Jussi Airas
Jenni Koskinen
Tuomas Nummelin
Timo Pekkala
Jani Starczewski

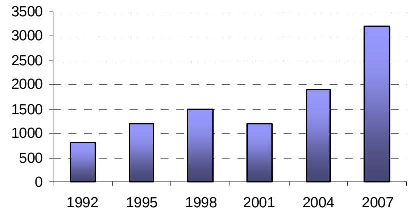
February 2008

PROJECT PLAN

REVISITING METHODS FOR AUTOMATED FX TRADING AND CONSTRUCTING A HYBRID MODEL

1 Introduction and motivation

The foreign exchange (forex, FX) market is a cash inter-bank or inter-dealer market, which is the biggest and most liquid market in the world. The average daily turnover in traditional FX markets has grown by an unprecedented 69% since April 2004, to $3.2 trillion in April 2007 (BIS, Triennal Central Bank Survey 2007). Daily averages in April for different years, in billions of US dollars, are presented on the chart below. At the same time, the currency composition of turnover has become diversified over the last three years, the share of the four largest currencies[1] fell and share of the emerging market currencies increased.



Source: BIS 2007

The ten most active traders account for almost 73% of trading volume (The Wall Street Journal Europe, (2/9/06 p. 20).[2] These large international banks provide the market with bid (buy) and ask (sell) prices. FX rates are affected by many factors such as economic and political conditions but most significant factors are interest rates, global trade, inflation, and political stability. Also psychological factors are important.

In this project we look into different formalised strategies of FX trading. The strategies are algorithms that manage an agent's currency position based on information about specific market indicators and time series data. By automating the trade with computers one can make the decision quickly and possibly analyse a wide range of external data to back up the decision. Typically, a decision about changing ones position has to be done in a fraction of a second. In our project, however, that restriction is lifted and the algorithm doesn not have to run in real time.

The general operating principle of an automated trading agent differs somewhat from a human trader. In this work we concentrate on building an algorithm that relies on the computer's number-crunching ability in normal and steady conditions. Thus we do not try to imitate the human trader's behavior and heuristics.

[1] 86 % for US dollar, 37 % for Euro, 16 % for Yen and 15 % for Pound Sterling in 2007 (BIS 2007).

[2] 17 % of volume for Deutsche Bank, 12,5 % for UBS, 7,5 % for Citigroup etc.

2 Objectives

The first objective of this work is to investigate the current literature to find information about algo-trading strategies that have been used up to the present. We will look into articles in this field published in the past ten years or so, and filter out the most promising and interesting ones. We will then categorise these methods and look into the general characteristics of each and present whether they have earlier been successful in forecasting exchange rates.

We will also take a broad look at which data inputs might be relevant in forecasting FX rates. We will investigate their dependencies and try to analyse the importance of the information each input contains.

We will then continue by taking a hands-on approach and by implementing number of selected algorithm ourselves. We will benchmark these algorithms on a number of currencies and their euro FX rates using an appropriate performance measure to differentiate on their performance.

Finally we will try to build a hybrid model to take advantage of the predictive power of the different algorithms. The hybrid should use the predictions of the underlying sub-models to generate an aggregate decision on how to manage the currency position. Here the decision-making algorithm is of special interest, as the weighing of different sub-model predictions may allow for considerable agility in changing situations. In general, a hybrid model should reduce the prediction error compared to the individual models.

In sum, research questions of the current study are the following:

- **What are the relevant models** to best forecast FX rate dynamics – for the given currencies [and thus to best maximize a speculative agent's profitability or risk-adjusted return (eg Sharpe ratio) related to conduct of currency transfers (buy, sell, neutral) and why **– pros and cons**?
- **How to improve** and combine the best techniques so that they can be incorporated into one?
- How would a hybrid model perform under certain conditions and why (demonstrate that certain information is able to enhance the performance of automated FX trading systems)?

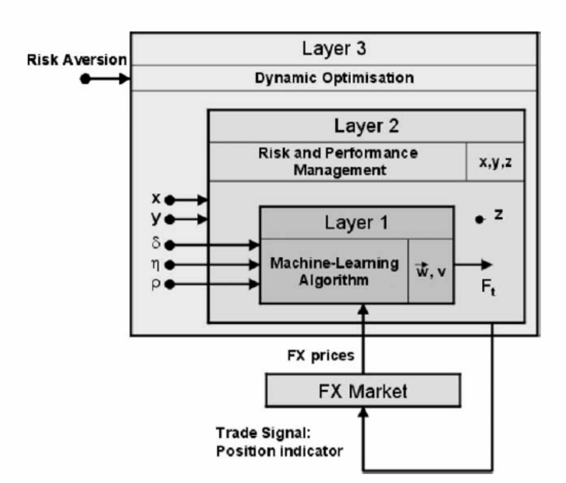
3 Action plan

3.1 Models

Literature discusses (a comprehensive list of references can be obtained from the team, will be attached to further documentation) several categories of forecasting techniques. Based on an extensive literature review, we choose the most promising models for the current forecasting task, re-construct and test some of them (categories 1, 3 and 5 below) in turn and attempt to provide an own hybrid model which incorporates the best aspects/techniques of the tested models. At the initial phase, the most promising models include the following, explanations and reasoning are bulleted.

1. **Linear models, GARCH etc.** (non-linearities exist in FX rates); improve upon random walk model.
   - Parametric non-linear models have been applied to FX forecasting, good for some but not all market conditions because not general enough to capture non-linearities in the data, however, some extensions of this category take account of non-linear structures and the changing characteristics of time series.
   - We use this category as a benchmark model (among the easiest to implement).

2. **Methods based on the Markov property** (promising but can prove too challenging to construct, given our constraints)**.**
3. **Artificial neural networks, ANNs,** may capture functional relationships within data, used eg in pattern recognition.
   - Pros and cons in this context: data-driven, self-adaptive and non-linear
   - Better than random walk models in forecasting FX rates.
4. **Support Vector Machine, SVM.**
   - recently researched as an alternative to the ARIMA and ANN approaches
   - "SVMs perform by nonlinearly mapping the input data into a high dimensional feature space by means of a kernel function and then do the linear regression in the transformed space." (Kamruzzaman, Sarker, Ahmad)."
5. **Other supporting techniques: Genetic algorithm and recurrent reinforcement learning (RRL).**
   - The following s**chematic illustration of the automated trading system (Dempster et al 2006)** captures the ideas of LLR by adding risk-management layer and dynamic optimization layer to a known algorithm. The middle (risk management, 2) layer stops trading if deemed necessary. The top layer (global trading performance optimization, 3) takes account of trade's appetite for risk and tunes system's hyperparameters.



3.2 Input data

The data set is divided into two: the **training set** and the **test set**. The training set is used for (eg ANNs) model development and the test set is used to evaluate the forecasting ability. Some of our techniques (eg, genetic programming) will discover automatically dependencies among the factors affecting the market and thus select the relevant variables to enter the model. This naturally is an advantage compared to more traditional (and popular) approaches, eg GARCH. We remain undecided for the time being whether a third set, the validation set, should be used to avoid the overfitting problem or to determine the stopping point in the training process. We distinguish between input (and modeling) for steady state and extreme market conditions and categorise data as follows.

- <u>FX market</u>: EUR/USD, EUR/SEK and EUR/RUB buy and sell rates + some other currency pairs. **Historical FX data is chosen for our key input**, the frequency is 1 min and the data spans a period from 10 Jan 2008 until 19 Feb 2008.

- Stock market: main stock index, on the relevant markets (eg MISEX, OMX , American Dow Jones Industrial Average and S&P 500 Index)
- Money market rates (eg O/N, short term interest rates) and key policy interest rates (central bank interest rates).
- Oil and basic metals (eg gold)

In addition, eg, ANNs can use fundamental input (eg foreign reserve, GDP, export and import volume, interest rates etc) and technical input (eg delayed time series data, moving average and relative strength index). Also individual forecast results can be used.

3.3 Resources

The project team has set-up two sub-groups for the better fulfillment of its ongoing task. Preference is given to ad hoc meetings with full or partial participation in case of need. The sub-groups report to the head of project. Main activities will be co-ordinated in such a way that people can assist each other whenever required.

The main activities and responsibilities are as follows.

Research (literature, available techniques, drafting, reviewing): Jani, Jenni and Timo

Development (programming and other tasks): Tuomas and Jussi

Head of project (documentation, administration incl. scheduling and follow-up, co-ordination/communication, data): Jenni

Other key issues (data): Antti Aho (Nordea)

Supervisors (ad hoc meetings on demand basis): Ahti Salo, Max Sundström

3.4 Project schedule and tasks

Given the three deadlines of the course, our project is divided into three stages as follows.

Stage 1. Kick off and preliminary work
Literature review
Selection of input variables
Techniques (that are to be adopted) determined; simple models simulated
Project plan 29 Feb 2008
Stage 2. Semi-final 28 March, 2008
Stage 3. Final 2 May, 2008

Detailed actions and calendar of the project are provided in the annex.

6 Potential risks

This type of work requires a good understanding of FX market issues as well as a knowledge of analytical work. Most critical project related risks can be found in the area of implementing certain models and constructing an own hybrid model so that it adds value to existing algorithms. Since the topic has been extensively studied – and it is widely accepted that it is difficult to make FX rate forecasting efficiently, we are confronted with a challenge when we attempt to beat the random walk model. At the same time, there is a minor risk that we fail to detect some important strand of algo-trading literature.

We are also aware that input selection is an important phase, which is highly critical for valid results. It should be noted that the already established efficient communication window with the data provider (Nordea) helps to mitigate data related risks.

Finally, taking into account the ambitious nature of the project and the time period, organization of this co-operative effort need to and will be efficiently managed. We are aware that our objectives need to be further defined so that we can work well and consistently with the view of finalising the project by 2 May 2008.

Annex. Actions detailed

Development Implementation Testing	Responsible persons	Deadline	Follow-up	Status Pending, +, ++ or +++
Constructing LLR/genetic algorithm based models	Tuomas, Jussi	Week 8	Week 9	++
Testing above models	Tuomas, Jussi	Week 10-11		+
Data (learning)	Jenni	Week 8		+++
Literature review (1st set of models)	Timo	Week 18	Every week	+
Literature review (2nd set of models)	Jenni	Week 18	Every week	+
Version of a plan	Jenni	Week 9		++
Literature review (3rd set of models and topology)	Jani	Week 18	Every week	+
Implementing other models (ARMA, neural and Markov if MATLAB code available)	Tuomas, Jussi	Week 15		Pending
Data	Jenni	Week 15		
Testing above models	Tuomas, Jussi	Week 16		Pending
Version of semi-final document	Jenni	Week 13		Pending
Implementing and testing a hybrid model	Tuomas, Jussi	Week 17		Pending
Version of final document	Jenni	Week 18		Pending