

Forecasting Nigerian Foreign Direct Investment One Quarter Ahead: A Leakage-Free Comparison of Six Regression Models

Ofieroher Ufuoma Earnest

ORCID: <https://orcid.org/0009-0002-7468-8082>

Department of Finance and Banking, Faculty of Management Sciences, University of Port Harcourt, Nigeria

Abstract: This study compares six supervised regression models for forecasting Nigerian foreign direct investment one quarter ahead, using 67 quarterly observations from 2008Q2 to 2024Q4. Two feature sets are tested: a compact core set of six contemporaneous predictors, and an enhanced set adding lags, four-quarter rolling means and percentage changes, all shifted so that no future information enters. Both sets are scored on an identical evaluation window, which matters because an earlier version of this work compared them on different windows and reached the opposite conclusion. Models are tuned with four-fold expanding-window cross-validation and evaluated by expanding-window walk-forward prediction. Gradient boosting on the enhanced feature set performs best, with MAE 385.59, RMSE 523.37 and out-of-sample R-squared of 0.0257. It is the only one of twelve model-specification pairs to beat a holdout-mean benchmark. Naive persistence returns RMSE 691.51 under both feature sets, confirming that the two are being scored on the same observations. Enhanced features improve on core features in three of five model classes. The main finding is negative and, we think, useful: at this sample size and horizon, quarterly Nigerian FDI is close to unforecastable, and eleven of twelve approaches do worse than predicting the mean.

Keywords: Foreign Direct Investment; Time-Series Forecasting; Feature Engineering; Walk-Forward Validation; Nigeria.

JEL Classification: C53; C45; F21; O55.

1. Introduction

Central banks, finance ministries and investment-promotion agencies all plan around expected capital flows, so a forecast of FDI has obvious uses. Econometric models explain historical associations well, but a good in-sample fit says little about out-of-sample accuracy (Shmueli, 2010). Machine-learning regression can capture non-linear interactions that linear models assume away, and has done well on macroeconomic and financial series elsewhere (Goulet Coulombe et al., 2022; Medeiros et al., 2021; Gu, Kelly & Xiu, 2020).

Applying these methods to a short, volatile emerging-market series is harder than the general literature suggests. Two risks dominate. The first is leakage: random train-test splitting, unshifted rolling features, or scalars fitted before splitting all produce accuracy that cannot be achieved in real time (Bergmeir & Benítez, 2012). The second is comparison error. If two feature sets are scored on different observations, any difference between them mixes model performance with the difference in test windows.

This paper addresses both. Every engineered feature is shifted at construction, and both feature sets are evaluated on one common window with a naive benchmark retained throughout.

1.1 What this paper contributes

- It builds a leakage-free forecasting pipeline for Nigerian FDI and states the validation protocol in enough detail to be reproduced.
- It compares a compact and an enriched feature set on one common evaluation window, and uses the naive benchmark as a check that the window really is common.
- It documents and removes a data problem, interpolation before 2008, that inflates measured persistence and flatters any model relying on it.

- It reports a negative result plainly. At this sample size the series resists forecasting, and most approaches lose to the mean.

2. LITERATURE REVIEW

2.1 Machine learning in macroeconomic forecasting

The case for flexible learners is that they approximate non-linear structure linear models impose away. Goulet Coulombe et al. (2022) find the gains are real but conditional, showing up mainly at longer horizons and in data-rich settings. Medeiros et al. (2021) report substantial improvements for inflation forecasting, and Gu, Kelly and Xiu (2020) find similar gains in asset pricing. All three environments involve many predictors, long samples, or both.

The countervailing evidence is just as clear. The M4 and M5 competitions found simple statistical methods competitive with, and often better than, complex alternatives on short univariate series (Makridakis et al., 2018, 2020). The mechanism is variance, not bias. Flexible learners estimate more parameters from the same information, and when the signal-to-noise ratio is low the extra estimation variance swamps any reduction in approximation error. Sixty-seven quarterly observations is close to the worst case for that trade-off.

2.2 Validation and leakage

Standard k-fold cross-validation assumes exchangeable observations. Applied to time series it lets the model train on the future of its own test fold (Bergmeir & Benítez, 2012). Bergmeir, Hyndman and Koo (2018) show ordinary cross-validation can be valid for purely autoregressive models with uncorrelated errors, but that qualification does not cover the mixed-predictor, rolling-feature setting used here. Tashman (2000) sets out the rolling-origin design this study adopts.

Leakage usually enters through the feature construction, not the split. Rolling statistics computed without shifting, scalars fitted on the full sample, and target-derived features built across the whole series are all invisible in reported results and all inflate accuracy.

3. DATA AND METHOD

3.1 Sample

The dataset behind this paper was originally assembled as 144 quarterly observations running from 1990Q1. Inspection showed that everything before 2008 was linearly interpolated from annual figures: FDI, GDP, the exchange rate, inflation and the trade series all change by a constant amount each quarter and then jump at the year boundary. Interpolation of that kind imposes autocorrelation by construction, so the earlier period has been dropped. Trade openness was also excluded, because the available series divides nominal trade by real GDP and tests as I(2). GDP is rebased in 2025 with an unadjusted splice, so 2025 is dropped as well. The working sample is 2008Q2 to 2024Q4.

3.2 Feature sets

The target is FDI one quarter ahead, in US\$ million. The core set uses six predictors at quarter t : current FDI, exchange-rate volatility, inflation, GDP growth, oil price and the interest rate. The enhanced set adds one- and four-quarter lags of FDI, the exchange rate and oil price, four-quarter rolling means of the same, the rolling standard deviation of exchange-rate volatility, and the percentage change in FDI. All rolling and lagged features are shifted so each quarter- t row uses only information available at $t-1$, and the shifting happens before any split.

3.3 Evaluation

Both feature sets are restricted to the rows where every enhanced feature is available, giving a common window of 62 observations from 2009Q2 to 2024Q3. The first 80% (49 observations) is used for training and tuning; the last 13 quarters, 2021Q3 to 2024Q3, are held out. Hyperparameters are tuned by four-fold expanding-window cross-validation, with scaling fitted separately inside each training fold. Final evaluation refits each model at every step on all data up to that point and predicts only the next quarter.

Restricting both sets to a common window is the fix for a specific error. In an earlier version of this analysis the enhanced set lost extra leading observations to its rolling features, which moved the 80/20 boundary. Naive persistence, which ignores features entirely, then returned two different numbers for the two specifications, and the core-versus-enhanced comparison was not measuring what it claimed to measure.

4. RESULTS

Table 1. Walk-forward performance, common window 2009Q2–2024Q3, test period 2021Q3–2024Q3 (n = 13)

Feature set	Model	MAE	RMSE	R ²
Enhanced	Gradient Boosting	385.59	523.37	0.0257
Enhanced	Random Forest	483.88	597.13	-0.2683
Enhanced	SVR	530.41	639.04	-0.4526
Enhanced	Ridge	594.65	640.67	-0.4600
Core	SVR	514.16	653.34	-0.5183
Core	Random Forest	541.50	659.26	-0.5460
Core	Naive persistence	534.54	691.51	-0.7009
Enhanced	Naive persistence	534.54	691.51	-0.7009
Core	Ridge	618.74	731.92	-0.9055
Core	Gradient Boosting	648.20	791.47	-1.2282
Core	Linear	673.62	806.43	-1.3132
Enhanced	Linear	756.39	893.27	-1.8383

The two naive-persistence rows are identical at 691.51, which is the check that both feature sets were scored on the same observations. Only enhanced gradient boosting beats the holdout mean.

5. DISCUSSION

Enhanced gradient boosting is the only model of twelve with a positive out-of-sample R-squared, at 0.0257. That is a small number and should be described as such: the model explains about 2.6% of holdout variance more than predicting the mean would. Eleven of the twelve pairs do worse than the mean, several of them badly.

The honest reading is that one-quarter-ahead Nigerian FDI is close to unforecastable from these predictors at this sample size. The test window contains 13 quarters spanning the COVID recovery and the 2023 exchange-rate unification, and FDI moves from 1,561 to negative values within it. No model in this comparison anticipates those turns.

Enhanced features beat core features in three of five model classes, and the best model overall uses them. This reverses the earlier version of this analysis, which concluded that engineered features added complexity without signal. That conclusion came from comparing the two sets on different test windows. The correction matters beyond this paper: a feature-set comparison is only informative if the naive benchmark returns the same number under both sets, and reporting that benchmark is a cheap way to demonstrate it.

Gradient boosting appears at both ends of the table, best on enhanced features and third from bottom on core. Model-class rankings in short samples are specification-dependent, and any claim that a particular algorithm suits a particular macroeconomic problem needs more evidence than one comparison can supply.

6. CONCLUSIONS

Six models were compared under a leakage-free protocol with both feature sets scored on identical observations. Enhanced gradient boosting performed best and was the only approach to beat a mean benchmark. Eleven of twelve did worse than the mean. Anyone building a near-term FDI monitoring tool should treat these forecasts as weak signals, not planning inputs, and should report a naive benchmark alongside any accuracy figure so that readers can judge whether a model is adding anything.

7. LIMITATIONS

The test set is 13 quarters. Accuracy estimates from a holdout that short carry wide sampling error, and the ranking in Table 1 should not be read as settled.

No formal test of predictive accuracy differences is reported. A Diebold-Mariano test on the paired errors (Diebold & Mariano, 1995) would establish whether the gaps in Table 1 exceed chance, and should be added.

Only the one-quarter horizon is examined. Machine-learning gains are typically larger at longer horizons, so this may understate what these methods can do.

Forecasts are point predictions. For policy use, intervals matter more, and conformal prediction would supply them cheaply.

Trade openness is missing from the predictor set for the data reason given in Section 3.1, and it belongs there on theoretical grounds.

REFERENCES

- [1]. Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- [2]. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83.
- [3]. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [4]. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- [5]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
- [6]. Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5), 920–964. <https://doi.org/10.1002/jae.2910>
- [7]. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
- [8]. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
- [9]. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.
- [10]. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition. *International Journal of Forecasting*, 36(1), 54–74.
- [11]. Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment. *Journal of Business & Economic Statistics*, 39(1), 98–119.
- [12]. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [13]. Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310.
- [14]. Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy. *International Journal of Forecasting*, 16(4), 437–450.