IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Classification of participants in the foreign exchange market [1]

Amir Khatib,
Bank of Israel

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Classification of Participants in the Foreign Exchange Market

Amir Khatib[1]

Abstract

This Article presents an analysis of granular data in the Bank of Israel's Big Data database of foreign exchange market transactions. The aim of this study is to expand the existing information on this market by predicting a sector and subsector of new unidentified entities.

Supervised classification algorithms, such as Support Vector Machines (SVM), Logistic regression, and Random Forest, are used to identify real sector market participants as exporters and importers, as well as to allocate unknown foreign market participants to either the financial or nonfinancial sector.

The results of this research demonstrate the benefit of analyzing granular data and of the ability of the chosen algorithms to accurately predict sectors and subsectors of unidentified new entities. The findings of this study allow the Bank of Israel to expand information about entities in the foreign exchange market in order to monitor market activity.

[1] Bank of Israel, amir.khatib@boi.org.il

INTRODUCTION

Similar to other countries, the foreign exchange (forex) market in Israel is the largest financial market in terms of trading turnover. In Israel, the daily average turnover in this market is about US$11 billion. This is significantly greater than turnover in the equities market (US$0.4 billion) or in the bond market (US$0.9 billion). Moreover, the offshore shekel foreign market is almost twice the shekel turnover in Israel.

Similar to many other small open economies, the forex market in Israel plays a major role in the complex effects of monetary policy decisions. Its significance is due to two main factors: first, the relatively high pass-through level (0.25) from the exchange rate to the Consumer Price Index; and second, the deep effect of changes in the exchange rate on the tradable segment of the real economy. Fluctuations in the exchange rate can have significant implications for export-oriented industries, importers, and sectors that depend on the trade of others. Understanding the complex interrelationships between the exchange rate and the tradable segment is essential for policy makers wishing to cultivate stability and sustainable economic growth. (Abir, 2020)

The main participants in the Israeli foreign exchange market are the business sector (exporters and importers), institutional investors (pension funds, mutual funds, and insurance companies), and foreign institutions. Understanding the dynamics of these sectors is essential in understanding the implications of monetary policy decisions. Real-time information on the microstructure of the market and the flow of transactions provides input that is important for policymakers' considerations.

Recognizing the importance of this information, the Bank of Israel initiated reporting requirements for transactions in the forex and shekel interest rate markets in June 2016. The financial intermediaries are required to provide daily reports at the transaction level, which enable the central bank to monitor and analyze market participants' behavior by various sectors and subsectors, and even by individual participants. An in-depth understanding of market flows provides an infrastructure for decision-making with regard to monetary policy and improving macroprudential tools.

Each sector has its own unique motives and different sensitivities to changes in the exchange rate. Therefore, by harnessing the power of advanced machine learning and supervised algorithms on the reporting database, we can better understand the foreign exchange market's dynamics by expanding the sector identification of the existing variety of participants in the market.

This study identifies the sector of unidentified market participants through a methodical analysis of the database of market participants identified by sector. In particular, we expand the identification of participants from the business sector into importers and exporters, and foreign institution participants into financial and nonfinancial sectors.

Recognizing the fact that each sector operates with unique objectives and demonstrates varying sensitivities to changes in the exchange rate, the classification of participants becomes a central factor in understanding the complex dynamics of this market. Researchers and market analysts can gain deeper insights regarding behavior, risk, and the market's potential effects related to various sectors.

The findings of this study will provide an improved basis for market analysis, which will enable policy makers to make more informed decisions and to response efficiently to changes in the activity of the market's segments and the dynamics between them.

Description of the database

The database managed by the Bank of Israel's Information and Statistics Department contains data obtained pursuant to the reporting order implemented in June 2016. Within this framework, roughly 50 financial intermediaries from Israel and abroad report daily data on foreign exchange transactions involving the shekel. The reports include detailed information on the features of the transaction, such as time of transaction (down to the second), instrument, value amount, transaction rate, and details of the parties to the transaction. The information is organized and processed into trading volumes, conversions, and foreign exchange exposures by various sectors, which allows for a comprehensive analysis of developments in the sectors of the foreign exchange market.

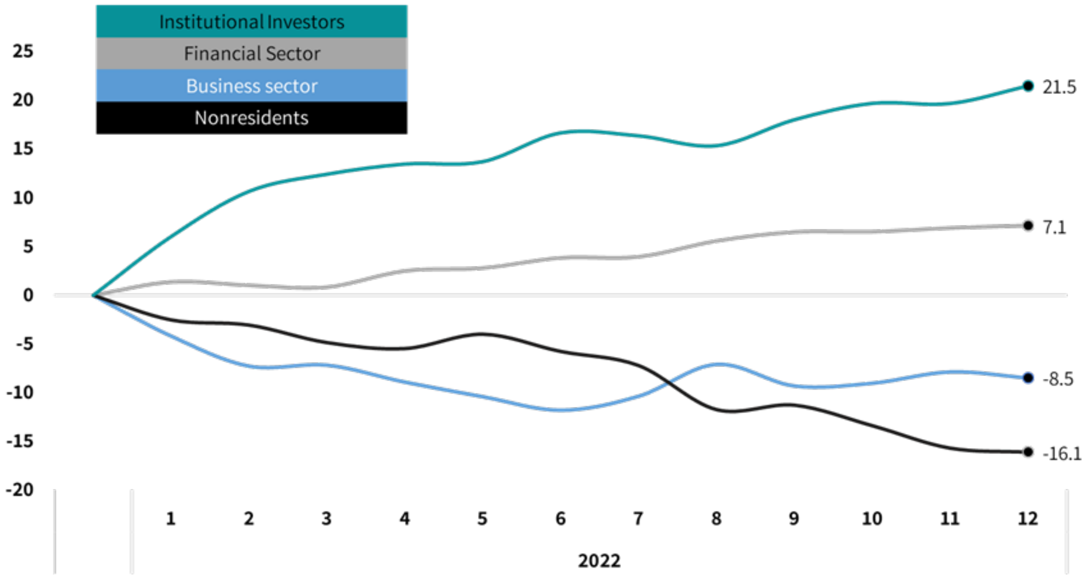
The Bank of Israel's Information and Statistics Department uses a meticulous data processing methodology in order to ensure the reliability and consistency of the information. The data undergo thorough controls, including logical and statistical tests, in order to verify their precision. The data serve to create various reports and products that offer important insights regarding the foreign exchange market. These reports are used as essential tools for decision makers at the central bank to monitor the market.

The data on transactions come from three main sources. Domestic banking corporations began reporting on their transactions in 2008, and foreign financial institutions and domestic financial intermediaries began reporting on their transactions in 2017. The data are obtained on a daily basis, and provide details of all transactions made during the previous day.

The database is comprised of about 50 reporting entities and tens of thousands of active participants in the OTC market. With millions of records received each year, the database contains a significant volume of trading activity. In order to ensure consistency and international standardization, the reporting format was adapted as much as possible to the guidelines set by the ISDA.

The Bank of Israel uses the forex database to produce various products, such as net cumulative foreign exchange purchases by main sector (Figure 1), daily trading volume by type of instrument, implied standard deviations in shekel/dollar options, and more. These products offer important insights regarding market trends, and help decision makers formulate efficient policy.

Figure 1: net cumulative foreign exchange purchases by main sector



Methodology

Data preparation

The data in the Bank of Israel's database undergo built-in logical and statistical controls in order to ensure precision and reliability. These controls are intended to identify any error or inconsistency, and to provide feedback to reporting entities.

Therefore, for the purpose of the analysis in this paper, the required data preparation is minimal. Thanks to the meticulous data preparation tool used by the Bank of Israel, the database for this analysis is already reliable and precise. This allows us to focus mainly on implementing controlled algorithms and content engineering techniques in order to extract valuable insights and efficiently classify forex market participants.

This section outlines the steps taken to prepare the database for analysis.

Selecting relevant features

We extract a group of key variables from the forex database, which will provide us with important insights regarding participants' behavior. These characteristics mainly include:

- Type of instrument: specific type of financial instrument involved in each transaction.
- Currency: The pair of currencies involved in the transaction.
- Financial intermediary: The reporting entity involved in the transaction.
- Transaction period: The duration during which each transaction took place.
- Value amount: Presents the value amount of the transaction.
- Purchase/Sale indicator: If the transaction includes the purchase or sale of shekels.
- Transaction rate: Reflects the exchange rate at which the transaction was made.
- The prevailing exchange rate at the time of the transaction.

One-hot encoding of categorical features

In order to efficiently integrate categorical features such as type of product, currency, and financial intermediary, our analysis implements one-hot encoding. This process transforms categorical variables into binary vectors, and enables the machine learning algorithms to precisely interpret them. A binary variable is assigned to each category in a feature, and a separate column is created for each unique category.

Calculating final features and grouping by market participant level

Following the selection of the relevant features and the start of one-hot encoding, we gather the data at the transaction level to the market participant level. This includes a calculation of various indices and grouping the data based on the unique identifiers of each participant. The compilation methods may include total transaction amounts, calculation of the average exchange rates, or deriving participant-specific indicators.

Transformation and Standardization

As part of the data preparation process, logarithmic transformation and standardization of features are employed to ensure optimal performance of the models. Transformation involves modifying the distribution or representation of the features to meet certain assumptions or improve interpretability. standardization, on the other hand, is applied to ensure that all features are on a similar scale and have comparable ranges. This is particularly important for models that are sensitive to the scale of features, such as SVM and KNN. Transformation and scaling techniques can enhance model performance, improve convergence, and prevent certain features from dominating the modelling process.

Feature engineering

Combining our in-depth knowledge of the forex market with insights from discussions with brokers and market participants over the years, we built additional features in order to further enrich the data array, help distinguish between the various sectors, and improve the models' proper forecasting.

For each of the following features, we first present the insight that we know from the market and the need to build it:

1. Median by net forex transaction days

   Importers generally engage in net forex purchases, while exporters tend to sell.

   In order to capture this behavior, we calculate the net foreign exchange transactions for each market participant over a defined period. By accumulating net daily buy/sell positions, we calculate a variable that is the median representing the main tendency of these transactions. This variable provides important insights regarding the overall trading activity of importers and exporters, and sheds light on their behavior and preferences.

2. Index of activity in anomalous depreciation

   Exporters frequently exploit periods in which the shekel is depreciated in order to sell large volumes of foreign exchange. In contrast, importers tend to make currency transactions based on demand, irrespective of the exchange rate.

   In order to capture this behavior, we calculate an index of activity under anomalous depreciation. This index is calculated by comparing the current shekel exchange rate with the moving monthly average of the exchange rate.

On days when there is an anomalous gap between the exchange rate level at the time of the transaction and the moving average in the previous month (significant depreciation of the shekel), we calculate the net total transactions (purchases/sales) of each entity, and divide that by the overall trading volume. This variable provides an insight regard the extent to which entities exploit the depreciation in order to make larger net transactions relative to their overall trading activity.

By integrating this variable, we can identify market participants that show a greater tendency to exploit the depreciation and make significant net transactions during such periods.

3. Spot foreign exchange purchase ratio

Importers have a greater tendency to purchase foreign exchange through spot transactions for current trade activity.

This variable presents the ratio between foreign exchange spot purchases and the total number of spot transactions for each market participant. Using this ratio, we can quantify the share of spot transactions that included foreign exchange purchases. Higher values of this ratio indicate a greater tendency to purchase foreign exchange via spot transactions.

These features give us a more comprehensive understanding of participants' behavior and enable us to more precisely classify entities in the forex market. By capturing the unique transaction patterns of exporters and importers, these features contribute to a deeper understanding of participants' strategies, approaches to risk management, and their potential effects on market dynamics.

4. Activity ratio for market participation

Financial participants are active in the market on more trading days

This variable measures the share of days on which the participant was active, relative to the total number of business days. It provides valuable information regarding the participant's consistent involvement and presence in the market over time.

In order to calculate this activity ratio, we first determine the total number of business days for each participant. This is calculated by taking the participant's first transaction recorded in the database and counting the number of business days from that point onward.

We then calculate the number of days on which the entity was active, by dividing the number of days on which the entity made transactions by the total number of business days. This ratio represents active participation in the market by the entity's rate within the given timeframe.

Additional Features

5. Average transaction period

This presents the time period between the transaction time and the agreed payment time. By averaging these repayment periods over all of the participant's transactions in the market, we derive that participant's average transaction period.

6. Forward foreign exchange purchase ratio

This variable presents the ratio of forward forex purchases to the overall number of forward transactions for each entity. This ratio is calculated by dividing the number of forward forex purchases by the total number of forward transactions.

Feature selection

In order to select the most relevant features from the first set of 31 features, we used a recursive feature elimination (RFE) used by Zhang, C., Li, Y., Yu, Z., & Tian, F. (2016) on the training data. The feature selection technique removes features in an iterative manner by building a model on the other features and evaluating their importance based on the model's precision.

For the purpose of our evaluation, we used a random forest model that is known for its ability to efficiently deal with categorical and numeric features. In order to ensure robustness, we carried out a 10-fold cross-validation repeated 5 times to evaluate the model's precision. This enabled us to obtain reliability evaluations of the model's performance.

```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

 Variables Accuracy  Kappa AccuracySD KappaSD Selected
         1   0.8976 0.7844    0.06068  0.1285
         2   0.8863 0.7612    0.06040  0.1261
         3   0.9045 0.7995    0.06122  0.1288
         4   0.9037 0.7980    0.06121  0.1285
         5   0.9107 0.8129    0.06783  0.1429
         6   0.9168 0.8259    0.06496  0.1360
         7   0.9194 0.8311    0.06448  0.1363
         8   0.9220 0.8364    0.06205  0.1309        *
         9   0.9185 0.8293    0.06429  0.1350
        10   0.9220 0.8364    0.06267  0.1322
        11   0.9186 0.8288    0.06305  0.1334
        31   0.9194 0.8308    0.06325  0.1331
```

In addition to the RFE, we used two filtering criteria to sharpen the feature selection process. First, we filtered features based on variance inflation factor (VIF) values, eliminating each feature with a VIF of greater than 5. This stage helps handle problems of multi-colinearity that may be due to highly-correlated predictions.

Second, we weighted the Pearson correlation between predictions and removed feature pairs with a correlation coefficient of greater than 0.7. This further ensures that the selected features are not too excessive or highly correlated with each other.

In conclusion, by using RFE on classification of participants from the business sector into importers and exporters, we identified a subgroup of 8 features that generated the highest average accuracy on the validations folds data through cross-validation within the training set and that met the selection criteria. However, one feature was not included due to a VIF value greater than 5. We thus obtained a final selection of 7 features.

Same procedure applied on classification of Nonresidents to either financial or Non-financial sector, we obtained a final selection of 12 features.

Training, parameter tuning, and model selection

At this stage, we trained and tuned various supervised classification algorithms using the training set, which constituted 70 percent of the available data. The main objective was to identify the best hyperparameters for each model. We obtained this by using 5-folds cross-validation (CV).

The following algorithms were selected for training and tuning: random forest, support vector machines (SVM), logistic regression, K-nearest neighbors (KNN), and linear discriminant analysis (LDA). Using the CV, we methodically evaluated various combinations of hyperparameters for each algorithm in order to determine the optimal form that would generate the highest performances.

After obtaining the best hyperparameters for each model, we continued to evaluate their performances on a separate validation set that included 20 percent of the data. This validation set enabled us to choose the model that demonstrated the most promising results in terms of precision. The aim was to choose the model that showed the best performance on unobserved data.

In order to ensure the stability and inclusion of the chosen model, we subjected it further to the array of tests that included the remaining 10 percent of the data. By evaluating the model's performances in this testing array, we can evaluate its ability to make precise predictions on completely unobserved data. This step provided a decisive index of the model's stability and implementation potential in the real world.

The combination of CV and validation set evaluation ensured that we chose a model that not only demonstrated excellent performances during the testing, but also presented good performances on unobserved data, thereby providing us with certainty of its classification abilities.

Results

The analysis on importers and exporters led to the selection of K-nearest neighbors (KNN) model as the most appropriate for classification within the business sector. The model was chosen because it had the highest precision in the validation set and achieved 0.938 accuracy. In order to validate the model's performance and avoid over-fitting, the accuracy was also on the training set, which produced a similar precision score of 0.935.

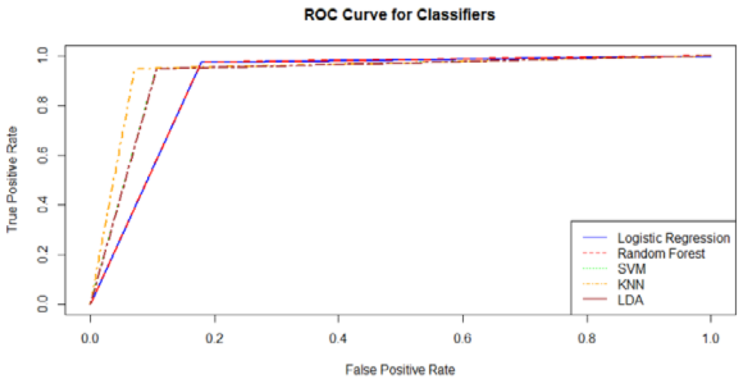
In order to further evaluate model's generalizability, it was tested on the holdout array or on a holdout dataset or test set (10%). The KNN model demonstrated a high level of precision on the test array as well, achieving a precision score of 0.939. and random forest This indicates that the model is stable and has the ability to precisely classify importers and exporters in the business sector in the forex market.

The optimal hyperparameter chosen for the KNN model in the training environment was k=5. This indicates that the model takes into account the five nearest neighbors when making classification decisions. In total, the KNN model proved itself as the most precise and reliable algorithm for classifying forex market participants, while demonstrating its matching to this specific analysis.

Moving on to the classification of nonresidents into financial and non-financial sectors, the Random Forest model was found to be the optimal choice for this classification task. The Random Forest model proved to be the most suitable algorithm for classifying nonresidents into the financial and non-financial sectors within the forex market. With its accuracy of 0.789 on the validation dataset, 0.787 mean accuracy through cross-validation, and 0.785 accuracy on the holdout test set, the model demonstrated its ability to effectively differentiate between the two sectors.

Importers/Exporters Classification model selection results

model	Accuracy in validation dataset	Mean Accuracy in train set
KNN	**0.938**	**0.935**
SVM	0.92	
LDA	0.92	
Random Forest	0.91	
Logistic Regression	0.91	

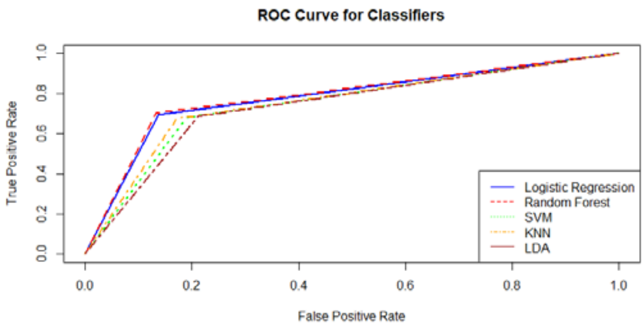


Selected Model

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
KNN	**0.938**	**0.939**	**k=5**

Nonresident – Financials/non-financials Classification model selection results

model	Accuracy in validation dataset	Mean Accuracy in train set (cv)
Random Forest	**0.789**	**0.787**
Logistic Regression	0.770	
KNN	0.754	
SVM	0.748	
LDA	0.739	



Selected Model

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
Random Forest	**0.789**	**0.785**	**Mtry=3,ntrees=500**

Conclusion and future work

In conclusion, the analysis conducted on both importers and exporters, as well as nonresidents classified into financial and non-financial sectors, yielded promising results using the K-nearest neighbors (KNN) model and Random Forest model, respectively. These models demonstrated high accuracy in classifying market participants within their respective subsectors in the forex market.

The utilization of these models for forecasting purposes provides valuable insights and enhances the central bank's ability to monitor and analyze different market segments. By accurately identifying importers, exporters, and distinguishing between financial and non-financial participants among nonresidents, a more comprehensive and detailed picture of the subsector-level market dynamics is obtained.

This deeper understanding of market flows serves as a crucial input for informing monetary policy decisions and developing effective macroprudential tools. The precise classification of market participants aids in assessing the overall health and stability of the forex market, enabling policymakers to make informed choices regarding interest rates, exchange rate management, and other regulatory measures. Additionally, it assists in identifying potential risks and vulnerabilities within specific subsectors, allowing for the implementation of targeted policies to mitigate these risks and ensure financial stability.

Future work in this area could involve expanding the analysis to include other subsectors or market segments, such as specific industries or types of financial institutions. By incorporating a wider range of variables and labeled data, the models' accuracy and predictive power could be further enhanced.

REFERENCES

1. Abir, A. (2020). "Developments in Israel’s foreign exchange market and the implications for the conduct of monetary policy and financial stability". *BIS Papers*, *113*, 163-171.
2. Schreiber, B. Z. (2014). Identifying speculators in the FX market: A microstructure approach. *Journal of Economics and Business*, *73*, 97-119.
3. Kumari, R., & Srivastava, S. K. (2017). Machine learning: A review on binary classification. *International Journal of Computer Applications*, *160*(7).
4. Zhang, C., Li, Y., Yu, Z., & Tian, F. (2016). "Feature selection of power system transient stability assessment based on random forest and recursive feature elimination". In *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)* (pp. 1264-1268). IEEE.

Classification of Participants in the Foreign Exchange Market

Amir Khatib

BoC-IFC seminar

15 July 2023

11

Foreign Exchange Market Turnover

Billion $ per day

The foreign exchange market in Israel plays a crucial role in the transmission of monetary policy, similar to other small open economies:

1. High pass-through of the exchange rate into inflation
2. Impact on tradable sector - Imports and Exports

Market participants, broken down by sector

Israeli residents


Nonresidents


Banking system


Financial sector


Households


Institutional investors


Business sector

Nonresidents:
- Financial
- Non-Financial

Business sector:
- Exporters
- Importers

Description of the database

Foreign exchange transactions– FXOption, Forward, Spot, FXSwap

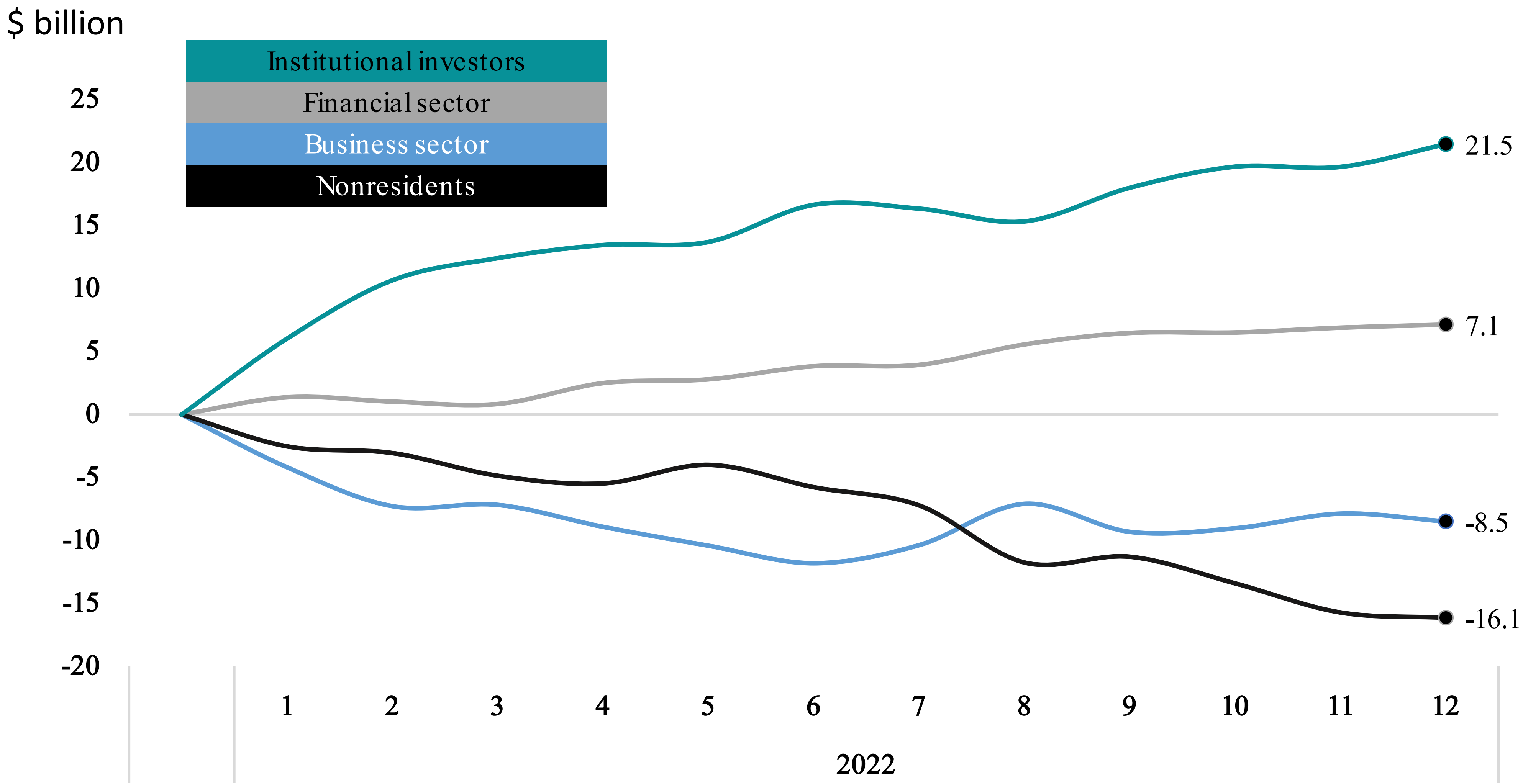
50 financial intermediaries in Israel and abroad

Daily reports

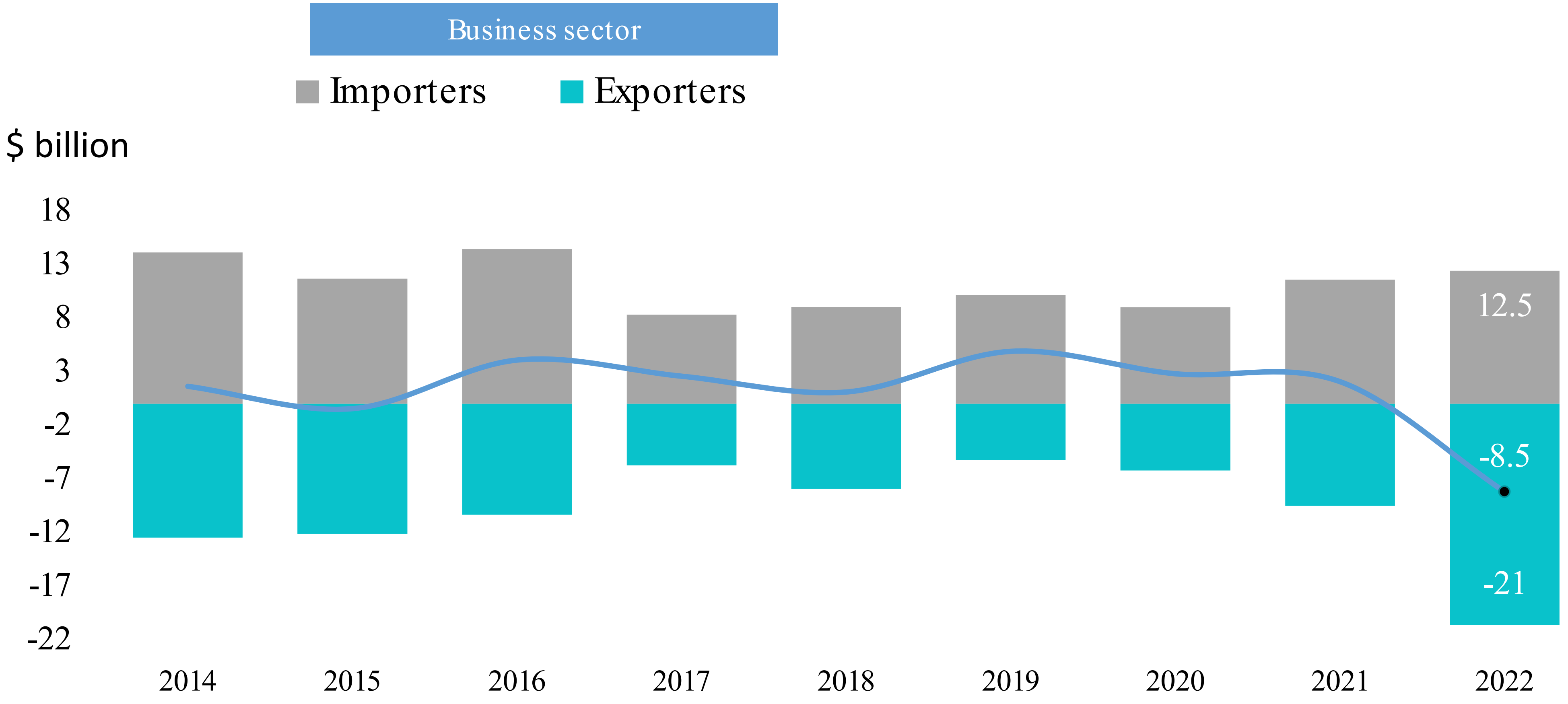
Detailed reporting at the transaction level

4 million transactions per year

Net cumulative foreign exchange purchases, by the main sectors



Net foreign exchange purchases



Problem statement

In certain cases, a market participant is not assigned to a subsector

Nonresidents

? ?

Financial Non-Financial

Israeli Business sector

? ?

Exporters Importers

Transaction date	Participant ID	Product	Currency basis	Notional amount	Buy/Sell NIS	Sector	Subsector
01/04/2018	52000117	Spot	USD/ILS	50,000,000	Buy	Business sector	**N/A**
05/04/2018	ADASFGFD	Forward	USD/ILS	30,000,000	Sell	Nonresident	**N/A**
07/04/2018	BANKESD12	Forward	EUR/ILS	1,000,000	Buy	Nonresident	Financial

The solution: Use of a **Supervised Classification** algorithm in order to classify participants to a subsector

Data preparation

- Relevant features selected from the database

  Product type, currency, financial intermediary, transaction period, sum, buy/sell indicator, and transaction rate.

- One-Hot Encoding of categorical features

cur_othr	cur_sek	cur_chf	cur_jpy	cur_gbp	cur_eur	cur_usd
0	0	0	0	0	1	0
0	0	0	0	0	1	0
0	0	0	0	0	1	0
0	0	0	0	0	1	0

- Group by Market Participant level, transformation and scaling of features

- Feature Engineering

- Partition data into Train, Validation, Test (70%,20%,10%)

Feature engineering

Importers/Exporters

(1) Importers generally engage in net forex purchases, while exporters tend to sell

Variable	Explanation
MED_Net	Median (by days) net forex purchases

(2) Exporters utilize periods in which the shekel is depreciated in order to sell foreign exchange at large scales, as opposed to importers, who buy in response to demand, without regard to the exchange rate

Variable	Explanation
LN_NETMA1	On days in which there is an anomalous depreciation in the shekel (relative to a 1-month moving average), net total forex purchases of the participant divided by total trading volume

The main variables that we added based on our familiarity with the foreign exchange market and conversations with the intermediaries and market participants over the years

Feature engineering

Nonresidents – Financial/Non Financial

(3) Financial participants are active in the market on more trading days

Variable	Explanation
Activity ratio	The share of days on which the participant was active, relative to the total number of business days

(4) Financial participants trading volume is higher than Non-Financial participants trading volume

Variable	Explanation
LN_VOL	The daily average trading volume

The main variables that we added based on our familiarity with the foreign exchange market and conversations with the intermediaries and market participants over the years

Feature selection - Recursive Feature Elimination (RFE)

```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

 Variables Accuracy  Kappa AccuracySD KappaSD Selected
         1   0.8976 0.7844    0.06068  0.1285
         2   0.8863 0.7612    0.06040  0.1261
         3   0.9045 0.7995    0.06122  0.1288
         4   0.9037 0.7980    0.06121  0.1285
         5   0.9107 0.8129    0.06783  0.1429
         6   0.9168 0.8259    0.06496  0.1360
         7   0.9194 0.8311    0.06448  0.1363
         8   0.9220 0.8364    0.06205  0.1309        *
         9   0.9185 0.8293    0.06429  0.1350
        10   0.9220 0.8364    0.06267  0.1322
        11   0.9186 0.8288    0.06305  0.1334
        31   0.9194 0.8308    0.06325  0.1331
```

Importers/Exporters

In addition,
VIF <5 , pearson correlation between predictors <0.7

7 features selected

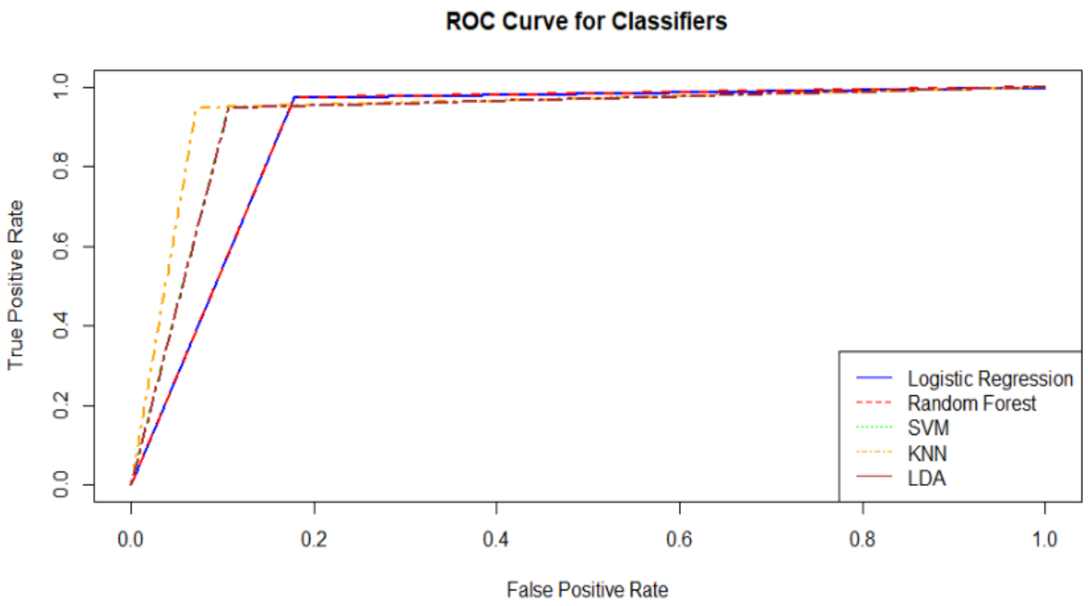
Train various Supervised Classification algorithms on train set (70%) and tune parameters (5-fold CV)

1. Random Forest
2. Support Vector Machines (SVM)
3. Logistic regression
4. K-Nearest Neighbors (KNN)
5. Linear Discriminant Analysis (LDA)

Model selection – on Validation dataset

Importers/Exporters

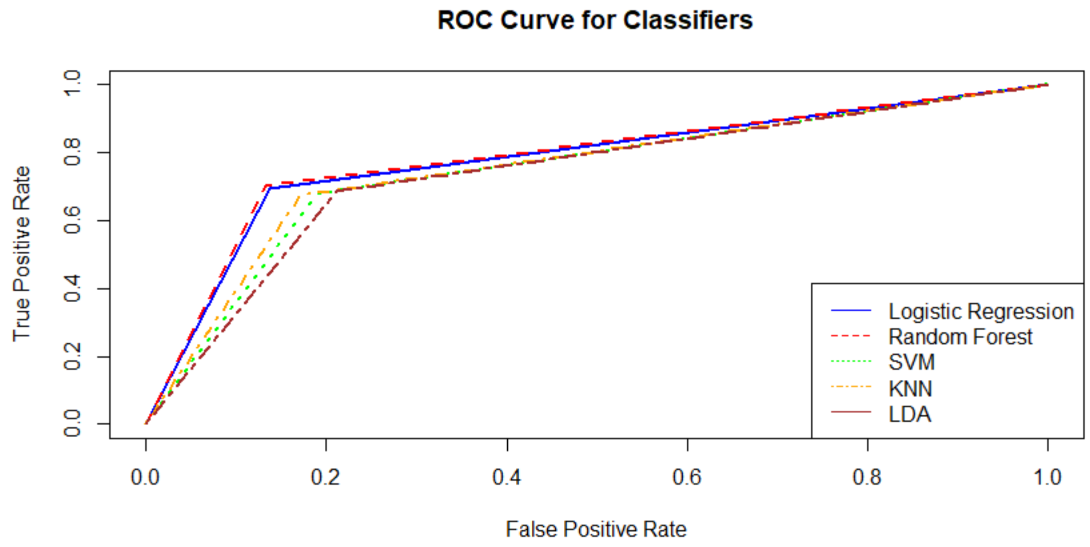
model	Accuracy in validation dataset	Accuracy in train set
KNN	**ě┬Ň38**	**ě┬Ň39**
SVM	ě┬Ň2	
LDA	ě┬Ň2	
Random Forest	0.91	
Logistic Regression	0.91	



Model selection – on Validation dataset

Nonresidents – Financial/Non Financial

model	Accuracy in validation dataset	Mean Accuracy in train set (cv)
Random Forest	**0.789**	**ĕ┬ŃńŃ**
Logistic Regression	0.770	
KNN	ĕ┬ŃĿľ	
SVM	ĕ┬Ń’ń	
LDA	ĕ┬ŃĽŇ	



Selected models

Importers/Exporters

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
KNN	0.938	0.939	k=5

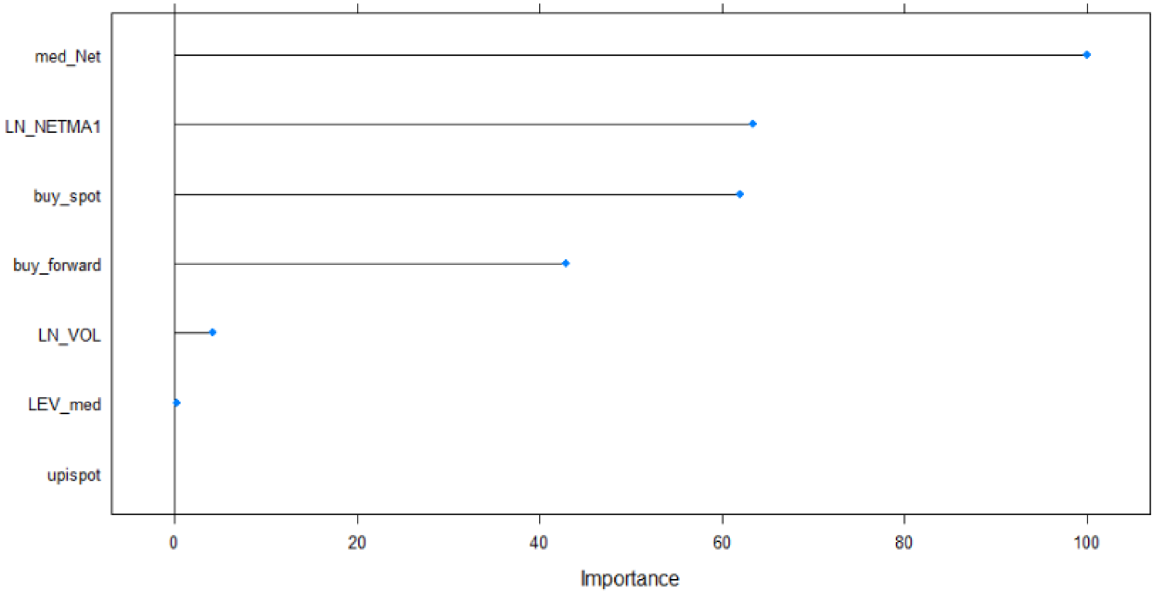
Nonresidents – Financial/Non Financial

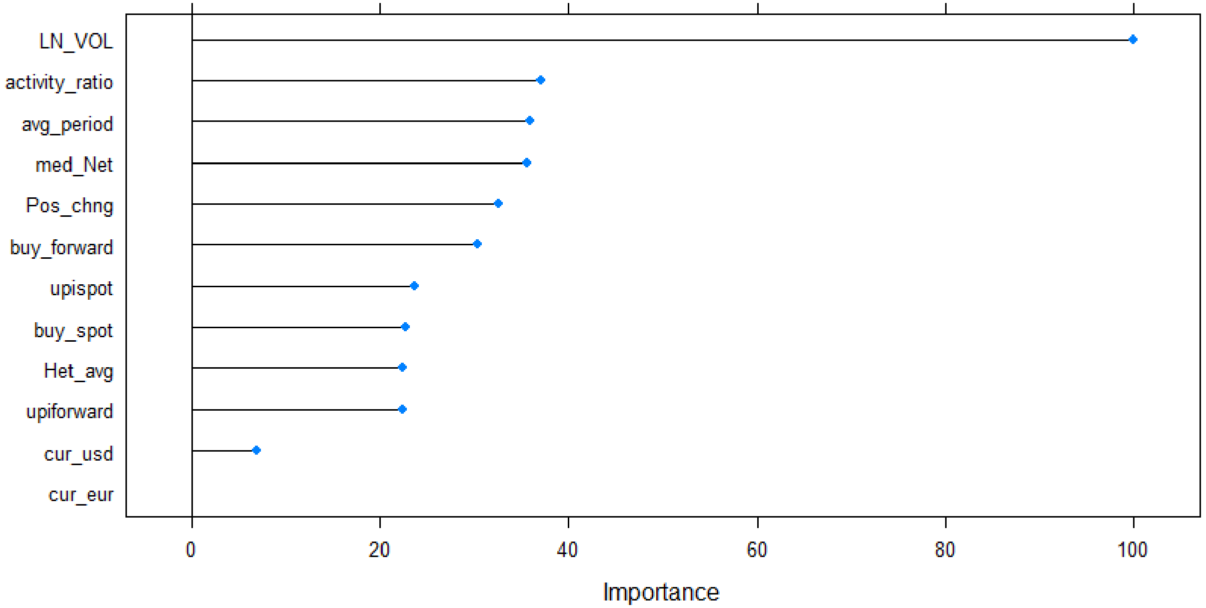
model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
Random Forest	ě┬ŃhŇ	0.785	Mtry=3,ntrees=500

Data interpretation

Importers/Exporters

Nonresidents – Financial/Non Financial





Summary

Israeli forex market significance

Granular data reporting from market participants

Accurate subsector predictions using supervised classification models

Enhanced central bank market analysis capabilities

Thank you for your attention.