

Proposal: Using Machine Learning to Set Exchange Rates for Medium Term Contracts

Eduardo Lorie
Georgia Institute of Technology
edlorie@gatech.edu

Matthew Robinson
Georgia Institute of Technology
mrobinson72@gatech.edu

1. INTRODUCTION

One of the biggest risks involved with doing business internationally is changes in the exchange rate between the time of the signing of the contract and the times of each payment. Consider the case when a manufacturer in the United States has a contract to build an airplane for a customer in Mexico. When the exchange rate between the Mexican Peso and the US Dollar changes, one of the parties will lose money, since the payment is made in dollars, and the customer must convert a different amount of pesos into dollars to make the payment. In order to avoid this, companies need to have ways to accurately predict future exchange rates in advance and include their findings in contracts.

The proposed project addresses the need to accurately predict future exchange rates by testing different machine learning models on financial data. The models consist of methods previously used to predict exchange rates, as well as methods which have not been previously applied to exchange rates. The focus of this project is to determine which combination of features and machine learning techniques gives the most accurate results when predicting exchange rates up to one year in advance.

2. DATA

The data used in this project considers changes in exchange rates between the United States and five of its biggest trading partners. The five trading partners chosen are Europe, Canada, Mexico, Japan, and South Korea. China is excluded from this study because it fixes its exchange rates. Europe refers to the 19 countries that make up the euro-zone and are treated as a single unit in this project because they share the same currency.

The data used was obtained through Quandl, an economic data repository. All data used was collected by central banks. The primary variables of interest include the inflation rate, balance of trade, foreign exchange reserves, interest rate, gross domestic product growth rate, current account balance, and foreign direct investment. Each variable is

measured for the United States and each of the trading partners chosen. The monthly percent change in exchange rate between the United States and each these trading partners was also recorded. Some variables, such as GDP growth, current account balance, and foreign direct investment were recorded quarterly. Cubic splines were used to interpolate these variables so they can be aligned with the rest of the data.

3. QUESTIONS

This project attempts to determine whether or not a unique combination of features and machine learning methods can outperform previously presented techniques in predicting future exchange rates. In particular, this project seeks to determine if lag variables and ensemble methods, such as random forests, can achieve more accurate results than a random walk, which predicts no change in the exchange rate. The methods tested consist of different forms of regression, as well as different tree based methods. There is also an assumption that looking at lag variables can improve results, so lag variables are recorded up to nine months back.

4. METHODS AND MODELS

The regression models tested in this project are linear regression, ridge regression, LASSO regression, and principal component regression. The regression models will be tested using the full set of variables, with and without the lag variables. The variables will also be chosen for model using forward stepwise variable selection with AIC as the selection criteria. The tree based methods will consist of decision trees using all the variables (including lag variables), trimmed decision trees, decision trees using only variables chosen with AIC, and random forests using both all variables and variables chosen with AIC.

Each model tested will be evaluated by first splitting the data into testing and training subsets. 80% of the data is going to be used to train the models and the remaining 20% will be used to test the models. The data used in predicting exchange rates is highly correlated with time. In order to make sure this feature does not contribute to overfitting, the testing data will consist of a continuous block of time and the data the six months before and after the start of the test block are not used in testing.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-2138-9.

DOI: 10.1145/1235