Exchange rate Analysis

TEAM Z

PHILIP WILSON, NUNZIO GATTI, SUSMITA GANGOPADHYAY, SUMESH RAMACHANDRA, and BANG DU

An exchange rate is the rate at which one country exchanges currency with another.In our application we have tried to capture the exchange rate of one country and found out other exchange rates which are most closely correlated with its behavior.We have considered G20 countries from a time frame of Dec 1998 to Dec 2017.

ACM Reference Format:
Philip Wilson, Nunzio Gatti, Susmita Gangopadhyay, Sumesh Ramachandra, and Bang Du. 2018. Exchange rate Analysis: TEAM Z. 1, 1 (January 2018), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

As team we had first decided on a different topic "Will I live longer if I cycle to work?" . But in the process of trying to collect data it was seen that there was a lot of scientific studies associated with the topic but we struggled to find a way of meaningfully combining different life expectancy models. Most of the data we found was already clean and we did not have much processing to do[5]. We could not find out any scientific method that directly links the factors we were considering(amount of oxygen intake, pollution,BMI etc)to the increase in life expectancy[3].Combining all these factors into one prediction model became a challenge.
So ,after spending almost 4 weeks in trying to collect data we had to shift our focus to some other topic where we could get a significantly high volume of data. Now we are trying to predict the top three indicators affecting the currency of that country. After a considerable amount of research We have taken into account the following factors exchange rates, interest rates, employment, population, import, export, GDP and inflation[6]. The annual factors, like the employment, population, trade networks, GDP and inflation are taken in the time range of 1991-2016.

2 PROBLEM STATEMENT

2.1 High Frequency Exchange Rate Data

Our aim was to investigate how exchange rates influence each other. As we were able to gather nearly a 1000 samples we where able to perform a number of different machine learning techniques to perform this analysis.

2.2 Low Frequency Economic Indicators

From investigating economic research we were directed towards a number of different economic indicators which are understood to influence a currencies value. These indicators are only reported on a annual basis and due to a number of factors was only available from 1999 to 2015. This means we only had 17 samples. This limited the type of analysis we where able to perform. So we decided to find the economic which where most correlated to the performance of a specific exchange rate.

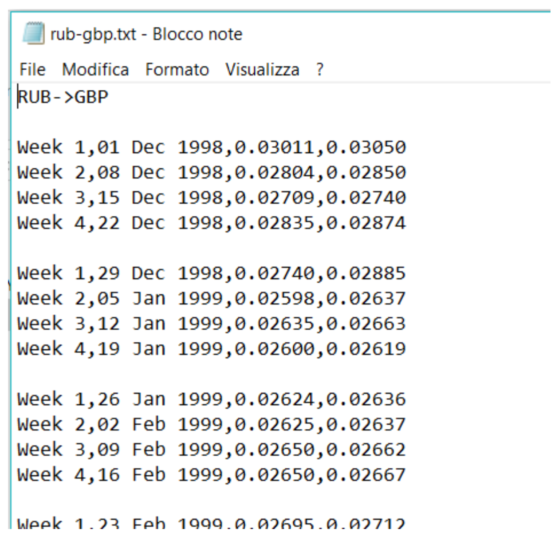
Authors' address: Philip Wilson; Nunzio Gatti; Susmita Gangopadhyay; Sumesh Ramachandra; Bang Du.

3 IMPLEMENTATION

3.1 High Frequency Exchange Rate Data

3.1.1 . Cleaning with python
The data that we have collected is from oanda.com, Bank of England and World Bank websites.

First step, Download : For the currency exchange data, it was about weekly average exchange rates from December 1998 to December 2017. We considered all currency of the G20's countries and downloaded all data considering the GBP as base currency[1]. Every file downloaded was made in this way (see Fig1):


rub-gbp.txt - Blocco note
File Modifica Formato Visualizza ?

```
RUB->GBP

Week 1,01 Dec 1998,0.03011,0.03050
Week 2,08 Dec 1998,0.02804,0.02850
Week 3,15 Dec 1998,0.02709,0.02740
Week 4,22 Dec 1998,0.02835,0.02874

Week 1,29 Dec 1998,0.02740,0.02885
Week 2,05 Jan 1999,0.02598,0.02637
Week 3,12 Jan 1999,0.02635,0.02663
Week 4,19 Jan 1999,0.02600,0.02619

Week 1,26 Jan 1999,0.02624,0.02636
Week 2,02 Feb 1999,0.02625,0.02637
Week 3,09 Feb 1999,0.02650,0.02662
Week 4,16 Feb 1999,0.02650,0.02667

Week 1,23 Feb 1999,0.02695,0.02712
```

Fig. 1. Raw currency data

For currency exchange data below technique was observed

- Size of table : 1240 x 4 columns
- Col[0]= Week(1,2,3,4)
- Col[1]= Month-Year(i.e may 2002)
- Col[2]= Bid (Bid is the price a buyer is willing to pay for a security)
- Col[3]= Ask (Ask is the price a seller is willing to accept for a security)

Second step, Aggregation : We had 16 files(16 because in the G20 group France, Italy, Germany and European Union has currency

Euro). Using python we gave in input a file formed by 1240(included blank space between rows) x 4 columns and received the output below in Fig 2.

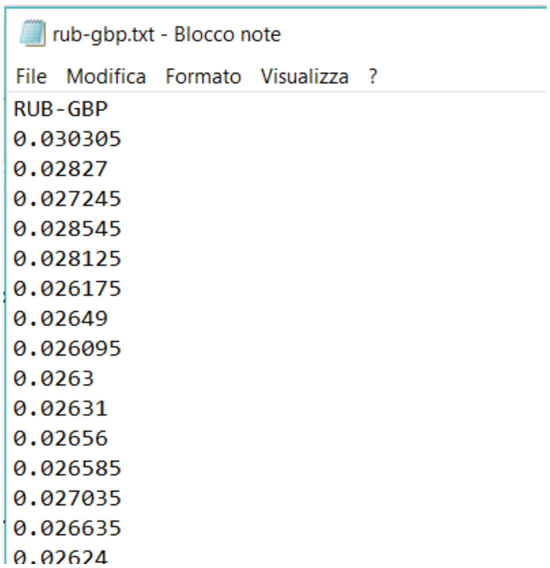


Fig. 2. Intermediate Processed Data

All spaces were removed using python, saved in one different file, and an average between bid and ask was calculated.
Third step, Creation of a complete matrix : We wrote another python script called complete Matrix.py that was used to create all possible currency pair. Remembering that the currency were 16 , we had a output matrix with 16x16 column and 994 rows (see Fig3).

3.1.2 Modeling. The target for this assignment was to find, giving in input one currency exchange, the **three** most relevant currency exchanges that most influences the input. To do that we considered the matrix containing all possible combinations of currency exchanges and we used Matlab for the analysis. We used linear regression technique to create a model to find if there is a clear relationship between one exchange rate and the other G20 exchange rates. We then look to find what are the top 3 currencies that influence a specific exchange rate.
Procedure: To find the most relevant exchange rates for the input, it was necessary to write a Matlab script using CVX tool. CVX is a modeling system for constructing and solving disciplined convex programs (DCPs)[7].

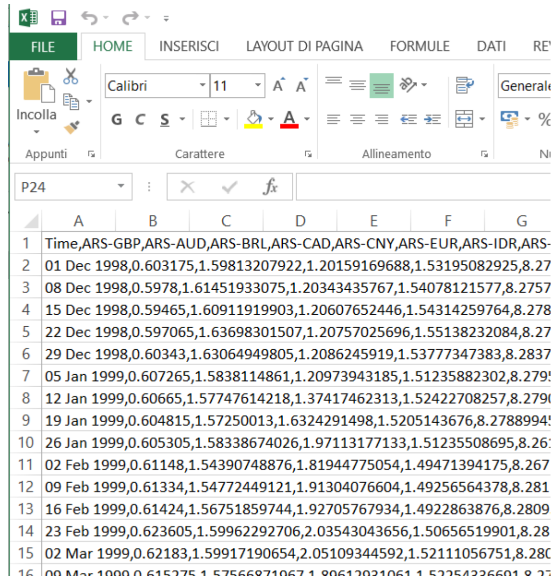


Fig. 3. Extrapolated exchange rate data

```
gamma=10;
for k=1:100
    cvx_begin quiet
        variable w2( p+1 )
        minimize( norm(Y*w2-f) + gamma*norm(w2,1) );
    cvx_end

    [iNzero] = find(abs(w2) > 1e-5);
    length(iNzero')
  if length(iNzero')<=3
      if length(iNzero')==3
         %check the accuracy of themodel calculating the mean
         %squared error on the test set
      end
      if rangeTooBig==1
         incr=incr-0.3;
         rangeTooBig=0;
      end
      gamma=gamma+incr;

  end
end
```

Explanation of the procedure :

- cvx_begin : Must be written as the first instruction of a CVX model
- cvx_begin quiet : Prevents the model from producing any screen output while it is being solved.
- cvx_end : Must be the last instruction of the CVX procedure
- variable : It is used to declare the variable, it includes the name of the variable, an optional dimension list, and one or more keywords that provide additional information about the content or structure of the variable.
- minimise : It is a command used to declare an objective function (can be also maximise) N.B. The objective function in a call to minimise must be convex; the objective function in a call to maximise must be concave. In this case Y*w2-f is convex.

Goal :To find the best weight vector that minimises the error

- +gamma(w2) : We used gamma for the L1 regularisation like ***Lasso***. This technique is normally used to solve the overfitting problem in statistical models. **In this particular case we thought that it was a good idea since we built a model using 211 different variables and so the model was complex and the risk of overfitting was high**
- find abs(w2)>(1e-5) : we found which weights are not switched off by the regulariser, that are the most relevant variables.

After obtaining the most relevant values, we split all the data in training set and test set and calculated the MSE (Means Squared Error) on the test set. After executing the script 15 times ,we had an average error of 20 on the test set[9].

Selection of gamma value:
One of the most difficult thing was about setting gamma for the regularisation. After a considerable amount of research we understood that it is almost impossible set a good gamma because it is totally dependent on both the training set and all parameters that were used[8].

- Gamma is dependent on both the training set and the other parameters you use.
- There is no âĂIJgood GammaâĂİ for any data set alone
- Mathematically you call âĂIJGammaâĂİ the âĂIJLagrangian multiplierâĂİ (complexity control).
- The higher Gamma is, the higher the regularisation. Increasing Gamma results in less overfitting but also greater bias.
- Gamma values around 20 are extremely high, and should be used only when you are using high depth or if you want to directly control the features which are dominating in the data set (i.e too strong feature engineering).

To find the most relevant values, based on the dataset, gamma was changed automatically in every iteration, and so it was increased until exactly three variables were not switched off.

Create a model:
The second analysis that we did using the high frequency data values, was about creating a model that was able to derive the currency exchange trend establishing the top three influential currency exchange. To do that we still used the CVX tool, regularising the function but in this case gamma was fixed. We made researches and we saw many examples, moreover we run practical testing with different values of gamma and at the end we decided to set gamma = 8.0.

The dataset was divided in training and test using the proportion 75% and 25% respectively. Then the script was run using CVX tool.

Justification of the model:
In the beginning there were some doubts about using regression technique or an ANN(Artificial Neural network)[7]. We considered both merits and de-merits but in the end we chose to use linear regression principally for two reasons :

- The main reason was that the ANN is a black box method and it is very difficult to find any relationship between variables, on the contrary these relationships can easily be shown by regression models.

```
 %model
  ii=randperm(N);
  %splitting in training and test set
  training=Y(ii(1:(N*0.75)),:);
  test=Y(ii((N*0.75)+1:N) ,:);
  fTraining=f(ii(1:(N*0.75)),:);
  fTest=f(ii((N*0.75)+1:N) ,:);


gamma=8.0;

cvx_begin quiet
    variable w2( p+1 )
    minimize( norm(training*w2-fTraining) + gamma*norm(w2,1) );
cvx_end

%storing all errors of all currencies in a vector to check
%the highest and the lowest
e(z)=sum((((test*w2)-fTest).^2)/(N/2))*100;
```

Fig. 4. Error Calculation

- The method of least squared regression converge much faster than a neural network, and this means a saving of resources and time

3.2 Low Frequency data

The low frequency data included the data of G-20 population, employment, trade network, GDP, inflation and interest rate. Every files downloaded was made in this way (see Fig 5):

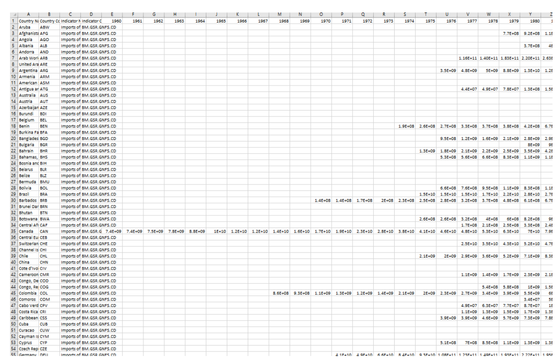


Fig. 5. Raw economic indicators data

The size of each factor of G-20 dataset table is 265 rows time 61 columns and each row is a country's factor data from 1960 to 2016. Firstly, we used python code to select the data of G-20 and used append function to put the G-20 data into a new array. Then, in order to clean out the data suitable for using and analysis, we used the zip function to transpose the array. Next, Due to the low frequency data were analysed from 1991 to 2016, we filtered the data and put it into the new csv file. Furthermore, we used replace function to remove the space, which may cause the Index error. Using this python script all low frequency data were filtered and we put the all cleaned data into one file (see Fig 6.

Country Code	ARG	AUS	BRA	CAN	CHN	DEU	EUU	FRA	GBR	IDN	IND	ITA
1991	11566000000	52679885811	28251000000	1.52E+11	54297000000	4.52E+11	1.98E+12	2.83E+11	2.51E+11	31398000000	27031902767	2.13E
1992	19319500000	55168883181	27984000000	1.57E+11	73819000000	4.85E+11	2.11E+12	2.98E+11	2.67E+11	34874000000	29665639166	2.32E
1993	21975400000	56603069409	34856000000	1.68E+11	98349000000	4.20E+11	1.86E+12	2.61E+11	2.55E+11	38222000000	30604948802	1.85E
1994	27273600000	66750323102	43495000000	1.83E+11	1.12E+11	4.64E+11	2.06E+12	2.81E+11	2.84E+11	43738000000	37872390417	2.05E
1995	26035117000	75423012738	63293000000	1.99E+11	1.35E+11	5.59E+11	2.51E+12	3.34E+11	3.27E+11	54461000000	48225107457	2.45E
1996	30212204066	80376140345	66018000000	2.09E+11	1.54E+11	5.53E+11	2.59E+12	3.34E+11	3.55E+11	59379000000	54959987568	2.50E
1997	37511586709	82747225907	75139000000	2.37E+11	97709000000	5.37E+11	2.59E+12	3.23E+11	3.80E+11	62830000000	58172791088	2.54E
1998	38797517537	79601560689	74415000000	2.41E+11	97527000000	5.65E+11	2.74E+12	3.46E+11	3.96E+11	44030357373	59367859754	2.66E
1999	32903076781	84902583708	62807000000	2.59E+11	1.19E+11	5.79E+11	2.84E+12	3.65E+11	4.19E+11	42974513498	62827495216	2.69E
2000	33068803787	88268064343	71576500702	2.87E+11	1.61E+11	5.95E+11	2.98E+12	3.77E+11	4.40E+11	56002463130	73075192253	2.84E
2001	27594942433	79477681918	71626048627	2.68E+11	1.80E+11	5.87E+11	2.97E+12	3.74E+11	4.39E+11	50548622609	71311160936	2.85E
2002	13337220000	89216900918	60778949499	2.71E+11	2.10E+11	5.89E+11	3.11E+12	3.91E+11	4.71E+11	52696753633	75741490128	3.02E
2003	18724510000	1.09E+11	62707203451	2.95E+11	4.10E+11	7.26E+11	3.72E+12	4.61E+11	5.29E+11	56946585710	92959121914	3.61E
2004	27823640000	1.34E+11	78995242018	3.37E+11	5.54E+11	8.58E+11	4.50E+12	5.42E+11	6.25E+11	67472159281	1.31E+11	4.23E
2005	34796990000	1.52E+11	96610925191	3.85E+11	6.49E+11	9.34E+11	4.95E+12	5.93E+11	6.86E+11	86268317494	1.82E+11	4.60E
2006	41261872520	1.70E+11	1.19E+11	4.30E+11	7.83E+11	1.08E+12	5.66E+12	6.58E+11	7.84E+11	87614055150	2.25E+11	5.28E
2007	53551718520	2.06E+11	1.58E+11	4.71E+11	9.49E+11	1.25E+12	6.64E+12	7.60E+11	8.41E+11	1.01E+11	2.79E+11	6.14E
2008	68242684098	2.46E+11	2.20E+11	5.08E+11	1.15E+12	1.41E+12	7.45E+12	8.69E+11	8.67E+11	1.36E+11	3.79E+11	6.69E

Fig. 6. Cleaned Economic data

3.3 Web Application Design

3.3.1 . Framework We created a simple web application using the python flask framework.

3.3.2 . Displaying Result The final analysis and displaying of results was implemented in JavaScript and D3.

The when viewing the low frequency economic indicator data we also load the high frequency (weekly samples) exchange rate data, this is then aggregated down to yearly samples for analysis with the economic indicators.

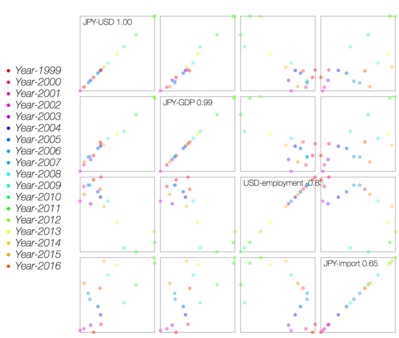


Fig. 7. Covariance Scatter plot of Economic Indicators with respect to a specific exchange rate

The web app has the following two functions:
1) Finding Strongest Correlation :When this button is pressed the web app iterates through each exchange rate and calculates the correlation coefficient between it and each related economic indicator. This sorted in order of strength (closest to either 1.0 or -1.0) and then displayed. This allows us to find what are the strongest and weakest correlations. The app then also aggregates the absolute value of each correlation coefficient (to prevent positive and negative correlations cancelling each other out and allowing us to find the average strength of the correlation) with respect to the type of economic indicator. This then allows us to find the type of economic indicator which is most correlated with exchange rates.
2) Go!:This button takes a selected exchange rate and then generates a scatter plot for each of the economic indicators that are most correlated with that exchange rate. These are present in a 4 x 4 grid (as seen in Fig 7, the selected exchange rate is in the top right corner, and each economic indicator is plotted on the diagonal, each labelled with their respective correlation coefficient related to the exchange rate. The user is then able to see a covariance scatter plot for each variable by tracing the horizontal and vertical intersection.

4 RESULTS

4.1 Observation

As the data we have used is not very huge, we chose to find the strongest correlation between the exchange rates and the indicators which led to this correlation. We have used Pearson's correlation coefficient as the mathematical approach since the data was not suitable for any mathematical regression or any other kind of predictive analysis. The result that we have found can be summarised below showing the ranking of the indicators in terms of those that most strongly correlate to exchange rate data averaged out over all exchange rates.

- Population = 0.62
- GDP = 0.60
- Imports = 0.59
- Exports = 0.58
- Foreign Trade = 0.50
- Inflation GDP Deflator = 0.40
- Employment = 0.40
- Interest Rate = 0.40
- Inflation Consumer Prices = 0.38

However in some particular cases we have seen a very strong correlation which are worth mentioning

Table 1. Strongest correlation and related indicator

Currency Code	Indicator	Value
JPY-USD	JPY GDP	0.99
JPY-SAR	JPY GDP	0.99
SAR-JPY	JPY GDP	-0.99
MXN-CNY	MXN Population	-0.98
AUD-TRY	TRY Population	0.98
CAD-TRY	TRY Population	0.98
SAR-CNY	SAR imports	-0.98
AUD-IDR	IDR Population	0.98
AUD-TRY	AUD Population	0.98
MXN-CNY	CNY Population	-0.98
RUB-CAD	CAD Population	-0.98
CAD-TRY	CAD Population	0.98
CNY-SAR	SAR Imports	0.97
TRY-MXN	MXN Inflation	0.97
RUB-AUD	AUD Population	-0.97
AUD-IDR	AUD Population	0.97
AUD-ZAR	AUD Population	0.97
AUD-ZAR	ZAR Population	0.97

From the table above, a few things can be concluded

- GDP, Population, Imports and Inflation have been the strongest indicators in the past[4]. These factors have played the most important role in determining the exchange rates for the two currencies
- A high correlation value indicates that the corresponding indicator is the strongest factor in determining the exchange rates between those two countries[2]. For example, We find Japanese GDB has a very strong correlation value (0.99) between Japanese Yen and US Dollar and also between Japanese Yen and South African Rand.

4.2 Analysis

4.2.1 . Currency Exchange Trends If we look at the currency exchange rates between Argentina (ARS) and USA (USD) we observe that the until 2001 the currency exchange value is 1 and then we see a sudden dip in the value. What could be the reason for this dip? Well, that was because each peso was index-linked to USD at 1ARS= 1USD. However, after the financial crisis of 2001, the fixed exchange rate system was abandoned. Since 2002, the exchange rate started to fluctuate, keeping the exchange rate at between 2.90 and 3.10 pesos per US dollar at that time. This is the same case in terms of Saudi Riyal, where even today it is index linked to the USD @ 1 USD = 3.75 SAR

4.2.2 . Correlation Coefficient If we look at the currency exchange rates JPY-USD, we observe that it has a .99 correlation value against the GDP of Japan. Yes, without any arguments we can agree that the GDP of a country has a direct impact on its currency value. Where we see that the overall influence (correlation coefficient value) of GDP to its countries currency is calculated to be .6. But in this case, we see a value of .99.
If we look at the history for the currency of Japan (YEN) we see that. Following World War II the Yen lost much of its value. To stabilize the Japanese economy the exchange rate of the yen was fixed at Âě360 per 1USD as part of the Bretton Woods system. When that system was abandoned in 1971, the Yen became undervalued and was allowed to float. The Yen had appreciated to a peak of Âě271 per 1 in 1973, then underwent periods of depreciation and appreciation due to the 1973 oil crisis, arriving at a value of Âě227 per 1USD by 1980. Since 1973, the Japanese government has maintained a policy of currency intervention, and the yen is therefore under a "dirty float" regime. This intervention continues until today and that is the reason we see such a tight correlation between the currency exchange of JPY-USD against the GDP of Japan.
But the question is does it stand good for all cases? Well, yes it does? For instance, if we look at the currency of China(CNY) and USA(USD) we see that both these two countries have a relatively huge GDP values, which does have an impact on their respective currencies. But when we look at the exchange rate which is a copula of both these countries currency. There is a possibility that the influence of the GDP values on the exchange rate tend to be slightly lower (Correlation coefficient of ChinaâĂŹs GDP and US GDP against the exchange rate of CNY-USD is .92 and .87 respectively) but still have a significant impact the exchange rates.

5 CONCLUSION

6 LIMITATIONS AND FUTURE WORK

- Our application works on static data .If we want to analyse new data, we need to upload new file into the application.
- Most of the currencies were index linked to USD prior to 1999, for which we had a limited availability of currency data.
-

REFERENCES

[1] Angela Abbate and Massimiliano Marcellino. 2018. Point, interval and density forecasts of exchange rates with time varying parameter models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181, 1 (2018), 155–179.

[2] Ariel Burstein, Martin Eichenbaum, and Sergio Rebelo. 2005. Large devaluations and the real exchange rate. *Journal of political Economy* 113, 4 (2005), 742–784.

[3] Ryan D Edwards and Carl N Mason. 2014. Spinning the wheels and rolling the dice: Life-cycle risks and benefits of bicycle commuting in the US. *Preventive medicine* 64 (2014), 8–13.

[4] Sebastian Edwards. 2006. *The relationship between exchange rates and inflation targeting revisited.* Technical Report. National Bureau of Economic Research.

[5] Michael Iacono, Kevin Krizek, and Ahmed M El-Geneidy. 2008. Access to destinations: how close is close enough? Estimating accurate distance decay functions for multiple modes and different purposes. (2008).

[6] Nuwan Kuruwitaarachchi, MKM Peiris, CN Madawala, KMAR Perera, and VUN Perera. 2018. Design and Development of an Algorithm to Predict Fluctuations of Currency Rates. *World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering* 5, 2 (2018).

[7] Mark T Leung, An-Sing Chen, and Hazem Daouk. 2000. Forecasting exchange rates using general regression neural networks. *Computers & Operations Research* 27, 11 (2000), 1093–1110.

[8] Adewole Adetunji Philip, Akinwale Adio Taofiki, and Akintomide Ayo Bidemi. 2011. Artificial neural network model for forecasting foreign exchange rate. *World of Computer Science and Information Technology Journal* 1, 3 (2011), 110–118.

[9] Barbara Rossi. 2013. Exchange rate predictability. *Journal of economic literature* 51, 4 (2013), 1063–1119.