Forecasting Devaluation Events in Latin American Countries Using Google Trends Data

Marcelo Emilio Ghia Salazar Ante

June 13, 2023

Abstract

This project aims to forecast devaluation events in Latin American countries using Google Trends data in order to analyse a decision-making problem. The study employs panel data extracted from Google Trends, including keywords related to economic crisis, default, devaluation, dollar price, inflation, and exchange rate. The data consists of the quantity of searches on a weekly basis for these keywords in nine different countries in Latin America (Argentina, Bolivia, Brazil, Chile, Colombia, Paraguay, Peru, and Uruguay). Additionally, the currency price in terms of USD is extracted for each week and the percentage variation in the price is calculated. A cutoff of 7.5% is used to identify "jump devaluations" exceeding the cutoff.

1 Introduction - Decision Making Problem

In this context, let's say a company based in Argentina buys raw materials in USD, where the timing to place an order depends on devaluation. In this case, the decision-maker can consider the different meanings for false positives (FP) and false negatives (FN) in their decision-making process.

The formula that represents the cost minimization where the timing to place an order depends on devaluation is:

Minimize $E[Cost_c] = CostDelay \cdot (FP \cdot E[FP_c] + TN \cdot E[TN_c]) + CostNoDelay \cdot (FN \cdot E[FN_c] + TP \cdot E[TP_c])$

FP (False Positives): This term represents the false positive prediction of a devaluation, where the company delays the purchase (Delayc) expecting a devaluation to occur, but it does not actually happen. The cost associated with false positives can include increased prices, storage costs, missed business opportunities, or any other costs incurred due to delaying the purchase based on an incorrect prediction.

E[FPc] (Expected Number of False Positives): This term represents the expected number of false positive predictions made by the company. It considers historical data, economic indicators, or any other factors that help estimate the likelihood of false positives occurring.

TN (True Negatives): This term represents the true negative prediction of no devaluation, where the company does not delay the purchase (NoDelayc) because no devaluation is predicted, and in reality, no devaluation occurs. The term TN captures the potential cost savings or avoided costs associated with correctly predicting the absence of devaluations.

E[TNc] (Expected Number of True Negatives): This term represents the expected number of true negative predictions made by the company. It considers historical data and other factors to estimate the likelihood of true negatives occurring.

FN (False Negatives): This term represents the false negative prediction of no devaluation, where the company does not delay the purchase (NoDelayc) despite the occurrence of a devaluation. The cost associated with false negatives can include additional costs due to unfavorable exchange rates, missed cost savings, or any other costs incurred by not delaying the purchase when a devaluation actually occurs.

E[FNc] (Expected Number of False Negatives): This term represents the expected number of false negative predictions made by the company. It considers historical data and other factors to estimate the likelihood of false negatives occurring.

TP (True Positives): This term represents the true positive prediction of a devaluation, where the company delays the purchase (Delayc) based on an accurate prediction, and a devaluation actually occurs. The term TP captures the potential cost savings or benefits associated with correctly predicting and taking advantage of devaluations.

E[TPc] (Expected Number of True Positives): This term represents the expected number of true positive predictions made by the company. It considers historical data, economic indicators, or any other factors to estimate the likelihood of true positives occurring.

2 Methodology

Data Extraction: Google Trends data is used to extract the quantity of searches for specific keywords. The keywords used are "economic crisis," "default," "devaluation," "dollar price," "inflation," and "exchange rate." Data is extracted for nine Latin American countries.

Data Preprocessing: The extracted data is stored in a panel data format. The currency price in terms of USD is added, and the percentage variation in the price is calculated. A new column, "Jump," is created to indicate whether there was a "jump devaluation" exceeding the 7.5% cutoff.

Target Variable Preparation: The target variable, representing the incidence of devaluation events, is created by shifting the "Jump" column t periods ahead. The maximum value for t periods forward is taken as the target variable. The ons_target_f variable is created to account for ongoing devaluations.

Data Slicing: The data is divided into three subsets: full model, text-only model, and history-only model. The full model includes all features, the text-only model includes only text-based features, and the history-only model includes only conflict history. Each subset is prepared separately for training and forecasting.

Model Building: Random Forest classifiers are trained using the prepared data slices. Grid search is conducted to optimize the hyperparameters of the classifiers. Separate models are built for the full, text-only, and history-only models.

Rolling Forecast: The models are used to generate fitted values and predict future devaluation events. The rolling forecast is conducted on a weekly basis, starting from a specific year and week. The predicted probabilities of devaluation events are stored in the fittedframe dataframe.

3 Results

3.1 Model Performance

Let's check the performance of the model. First, in the below table we can see AUC and Avg Precision of the Full Model, the History Model and the Text Model:

Model	AUC	AVG Precision
Full Model	0.77	0.77
History Model	0.76	0.75
Text Model	0.65	0.7

Table 1: Model Performance

As we can see in the performance result, History Model performs better than text model, which make sense. Perhaps, additional text features can improve the performance as I only create 6 columns to signal devaluation. Full model has the higher performance, which also make sense as I added text model to the history model to make the predictions. However, there is not a big difference between the full model and the history model, so we can say that historical data on price weight much more when forecasting a devaluation jump than text (in this case). To improve this results, I would probably need to have a bigger panel data, adding more countries. In addition, the Google searchs to signal devaluation could differ from country to country, so this is also something to analyse when building the panel data. On the other hand, let's see the ROC curve and the Precision-Recall curve as below:

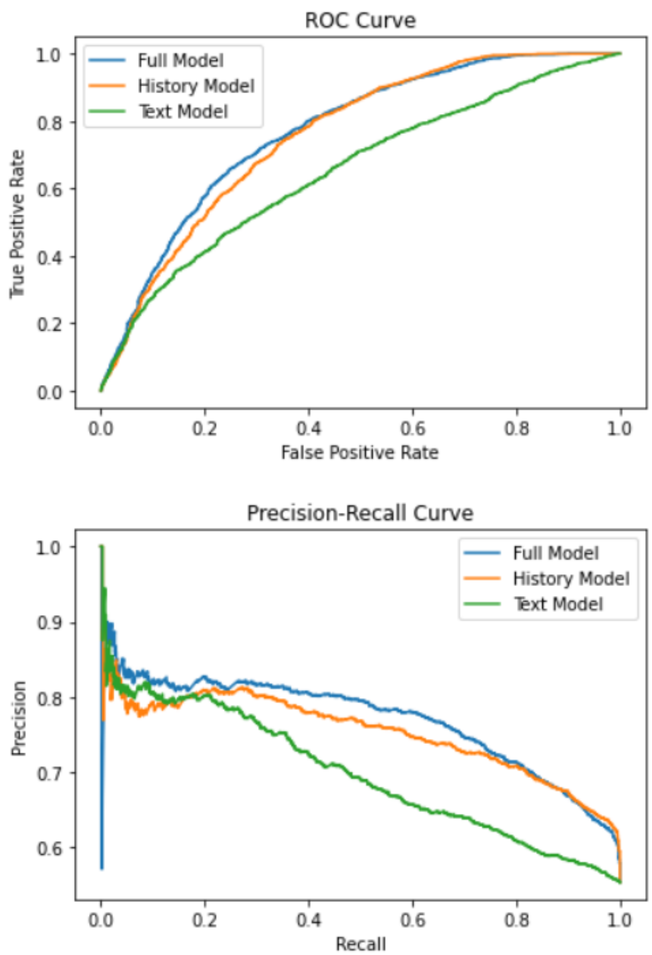

Figure 1: Perfomance Curves

And finally, the Full Model Separation Plot:

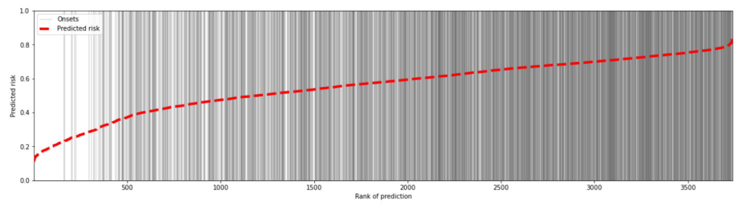

Figure 2: Separation Plot

By examining the plot, we can assess how well the model separates the two classes. Ideally, we would expect to see a clear distinction between the positive and negative classes. This means that the points associated with the positive class should be predominantly located at high predicted scores, while the points associated with the negative class should be predominantly located at low predicted scores. Such a clear separation indicates that the model is performing well in distinguishing between the two classes. On the other hand, if there is a significant overlap between the points representing the positive and negative classes, it suggests that the model may be struggling to accurately separate the classes. This could indicate potential issues with misclassification or confusion between the classes.

3.2 Decision Making Results

To compute and analyse the Decision Making Model first I establish a threshold 0f 0.5 for classifying the predictions and I have the following results:

Metric	**Value**
True Positives (TP)	1427
True Negatives (TN)	1109
False Positives (FP)	797
False Negatives (FN)	947

Table 2: Confusion Matrix

False Positive (FP) prediction: FP occurs when the model predicts a devaluation, but it does not actually happen. FP implies that the company might place an order for raw materials prematurely, expecting a devaluation that does not occur. The cost associated with FP can be related to the potential loss or extra expenses incurred due to early ordering, such as additional storage costs, currency exchange fees, or lost opportunities to benefit from a lower exchange rate.

False Negative (FN) prediction: FN occurs when the model fails to predict a devaluation, but it actually happens.FN implies that the company might delay placing an order, expecting no devaluation, but a devaluation does occur.

The cost associated with FN can be related to the potential loss or negative impact on the company's operations due to delayed ordering, such as higher purchasing costs, supply chain disruptions, production delays, or missed market opportunities.

Given this insights, to compute the decision making model I considered Cost-Delay = 40 and CostNoDelay = 10 because a delay is usually more expensive in a supply chain perspective, as it is usually higher in terms of costs a supply chain disruptions or production delays rather than additional storage costs or exchange fees. Then, I plot isocost curves looking at different thresholds:

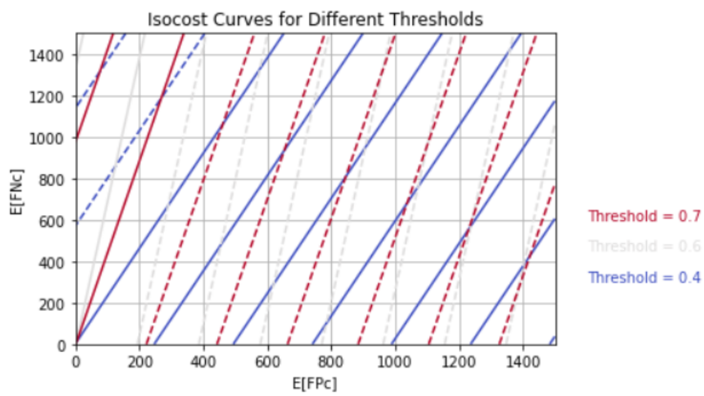

Figure 3: Isocost Curves

And calculate performance given each threshold:

Threshold	Accuracy	Precision	Recall	F1 Score
0.4	0.570093	0.582204	0.796546	0.672714
0.5	0.592523	0.641637	0.601095	0.620705
0.6	0.564486	0.728495	0.342460	0.465903
0.7	0.519626	0.802281	0.177759	0.291034

Table 3: Model Performance

And finally, the Expected Cost of each threshold:

Threshold	Expected Cost
0.4	207.809044
0.5	223.126284
0.6	240.110154
0.7	243.535641

Table 4: Expected Costs

As the threshold increases from 0.4 to 0.7, the expected cost also increases, indicating a higher overall cost for the model. The accuracy, precision, recall, and F1 score values vary across different thresholds, reflecting the trade-off between different performance metrics. Lower thresholds (0.4 and 0.5) result in higher recall values, meaning more positive instances are correctly classified, but at the expense of higher false positive rates and overall cost. Higher thresholds (0.6 and 0.7) lead to higher precision values, indicating a lower false positive rate, but at the cost of lower recall and overall model performance.

Considering these factors, a higher threshold would result in a higher false negative rate, meaning more positive instances would be incorrectly classified as negative. This would lead to higher costs because of the higher FNc value (cost of false negatives).

On the other hand, a lower threshold would result in a higher false positive rate, meaning more negative instances would be incorrectly classified as positive. This would lead to higher costs because of the higher FPc value (cost of false positives).

Therefore, from a cost perspective, it is likely that a higher threshold would result in higher costs, given the higher FNc value. However, the specific trade-off between false positives and false negatives should be carefully considered based on the context and requirements of the problem. It ultimately depends on the relative importance and associated costs of different types of prediction errors in the specific application. In this specific case, I would prioritize a higher recall (meaning being more flexible in terms of the probability of having a devaluation, so lower threshold) because of the inputs I chose. A delay is usually more expensive in a supply chain perspective, as it is usually higher in terms of costs a supply chain disruptions or production delays rather than additional storage costs or exchange fees.

4 Conclusion

In this study, I aimed to forecast devaluation events in Latin American countries using Google Trends data and analyze the decision-making problem associated with the timing of placing orders for raw materials based on devaluation predictions. I constructed a model based on panel data extracted from Google Trends, incorporating keywords related to economic crisis, default, devaluation, dollar price, inflation, and exchange rate. The model utilized historical search data and currency price variations to predict devaluation events.

The results indicate that the full model, which combines historical data with text-based features, performed better in terms of model performance compared to the history-only model and text-only model. However, the difference between the full model and the history-only model was not significant, suggesting that historical data on price weighed more heavily in forecasting devaluation jumps than text-based features in this specific case.

I evaluated the model's performance using metrics such as AUC, average precision, ROC curve, and precision-recall curve. The model exhibited reasonable performance, with an AUC of 0.77 for the full model, indicating a good ability to distinguish between positive and negative classes.

To analyze the decision-making aspect, I introduced a cost minimization formula that considered false positives (FP), false negatives (FN), true negatives (TN), and true positives (TP). By varying the threshold for classifying predictions, I computed the expected costs associated with different thresholds. Our analysis revealed that higher thresholds resulted in higher expected costs due to increased false negatives, while lower thresholds led to higher false positive rates and overall costs.

Ultimately, the choice of threshold depends on the specific context and requirements of the problem. In this case, considering the supply chain perspective and the higher cost associated with delays, I recommended a lower threshold to prioritize a higher recall, thereby reducing the likelihood of false negatives.

The study provides valuable insights into forecasting devaluation events and the decision-making process associated with timing orders based on such predictions. Further research could explore the inclusion of additional countries and fine-tune the model to improve its performance and practical applicability.