Predicting direction of change for Latin American currencies

By Emmett Sexton

Introduction

In this analysis, I test the feasibility of using accessible machine learning algorithms and public data to predict daily movements in Latin American currencies. Utilizing free daily financial data spanning 2010-2022, the models are trained to predict whether each currency will appreciate or depreciate the following day against the US dollar (USD).

Figure 1 presents the value of the four currencies used in this analysis. The selected currencies are the Brazilian Real, Colombian Peso, Mexican Peso, and Chilean Peso. Each of these currencies has exhibited medium term volatility around a generally upward trend. Note that increases in the graphs represent depreciations in the value of the currencies. Each currency has lost substantial value since 2010, with Mexico seeing a roughly 30% decline compared to declines of more than 50% in Brazil and Colombia.

Figure 1: Exchange rates against USD, Jan 2010–Oct 2023

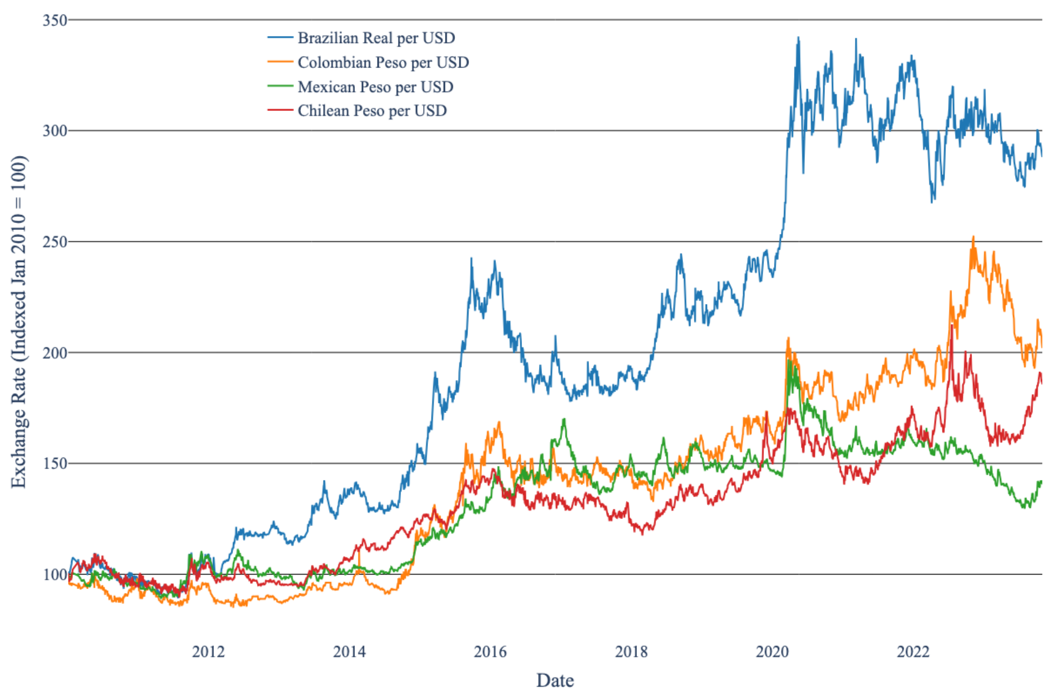
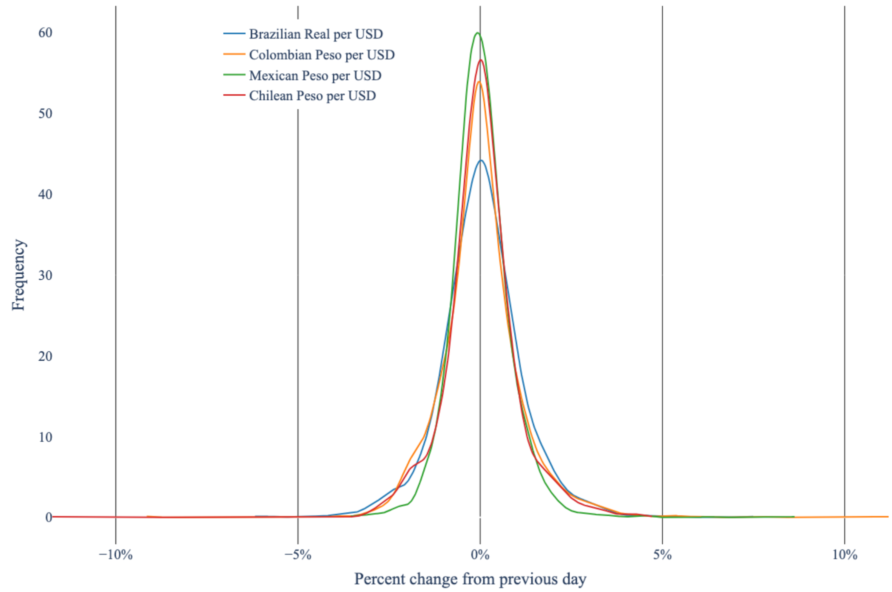


Figure 2 highlights the relative uniformity in the daily returns of the four currencies. Daily returns are centered near 0% and are larger in magnitude than 5% for less than 1% of the daily observations. This helps visually confirm that the target variable has relatively good class balance.[1]

[1] The mean value of the target binary variable ranged between 0.47 and 0.51 across the four currencies' train data sets.

Figure 2: Density plots for daily exchange rate changes against USD, Jan 2010–Oct 2023



Data

I limit data collection to data which is freely available online at a daily frequency.[2,3] The data consists of four main categories: exchange rates against the US dollar, primary equity indices by country, global commodity futures prices, and global and US specific controls.[4] Note that the index for Chile's main stock exchange and the commodity price for iron were not available. I used two ETFs for this period as proxies for those features.[5] For data availability reasons I restrict my analysis to the four currencies outlined above. I collect day closing data for model training and cross-validation for the period 2010–2022 and data for final model validation for the period from the beginning of Jan 2023 through the end of Oct 2023.

Data missingness is highly correlated between variables. Instead of interporalting or filling in missing values I choose to keep only complete observations. This reduces the number of total observations from 3,607 to 3,166 but it avoids training my models on observations that would have been almost entirely generated during data processing. I additionally visually identify a handful of obvious outliers in the exchange rate data that I drop.[6] For all features used, I perform two preprocessing actions: (1) I calculate the percent change from the previous day and then (2) apply a MinMax scalar to each feature. The decision to use percent changes for the features was made after initial training was less successful using

[2] Note that I experimented with more canonical variables from exchange-rate regressions (e.g. government interest rates, consumer prices, exports, imports, industrial production, etc.) at the monthly level, but these models failed to yield strong predictive power on out of sample testing data due to the lower sample sizes in the training data.
[3] I used Yahoo Finance's python API to download the data on 2023/11/20.
[4] See the Data Appendix for specific variables used along with their tickers on Yahoo Finance.
[5] Ticker ECH is the iShares MSCI Chile ETF and ticker XME is an ETF that tracks the S&P Metals and Mining industry index. Both were available for the full period of analysis.
[6] 13 observations in total were dropped in this fashion because they deviated more than four standard deviations from the previous day's exchange rate.

levels and the scaling is done to accommodate the models which are sensitive to differences in feature ranges.

For each of the four currencies, I create a features data set that contains all of the collected data, excluding only that target currency.[7] I then apply the percent changes and feature scaling outlined above. For the target variable, I create an indicator that takes a value of 1 if a target currency loses value relative to the USD compared to the day before and a value of 0 otherwise, resulting in a binary classification problem. Next, I shift the target indicator backward one period so that predictions are made using only data seen prior to the prediction day. The resulting predictions on the direction of the currency value at time t_0 are made using the percent change in feature values from time t_{-2} to time t_{-1}.

Methodology

In total, I train 16 models, one for each combination of my four currencies and four machine learning classification methods: Elastic net (EN), Random Forest (RF), Gradient Boosting (GB), and Multi-Layer Perceptron (MLP). Amongst the more accessible techniques in machine learning, I believe that together these models provide a suitable set of approaches for testing the feasibility of predicting next-day currency appreciation and depreciation without proprietary economic data or access to advanced compute resources. In particular, EN can efficiently model potential linear relationships and perform automatic feature regularization, while the tree-based ensemble methods (RF and GB) can model potentially complex nonlinear relationships and interactions between features. Lastly, the neural network architecture of the MLP enables modeling of diverse functional relationships between features and exchange rate movements that may evade detection in linear and tree ensemble methods.

Each model is trained separately for each currency using data from 2010–2022 with cross validation to reduce overfitting and tune hyperparameters.[8] During cross validation I perform hyperparameter tuning using random search across appropriate parameter grids specific to each model.[9] Because outcome classes (appreciation or depreciation) are balanced, best hyperparameters are chosen based on model accuracy.[10]

Next, I perform a final validation of each model's reliability in the context of unseen data by using the best hyperparameters for each currency-model combination to predict daily appreciation or depreciation for the period Jan 2023–Oct 2023. I then calculate the accuracy of the model for this final validation period, along with several other classification metrics.

[7] I keep equity indices and exchange rates from the non-target currency countries in the data as features for the target currency data set to capture potential shared regional movements or bilateral dependencies. I experimented with different feature selection rules, none of which outperformed retaining all of my collected features and allowing the models to perform their implicit feature selection or feature importance weighting.

[8] Scikit-learn's TimeSeriesSplit was used to ensure all testing folds followed a given training fold, thereby respecting the time dependent nature of the data.

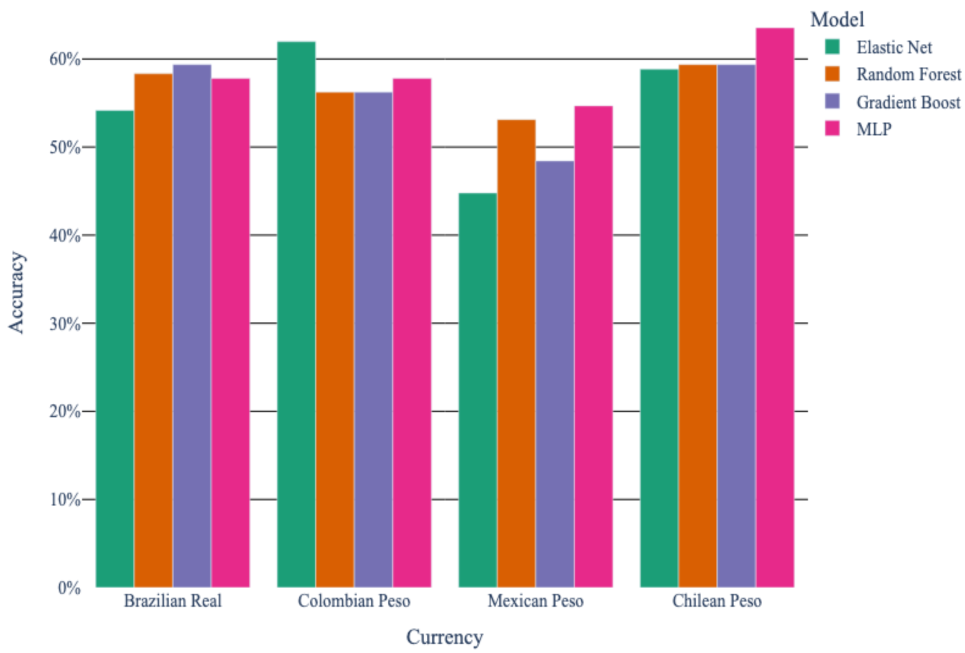
[9] I experimented with a random search followed by a grid search centered on the best parameters from the random search and found no noticeable improvement in accuracy, so for conciseness and reduced runtime, I retained only the random search. To reduce runtime further, if a hyperparameter was the same across all four currencies, I fixed the value of that hyperparameter in the random search grid for that model.

[10] Additionally, in the context of market participants, one can bet as easily on an increase as on a potential decrease in a given currency. Here, the cost of false positives and false negatives are equal to a trader.

Results

Focusing on the validation set from Jan 2023–Oct 2023, results varied most across currencies, opposed to across models. Figure 3 shows that the Chilean Peso models had the highest average accuracy at 60%, compared to averages of approximately 55% for the Brazilian and Colombian Pesos. The Mexican Peso models performed the worse with an average accuracy of 50%.

Figure 3: Accuracy on validation set, Jan 2023–Oct 2023

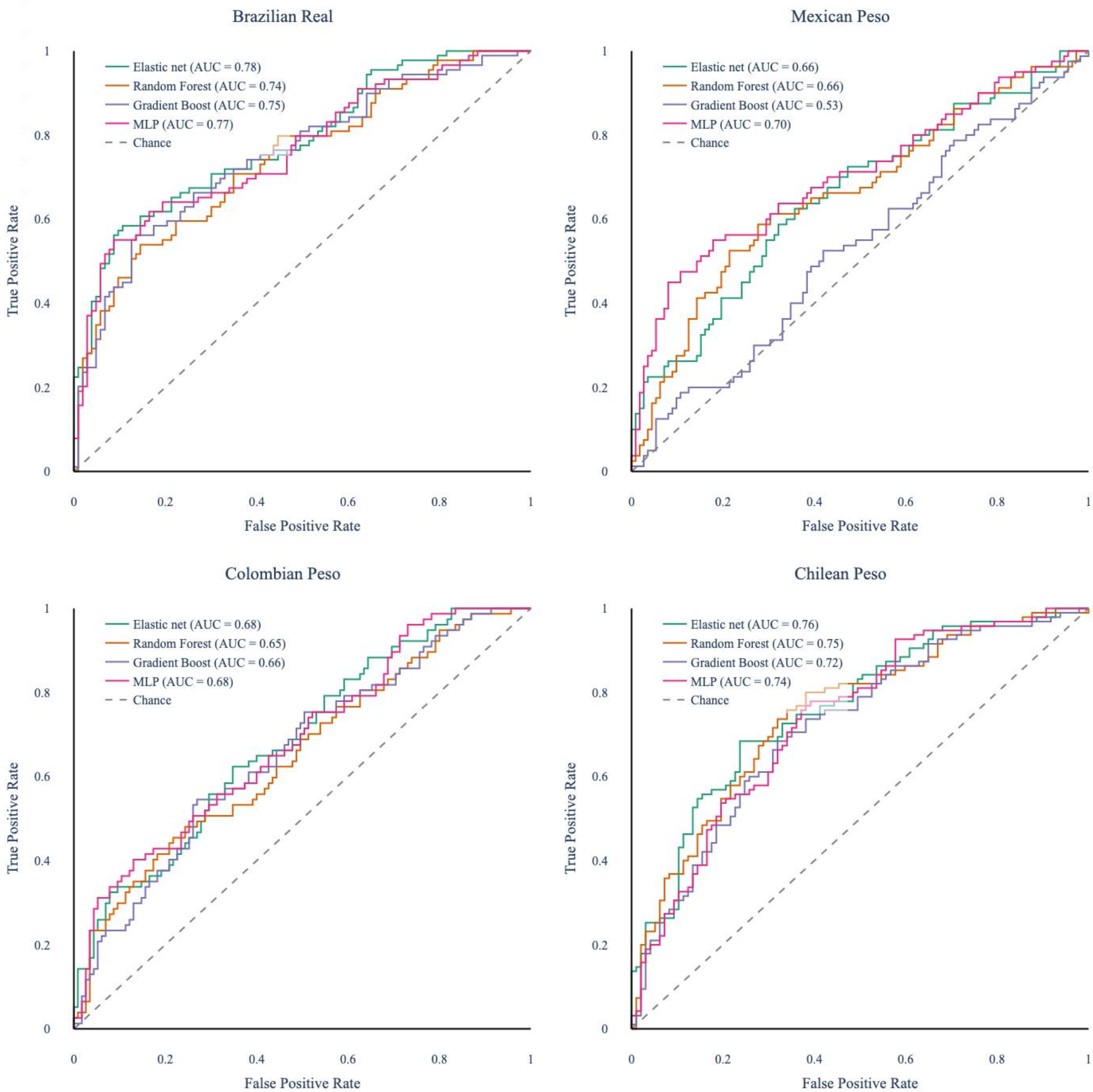


No model was consistently superior to the others across currencies. The EN model had the best accuracy for the Colombian Peso, the MLP had the best accuracy for the Mexican and Chilean Peso, while the GB model performed the for the Brazilian Real. This can also be seen in the ROC curves from Figure 4. The AUCs for the four models are relatively consistent within currencies, with the exception of the GB model for the Mexican Peso which far underperformed the other three models for that currency.

Across models, the average AUCs were significantly higher for the Brazilian Real and Chilean Peso relative to the Colombian and Mexican Pesos. This was to be expected because a disproportionate number of the freely available feature variables used in training were commodity prices and Chile and Brazil both have economies that rely more heavily on commodities exports relative to Mexico and Colombia. Chile in particular has a high value of exports relative to its GDP, and roughly half of those exports rely on copper.[11] Unsurprisingly, when analyzing feature importances across models for the Chilean Peso, the copper futures price was the second most important feature.[12]

[11] https://oec.world/en/profile/country/chl
[12] This was for the RF and GB models.

Figure 4: ROC curves for validation set, Jan 2023–Oct 2023



More surprisingly, the ETF I used to proxy for Chilean equity prices was amongst the most important features for all currencies. This may be driven by the fact that the ETF is priced in USD and is picking up signals of broader foreign investor appetite that may cause movements in Latin American currencies.

Finally, while the average accuracy on my validation set was significantly greater than 50% for three of the four currencies, my models in general suffered from a poor balance between precision and recall. Precision values ranged from 42–58%, while recall values ranged from 66–100%. For robustness, I trained all models using their F1 score instead of accuracy, but I saw no significant differences in validation set performance. To have more confidence that the models are picking up real signals opposed to noise, I will need to focus future research on striking a greater balance between precision and recall.

Data Appendix

Exchange rates	Equities indices	Global-US controls	Commodity futures prices
Brazilian Real (USDBRL=X)	Brazil (^BVSP)	Emerging market bond index (EMB)	Gold (GC=F)
Colombian Peso (USDCOP=X)	Mexico (^MXX)	Dollar strength index (DX-Y.NYB)	Silver (SI=F)
Mexican Peso (USDMXN=X)	Colombia (BVC.CL)	US equity volatility (^VIX)	Platinum (PL=F)
Chilean Peso (USDCLP=X)	Chile ETF (ECH)*	S&P 500 (^GSPC)	Copper (HG=F)
		13 week US Treasury (^IRX)	Palladium (PA=F)
		5 year US Treasury (^FVX)	Crude Oil (CL=F)
		10 year US Treasury (^TNX)	Heating Oil (HO=F)
		30 year US Treasury (^TYX)	Natural Gas (NG=F)
			Corn (ZC=F)
			Oat (ZO=F)
			Wheat (KE=F)
			Rough Rice (ZR=F)
			Soybean Oil (ZL=F)
			Soybean (ZS=F)
			Lean Hogs (HE=F)
			Live Cattle (LE=F)
			Cocoa (CC=F)
			Coffee (KC=F)
			Cotton (CT=F)
			Sugar (SB=F)
			Metals & mining ETF (XME)*

*ETFs were used when standard equity indices or commodities were unavailable for the full duration of analysis.