

SIADS 696

Project: Currency Movement Forecasting; Leveraging Regional High-GDP Country Interdependencies

Team: Spencer Cheung, Phuc Nguyen, Joey Higgins

Overview: The purpose of this project is to answer the question of whether using strong anchor currencies improves forecasts of weaker neighbors. Our goal is to use both unsupervised (clustering, PCA) and supervised (ARIMA, VAR, XGBoost, Granger causality) methods to explore the relationship between a dominant regional currency (anchor) and its weaker neighbors. Specifically, we aim to investigate if the movement of a strong currency, currency of nations with large GDP and trading partners, can be used to help explain and predict the behavior of weaker currencies within the same region. We would define the region based on trade linkage rather than purely on geography. We plan to use DeepNote (and GreatLakes) for collaboration efforts and reproducibility. Key challenges we anticipated are stationarity, causality interpretation, and excluding pegged currencies.

Similar Project: <https://github.com/kushg18/currency-exchange-rate-prediction>

- This project focuses on predicting the exchange price of currencies for two randomly selected countries
- This project uses a different training dataset (not matching the IMF's formatting/standards)
- Our project differs from this one, as we're focusing on anchor currency-neighbor currency causality and examining if adding these strong currencies as anchors helps our ability to forecast their weaker neighbors.

Training Data:

- [https://data.imf.org/en/Data-Explorer?datasetUrn=IMF.STA:ER\(4.0.1\)](https://data.imf.org/en/Data-Explorer?datasetUrn=IMF.STA:ER(4.0.1))
This dataset will be used to retrieve exchange rates of currencies to calculate returns. This will be used to develop the prediction models.
- [https://data.imf.org/en/Data-Explorer?datasetUrn=IMF.STA:IMTS\(1.0.0\)](https://data.imf.org/en/Data-Explorer?datasetUrn=IMF.STA:IMTS(1.0.0))
This data set will be used to attain trade flow between countries. We will use this to help define anchor-neighbor relationships between countries.
- <https://www.investing.com/currencies/>
This dataset will be a supplementary resource for exchange rates of currencies to help with missing data points and help with consistency in time series between anchors and their neighbors.
- <https://fred.stlouisfed.org/series/NYGDPMKTPCDWLD>
This dataset will be used to help us find anchors, as they will be identified by large GDP and strong trade dominance.
- All currency pairs will be captured against USD (XXX/USD) for consistent comparison across regions. USD itself will not be treated as an anchor for this exercise, for this reason, as all pairs are already in USD denomination. We will put our focus on non USD anchors like EUR and CNY.
- Pegged currencies (HKD, AED, SAR, etc) will be excluded from the dataset to avoid artificial causality in our models.

Supervised Learning Model(s):

- We plan to use ARIMA for time series analysis of neighbors' returns without the added anchor factors as a control group.
- VAR will be used for time series analysis to model returns using both anchor and neighbors as our treatment group.

- XGBoost (gradient boosting) to examine the comparison with a more complex and non linear model. We will compare XGBoost model with neighbor returns only with XGBoost model with both anchor and neighbor returns
- Due to the fact that all series are in XXX/USD, movements in USD can affect all series at the same time. To ensure we are isolating the anchor-neighbor effect, we will test whether using anchor/USD as additional input would improve our model's ability to forecast neighbor/USD.
- We plan to use Granger causality test to explore the predictive power of strong regional currencies over weaker ones.
- We will evaluate using Root Mean Squared Error (RMSE) as it is less favorable to larger errors (we need the predicted returns to be as close to actual as possible)
- Key visualizations will be line plots for epochs of training data, and bar charts for evaluation

Unsupervised Learning Model(s):

- We plan to use pairwise correlation heatmap to find correlation of daily returns between all countries to help with grouping effort.
- We plan to use k-means clustering on collected currency return data to identify natural cluster of currency
- PCA will be used to simplify data and visually group currencies in multiple dimensions and verify if the anchors and neighbors are grouping together.
- We will evaluate our generated clusters by silhouette score, check whether smaller countries cluster with their expected trade anchors (based on IMF trade flow data).
- Key visualizations will contain pairwise correlation heatmap, PCA plots, elbow charts for cluster selection, and charts of silhouette score for evaluation

Team Planning:

- Spencer Cheung: Handle dataset integration, data processing, assist with supervised/unsupervised models, visualizations, report writing and anything in between.
- Phuc Ngyuen: Data gathering, data cleaning, supervised/unsupervised models development, creating visualization, report writings.
- Joey Higgins: Determine shared environment, determine technology stack, develop pre-processing code, develop supervised learning model(s), develop unsupervised learning model(s), visualize results, summarize output in final report, assist as scrum master in standups

Timeline:

- 9/8/2025: Submit proposal draft
- 9/6/2025: Team meeting @ 1-2p EST
- 9/9/2025: Team meeting @ 8-9p EST
- 9/13/2025: Team meeting @ 1-2p EST
- 9/15/2025: Submit proposal final
- 9/22/2025: Begin development
 - Solidify shared platform
 - Import dataset(s)
 - Complete preprocessing code