A machine learning-based framework for exchange rate analysis and prediction

Chien-Yi Huang [a], Marvin Ruano [b,*] , Marco Tulio Espinosa Herrera [a] and Ricardo Neftali Pontaza Rodas [c]

* Correspondence: marvin.ruano100@gmail.com
[a] Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei, Taiwan
[b] Department of Management, National Taipei University of Technology, Taipei, Taiwan
[c] Department of Computer Science Engineering, National Taiwan University, Taipei, Taiwan

Abstract

Interests towards exchange rate (ER) analysis and prediction have increased in past decades. Theoretical and empirical literature support the relationship between ER and trading goods activity (imports and exports). However, there has not been any approach identifying the most important trading goods in order to forecast the ER based on these goods. Through a case study using the Guatemalan quetzal against the US dollar, this paper aims to predict the ER based on relevant trading goods and provide an organised database of the desired results. Using principal component analysis (PCA), we predict the ER with a low approximation error and determine which production sector should be the focal point for the country. We also provide theoretical foundations for a proper ER forecasting implementation using genetic algorithm (GA), lower–upper decomposition, and Lagrange and spline interpolation. The empirical results show manufacturing industries and exports to El Salvador and Honduras as the most relevant trading goods. This work could be beneficial and useful to national institutions, policy makers, and other academics or analysts who may utilise the data for further analysis.

Published by IJRP.ORG. Selection and/or peer-review under responsibility of International Journal of Research Publications (IJRP.ORG)

Keywords: exchange rate; forecasting; international trade; prediction; principal component analysis

1. Introduction

Worldwide interests towards exchange rate (ER) analysis and prediction have increased rapidly over the past decades, especially in areas of economics and finance (Fang and Kwong, 1991). ER has become a representation of a country's economic stability and a trigger factor for many financial and commercial decisions. However, the lack of public information and proper structure of the data under consideration can greatly affect investors' decision-making. Thus, accurate ER calculations and predictions are necessary for investors to ensure accountability of outcomes.

In countries with a free economy, ER is influenced by a series of factors (Tsen Wong, 2013). The trading sector is considered to have the biggest impact on the ER value, which is directly affected by the decrease in trading goods activity (imports and exports) (Tarasova and Coupe, 2009). Past studies use cointegration tests to prove the relationship between ER and trading goods activity. Hussain, Ahmad, and

Masud (2017), for example, find a significant long-term relationship between ER and international trade in Pakistan, through cointegration tests, concluding that a higher rate in the Pakistani rupee against the US dollar directly results in a higher level of exports. Alam (2010) also uses cointegration tests to explain the relationship between changes in the ER and revenues in Bangladesh. Chebbi and Olarreage (2019) use cointegration techniques to determine the long and short-term impact of changes in the ER on the net agricultural trade balance in Tunisia.

To predict the ER, we must first identify and understand the goods which influence the ER value. By identifying these relevant goods, we automatically remove the insignificant ones, allowing more focus on the most important ones. However, there has not yet been any existing approach which identifies the most relevant trading goods and forecasts the ER based on these goods. Current approaches are either too complicated to follow or not being used properly. Also, the majority of these approaches merely focus on finding a relationship between ER and international trade.

ER is a continuous variable (or nonlinear function) and a very complex mechanism (Federici and Gandolfo, 2002). There are thousands of transactions and hundreds of goods to take into account when the ER value is influenced by trading goods activity, producing a huge and complicated database. Thus, identifying the relationship between each trading good and the ER is very time consuming. Moreover, the complexity of the database cannot be analysed through simple mathematical approaches or solved by simply reducing the amount of data, because this would lead to inaccurate results. In addition, the presence of 'unwanted' values further increases the overall complexity of the database.

This paper applies principal component analysis (PCA) as a strong analytical tool for big data analysis to: (1) reduce the dataset dimension, (2) determine the trading goods affecting the ER value by degree of importance, (3) identify the most relevant trading goods, (4) predict the ER based on those relevant trading goods, and (5) analyse the obtained results. The main objectives are to predict the ER using all the available trading goods activity data for the country (Guatemala in Central America) and to provide an organised, filtered, and complete database of the desired trading goods, industries, countries, and regions, in relation to the ER value.

In recent years, researchers have been relying on artificial neural networks (ANNs) as a major analytical tool for prediction (Bal and Demir, 2017). Galeshchuk (2016) provides different ways of predicting the ER using ANNs, through a study measuring the performance of ANNs based on daily, monthly, and quarterly predictions. A comparison with different currencies is made, demonstrating how this particular approach can be applied in different countries. Galeshchuk (2016) also highlights the importance of analysing ER-ANNs in specific economic sectors, such as agriculture.

Although there are many other techniques for ER prediction, most studies use backpropagation algorithm (BP), which works with a wide range of algorithms, including: Levenberg–Marquardt (LM), Broyden–Fletcher–Goldfarb–Shanno (BFGS), quasi-Newton, resilient backpropagation (Rprop), and scaled conjugate gradient (SCG) (Werbos, 1974). Other approaches include feedforward neural networks (FFNNs) and the random walk model. FFNNs identify the previous sequence to provide a future value forecast, whereas the random walk model uses its 'step-ahead' data prediction ability to provide a more stable daily value forecast (Tyree and Long, 1995). Bal and Demir (2017) use FFNNs to perform a study on the ER in Turkey consisting of 331 observations (TRY/USD[1]), where 281 are taken as training data and 50 are used for testing. The results show the feasibility of a comparison between predicted and real values. Fang and Kwong (1991) use FFNNs and the Box–Jenkins model in a different study on the ER in the UK, with 60 observations

[1] Turkish lira against the US dollar

(US/GBP[2]). After comparing FFNNs, the Box–Jenkins model, and other econometric models, they concluded that the Box–Jenkins model provides the best forecasting results, based on mean absolute error, mean squared error, and mean absolute percentage error. However, although the aforementioned techniques are all powerful and widely used, they do not fulfil the requirements for the present paper, because BP once deployed has a 'forgetting' problem, so there is still a need for a genetic algorithm (GA). Moreover, the Box–Jenkins model is only reliable for short-term prediction, which is unsuitable for the proposed model. Finally, although most of these studies have shown positive results, the majority focus on pervious ER observations only, without considering other factors like trading, inflation, or production levels.

Previous research has proven that PCA is a great analytical tool for big data analysis and prediction. Skittides and Fruh (2014), for example, use PCA to conduct a wind-forecasting study, predicting the wind speed using past data with an ensemble of dynamically similar past situations. This shows the wide range of PCA application for prediction purposes. However, although PCA may be a great analytical and prediction tool, it lacks the ability to develop neural networks (NNs). NNs allow the development of machine learning capable of predicting future trends and values and provide more proper and accurate analysis (Philip, Taofiki, and Bidemi 2011). Previous approaches combine PCA with support-vector machines (SVMs) to train NNs for data identification. SVMs are based on statistical learning theory and are considered one of the best classification techniques for their mathematical background (Arsalane et al., 2018). Hu and Cui (2019) state that PCA plays a crucial role in image processing for digital recognition, as it uses a low dimensional space description of the input image, without losing important information for recognition precision. In addition, PCA reduces the input variables needed for pattern identification, providing a boost for image recognition accuracy. Hu and Cui (2019) also explain how SVMs are similarly important for image recognition in terms of system improvement. Jing and Hou (2015) apply PCA and SVMs as the principal analytical tool in the industrial processes sector, where the combination performs well and delivers trustworthy results. Another relevant application of PCA-SVMs is in the financial market, where Yu et al. (2014) propose a PCA-SVM stock selection model. Finally, Zhi and Liu (2019) combine PCA-SVMs with GA to develop a face recognition model. The results are up to standard, showing that the GA optimises the process of feature search, PCA reduces the dimension of features, and SVMs act as a classifier. However, while PCA-SVM systems may have a high accuracy rate, they merely fall into the classification segment of machine learning. ER prediction and relevant trading goods identification require more than just classification, as the proposed framework looks for a specific value. Thus, SVMs are unsuitable for the present study, instead we apply lower–upper (LU) decomposition (factorisation) as a better fit for large amounts of data demanding high computational power. LU decomposition is known for its versatility and performance in solving equations. It breaks down a matrix into two, reducing the memory needed for calculation and making it easier to solve (Ford, 2014). The remainder of the present paper is segmented as follows: Section 2 details the preliminary analyses, Section 3 focuses on the methodology, Section 4 presents the application and findings, and Section 5 details the discussion and conclusion.

2. Preliminary Analysis

This paper is part of preliminary research work now being applied to the proposed framework. The input data for the GA comes from two sources: one is obtained using the LU decomposition and the other comes from the PCA. Figure 1 shows the steps performed in the preliminary case study. The organised data

[2] US dollar against the British pound sterling

per country is divided into imports and exports. Matrix MAV contains the values obtained from the National Bank of Guatemala (Banguat) and the National Stock Exchange of Guatemala (BVNSA).

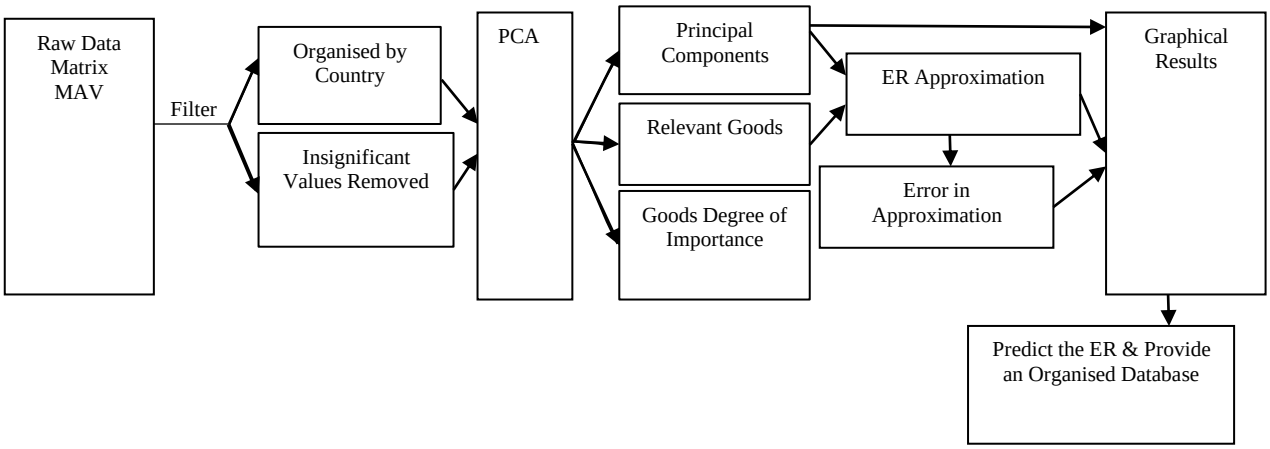


Fig. 1. Previous work – PCA steps

2.1. Preliminary analyses empirical example

The case study includes 16 countries, and the total amount of original data is 460,800. Table 1 shows the dataset after removing the inconsistent data, duplicated goods, and insignificant zero values. However, the total amount of filtered data is still sufficient. The values obtained in Table 1 become the new input matrix for the PCA. Six cases are taken into consideration, as part of the preliminary work, which provide a positive result with a low error coefficient. One of those six cases is provided as an example in the following subsection.

Table 1. Preliminary work example – filtered data values

Country	Goods (Imports and Exports)	Total Data (G*M*Y)[3]
United States	110	21,120
Netherlands	11	2,112
Canada	13	2,496
China	21	4,032
Costa Rica	48	9,216
El Salvador	86	16,512
Germany	19	3,648
Honduras	53	10,176
Italy	10	1,920
Japan	12	2,304

[3] G = Goods, M = Months, Y = Years

Mexico	65	12,480
Nicaragua	36	6,912
Panama	31	5,952
South Korea	18	3,456
Taiwan	11	2,112
United Kingdom	7	1,344
	Total	105,792

2.2. Preliminary analyses case study

In this model, the input data includes imports and exports from El Salvador, Honduras, Nicaragua, Costa Rica, and Panama, calculated against the ER in Guatemala. The region consists of countries with long-lasting trade relations with Guatemala and those nearest to Guatemala. Thus, the ER error was estimated to be low. The results obtained after the PCA are shown in Table 2, section 1 (Preliminary Work). After comparing the monetary values of the 254 trading goods with the ER in Guatemala (GTQ/USD[4]), the amount of relevant trading goods is 64, which accounts for 25.2% of the total number of trading goods. The monthly ER data is secondary time series data, obtained through daily observations, for a period of 16 years (January 1st, 2002 to December 31st, 2017). Moreover, Table 3 shows the degree of importance for 12 of the relevant trading goods with values reaching a 0.50 significance level (for example purposes only).

Table 2. Case study – Central America

	Description	Section 1: Preliminary Work	Section 2: Improved with GA
Criteria	Region	Central America	Central America
	Data Mapping	97.5%	97.5%
	Significance	0.80	0.80
	Technique	PCA	PCA-GA
Inputs	Number of Variables	254	254
	Total Trading Goods Data	48,768	48,768
	Exchange Rate Data	192	192
Outputs	Relevant Trading Goods	64 out of 254	64 out of 254
	PC 97.5% Mapping	16	14
	Value of First PC	80.886	80.886
	Exchange Rate Approximation Error	0.01417	0.01622

Table 3. Relevant trading goods for case study – Central America

Goods[5]	Description	Degree of Importance	Accumulative Value
MI (SAL)	Exports	0.07674	0.076738
MI (HND)	Exports	0.07404	0.15078

[4] Guatemalan quetzal against the US dollar
[5] SAL = El Salvador, HND = Honduras, NIC = Nicaragua, CR = Costa Rica, PTY = Panama, EI = Extractive Industries, MI = Manufacturing Industries, PP = Pharmaceutical Products

MI (SAL)	Imports	0.07038	0.22116
MI (NIC)	Exports	0.05736	0.27851
MI (CR)	Imports	0.04107	0.31959
MI (CR)	Exports	0.03728	0.35686
MI (PTY)	Imports	0.03137	0.38823
EI (SAL)	Exports	0.03066	0.41889
MI (PTY)	Exports	0.02618	0.44507
MI (HND)	Imports	0.02321	0.46828
PP (PTY)	Imports	0.02311	0.49138
EI (HND)	Exports	0.01911	0.51049

Figure 2 shows the ER approximation to a 97.5% data mapping, with the value in the graph being accumulative of each variable. Agricultural products are not included in the most important trading goods, because the Central American region has very similar agricultural sectors.

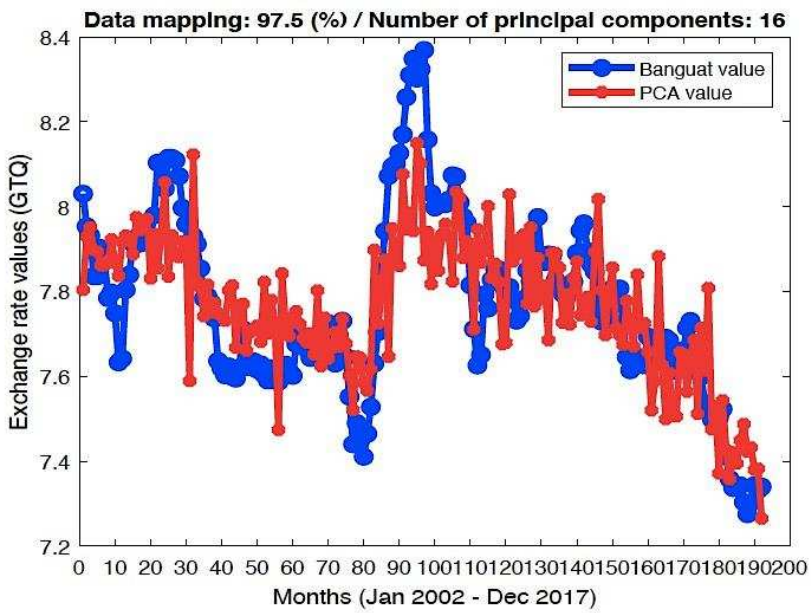
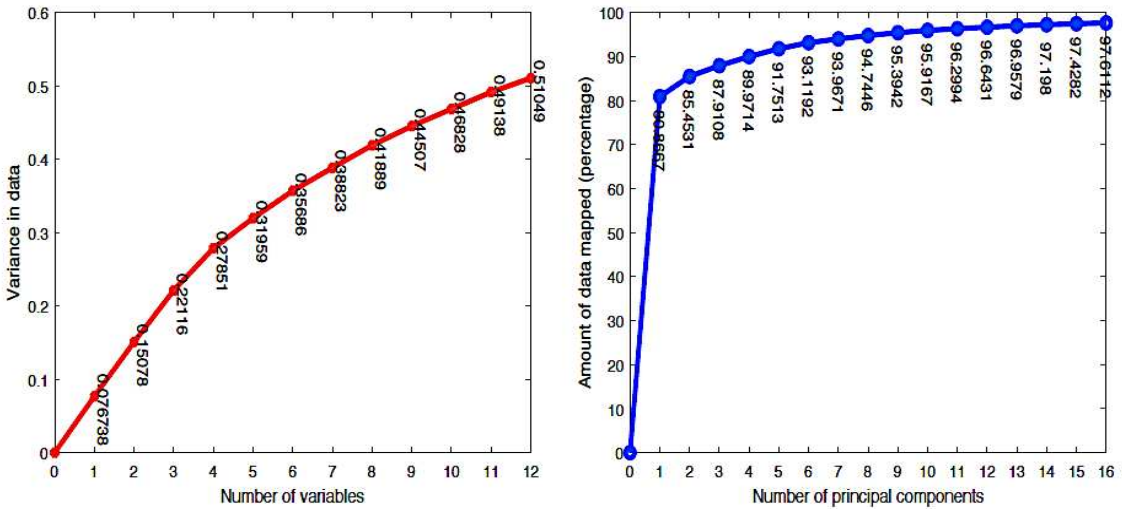


Figure 2. Obtained vs. original values

Figure 3 illustrates the PC required to map the data up to 97.5%, where Figure 3(a) shows the variance in data to a 0.50 significance level (all 64 values cannot be displayed appropriately) and Figure 3(b) shows the values of each PC needed to obtain the 97.5% forecast.



(**a**) (**b**)

Figure 3. Case study results for Central America: (a) relevant trading goods and (b) PC values

3. Materials and Methods

3.1. Framework structure and empirical and computational procedures

The proposed framework works under LU decomposition-GA, PCA-GA, and Lagrange and spline interpolation to determine the importance of each influential factor under consideration in relation to Guatemala's international trade. The original matrix, known as matrix A, goes through a series of steps shown in Figure 4 to obtain the desired result (an ER forecasting with a low error coefficient). Matrix A includes the raw data, which can be analysed in four different ways. (1) PCA – data goes through the process shown in Figure 5 and provides the outputs which are later combined with the GA. (2) LU decomposition – data is decomposed into two matrices, matrix L and U. (3) GA – outputs from the previous steps become the input for the GA and data is analysed using random variation and propagation to grandchildren variation, with the decision based on user discretion. (4) Lagrange and spline interpolation – after the data is analysed, the output is a unique function capable of interpolating the obtained function at any given point in time. Finally, with the obtained values from the GA, it is possible to construct NNs capable of analysing the data mathematically and graphically. This paper provides visual representations of the NNs, errors, interpolations, and real time NN updates, for a better understanding and usefulness of the proposed framework, which becomes ready for further implementations.

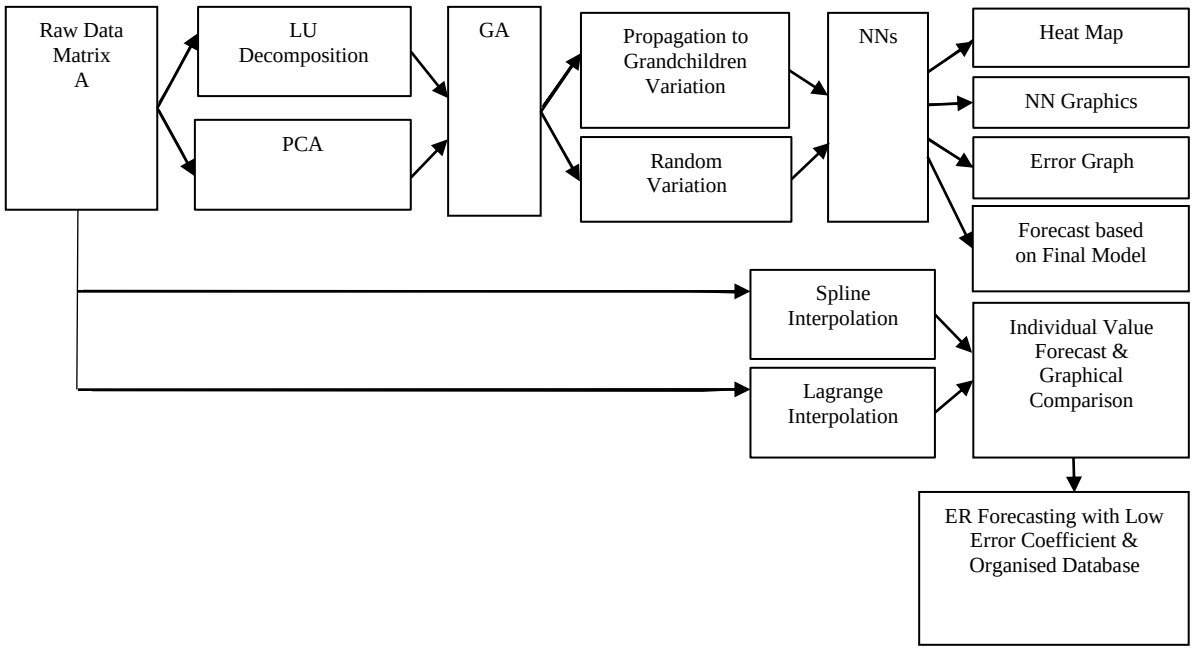


Figure 4. Framework summary

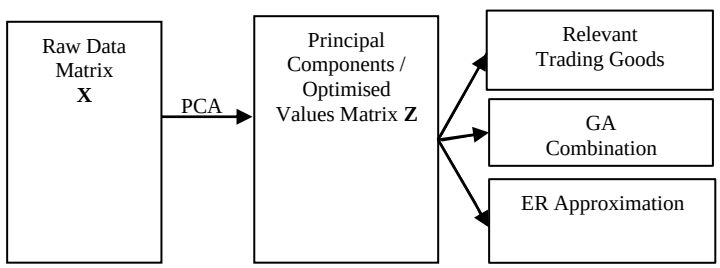


Figure 5. Summary of PCA

MATLAB is used for the analysis, organisation, and graphical plotting of the data. Difficulties faced in the development of the proposed framework include unorganised raw data, inconsistent ER values with some days missing, and useless data (zero or negative values), which could affect the analysis in the long run. However, the framework still provides the desired outputs, by organising the data and deploying the requested analysis.

3.2. Principal Component Analysis (PCA)

PCA is applied in the present study for its main characteristics of dimension reduction and conservation of data variability. The general purpose of PCA is to move from matrix X, a large database containing the original values or raw data, to matrix Z, a summarised dataset with the new or optimised variables known as the principal components (PC), while still retaining up to 90% of the original or important information (Vidal et al., 2002). This consistency of data achieved through the PCA gives stability to the proposed framework.

Moreover, the ER prediction is based on the most relevant trading goods obtained from the PCA values, which are the PC. PCA optimises the data to show the relevant components only, then it defines the predicted values through the NNs in combination with the GA, which improves the overall accuracy of the prediction and prepares the data for further analysis.

3.3. Neural Networks (NNs)

NNs are implemented in the present study for their pattern recognition and adaptive ability, which makes them useful for ER prediction. This study works with a large database, which is what NNs are typically used for (Refense et al., 1993). NNs break down the database into stages, enhancing the forecasting approach and predicting the ER with a low error coefficient. The general NN structure consists of three layers: input, hidden, and output layer (Philip et al., 2011). The number of hidden layers depends on the complexity of the system. In general, each layer can apply any given function obtained from the previous layer. The hidden layers are there to transform the values obtained into something that the output layer can understand and to provide the best possible outcome. The PCA-GA model goes as follows: the input layer includes the ER value and the relevant trading goods obtained from the PCA, the hidden layer (the filter) includes the results obtained from the previous layer, and the output layer provides the final result from the NNs (the predicted ER value based on the relevant trading goods).

3.4. Empirical study model approach

Let X_n be the n – th measurement and let F be stated as $F(X_{n-1}) = X_n$X

Theorem 1. Let F be as stated in the first model. If there are two measurements X_n, X_m such that $n \neq m$ and $X_n = X_m$, then either F does not exist or $(X_n)_{n\in\mathbb{N}}$ has a period.

Proof of Theorem 1. Let F be as stated in the first model and X_n, X_m be such that $X_n = X_m$ and $m \neq n$. Assume, without loss of generality, that m > n. Then:

$$\left.\begin{array}{l} X_{m+1} = F(X_m) \\ X_m = F(X_{m-1}) \\ X_{m-1} = F(X_{m-2}) \\ \vdots \\ X_{n+1} = F(X_n) \\ X_n = F(X_{n-1}) \end{array}\right\} \quad (1)$$

Because $X_n = X_m$, then $X_{n+1} = F(X_n) = F(X_m) = X_{m+1}$ and in general $X_{n+i} = F(X_{n+i-1}) = F(X_{m+i-1}) = X_{m+i}$, hence $X_{n+i} = X_{m+i}$ for some $i \in \mathbb{N}$, hence F has a period i; thus, $(X_n)_{n\in\mathbb{N}}$ has a cycle. Finally, if the measurements do not present a cycle, then F cannot exist.

4. Application and Results

The ER prediction results obtained from the PCA are all-around positive. They show the trading goods and sectors, with a closer relation to the ER value, and the strongest region. Each model is performed to provide the graphical and numerical results obtained from the proposed framework, with the criteria for each model based on user discretion.

4.1. Framework model 1

This model is performed using the LU-GA technique, which allows the system to obtain the desired values through a series of 'generations' in order to reach the lowest error in final generation. Using random variation (RV), set by a value 'epsilon' (in this case 0.1), this method starts by calculating each node according to the previous node and so on, until it reaches the final generation. Table 4, section 1 (Framework Model 1) provides a summary based on the criteria, inputs, and outputs for this specific model, using the proposed framework. The error values get closer to zero after each generation, which means that the system is approaching the expected results as the values learn with each generation.

Table 4. Framework models 1 and 2

	Description	Framework Model 1	Framework Model 2
Criteria	Epsilon Technique	0.1 LU-GA, RV	0.1 LU-GA, PGC
Inputs	Exchange Rate Data	204	204
Outputs	Number of Generations Exchange Rate Min Error Exchange Rate Max Error Exchange Rate Avg Error	203 0 $3.908 \times \sqrt{10^{-14}}$ $6.209 \times \sqrt{10^{-15}}$	203 0 $8.88 \times \sqrt{10^{-15}}$ $1.597 \times \sqrt{10^{-15}}$

Each model has a 'heat map' to illustrate how far behind the values are from the desired results. Appendix 1(a) shows different error values for each generation, for example, generation 71 has a value of 8.882 × 10 ^-16, while the next generation has a value of 7.63. This is due to the RV behaviour which calculates generation per generation. Whereas, Appendix 1(b) reaches the desired level, showing values closer to zero, with a max error value of 8.53 × 10 ^-14 and a minimum of 0. These results are obtained through the NNs generated with the LU-GA technique and RV.

4.2. Framework model 2

This model is also performed using the LU-GA technique, along with propagation to grandchildren variation (PGC), which is also set by a value epsilon (in this case 0.1). PGC decreases the number of nodes displayed. It starts by calculating all the values at once, which may not be accurate at the beginning, but then corrects itself with each generation. The error term gets to a maximum point and then starts decreasing (getting closer to zero), meaning that the system is approaching the expected results. Table 4, section 2 (Framework Model 2) provides a summary based on the criteria, inputs, and outputs for this specific model using the proposed framework. Appendix 2(a) shows how PGC readjusts itself with each generation, for example, generation 71 has a value of 0 and the next generation has a value of -3.254. Whereas, Appendix 2(b) reaches the desired level, where the values are closer to zero, with a max error value of 8.53 × 10^-14 and a minimum of 0. These results are obtained through the NNs generated with LU-GA and PGC.

4.3. Interpolation: framework model 3

This model is performed using Lagrange interpolation, which allows the system to obtain the desired values at any given point in time. This approach increases the versatility of the framework, as it provides faster and more accurate results. Lagrange shows an elapsed time of 0.001696 seconds for 16 models, with 1,536 bytes and ER input data of 192 (for the year 2015). Figure 6 shows the exact values of the Lagrange interpolation passing through the real or original values.

4.4. Interpolation: framework model 4

This model is performed using spline interpolation, which also allows the system to obtain the desired values following an interpolation. However, spline differs from Lagrange in its adaptability to real world problems, as it is 'smoother' and does not fluctuate as much. Moreover, spline takes longer, with an elapsed time of 2.4884 seconds for one model (with ER input data of 192 for the year 2015). Finally, spline is 1.25 times 'heavier' than Lagrange, with 1,920 bytes. Figure 6 also shows the exact values of the spline interpolation passing through the original values.

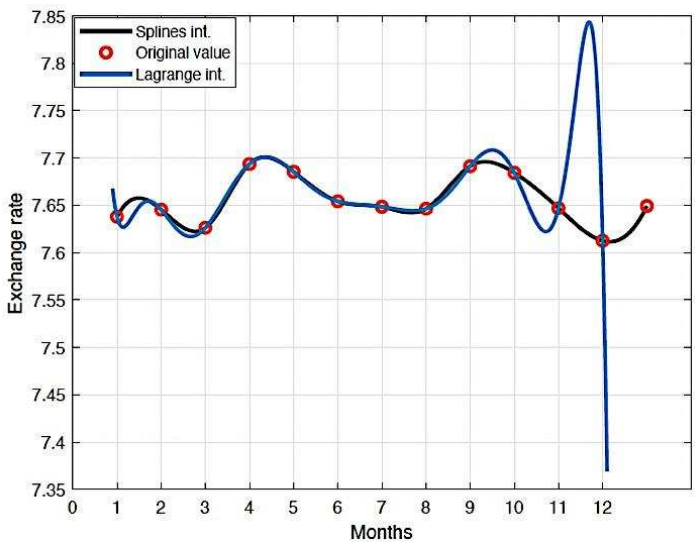


Figure 6. Lagrange vs. spline interpolation.

4.5. Framework model 5

This last model uses the PCA-GA combination as its main technique. The Central American region is the closest to Guatemala; thus, many of the goods produced in this region are very similar. However, countries within this region still maintain important and constant trading relationships with each other. The results obtained from the PCA-GA are shown in Table 2, section 2 (Improved with GA). It is important to note that the output obtained from the PCA is the input for the GA. This output is shown in Table 3. Figure 7 shows the PC and Banguat values, with the value in the graph being accumulative of each variable. The most relevant trading goods remain the same, with manufacturing industries and exports to El Salvador as the most relevant, with a 0.07674 degree of importance, followed by manufacturing industries and exports to

Honduras, with a 0.07404 degree of importance (as shown in Table 3). Figure 8(a) shows the relevant trading goods, while Figure 8(b) shows the values of each PC needed to obtained the 97.5% forecast.

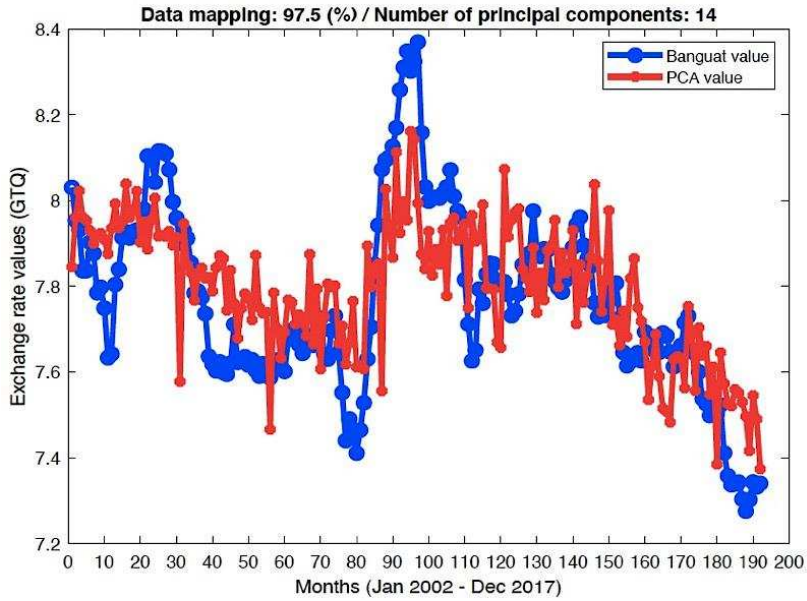


Figure 7. Obtained vs. original values

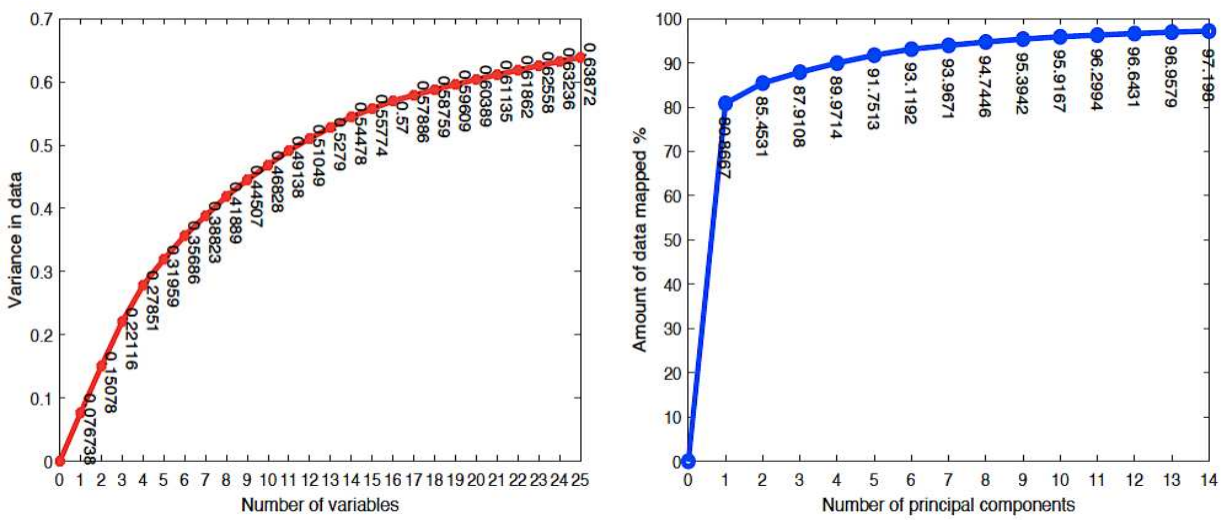


Figure 8. Framework model 5 – Central America case study: (a) relevant trading goods and (b) PC values

Table 5 shows a summary of the different case studies, highlighting critical factors taken into

consideration when analysing the most important region. It is clear that Central America is the most important region, as it has the highest number of relevant trading goods and the lowest error coefficient.

Table 5. Summary of case studies

Description[6]	USA	C.A.	N.A.	Asia	Europe	World
Number of Goods	110	254	188	62	40	544
Important Goods	16	64	31	14	9	98
PC 97.5% Mapping	9	14	12	8	12	26
First PC	65.739	80.886	67.14	79	48.05	67.256
ERA Error	0.01889	0.01417	0.01511	0.01984	0.0208	0.01302

5. Discussion

Many developing countries, like Guatemala, may be suffering from huge deficits; thus, identifying which trading goods have the biggest effect on the ER value, and to what extent, is crucial. This can be achieved through the PCA-GA technique, which also determines which products the country should focus on and whether or not the country is using its resources properly. Moreover, the interpolations not only provide a deeper look into the ER behaviour, but also increase the usefulness of the framework, making a huge difference opposed to only using the combination techniques, which cannot provide such detailed predictions. This paper also points out that classification techniques are not optimal for ER prediction. The models taken into consideration are based on performance after activation and the usefulness of the four techniques depends on the purpose of the forecast.

- If the analyst requires specific daily or hourly data, Lagrange and spline interpolation are the best option, as LU decomposition-GA and PCA-GA cannot provide this detailed information.
- Spline interpolation has a ‘smoother’ fit with real time values than Lagrange, but it takes longer and needs more computational space.
- Real time updates of the NNs, provided by the LU decomposition-GA and PCA-GA combinations, can help users keep track of movements when new data is being input.
- PCA-GA has a great impact on the proposed framework, as it provides information on the relevant trading goods and predicts the ER based on these values, creating a faster process (with less data and less memory) and enhancing the understanding of the data. However, this combination requires two or more sub-variables related to a common variable.
- PCA shows that USA is Guatemala’s strongest trading partner, because Guatemala has more trading goods with the USA than any other region (more than Asia and Europe combined). The amount of trading goods with the USA accounts for 58.5% of the data in North America, with a lower ER approximation error than Asia or Europe.

The system is not limited to only GTQ/USD. Through a simple input of the desired data in the required format, it can also be implemented in other countries where understating the ER behaviour is very critical. However, the quality and accuracy of the data may be affected by different factors, such as: (1) political instability – without a proper structure, it is impossible to obtain accurate data, especially in financial areas lacking transparency; (2) lack of data availability – Guatemala suffers from budget inequity; thus, some sectors are left without the necessary funds to perform proper data collection; and (3) no data updates – there

[6] C.A. = Central America, N.A. = North America, PC = Principal Component, ERA = Exchange Rate Approximation

were some numbers that have not been updated since 2013. Thus, to avoid these problems, it is necessary to support the study with a well-structured system of data collection and ensure that the data is available to everyone in the same structure (regarding formatting, nomenclature, description, and values). Lastly, due to problems such as natural disasters, international tensions, etc., it is also important to keep track of sectors that need improvement and goods which can be produced domestically, to reduce imports.

6. Conclusions

This paper predicts the ER based on relevant trading goods and provides a filtered and organised database of the desired results. We apply different combinations of machine learning techniques and predict the ER using Guatemala's import and export data, with empirical results supporting the accuracy of the application and reflecting a low approximation error, showing manufacturing industries and exports to El Salvador and Honduras as the most relevant trading goods. Through the visible data reduction and dynamic graphical analysis, we enhance the understanding of the data and its structure, providing a well-structured and stable approach to ER prediction.

Furthermore, under the general objectives is the new approach with the addition of relevant trading goods identification and ER prediction based on principal components. The empirical results and procedures provided could be beneficial or useful to the government, national institutions, policy makers, or other academics and analysts, who may utilise the data for further analysis. Although we only focus on ER prediction, the framework has the theoretical foundations and computational background for further development and implementation in other economic elements containing time series data, such as production, quality control, sales, logistics, semiconductor IC design (FPGAs), crop farming, or stock investment.

Acknowledgements

The authors would like to thank the reviewers for their invaluable support and contribution.

References

Alam, R. (2010). The link between real exchange rate and export earning: A cointegration and granger causality analysis on Bangla desh. International review of Business Research papers, 6(1), 205–214.

Arsalane, A., El Barbri, N., Tabyaoui, A., Klilou, A., Rhofir, K., and Halimi, A. (2018). An embedded system based on DSP platform and PCA-SVM algorithms for rapid beef meat freshness prediction and identification. Computers and electronics in agriculture, 152, 385–392.

Bal, C., and Demir, S. (2017). Forecasting TRY/USD exchange rate with various artificial neural network models. TEM Journal, 6(1), 11.

Chebbi, H. E., and Olarreaga, M. (2019). Investigating exchange rate shocks on agricultural trade balance: The case of Tunisia. *The Journal of International Trade & Economic Development, 28*(5), 628–647.

Fang, H., and Kwong, K. K. (1991). Forecasting foreign exchange rates. The Journal of Business Forecasting, 10(4), 16.

Federici, D., and Gandolfo, G. (2002). Chaos and the exchange rate. The Journal of International Trade & Economic Development, 11(2), 111–142.

Ford, W. (2014). Numerical linear algebra with applications: Using MATLAB. Academic Press.

Galeshchuk, S. (2016). Neural networks performance in exchange rate prediction. Neurocomputing, 172, 446–452.

Hu, L., and Cui, J. (2019). Digital image recognition based on fractional-order-PCA-SVM coupling algorithm. Measurement.

Hussain, A., Ahmad, B., and Masud, S. (2017). International trade and exchange rate behaviour. European Online Journal of Natural and Social Sciences, 6(4), pp–659.

Jing, C., and Hou, J. (2015). SVM and PCA based fault classification approaches for complicated industrial process. Neurocomputing, 167, 636–642.

Philip, A. A., Taofiki, A. A., and Bidemi, A. A. (2011). Artificial neural network model for forecasting foreign exchange rate. World of Computer Science and Information Technology Journal, 1(3), 110–118.

Refenes, A. N., Azema-Barac, M., Chen, L., and Karoussos, S. (1993). Currency exchnage rate prediction and neural network design strategies. Neural Computing & Applications, 1(1), 46–58.

Skittides, C., and Fruh, W.-G. (2014). Wind forecasting using principal component analysis. Renewable Energy, 69, 365–374.

Tarasova, I., and Coupe T. (2009). Exchange rate and trade: an analysis of the relationship for Ukraine. Journal of International Money and Finance.

Tsen Wong, H. (2013). Real exchange rate misalignment and economic growth in Malaysia. Journal of Economic Studies, 40(3), 298–313.

Tyree, E. W., and Long, J. (1995). Forecasting currency exchnage rates: neural networks and the random walk model. In City university working paper, proceedings of the third international conference on artificial intelligence applications.

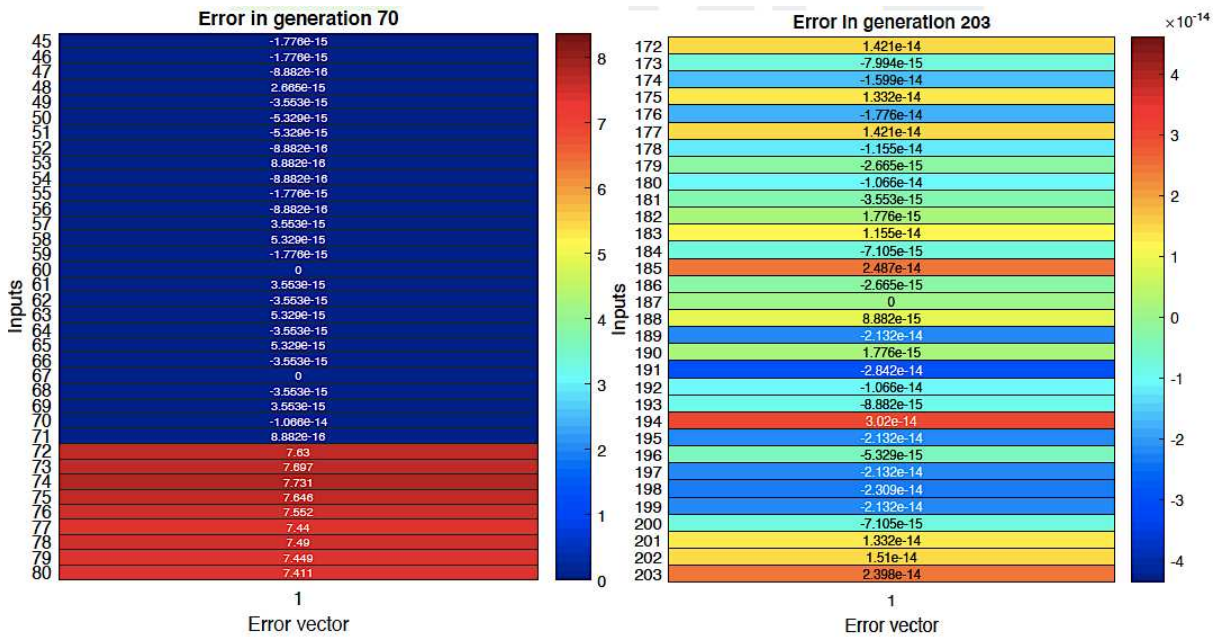
Vidal, C., Malassidis, E., García-Gómez, J., Martí-Bonmatí, L., Robles, M., and Mollet, J. (2002). El análisis de componentes principales como método de clasificación y visualización de tumores de partes blandas. In IX congreso nacional de informática médica.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioural sciences. Ph.D. thesis, Harvard University, Cambridge, MA, 1974.

Yu, H., Chen, R., and Zhang, G. (2014). A SVM stock selection model within PCA. Procedia Computer Science, 31, 406–412.

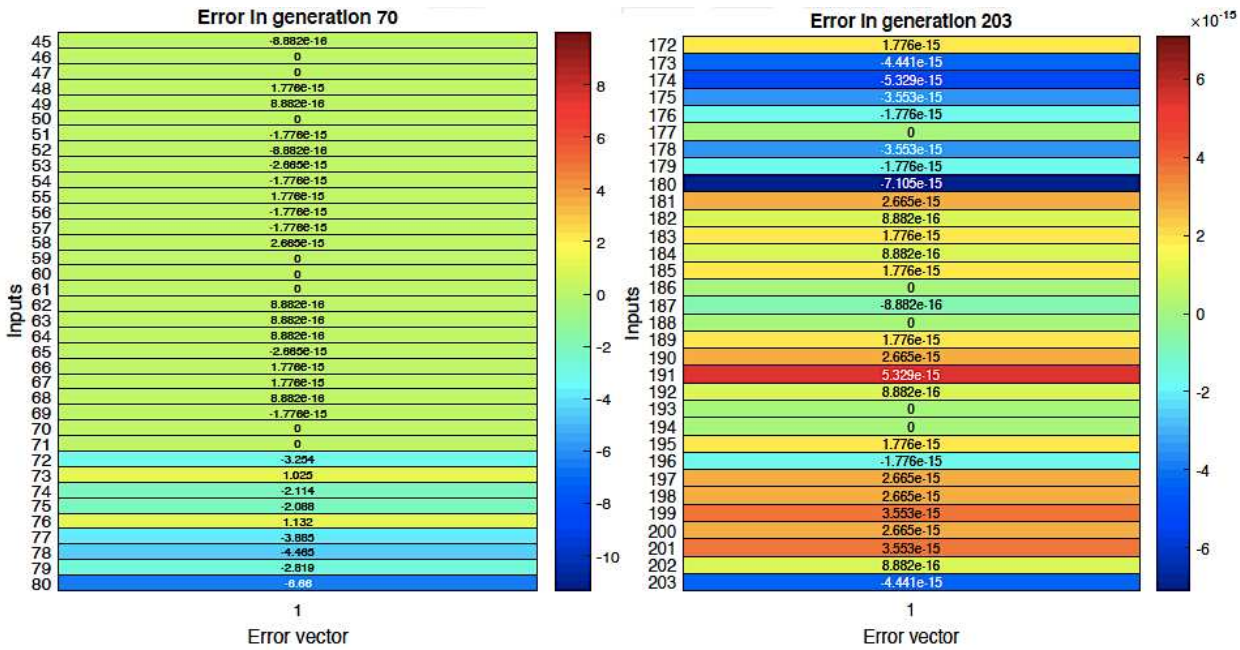
Zhi, H., and Liu, S. (2019). Face recognition based on genetic algorithm. Journal of Visual Communication and Image Representation, 58, 495–502.

Appendix A. LU-GA with RV



(a) **(b)**

Appendix B. LU-GA with PGC



(**a**) (**b**)