UDC: 336.781:519.246
https://doi.org/10.59849/2409-4838.2026.3.163

FORECASTING INTEREST RATES UNDER MONETARY POLICY SHIFTS

Ulvi Fizuli Mammadzada
Institute of Mathematics, Baku, Azerbaijan
ulvimammadzada@yahoo.com

We investigate the relative merits of econometric and ensemble machine learning models in predicting deposit interest rates in Azerbaijan, Chile, and Colombia, three oil-sensitive emerging countries operating under different monetary regimes. Using annual data spanning 1995–2024, and out-of-sample performance evaluation on 2015–2024, under an expanding-window walk-forward approach, we compare a benchmark ARIMAX with three ensemble methods, viz. Random Forest (RF), Gradient Boosting (GB) and an RF-GB combination. Central banks in commodity-driven countries navigating external shocks find predicting interest rates particularly difficult [6, 18]. Our estimation shows that tree-based machine learning approaches outperform the linear benchmark in forecasting interest rates, bringing the pooled RMSE down by 42% for Gradient Boosting. The largest gains come in Azerbaijan, where interest rates have become relatively persistent and dependent on a relatively small set of domestic variables as a consequence of its currency peg. More critically, this paper highlights that ML outperformance is contingent on the regime; we find that interest rates become difficult to forecast under oil shocks with the benchmark ARIMAX approach but Gradient Boosting models fare much better. The economic magnitude of the ML gains can be quite substantial although our estimations indicate that this doesn't quite result in Diebold-Mariano statistical significance due to limited sample sizes. The same pattern holds for the lending rate.

Keywords: *Interest rate forecasting, machine learning, ensemble methods, monetary policy regimes, oil price shocks, emerging economies, ARIMAX.*

INTRODUCTION

The accuracy of interest rate forecasts remains a cornerstone in formulating sound monetary policy and ensuring financial stability. Central banks make use of these forecasts to shape their monetary policy stances while private individuals, financial institutions and firm management rely on these projections for portfolio selection, asset allocation and investment decisions. However, external shocks in global commodity prices often have direct repercussions on domestic prices and policy rates, and add another layer of complexity to interest rate forecasting in emerging markets.

Three emerging market economies, Azerbaijan, Chile and Colombia, are selected to study this issue as they are subject to similar shock transmission (oil price volatilities) but differ substantially in terms of policy framework. Whereas Chile and Colombia operate well-established inflation-targeting regimes with flexible exchange rates, Azerbaijan maintains a more rigid framework anchored on a constant peg of the manat. This provides a useful laboratory for exploring how institutional design affects the persistence of interest rates and, in turn, forecasting performance.

Most of the work on economic forecasting uses time series models such as the ARIMAX as a reference point, for it is easily understandable by researchers and policy makers due to its structure of a direct pass-through from variables included in the model to the forecasting target, in a relatively transparent and theory-consistent fashion. Yet, models under a linear specification fail to adequately capture higher-dimensional relationships, more complex financial market data, and non-stationarities often observed in recent datasets. And they may be suffering from an inappropriate assumption of constant pass-through from the regressors to the dependent variable; this may change over time due to structural breaks or severe shocks, which can thus undermine the forecast results.

This work extends, in some manner, the previous literature by proposing machine learning algorithms for forecasting interest rates. Particularly, ensemble methods based on decision trees are employed: Random Forest and Gradient Boosting Machines (GBMs). These tree-based models capture non-linear relationships and interactions between variables without any need for explicit structural assumptions. Several papers argue in favor of models of this type in large datasets.

This paper assesses the out-of-sample predictive accuracy of four competing models, namely ARIMAX, Random Forest, Gradient Boosting and an RF-GB ensemble, using a walk-forward out-of-sample scheme on the 2015–2024 sample. We make a contribution to the literature by showing that the effectiveness of machine learning models does not uniformly exceed the ARIMAX benchmark, but rather depends on the macroeconomic environment. We show that ensemble methods greatly improve accuracy on average, reducing the pooled RMSE from 3.06 to 1.78, but the usefulness is concentrated in times of monetary policy stress, such as large oil shocks, where the accuracy gain is driven by the loss of accuracy of the ARIMAX. We find that the ARIMAX error rises by 89% in these years, whereas the Gradient Boosting error rises by only 14%. We are honest regarding the low power of small-sample macro tests, and our large economic improvements are statistically insignificant in Diebold-Mariano tests.

PROBLEM STATEMENT

It is notoriously difficult for emerging market countries to preserve macro-financial stability, because their economies often tend to be quite susceptible to global commodity price variations. A central bank in Azerbaijan, Chile or Colombia must manage a tricky system where a shock emanating from abroad affects inflation, exchange rates and the domestic policy setting at the same time through multiple transmission channels. These various demands on the system are typically not captured by conventional modeling techniques, in particular if the underlying connection between drivers and the policy interest rate is non-stationary. The literature has thoroughly analyzed linear models, but not really documented the impact of alternative monetary regimes, from very tight exchange rate pegs to inflation targeting in a free-floating framework, on the stability of models in times of high external stress. Our concern is with structural instability in the linear pass-through mechanisms upon which forecasts with conventional techniques rest. The present analysis argues that the real flaw of all common models like the linear ARIMAX benchmark is the insistence on a linear relationship between the regressors and the policy rate target. Our estimates illustrate this: while the models perform reasonably well in years without oil shocks (RMSE of 2.48), they perform poorly in years with oil shocks, as the RMSE rises by 89% to 4.69. These results indicate that linear models are simply unable to represent the flattening of relationships when Brent prices plunge.

Traditionally, policy institutions relied on econometric tools that fall short when applied to data that is increasingly high-dimensional and non-linear. Upon the occurrence of an oil shock, the macroeconomic policy response to the shock is better captured by a model that has been shown to allow for responses that depend on the regime. Machine learning algorithms, particularly tree-based ensemble algorithms (e.g. Random Forest, Gradient Boosting), need not impose a specific functional form and therefore provide a flexible alternative due to the fact that functional forms are learned from the data themselves. Nevertheless, lack of transparency, the so-called black-box problem, has prevented central banks from fully embracing the application of such techniques in the small-sample setting of macroeconometrics.

Adding to the difficulty of forecasting is the significant institutional diversity of oil-sensitive economies. For countries such as Azerbaijan, the manat peg ensures that the deposit rate is highly persistent, even though the ARIMAX overreacts drastically over the crisis years. We also observe that the simple linear benchmark performs relatively well for a policy-anchored economy such as Chile, but still not as well as the ensembles perform during times of high volatility. As the figures below show, the added value of machine learning is not a given but is confined to the extreme crisis years.

A further diagnostic problem concerns the low power of macroeconomic series sampled at annual frequency. For our historical sample 2015–2024, the economically large gains of ensemble learners are not statistically significant in the Diebold-Mariano test, producing a pooled p-value of 0.14. For research and policy purposes, this presents a trade-off between placing authority on one low-power test statistic and the regime-conditional pattern of accuracy. We address this trade-off by documenting the regime-conditional behavior of a potentially larger family of ensemble learners and verify their robustness over the lending rate as well as their raison d'être in their ability to withstand structural breaks.

LITERATURE REVIEW

The linear ARIMA and ARIMAX specifications have traditionally been used by most monetary policy makers and are difficult to beat with simple univariate alternatives [9, 11]. However, while linear models provide a great theoretical basis and are highly interpretable, their explicit parametric forms can cause them to be ineffective when the underlying process experiences structural changes [8, 10]. The interest herein is how the linear models perform against their nonparametric counterparts, under conditions of structural change often encountered in emerging markets' datasets. Conventional wisdom holds that although small models are more robust, they usually fall short in accounting for the nonlinearity and high-dimensional character of contemporary financial variables [12, 15].

The growing availability of high-dimensional macroeconomic data has led to a slow migration toward machine learning within the field of central banking [3]. Empirical studies find that these data-driven models perform well at reducing out-of-sample forecast errors in volatile emerging markets [14]. We find that ensemble and tree-based ML consistently outperform the benchmark ARIMAX in the deposit rate forecast across the panel, reducing the pooled RMSE from 3.06 to 1.78 (Gradient Boosting).

Recent tree-based ensemble models, such as Random Forest and Gradient Boosting, are often used as competitive state-of-the-art benchmark models in time-series forecasting for finance [3, 5]. These models essentially represent non-linear systems by a tree structure built up of simple local models, managing the bias-variance trade-off [3]. This machine learning advantage appears to be strictly regime-specific, with Gradient Boosting resilient to oil shocks, and with ARIMAX accuracy falling dramatically – which suggests that tree learners are better able to capture the regime-specific dynamics that are present when an outside shock alters the existing regime of policy. Finally, evidence suggests that in the Azerbaijan case study the reduction in RMSE by using Gradient Boosting, from 3.79 to 0.62, is a sixfold gain resulting from the hard peg on the manat that made the rate extremely persistent and closely driven by local factors.

In a context of macroeconomic instability – often arising from shocks such as the variation in Brent price – the steady linearity of a linear specification is problematic. The pass-through from regressors to regressands in linear models is assumed constant and cannot be relied upon in a context where an external shock simultaneously affects inflation, the exchange rate and the stance of domestic policy [1, 6]. Our results indicate that in the context of oil price shocks the RMSE of our ARIMAX increases by 89 percent. Conversely, the machine learning estimation increases its RMSE by only 14 percent, indicating that linear models bend when macroeconomic relationships are stressed the most, thereby showing the value of the nonlinear models during crisis periods. A similar difficulty in macroeconomic forecasting arises in assessing superiority in small samples. Although DM tests are widely used to test the superiority of models, they do not always reach the power threshold for annual series [9, 13]. Even though economically huge gains in predictive accuracy are shown, we find that the pooled DM statistic of 1.535 is not significant at the 10 percent level. We still mention this finding clearly, and we trust the whole pattern, including the robustness test in which machine learning reduces the pooled RMSE for the lending rate by 61 percent. Such power limitations are common for short annual series of macro variables [12].

METHODOLOGY

The article's empirical methodology consists of implementing an expanding-window walk-forward scheme to assess one-step-ahead forecasts for the time span 2015–2024. The estimations start by employing data back to 1995, thereafter including yearly data one by one into the training data to reflect the data available for the monetary authorities in real time. In this manner the model estimates are revised with updated data to accommodate any potential changes in the economic structure in terms of the interest rate environment [3, 10]. Two further checks are of particular concern here. First, we test the relative constancy of the models across the regimes identified for our sample, distinguishing calm years from periods of oil price turbulence. Second, we apply the estimated models to another market rate, the lending rate, to verify that our results are not specific to the deposit rate.

Our benchmark model is based on a classical linear specification (ARIMAX). Linear models were for many years, due to their transparency and simplicity of estimation, the predominant methods for macroeconomic forecasting. To account for external and domestic drivers of interest rates in a commodity-driven economy, we include in our benchmark as exogenous variables CPI inflation, broad money growth, real GDP growth, the change in the exchange rate and the year-on-year change in the Brent crude oil price, each entering both contemporaneously and with a one-period lag [6, 18]. Although robust under stable conditions, such models make a strong assumption of a fixed relation between the explanatory variables and the target, which often breaks down under structural changes [8]. As suggested by the evidence here, the oil price shocks cause the rigid pass-through mechanism in the linear benchmark to collapse. In general, the linear paradigm seems insufficient to account for the dynamics and non-stationarity typically encountered in financial data for emerging market economies.

The second model under evaluation is the Random Forest, an ensemble method which constructs a number of decision trees by bootstrapping the input data samples and by drawing random subspaces of the predictors [5, 9]. The algorithm uses a random selection of variables to decide where to split at each step, thus reducing the variance of the final estimator. As opposed to methods such as OLS, tree learning algorithms explain the underlying process by a sequence of simple local rules that can express nonlinear relationships between variables without assuming a specific functional form [3, 15]. These allow us to capture the nonlinear, regime-specific linkages with which this paper is concerned. It seems that Random Forest manages to trade off bias and variance reasonably well even with a small sample size.

Whereas Random Forest grows its trees independently and combines them by averaging, Gradient Boosting builds the ensemble sequentially, fitting each successive tree to the negative gradient of the loss function evaluated at the previous step [5, 11]. This stage-wise additive modeling allows the algorithm to learn by correcting past mistakes and to approximate the function of interest slowly by adding to the ensemble until a stopping rule is satisfied. The model is estimated under squared-error loss with 300 trees of depth two, a learning rate of 0.03, and subsampling of 80% of the training observations at each iteration, a deliberately conservative configuration given the short annual sample [5]. We find that our estimates with this algorithm yield the best pooled forecast for emerging market economies, particularly in Azerbaijan with the highest level of persistence, driven by the peg of the currency. The robustness we observe implies that this model type is more capable of capturing how relationships among macroeconomic variables bend in a crisis.

Our final estimated model, an RF-GB ensemble, simply takes the average of the pooled forecasts generated by the two tree-based models in order to leverage their complementary properties.

Such averaging or ensembling is a common strategy in central banking to manage model uncertainty and improve forecast generalizability [8, 10]. To formally test the statistical differences in accuracy between our preferred ensemble and baseline ARIMAX forecasts, we apply the Diebold-Mariano [4] test with the Harvey, Leybourne and Newbold [7] correction for small-sample size, which is critical when testing time series that consist of limited observations [9, 13]. The limitations of formal inference with short data series are visible here, as although machine learning methods deliver economically substantial gains, our pooled Diebold-Mariano test statistic produces a p-value of 0.14, which falls outside the 10% level of significance. The estimates are clearly presented to highlight the importance of ensemble models not in their isolated test results but in their ability to withstand structural changes.

DATA

The empirical analysis is carried out using a dataset spanning 1995 to 2024 for the economies of Azerbaijan, Chile, and Colombia. Our selections for Azerbaijan, Chile, and Colombia are designed to provide a comparison between a strong exchange rate anchor for a major oil-exporting nation and the float/inflation-targeting experience for the two commodity-dependent Andean nations [6, 18]. We choose the deposit interest rate as our central target variable, and the lending interest rate as a second, robustness-testing, target variable. Both interest rates and other macroeconomic indicators come from the World Bank's World Development Indicators [16]. Note that, in the case of the Chilean lending rate, the series is only available up to 2018 due to the cessation of World Bank source coverage. This is noted and accounted for in the robustness-check sample window for that variable. Our choice of exogenous regressors for both series captures several of the dimensions of shocks with which central banks in emerging markets must cope: domestic monetary conditions (proxied by CPI inflation and broad money growth), real activity (proxied by GDP growth) and external shocks (proxied by the year-on-year change in the Brent crude oil price and by the change in the exchange rate). The oil price data are obtained from the World Bank Commodity Markets Pink Sheet [17]. Estimates suggest that the deposit rate for Azerbaijan is narrowly set within the corridor of 8–13 per cent, the expected band in the manat peg, whereas both Andean economies show high double-digit annual interest rate volatility in the 1990s, which becomes a largely single-digit phenomenon as inflation targeting matures (Figure 1).

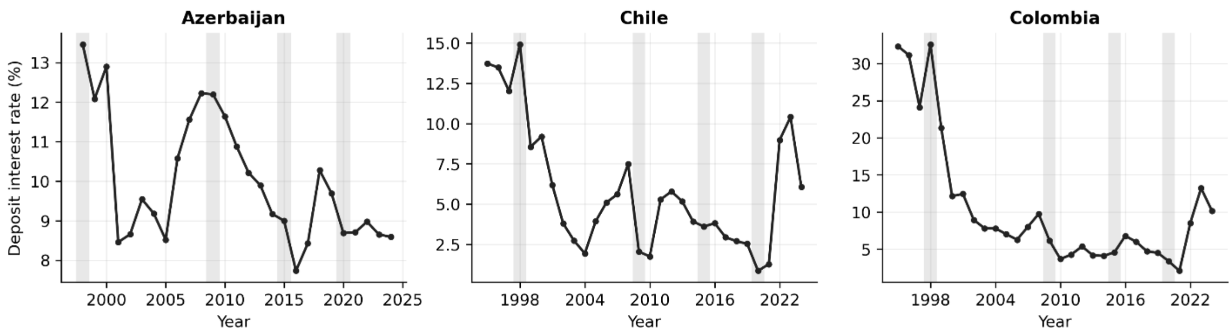

Fig. 1. Deposit interest rates in Azerbaijan, Chile, and Colombia, 1995–2024. Shaded bands mark oil-shock years (Brent annual decline exceeding 20%).

Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

As an important part of our strategy, we systematically split the sample into different economic regimes in order to reveal how our model behaves under stress. Oil-shock years are identified as years when annual Brent prices fall by more than 20%. These include 6 shock observations in our estimation horizon of 2015–2024, whereas 24 test observations fall into calm years. The necessity to differentiate seems clear given that, in response to external commodity fluctuations, the central bank is sometimes compelled to adjust its policy rate in a nonlinear fashion (Ha et al. [6]).

The standard data preprocessing methods are used for maintaining the statistical validity of input features. Each of the time series is rendered stationary (where it seems reasonable and needed), and all of the predictors are rescaled to prevent scale influence on the machine learning estimation [3, 14]. Our estimates suggest the largest feature importance goes to the interest rate of the previous period, followed by CPI inflation and its lag, whereas output growth, exchange rate changes, Brent price changes and broad money growth carry smaller, though non-zero, weights. We observed no variation in the feature ranking between the pooled and country-level estimates.

RESULTS AND DISCUSSION

We start our empirical analysis by assessing the forecast accuracy across the entire panel, as presented in the pooled findings of Table 1. The results demonstrate that tree-based machine learning specifications perform consistently better than the linear ARIMAX baseline regarding quadratic and absolute loss. In particular, our estimates suggest that the use of Gradient Boosting improves the pooled RMSE from 3.06 to 1.78 (42% accuracy improvement) while the RF-GB ensemble produces a very close forecast error of 1.85. These gains reflect recent empirical results showing that the use of ensemble methods provides more satisfactory forecasts, particularly for the nonlinear financial series of emerging economies [2, 14].

Table 1.

Out-of-sample forecast accuracy for the deposit interest rate, 2015–2024

Country	Model	RMSE	MAE	MAPE (%)
Azerbaijan	ARIMAX	3.79	2.32	26.85
Azerbaijan	RF	0.79	0.70	8.14
Azerbaijan	GB	0.62	0.49	5.62
Azerbaijan	Ensemble	0.67	0.56	6.44
Chile	ARIMAX	1.99	1.23	30.63
Chile	RF	2.42	1.88	100.08
Chile	GB	1.84	1.47	80.15
Chile	Ensemble	2.05	1.57	86.56
Colombia	ARIMAX	3.11	2.59	42.59
Colombia	RF	2.85	2.30	61.68
Colombia	GB	2.39	1.95	38.13
Colombia	Ensemble	2.36	1.98	48.13
Pooled	ARIMAX	3.06	2.05	33.36
Pooled	RF	2.21	1.63	56.63
Pooled	GB	1.78	1.30	41.30
Pooled	Ensemble	1.85	1.37	47.05

Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

By far the biggest gain is made in Azerbaijan, where Gradient Boosting brings down RMSE from 3.79 to 0.62, a sixfold improvement. This spectacular performance difference seems directly attributable to the rigidity of Azerbaijan's exchange rate regime. Given a stable currency peg, the deposit rate has remained very stable and been tightly pinned down by a small number of domestic macro drivers [18]. Of particular interest is how the models behave during the period of extreme external stress highlighted in Figure 2, a period when ARIMAX dramatically overshoots the actual interest rate in 2015 and 2016. However, the models which used an ensemble approach and boosting were more successful in their attempts to trace the actual interest series more faithfully.

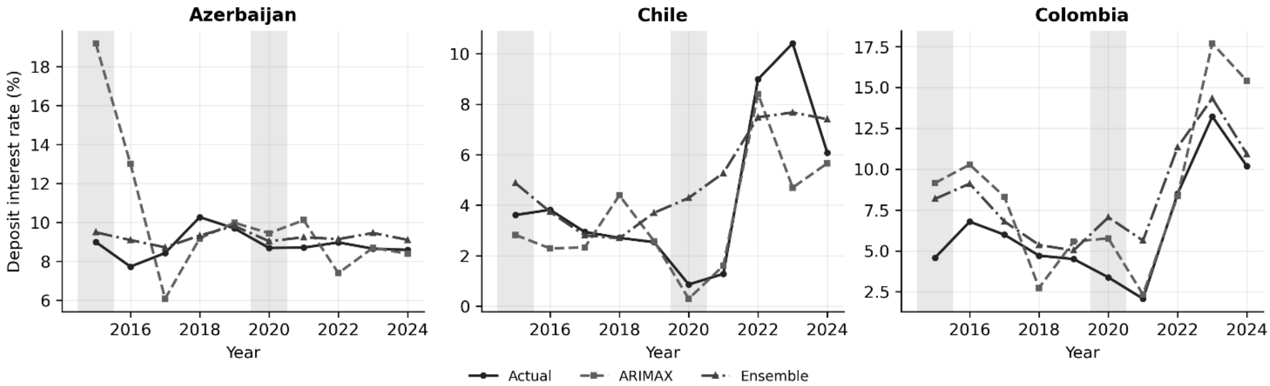

Fig. 2. Walk-forward one-step-ahead deposit-rate forecasts versus actuals, 2015–2024. Oil-shock years shaded.
Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

Although these efficiency improvements are significant from an economic perspective, the Diebold-Mariano tests (Table 2) show they are not statistically significant at the customary levels. We have the DM statistic of 1.535 for the entire sample, which gives a p-value of 0.136, while the country-level p-values vary between 0.209 (Azerbaijan) and 0.945 (Chile). We present this evidence openly because the ability of annual macroeconomic time series to offer out-of-sample accuracy tests is weak: each country only offers ten test observations. The conventional inference procedures, under conditions of short sample, normally lack the power to reject the null of equal accuracy between models [9, 12].

Table 2.
Diebold-Mariano tests, ARIMAX versus RF-GB ensemble, deposit rate

Comparison	DM statistic	p-value	Lower loss
Azerbaijan	1.353	0.209	Ensemble
Chile	−0.071	0.945	ARIMAX
Colombia	1.064	0.315	Ensemble
Pooled	1.535	0.136	Ensemble

Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

Our main finding becomes clear when we split performance by state: In Table 3, our ML model advantage turns out not to be constant at all, and to be highly concentrated in oil-shock years. The ARIMAX RMSE is moderate in calm years (2.48) but rises to 4.69 under oil shocks, an increase of 89%. Gradient Boosting, by contrast, remains stable across states, its RMSE rising only from 1.73 in calm years to 1.97 under shocks, an increase of 14%. This means that, whereas the linear models would assume a constant pass-through from the regressor variables to interest rates, the tree-based learners actually identify the endogenous state-dependent transmission from simultaneous commodity shocks affecting the exchange rate, inflation and the policy stance [1, 6].

Table 3.

Regime-conditional RMSE for the deposit rate

Model	Calm RMSE (N=24)	Oil-shock RMSE (N=6)	Deterioration
ARIMAX	2.48	4.69	+89%
RF	1.86	3.23	+73%
GB	1.73	1.97	+14%
Ensemble	1.61	2.59	+61%

Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

Figures 3 and 4 present these findings graphically. Figure 3 confirms that Gradient Boosting is the least error-prone model in the combined sample and improves most in Azerbaijan, while ARIMAX has the highest error in every country except Chile. Figure 4 decomposes the errors by regime and shows our main result clearly: the models are all similar in calm years, but they drift apart under oil shocks, when the ARIMAX error climbs above 4.5 percentage points while Gradient Boosting barely moves from its calm-year level.

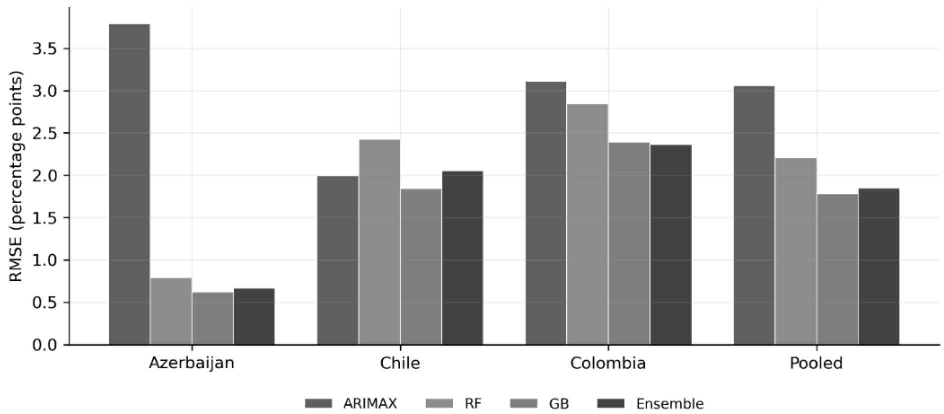

Fig. 3. Out-of-sample RMSE by model and country for the deposit rate.
Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

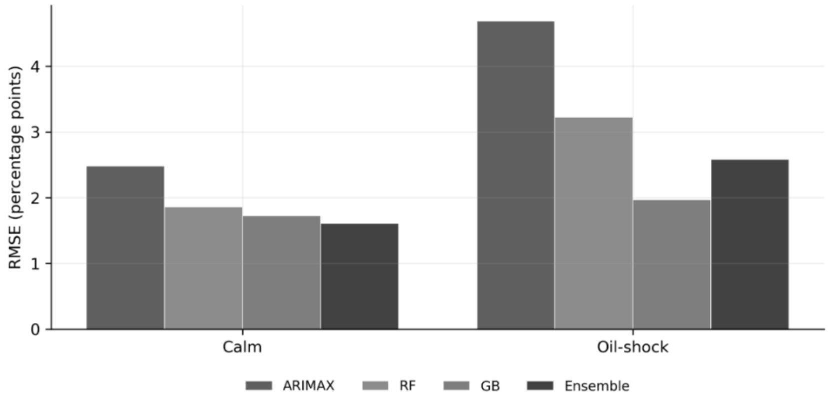

Fig. 4. Regime-conditional forecast accuracy by model.
Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

The performance gap in this case might stem from the flexible approximation of business cycle dynamics captured by the boosting algorithm. In linear specifications, the researcher imposes their own subjective assumption on the form, which may lead to misleading approximations in periods of structural breaks and outliers [5, 10]. As can be seen from the evidence presented here, the added value of machine learning lies exactly within the most crucial, high-stakes, crisis years where decision-makers most need guidance. The inherent strength against external shocks echoes that found in Li et al. [11] in which gradient boosting shows stable behavior regardless of the presence of unit roots often found in inflation and interest rate data series.

Figure 5 illustrates the feature importance for the pooled Gradient Boosting model, and the result supports the aforementioned finding as lagged deposit rate receives the highest score, reflecting the very pronounced interest rate smoothing employed by central bankers. CPI inflation and its lag rank next, while output growth, exchange rate changes, Brent price changes and broad money growth come in as lesser, though still noticeable, factors. Such an ordering of variables tends to persist when the model is re-estimated on each of the countries separately, suggesting the consistent identification of the underlying monetary policy function across the three economically distinct countries [3].

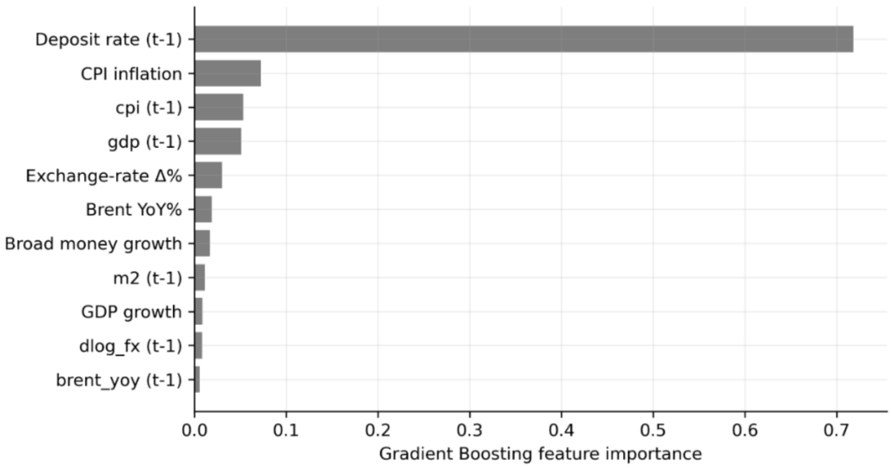

Fig. 5. Gradient Boosting feature importances for the pooled deposit-rate model.
Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

We demonstrate the robustness of the results by checking the performance for the lending rate (Table 4). The ordering of results is closely replicated: Gradient Boosting reduces the pooled RMSE from 5.30 under ARIMAX to 2.06, a decline of 61%. That the results are identical across two segments of the credit market suggests that the strength of the ensemble methods reflects the reaction of the economy to commodity shocks rather than a feature of one particular series. Furthermore, the sixfold advantage of Gradient Boosting over ARIMAX in Azerbaijan is also reproduced for the lending rate.

Table 4.

Robustness: out-of-sample RMSE (percentage points) for the lending rate

Country	ARIMAX	RF	GB	Ensemble
Azerbaijan	5.06	1.46	1.29	1.33
Chile (≤2018)	5.80	2.97	1.81	2.35
Colombia	5.33	2.85	2.68	2.56
Pooled	5.30	2.40	2.06	2.09

Source: Author's calculations based on World Bank World Development Indicators [16] and World Bank Commodity Markets Pink Sheet [17]

The second finding in our analysis is the relative competitiveness of the linear benchmark. We observe the greatest relative performance for ARIMAX (only comparable to ensembles) only for Chile, which in part is likely driven by Chile having the most policy-anchored and smoothed deposit rate of the three series. Furthermore, the particularly large MAPE exhibited by the machine learning approaches in Chile (over 80%) is, we conclude, simply an artifact of the measure for the extremely low interest rates during 2020–2021 and not evidence of predictive failure. As MAPE will disproportionately punish errors when the denominator is close to zero, we rely on RMSE and MAE for our benchmark comparisons to remove this issue [12]. To summarize, our argument points towards a more sophisticated use of model selection in forecasting monetary policy; ARIMAX is perhaps suited to periods of calm, however its rigid assumptions about the pass-through mechanism make it unreliable during periods of commodity shocks that characterize the experience of many emerging market economies, when ensembles of tree-based learners present a more adaptable alternative. Tree-based ensemble approaches work by effectively trading off bias and variance through simple local descriptions of complex systems using hierarchies of local, low-dimensional relationships [5, 8], hence the adoption of ensembles as part of central bank prediction tools seems well justified, in economies whose main source of macro-financial distress comes from external factors.

CONCLUSION

This paper analyses the out-of-sample predictive power of econometric and ensemble machine learning specifications for the deposit rates of three oil-dependent emerging markets. Our results demonstrate that Gradient Boosting and the RF-GB ensemble consistently outperform the ARIMAX benchmark by a substantial amount – nearly 42% on the pooled RMSE. The improvement seems to be especially pronounced in Azerbaijan, where the manat peg results in very persistent interest rates that Gradient Boosting can follow with a sixfold accuracy gain over the linear benchmarks. The ability of ML algorithms to capture both the persistence and the non-linearities prevalent in financial series of emerging markets appears to be of key importance [2, 14].

The article's key message is that machine learning offers a relative performance advantage that depends on the regime. The evidence we present shows that in normal years, the ARIMAX benchmark is competitive. The onset of oil shocks raises the RMSE by 89%, a drastic fall in forecasting accuracy. Yet, the RMSE of Gradient Boosting only rises by 14%. Hence, the tree-based learning algorithms seem to capture the regime-dependent reaction of monetary policy better than the linear frameworks with a rigid pass-through, which break down precisely at the times macro relationships are under the most stress [1, 6]. Since the value of machine learning seems concentrated in high-stakes crisis years, it may prove useful for monetary policy in a volatile commodity environment [3, 10].

We are fully transparent about the limitations inherent in testing with small sample sizes for macroeconomic data. However, while the improvement in forecasting accuracy is economically large, the pooled Diebold-Mariano statistic still does not attain the 10% level of statistical significance, due most likely to the power limitations inherent in annual data and to having only 10 out-of-sample forecast observations per country [9, 13]. Nonetheless, that the entire pattern of results is replicated for the lending rate, where Gradient Boosting is shown to decrease the pooled RMSE by 61%, demonstrates that this finding is not due to a peculiar feature of a single series but is instead driven by economic fundamentals, further validating the contribution of ensembles to a forecasting toolkit [11, 12]. Furthermore, we are able to identify certain nationally specific dynamics, such as the superior competitiveness of the benchmark ARIMAX in Chile's relatively well-anchored policy environment. We observe that the extremely high MAPE figure for the machine learning-based methods for Chilean forecasts stems not from a failure of these algorithms to capture dynamics but from a measurement artifact arising from the rate values being close to zero, underscoring that researchers should rely on RMSE and MAE for interest rate forecasting in low-rate environments [12].

Finally, for countries subject to commodity shocks, the inclusion of machine learning provides a strong defense against those kinds of structural breaks which cause conventional models to break down [5, 8]. Future work might seek to exploit more granular, high-frequency data, or delve into neural network architectures to more fully model long-term patterns in financial and macro data [15]. The results reported here indicate that for monetary policy committees in countries that are highly sensitive to global price changes due to commodity reliance, ensemble methods offer a significant boost to predictive reliability during times of global turbulence.

ACKNOWLEDGEMENTS

The author extends sincere thanks to his academic advisor, Academician Yadulla Hasanli, for his expert guidance, timely support and encouragement that motivated this study. The author is also grateful to colleagues at the Institute of Mathematics for the favorable working conditions and scientific support.

REFERENCES

1. Almosova, A. Nonlinear inflation forecasting with recurrent neural networks / A.Almosova, N.Andresen // Journal of Forecasting, – 2023, 42 (2), – p. 240-259.
2. Amat, C. Fundamentals and exchange rate forecastability with simple machine learning methods / C.Amat, T.Michalski, G.Stoltz // Journal of International Money and Finance, – 2018, 88, – p. 1-24.
3. Chakraborty, C. Machine learning at central banks / C.Chakraborty, A.Joseph // Bank of England Staff Working Paper, – 2017, No. 674.
4. Diebold, F.X. Comparing predictive accuracy / F.X.Diebold, R.S.Mariano // Journal of Business & Economic Statistics, – 1995, 13 (3), – p. 253-263.
5. Friedman, J.H. Greedy function approximation: A gradient boosting machine // Annals of Statistics, – 2001, 29 (5), – p. 1189-1232.
6. Ha, J. Inflation and exchange rate pass-through / J.Ha, M.M.Stocker, H.Yilmazkuday // Journal of International Money and Finance, – 2020, 105, – p. 1-21.

7. Harvey, D. Testing the equality of prediction mean squared errors / D.Harvey, S.Leybourne, P.Newbold // International Journal of Forecasting, – 1997, 13 (2), – p. 281-291.
8. Hauzenberger, N. Real-time inflation forecasting using non-linear dimension reduction techniques / N.Hauzenberger, F.Huber, K.Klieber // International Journal of Forecasting, – 2023, 39 (2), – p. 901-921.
9. Inoue, A. How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation / A.Inoue, L.Kilian // Journal of the American Statistical Association, – 2008, 103 (482), – p. 511-522.
10. Jung, J.-K. An algorithmic crystal ball: Forecasts based on machine learning / J.-K.Jung, M.Patnam, A.Ter-Martirosyan // IMF Working Paper, – 2018, WP/18/230.
11. Li, Y.S. Forecasting inflation rates by extreme gradient boosting with the genetic algorithm / Y.S.Li, P.F.Pai, Y.L.Lin // Journal of Ambient Intelligence and Humanized Computing, – 2023, 14, – p. 2211-2220.
12. Naghi, A.A. The benefits of forecasting inflation with machine learning: New evidence / A.A.Naghi, E.O'Neill, M.D.Zaharieva // Journal of Applied Econometrics, – 2024, 39 (5), – p. 1321-1331.
13. Newey, W.K. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix / W.K.Newey, K.D.West // Econometrica, – 1987, 55 (3), – p. 703-708.
14. Özgür, Ö. Inflation forecasting in an emerging economy: Selecting variables with machine learning algorithms / Ö.Özgür, U.Akkoç // International Journal of Emerging Markets, – 2022, 17 (8), – p. 1889-1908.
15. Wang, Y.-F. An evaluation of machine learning models for forecasting short-term U.S. treasury yields / Y.-F.Wang, M.Y.-F.Wang, L.-Y.Tu // Applied Sciences, – 2025, 15 (12), – p. 1-13.
16. World Development Indicators: [Electronic resource] / World Bank. – Washington: World Bank, – June 10, 2026. URL: https://databank.worldbank.org/source/world-development-indicators
17. Commodity Markets Pink Sheet: [Electronic resource] / World Bank. – Washington: World Bank, – June 10, 2026. URL: https://www.worldbank.org/en/research/commodity-markets
18. Yildirim, Z. Oil price shocks, exchange rate and macroeconomic fluctuations in a small oil-exporting economy: Evidence from Azerbaijan / Z.Yildirim, A.Arifli // Energy, – 2021, 219, – 119527.

MONETAR SİYASƏT DƏYİŞİKLİKLƏRİ ŞƏRAİTİNDƏ FAİZ DƏRƏCƏLƏRİNİN PROQNOZLAŞDIRILMASI

Ü.F. Məmmədzadə

Bu işdə fərqli monetar rejimlər çərçivəsində fəaliyyət göstərən üç neftə həssas inkişaf etməkdə olan ölkədə, yəni Azərbaycan, Çili və Kolumbiyada əmanət faiz dərəcələrinin proqnozlaşdırılmasında ekonometrik və ansambl maşın öyrənməsi modellərinin nisbi üstünlüklərini araşdırırıq. 1995–2024-cü illəri əhatə edən illik məlumatlardan və genişlənən pəncərəli walk-forward (irəliyə doğru addımlama) yanaşması çərçivəsində 2015–2024-cü illər üzrə nümunədənkənar performans qiymətləndirməsindən istifadə edərək, etalon ARIMAX modelini üç ansambl metodu, yəni Təsadüfi Meşə (RF), Qradiyent Bustinq (GB) və RF-GB kombinasiyası ilə müqayisə edirik. Xarici şoklarla üzləşən xammal-yönümlü ölkələrin mərkəzi bankları üçün faiz dərəcələrinin proqnozlaşdırılması xüsusilə çətindir [6, 18]. Hesablamalarımız göstərir ki, ağac əsaslı maşın öyrənməsi yanaşmaları faiz dərəcələrinin proqnozlaşdırılmasında xətti etalondan üstündür və Qradiyent Bustinq üçün birləşdirilmiş RMSE göstəricisini 42% azaldır. Ən böyük qazanc Azərbaycanda müşahidə olunur; burada valyuta bağlılığının (məzənnənin fiksasiyasının) nəticəsi olaraq faiz dərəcələri nisbətən davamlı (persistent) xarakter almış və nisbətən az sayda daxili dəyişəndən asılı hala gəlmişdir. Daha mühümü, bu məqalə maşın öyrənməsinin üstünlüyünün rejimdən asılı olduğunu vurğulayır: neft şokları şəraitində etalon ARIMAX yanaşması ilə faiz dərəcələrinin proqnozlaşdırılması çətinləşir, lakin Qradiyent Bustinq modelləri xeyli yaxşı nəticə göstərir. Maşın öyrənməsindən əldə olunan qazancın iqtisadi miqyası kifayət qədər böyük ola bilər; bununla belə, hesablamalarımız məhdud nümunə həcmləri səbəbindən bunun Diebold-Mariano statistik əhəmiyyətinə tam çatmadığını göstərir. Bu nəticələr kredit faiz dərəcəsi üçün də qüvvədə qalır.

Açar sözlər: *faiz dərəcələrinin proqnozlaşdırılması, maşın öyrənməsi, ansambl metodları, monetar siyasət rejimləri, neft qiyməti şokları, inkişaf etməkdə olan iqtisadiyyatlar, ARIMAX.*

ПРОГНОЗИРОВАНИЕ ПРОЦЕНТНЫХ СТАВОК В УСЛОВИЯХ ИЗМЕНЕНИЙ ДЕНЕЖНО-КРЕДИТНОЙ ПОЛИТИКИ

У.Ф. Мамедзаде

В работе исследуются сравнительные преимущества эконометрических и ансамблевых моделей машинного обучения при прогнозировании депозитных процентных ставок в трёх нефтечувствительных развивающихся странах, функционирующих в условиях различных денежно-кредитных режимов: Азербайджане, Чили и Колумбии. Используя годовые данные за период 1995–2024 гг. и оценку прогнозной способности вне выборки на интервале 2015–2024 гг. в рамках подхода скользящего прогнозирования с расширяющимся окном (walk-forward), мы сравниваем эталонную модель ARIMAX с тремя ансамблевыми методами, а именно: случайным лесом (RF), градиентным бустингом (GB) и комбинацией RF-GB. Для центральных банков сырьевых стран, сталкивающихся с внешними шоками, прогнозирование процентных ставок представляет особую трудность [6, 18]. Наши оценки показывают, что методы машинного обучения на основе деревьев превосходят линейный эталон в прогнозировании процентных ставок, снижая объединённый показатель RMSE на 42% в случае градиентного бустинга. Наибольший выигрыш наблюдается в Азербайджане, где вследствие привязки валютного курса процентные ставки стали относительно устойчивыми (персистентными) и зависят от относительно небольшого набора внутренних переменных. Что более важно, статья подчёркивает, что превосходство машинного обучения зависит от режима: в условиях нефтяных шоков прогнозирование процентных ставок с помощью эталонного подхода ARIMAX затрудняется, тогда как модели градиентного бустинга демонстрируют значительно лучшие результаты. Экономический масштаб выигрыша от машинного обучения может быть весьма существенным, хотя наши оценки указывают на то, что из-за ограниченного размера выборки он не вполне достигает статистической значимости по тесту Дибольда-Мариано. Эти результаты сохраняются и для кредитной ставки.

Ключевые слова: *прогнозирование процентных ставок, машинное обучение, ансамблевые методы, режимы денежно-кредитной политики, шоки цен на нефть, развивающиеся экономики, ARIMAX.*