certain inputs and outputs, which in our case is the time series of the foreign exchange chosen. The process of when the ANN is trying to fit itself to the input is commonly referred to as training the ANN. In our case, what happens during the training is that, with the inputs and outputs at hand, the ANN decides what multiples, or numerical weights, the different connections between the different neurons should have to best fit the ANN to the certain input and output.

To really get a basic understanding of this, it is helpful to look at figure 1 on page 19. The inputs start in the input layer, where every input travels forward on the connections between the so called input layer to the hidden layer. While traveling on these connections, the different inputs get a different numeric weight depending on what neuron it is traveling to. Finally, the inputs from the input layer are all mixed up and become new values in the hidden layer, whereas they travel, on new connections and get new numerical weights before arriving at the output layer. In the output layer they get added up, preferably to the value of the output, which in our case is the next day’s value of the currency.

Now, so far, the inputs have only been referred to as inputs, not to what the input of the model actually is. With this paper, in an attempt to help the ANN predict the correct next value, we separate the input time series into different time horizons, using something called wavelets.
As this appendix is written as an attempt to explain and give an understanding of the main concepts of this paper, a very simplified explanation of what the wavelet is will be given.

To be able to understand the concept of wavelets it can be useful to take a look at figure 10 on page 33 as an example. To begin with one has a time series. This main time series is in this paper referred to as the raw signal. The raw signal contains all information about the fluctuations and changes in the value of the time series. It is important to understand that this time series could be any kind of time series: a stock price, the weather in Linköping, the number of car accidents in Germany or, as in our case, a foreign exchange etc.

As there are many different operators on the foreign exchange market, some of whom buy and sell currency for short-term purposes and some who does it for longer-term purposes, the idea of this model is to try and separate these different operators into different groups. This is done separating the raw signal, into signals which only contain the information of the market operators that buy and sell on that specific time horizon, be it for 2 - 4 days, or 256 - 512 days. For a list of the different time horizons used in this paper, look at table 5 on page 28. For example, if you are a currency trader, and buy a currency one day, selling it two weeks later, your trade would be on level 3 in figure 5, grouped together with other trades done with a time horizon of 8-16 days. For an example of how the plot of this group of trades would look like for Won, look at figure 10.

The process of separating the raw signal into these different groups is called to decompose the signal. The groups that has been referred to so far, can instead be referred to as the wavelets. For this simplified explanation of wavelets there will not be any explanation on how the decomposition actually is executed.

B.3 WAVELETS AND THE ANN

Now that the decomposition is done, the different time series, or wavelets, obtained is in the approach used in this paper fed as inputs to the ANN. Once the ANN is trained and the weights on the connections between the different neurons are obtained, feeding the input time series to the ANN should give an output. In the case of this paper, that is the value the ANN believes that the foreign exchange time series will have the next day, whether it is the volatility or the return of the foreign exchange.


Upphovsrätt

Detta dokument hålls tillgängligt på Internet – eller dess framtida ersättare – under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/.

Copyright

The publishers will keep this document online on the Internet – or its possible replacement – for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/.

© Olle Dahlén, Anton Ask Åström