



Beyond the Aggregate: Provenance, Communication and Scope in Machine-Learning FDI Forecasting for Policy

Ofieroher Ufuoma Earnest

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

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

Corresponding Author

Abstract: Machine-learning forecasts of national-aggregate FDI are increasingly offered to policymakers as decision support. This paper identifies three risks that accompany such forecasts and are rarely stated alongside the accuracy figures, and grounds each in a documented case. The first is provenance. A forecast assembled from central bank, statistical agency and multilateral sources inherits their definitional choices and revision practices, and downstream users usually cannot trace which source supplied which observation. In the case examined here, half the assembled series was linearly interpolated from annual figures and two further construction defects were present, none of which was documented or detectable from the published output. The second is communication. An out-of-sample R-squared says nothing about which quarters a model handles well; the best model in the companion forecasting study achieved 0.0257 and was the only one of twelve to beat a mean benchmark, a context that a headline figure omits. The third is aggregation. National-aggregate FDI in Nigeria is dominated by a small number of large transactions, so a forecast validated at that level supports no inference about sectoral or regional dynamics. We propose three safeguards: a provenance table including construction method, mandatory disclosure of failure modes alongside any accuracy figure, and explicit scope limits restricting aggregate forecasts to aggregate questions.

Keywords: Data Governance; Model Documentation; Forecast Communication; Policy Use; Nigeria.

JEL Classification: O38; C53; K24; F21.

1. INTRODUCTION

The technical literature on macroeconomic forecasting with machine learning has developed quickly, with growing attention to leakage-free validation and explainability. Much less has been written about what happens after a validated forecast leaves the pipeline and reaches a non-technical audience: which assumptions travel silently with the number, what the number cannot say, and whether the level at which a model was validated matches the level at which someone intends to use it.

This paper treats those questions as the object of study, using a documented Nigerian FDI forecasting exercise as the case. The question is not whether that pipeline was technically sound. It is what should govern the use of its output.

1.1 What this paper contributes

- It separates three distinct risks rather than treating ethics as one undifferentiated concern.
- It grounds the provenance argument in a case where three undocumented construction defects survived a formal validation protocol and reversed the companion study's conclusion.
- It proposes safeguards specific enough to be implemented, not general calls for responsible practice.
- It is written for a policy audience and turns a modelling exercise into institutional guidance.

2. LITERATURE REVIEW

2.1 Documentation frameworks

The FAIR principles (Wilkinson et al., 2016) set criteria for research data. Gebru et al. (2021) propose datasheets recording provenance, composition and known limitations. Mitchell et al. (2019) propose model cards recording intended use, performance across relevant conditions, and known failure modes. Both frameworks were built for machine-learning artefacts and transfer directly to forecasting systems assembled from multiple institutional sources, where the composition question is rarely answerable from a published paper.



2.2 Communication and over-reading

Rudin (2019) argues against post-hoc explanation of black-box models for high-stakes decisions. Floridi and Cowls (2019) treat explicability as a cross-cutting requirement. The risk addressed here is narrower and less discussed: not that a model is uninterpretable, but that a correctly computed accuracy figure can be over-read by an audience without the background to know what it covers.

2.3 Aggregation

Barocas and Selbst (2016) show that systems trained on aggregate historical data reproduce that data's structural features. For resource-dependent economies the relevant structure is concentration: Corden and Neary (1982) on booming-sector effects, and Sachs and Warner (2001) and Frankel (2010) on resource dependence more broadly. A Nigerian aggregate is disproportionately a series about oil-linked flows.

3. METHOD

This is a qualitative, policy-analytic paper. It uses a documented technical case as an anchor and applies established governance and documentation frameworks to it. The case is a leakage-free pipeline forecasting Nigerian FDI one quarter ahead, evaluated on a 13-quarter holdout with a naive benchmark retained throughout.

4. RESULTS

Table 1. Three risks, with evidence from the case

Risk	Evidence from this case	Proposed safeguard
Provenance	Half the assembled series was linearly interpolated from annual data; a constructed ratio divided nominal trade by real GDP; GDP was rebased mid-sample and spliced unadjusted. None was documented. A formal validation protocol passed on all three.	Publish a provenance table giving, for each series: source institution, access date, native publication frequency, construction or interpolation method, and release vintage.
Communication	The best model achieved out-of-sample R^2 of 0.0257 and was the only one of twelve to beat a mean benchmark. Eleven did worse than predicting the mean. A headline figure conveys none of this.	Require a model card alongside any figure shown to policymakers, stating the benchmark, the number of competing specifications tried, and the conditions under which the model fails.
Aggregation	FDI correlates weakly with the broad macroeconomic aggregates (levels: -0.44 with the exchange rate, -0.43 with trade), consistent with being driven by a narrow set of large transactions.	Restrict aggregate forecasts to aggregate-level questions; require separate validation before any sectoral or regional application.

5. DISCUSSION

The provenance case is the strongest of the three, because it is documented rather than hypothesised. The companion study ran a validation protocol that checked chronological integrity, recalculated a derived variable to a stated tolerance, reconciled trade aggregates against their components, and reproduced end-to-end. It passed. Three construction defects survived it, and correcting them reversed the study's central finding on cointegration. A policy user reading the published output would have had no way to detect any of this.

That is the argument for provenance documentation in a sentence. The information needed was not secret or expensive. It was simply never recorded, and nothing in normal practice required it to be.

On communication, the specific number matters. An R-squared of 0.0257 means the model explains about 2.6% more holdout variance than predicting the mean. Presented as the best of twelve approaches, it invites confidence. Presented with the fact that the other eleven lost to the mean, it invites the correct conclusion, which is that this series resists



forecasting. Both statements are true. Only the second is useful to someone deciding how much weight to place on the forecast.

On aggregation, the evidence here is weaker and should be described as such. FDI's weak correlation with the broad aggregates is consistent with concentration in a few large transactions, but it does not measure concentration directly. The CBN and UNCTAD both publish sectoral FDI breakdowns, and reporting the share accounted for by the largest two or three sectors would replace an inference with a measurement. Until that is done, the aggregation argument rests on plausibility.

6. CONCLUSION

National-aggregate FDI forecasts carry three risks that accuracy figures do not address. Provenance is the most serious and the best documented: a formal validation protocol passed on a dataset with three construction defects severe enough to reverse the study built on it. Communication is next: a small positive R-squared in a field of negative ones is a weak signal, and presenting it without that context invites over-reading. Aggregation is a genuine concern for which this paper offers suggestive rather than direct evidence. The three safeguards proposed cost little and would have caught or contextualised each problem.

7. LIMITATIONS

Sectoral concentration is inferred from correlations among aggregates and not measured. Obtaining CBN or UNCTAD sectoral breakdowns would settle it.

The analysis is conceptual on the policy side. Interviewing officials at Nigerian investment-promotion agencies about how forecasts are currently received would ground the communication argument in observed practice.

The safeguards are proposed and untested. Piloting a provenance table and a model card with a policy audience, and measuring whether interpretation changes, would test whether they work.

The legal discussion is limited to provenance and attribution. Nigerian data-protection law, licensing terms attached to IMF and World Bank products, and liability where a public decision relies on a model output are all outside scope.

REFERENCES

- [1]. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
- [2]. Corden, W. M., & Neary, J. P. (1982). Booming sector and de-industrialisation in a small open economy. *Economic Journal*, 92(368), 825–848.
- [3]. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
- [4]. Frankel, J. A. (2010). The natural resource curse: A survey. NBER Working Paper 15836.
- [5]. Gebru, T., et al. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.
- [6]. Mitchell, M., et al. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.
- [7]. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.
- [8]. Sachs, J. D., & Warner, A. M. (2001). The curse of natural resources. *European Economic Review*, 45(4–6), 827–838.
- [9]. UNCTAD. (2025). *World Investment Report 2025*.
- [10]. Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.