Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Anchoring Machine Learning to Persistence: A Regime-Aware Residual-Learning Framework for Forecasting Ghanaian Headline Inflation

Isaac Nyame
Kwame Nkrumah University of Science and Technology
Baffour Osei
University of Skills Training and Entrepreneurial Development
Abigail Boatemaa
Transilvania University of Brașov
Gabriel Osei Forkuo
gabriel.forkuo@unitbv.ro

Transilvania University of Brașov

Research Article

Keywords: inflation forecasting, machine learning, ARIMA, residual learning, Markov switching, conformal prediction, Ghana

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11769962/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Anchoring Machine Learning to Persistence: A Regime-Aware Residual-Learning Framework for Forecasting Ghanaian Headline Inflation

Abstract

Machine learning models are often reported to forecast inflation well, yet nine standard models tested on Ghana's headline inflation through the 2022–23 surge, were all beaten by an ARIMA benchmark and a random walk. Here we ask why that happened and what can be done about it. We show that much of the failure comes from how the target is parameterized. A learner asked to predict the level of a highly persistent series has to rebuild that persistence from a short sample, and tree-based learners cannot forecast above the highest value they saw in training (41.95 percent here, against an actual peak of 54.1 percent). We propose persistence-anchored residual learning (PARL). An autoregressive anchor carries the persistent component, machine learning models are trained only on the anchor's one-step errors, and an adaptive weight, whose memory can be tied to a two-state Markov-switching regime model, decides how much of the correction to use. Prediction intervals come from locally scaled adaptive conformal inference. Using 280 months of Bank of Ghana data (January 2000 to April 2023), strictly one-step-ahead forecasts and a 42-month test window that contains the surge, we find that predicting the monthly change instead of the level cuts the test RMSE of the nonlinear learners by 64 to 80 percent, to values near that of the ARIMA anchor (2.15 percentage points). PARL lowers the RMSE by a further 3 percent (2.08), a gain that is not statistically significant against the anchor (Diebold–Mariano $p = 0.41$) but is against the random walk ($p = 0.03$). Letting the weight's memory depend on the regime gave no additional gain. Adaptive conformal intervals raised the coverage of nominal 90 percent intervals in the test window from 67 to 86 percent and cut the interval score by about a quarter. The results suggest that, for persistent macroeconomic series, how the target is written matters more than which learner is used.

Keywords: inflation forecasting; machine learning; ARIMA; residual learning; Markov switching; conformal prediction; Ghana

JEL Classification: C22, C53, C45, E31, E37, E52

1. Introduction

Central banks that target inflation need forecasts that hold up when conditions change, and Ghana's recent experience shows why. Headline year-on-year inflation was close to 10 percent in early 2022 and reached 54.1 percent in December 2022 (Bank of Ghana data used in this study). Exchange-rate pass-through, imported food and energy prices and domestic monetary conditions all feed into the Ghanaian price level (Acquah-Sam, 2017; Ohemeng et al., 2021), and a forecasting model has to cope with all of them at once.

Two forecasting traditions compete. The classical one relies on autoregressive and structural time-series models and is the default in many African central banks, for example in SARIMA studies of Nigerian inflation (Otu et al., 2014) and Kenyan inflation (Nyamao, 2014) and in the IMF's food and non-food framework for low-income countries (Andrle et al., 2013). The newer one uses machine learning and deep learning, with encouraging results for several African macro-financial series (Osei-Dwomoh & Forkuo, 2026), including an LSTM network for the cedi–dollar exchange rate (Adekoya et al., 2021) and a stacked ensemble for the drivers of Nigerian inflation (Akande et al., 2021). In data-rich settings the evidence for machine learning is real but uneven. Medeiros et al. (2021) found that random forests beat standard benchmarks for US inflation, while Naghi et al. (2024) later showed that the same random forest did poorly during the 2020–22 inflation episode.

We began with a standard comparison of nine models for one-month-ahead Ghanaian inflation and found that a plain ARIMA(3,1,0) and a random walk beat every nonlinear learner on a test window that spans the 2022–23 surge. Level-target Random Forest, XGBoost and SVR made errors of 7.6, 9.0 and 13.4 percentage

points, against 2.4 for ARIMA. A benchmarking result of this kind is informative, but it leaves open why flexible learners fail so badly on a persistent series. If the failure has a cause, the cause can be diagnosed and, to some extent, repaired. That is what this paper does.

Our diagnosis has two parts. First, when the target is the level of inflation, the learner must work out from about 200 observations that tomorrow's value is nearly today's value. A linear model with a lagged dependent variable can do this easily. A regression tree cannot, and it has a second handicap: its forecast is an average of training targets, so it can never exceed the largest inflation rate in the training sample. In our data that ceiling is 41.95 percent, while inflation reached 54.1. Second, the sample-size explanation offered by Cerqueira et al. (2019) is only partly right, because the shortfall of level-target learners shrinks with more data but does not disappear, whereas the same learners fitted to monthly changes are almost insensitive to the length of the training window.

The repair we propose is persistence-anchored residual learning (PARL). The persistent part of inflation is handed to an autoregressive anchor. Machine learning models are trained only to predict what the anchor gets wrong, an adaptive weight decides how much of that correction to apply, and the final forecast comes with an interval from adaptive conformal inference. The study makes four contributions.

(1) A diagnosis of why level-target learners fail on persistent series, with two testable mechanisms: the extrapolation ceiling of tree-based learners and the cost of re-learning persistence from a short sample. Both are checked against the data.

(2) A forecasting framework, PARL, that combines an anchor, an ensemble of learners trained on anchor errors, a recursive shrinkage weight with regime-dependent memory (RMAS, for regime-modulated adaptive shrinkage), and a regime-sensitive prediction interval. The first two components are the ones that matter in our data. We also report what did not help.

(3) Interval forecasts for Ghanaian inflation. Most comparisons of inflation forecasting models stop at point forecasts. We compare Gaussian, split-conformal and adaptive conformal intervals and score them with coverage and the interval score (Gneiting & Raftery, 2007).

(4) A wider evaluation: a Markov-switching AR benchmark, a rolling-origin record of 136 forecast origins, Diebold–Mariano and Clark–West tests, the Model Confidence Set, sub-period results and a learning-curve experiment.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the data, the framework and the evaluation design. Section 4 reports the results, Section 5 discusses them and their limits, and Section 6 concludes.

2. Related Work

2.1 Inflation Dynamics and Monetary Policy in Ghana and Sub-Saharan Africa

Ghana adopted a formal inflation-targeting framework in the 2000s. Acquah-Sam (2017) identifies money supply growth, exchange-rate depreciation and fiscal deficits as the main proximate drivers of Ghanaian inflation, in line with wider West African evidence on exchange-rate pass-through (Ohemeng et al., 2021). Cross-country evidence for the Economic Community of West African States points to slow and partial inflation convergence (Muhammed-Gani et al., 2022), which is one reason a Ghana-specific forecasting exercise is worthwhile. Kunawotor et al. (2021) show that extreme weather affects African inflation through agricultural supply, a channel that matters for a food-heavy consumer basket such as Ghana's.

Experience elsewhere in the region, including Sierra Leone (Jackson et al., 2020), Mauritius (Madhou et al., 2021), Rwanda (Ngarambe et al., 2016; World Bank, 2020b), Nigeria (George-Anokwuru, 2023; World Bank, 2020a) and Uganda (International Monetary Fund, 2013, 2018), suggests that the credibility of

inflation targeting, exchange-rate management and fiscal discipline interact in country-specific ways. Studies of South Africa (Ncanywa & Ralarala, 2020; Oladipo, 2017; World Bank, 2012), Brazil (Filho & Milan, 2018) and India (Islam & Ahmed, 2023) reach similar conclusions, and Svensson (2010) and Chugunov et al. (2019, 2021) give the general framework against which these implementations are usually judged.

2.2 Classical Time-Series Forecasting of Inflation

Autoregressive integrated moving average (ARIMA) models remain a workhorse for inflation in developing economies, because headline inflation is persistent and close to a unit-root process, which makes simple autoregressions hard to beat. Otu et al. (2014) and Nyamao (2014) report good short-horizon results for Nigeria and Kenya. The difficulty of beating such benchmarks is an old theme. Atkeson and Ohanian (2001) found that Phillips-curve forecasts of US inflation did no better than a naive forecast, and Stock and Watson (2007) traced the problem to changes in the size of the unpredictable component of inflation. Hendry (2006) explains why forecasts built on differenced, error-correcting structures are robust to location shifts, which is close to the reason an anchor on the previous observation works in a surge. Hamilton (1989) introduced Markov-switching models for series whose dynamics change between regimes, and Pesaran and Timmermann (2007) discuss how much history to use when breaks are possible. Charemza et al. (2014) add that inflation forecast risk is asymmetric, so point accuracy is not the whole story.

2.3 Machine Learning in Macroeconomic Forecasting

Medeiros et al. (2021) report that random forests and other machine learning methods with many predictors beat autoregressive and factor benchmarks for US inflation. Goulet Coulombe et al. (2022) ask where such gains come from and conclude that nonlinearity helps most when it comes with sensible target and feature transformations, and helps little when it is applied to raw levels. Naghi et al. (2024) replicate Medeiros et al., extend the sample to October 2022 and find that random forests did poorly in the pandemic and the high-inflation period that followed, while stochastic-volatility models and some boosting methods did better. Our results for Ghana point in the same direction.

For African series, Adekoya et al. (2021) train an LSTM network for the cedi–dollar exchange rate and Akande et al. (2021) use a stacked ensemble for Nigerian inflation components. Both targets differ from year-on-year headline inflation, and neither design includes a held-out period of the 2022 kind. Baidoo and Obeng (2024) apply machine learning to Ghanaian inflation directly and report favorable results. Cerqueira et al. (2019) show across many series that machine learning needs larger samples than statistical methods before it catches up. Tassi (2019) places this in the wider turn to big data in economic statistics.

Hybrid designs that let a statistical model carry the main structure and a learner model the rest have a long history. Zhang (2003) fits a neural network to ARIMA residuals, and Smyl (2020) won a major forecasting competition with a hybrid of exponential smoothing and recurrent networks. What these designs share with ours is the idea that the learner should be asked an easier question than the raw forecast.

2.4 Uncertainty Quantification for Forecasts

Policy committees need ranges as well as point forecasts. Gaussian intervals built from past errors assume stable variance, which fails when inflation accelerates. Conformal prediction (Vovk et al., 2005) yields intervals with finite-sample coverage under exchangeability, an assumption that time series break. Gibbs and Candès (2021) adjust the working miscoverage level online so that long-run coverage is maintained when the distribution shifts. We combine this with a local error scale, and evaluate the result with the interval score of Gneiting and Raftery (2007).

2.5 Positioning of the Present Study

Relative to the above, this paper does three things. It explains why machine learning failed in a specific, well-documented setting (Ghana, 2019–2023) instead of only reporting that it did. It proposes a reparameterization and hybrid design that removes most of the failure. And it supplies calibrated interval forecasts together with a broader set of tests than the point-forecast comparison we began with.

3. Data and Methodology

3.1 Data Sourcing and Cleaning

Data come from 19 monthly bulletin exports published by the Bank of Ghana (Monthly Interest Rates, Monthly Monetary Survey and Time Series Monthly series), several of which overlap or duplicate each other. We merged the files, removed exact duplicate rows and reshaped each from a wide year-by-month layout to a long format (Date, Variable, Value). Every variable's "0.00" entries fell in the trailing months of its series, which points to months not yet reported and not to true zeros, so we recoded them as missing. One real conflict remained: the Monetary Policy Rate for May 2018 appeared as both 17.00 and 18.00 percent in different files. The Monetary Policy Committee's decision history (18 percent held from March to April 2018, then cut to 17 percent in May) settled it in favor of 17.00.

We kept 24 core macro-monetary variables (Table 1) with continuous coverage from January 2000 to April 2023, 280 months in all. They cover headline inflation, policy and market interest rates, monetary aggregates, private-sector credit, three commodity prices and the Bank of Ghana Composite Index of Economic Activity. The larger 56-variable cleaned data set, described in the companion data-preparation workbook, has a common window of only 52 months and was not used.

Table 1. The 24 core macro-monetary variables, January 2000–April 2023

Category	Variables included
Inflation	Headline Inflation (YoY %) [target variable]
Interest rates	Monetary Policy Rate; 91- and 182-Day Treasury Bill discount and interest-rate-equivalent rates; Average Savings Deposit Rate; 3-Month Time Deposit Rate; Average Commercial Bank Lending Rate
Money supply and credit	Broad Money (M2); Narrow Money (M1); Total Liquidity (M2+); Reserve Money; Currency Outside Banks; Demand Deposits; Savings & Time Deposits; Foreign Currency Deposits; Net Foreign Assets (BoG); Private Sector Credit
Economic activity	BoG Composite Index of Economic Activity (Nominal, Real)
Commodity prices	International Brent Crude Oil, Cocoa and Gold prices (monthly average, US$)

3.2 Conceptual Framework: Persistence Anchoring

Write headline inflation as y_t and its monthly change as $\Delta y_t = y_t - y_{t-1}$. The series is highly persistent, so the variance of Δy_t is small compared with the variance of y_t. A learner that targets y_t must therefore spend most of its capacity reproducing y_{t-1}, and what is left over is what it can use for the genuinely forecastable part of Δy_t.

We separate the two jobs. An anchor model delivers a forecast $\hat{y}^A_t$ that uses only information up to t−1 and captures persistence. Its error is $e_t = y_t - \hat{y}^A_t$. A learner g is trained to predict that error from lagged predictors x_t, giving a correction $c_t = g(x_t)$. The PARL forecast is

$$\hat{y}_t = \hat{y}_t^A + w_t c_t, \quad 0 \le w_t \le 1 \tag{1}$$

where w_t is a weight estimated recursively from past performance of the correction (Equations 2 to 5 in Section 3.5). When the learner has nothing to add, w_t drifts to zero and PARL collapses to the anchor, so the design has a built-in "do no harm" property that a level-target learner lacks.

The ceiling argument is simple. A regression tree returns an average of training targets in a leaf, so any ensemble of trees (random forest, boosting) returns a value between the smallest and largest target seen in training. With a level target, that range is the range of inflation rates seen in training. With an anchored target the same bound applies to the correction c_t or the change Δy_t, and the level forecast equals the anchor (which can be any value) plus a bounded adjustment. Section 4.2 shows that this is not a hypothetical concern in 2022.

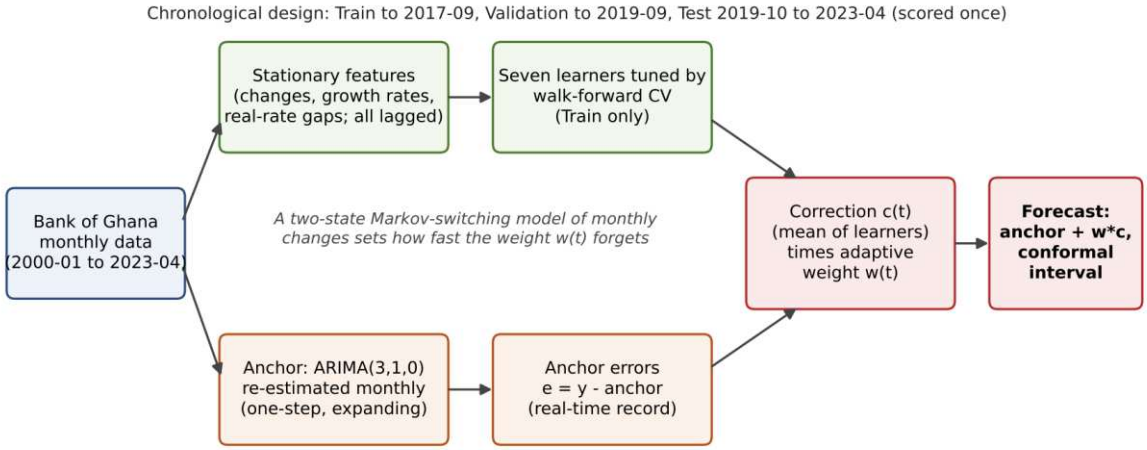


Figure 1. The PARL framework. An autoregressive anchor and a set of learners trained on the anchor's errors are combined through an adaptive weight; the forecast is accompanied by an adaptive conformal interval.

3.3 Feature Construction, Data Splits and Updating

The forecasting target is headline year-on-year inflation as published by the Bank of Ghana. Every predictor in the row dated t uses only information from t−1 or earlier. The anchored learners use a new, stationary feature set built from the same 24 variables: lagged monthly changes of inflation and changes over 3 and 12 months, the deviation of inflation from its 6- and 12-month means, six-month volatility of changes, changes and real-rate gaps (rate minus inflation) for the policy rate, the 91-day Treasury bill rate, the lending rate and the savings rate, the 182-day minus 91-day Treasury bill slope, 1-, 3- and 12-month log growth of M2, M1, reserve money, currency outside banks, foreign-currency deposits and private credit, changes in net foreign assets scaled by M2, 1-, 3- and 12-month log changes in oil, cocoa and gold prices, and the calendar month as sine and cosine. In total there are 50 features, listed in the supplementary dictionary. The residual learners add the anchor's last two errors and its own predicted change. We dropped the Composite Index of Economic Activity from this feature set because it is missing for December 2022 and April 2023. The baseline level-target models use a 65-feature set of levels and lags, which includes it, and that is why

January 2023 is absent from the baseline test window: the row had no valid lagged index. The new models can forecast January 2023, but to keep results comparable we score every model on the same 42 test dates.

The sample is split chronologically into three windows. Train runs from February 2001 to September 2017 (200 months) and is the only window used to tune hyperparameters. Validation runs from October 2017 to September 2019 (24 months) and is used to choose the gate parameters. Test runs from October 2019 to April 2023 and is scored once. It contains the pandemic period and the 2022–23 surge, in which inflation rose from about 10 percent to 54.1 percent. Hyperparameters of every learner come from five-fold expanding-window cross-validation (scikit-learn's TimeSeriesSplit) on Train rows, with RMSE as the criterion.

To produce real-time forecasts, every new model is re-estimated every month from January 2012 to April 2023 on all earlier rows (136 forecast origins), with hyperparameters held fixed after tuning on Train. This gives each model the same updating privilege as ARIMA. The baseline level-target learners, by contrast, are fitted once on Train and Validation and then held fixed over the test window, as is common in applied work. They are labeled "baseline" in what follows and are not refit monthly.

3.4 Models

Baseline specification. A random walk, a seasonal naive forecast, an ARIMA(3,1,0) fitted by maximum likelihood on Train and updated monthly, and six level-target learners (Ridge, Lasso, Random Forest, XGBoost, SVR and a one-hidden-layer MLP), with configurations selected by cross-validation on Train (Table 2).

Anchor. An AR(3) in first differences, equivalent to ARIMA(3,1,0) without drift, estimated by conditional least squares on all data from January 2000 up to $t-1$ and re-estimated every month. It needs 48 months of history, so its first forecast is for January 2004. We use least squares and not maximum likelihood so that the anchor can be reproduced with a few lines of numpy. It differs from the baseline ARIMA in the estimator and in the training history, and this matters: its test RMSE is 2.15 against 2.41. We report both.

Markov-switching AR(3). A two-state model for the monthly change, with state-specific intercept, autoregressive coefficients and variance (Hamilton, 1989), estimated by the EM algorithm at every origin. Its forecast mixes the two state-specific forecasts with the predicted regime probabilities. This adds a regime-switching benchmark to the comparison.

Anchored learners. Seven learners (Ridge, Lasso, Elastic Net (Zou & Hastie, 2005), Random Forest, gradient boosting (Friedman, 2001), SVR and an MLP trained with L-BFGS) predict Δy_t from the stationary features. The level forecast is y_{t-1} plus the predicted change. We use scikit-learn (Pedregosa et al., 2011) throughout. XGBoost was not available in the computing environment used for the anchored designs, so gradient boosting means scikit-learn's implementation here.

Residual learners and PARL. The same seven learners are trained on the anchor's error e_t. The correction c_t is the mean of the seven predictions. Taking the mean avoids having to pick one learner on a 24-month validation window. We call the unweighted version PARL-static ($w_t = 1$).

Table 2. Candidate models and selected configuration

Model	Target	Estimation	Selected configuration
Random walk	Level	None	Forecast = previous month
Seasonal naive	Level	None	Forecast = value 12 months earlier
ARIMA(3,1,0), baseline	Level	MLE, Train; updated monthly	Order chosen by AIC on Train

Model	Target	Estimation	Selected configuration
Anchor: AR(3) in differences	Change	Least squares, monthly, from 2000-01	Order fixed at (3,1,0), no drift
Markov-switching AR(3)	Change	EM, monthly	2 states; regime-specific mean, AR terms, variance
Ridge / Lasso, baseline	Level	Refit once on Train+Val	$\alpha = 100$ / $\alpha = 0.5$
Random Forest, baseline	Level	Refit once on Train+Val	400 trees; depth 5; min leaf 1
XGBoost, baseline	Level	Refit once on Train+Val	200 trees; depth 4; learning rate 0.1; subsample 0.8
SVR / MLP, baseline	Level	Refit once on Train+Val	$C = 1$, $\varepsilon = 0.1$ / 1 hidden layer of 32 units, $\alpha = 0.001$
Ridge, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	alpha: 1000 / alpha: 10000
Lasso, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	alpha: 3.0 / alpha: 3.0
Elastic Net, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	alpha: 3.0, l1_ratio: 0.5 / alpha: 3.0, l1_ratio: 0.5
Random Forest, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	depth: 2, max features: 0.33, min leaf: 5 / depth: 2, max features: 0.33, min leaf: 10
Gradient boosting, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	learning rate: 0.03, depth: 3, trees: 100, subsample: 0.7 / learning rate: 0.03, depth: 2, trees: 100, subsample: 1.0
SVR, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	C: 0.1, epsilon: 0.2, gamma: scale / C: 0.1, epsilon: 0.05, gamma: scale
MLP, anchored / residual	Change / anchor error	Monthly refit, from 2012-01	alpha: 1.0, hidden units: (16,) / alpha: 0.01, hidden units: (16,)
PARL-static	Anchor error	Monthly refit	Mean of the seven residual learners; $w = 1$
PARL-fixed / PARL-RMAS	Anchor error	Recursive weight	$\lambda = 0.98$, $\rho = 0.5$ / $\lambda_{calm} = 0.98$, $\lambda_{turb} = 0.95$, $\rho = 0.5$

Note: "baseline" marks the level-target comparison (references: Box & Jenkins, 1970; Hoerl & Kennard, 1970; Tibshirani, 1996; Breiman, 2001; Chen & Guestrin, 2016; Drucker et al., 1997; Rumelhart et al., 1986; Hyndman & Athanasopoulos, 2021). For the anchored learners the entry before the slash belongs to the change-target design and the entry after it to the residual design. Cross-validation picked the largest penalty on the grid for Lasso, Elastic Net and, in the residual design, Ridge, which sets all Lasso and Elastic Net coefficients to zero.

3.5 Regime-Modulated Adaptive Shrinkage

The weight w_t is the exponentially weighted least-squares slope of past anchor errors on past corrections, clipped to [0,1] and ridge-stabilized:

$$w_{t+1} = \min\left\{1, \max\left\{0, \frac{S_t^{ce}}{S_t^{cc} + \rho}\right\}\right\} \tag{2}$$

$$S_t^{cc} = \lambda_t S_{t-1}^{cc} + {c_t}^2 \tag{3}$$

$$S_t^{ce} = \lambda_t S_{t-1}^{ce} + c_t e_t \tag{4}$$

Each c_t used here is itself a real-time out-of-sample correction, so the weight is trained on honest errors. The forgetting factor is regime-dependent:

$$\lambda_t = \lambda_{calm}(1 - \pi_t) + \lambda_{turb}\pi_t \tag{5}$$

where π_t is the filtered probability of the high-variance state of a two-state Markov-switching model for Δy_t, re-estimated at every origin on data up to t−1. The motivation is that memory should shorten when the regime changes (Pesaran & Timmermann, 2007). The parameters (λ_{calm}, λ_{turb}, ρ) were chosen on Validation from a grid of 36 combinations. We call the resulting forecast PARL-RMAS. Two ablations are reported: PARL-fixed uses a single λ (no regime modulation), and PARL-static uses $w_t = 1$.

3.6 Prediction Intervals

For a point forecast $\hat{y}_t$ we compare four 80 and 90 percent interval rules. (G) Gaussian, with the standard deviation of all past one-step errors. (SC) Split conformal, with one fixed half-width equal to the empirical quantile of absolute errors up to the end of Validation. (LS) Locally scaled adaptive conformal, where the non-conformity score is

$$r_t = \frac{|y_t - \hat{y}_t|}{s_t} \tag{6}$$

and s_t is an exponentially weighted mean of past absolute errors (decay 0.90). The interval quantile is taken over the last 60 scores, and the working miscoverage level is updated online as (Gibbs & Candès, 2021)

$$\alpha_{t+1} = \alpha_t + \gamma(\alpha - \text{miss}_t) \tag{7}$$

with $\gamma = 0.05$ and miss_t equal to one if y_t fell outside its interval and zero otherwise. (RS) The same as LS but with the decay shortened to 0.70 in proportion to the regime probability. All four rules use information available at t−1 only.

3.7 Evaluation

Point accuracy is measured by RMSE (the main criterion), MAE, MAPE and R^2. We test equal predictive accuracy with the Diebold–Mariano (1995) test using the Harvey et al. (1997) small-sample correction and compare nested models with the Clark–West (2007) test. The Model Confidence Set (Hansen et al., 2011) is computed from squared errors with the T_{max} statistic and a moving-block bootstrap (block length 3, 5,000 replications). Intervals are scored by empirical coverage, mean width and the interval score. A learning-

curve experiment refits level-target and anchored learners on the most recent k months before the test window and scores them once on the test dates.

4. Results

4.1 Descriptive Statistics

Table 3 gives descriptive statistics. Headline inflation averages 16.45 percent over the 280 months, with a standard deviation of 9.07 points, a minimum of 7.50 and a maximum of 54.10 reached in the 2022–23 surge. Interest rates move broadly with inflation. Monetary aggregates and credit trend upward strongly, which is why we use log growth rates and not levels in the anchored feature set.

Table 3. Descriptive statistics of core variables, January 2000–April 2023 (n = 280 months)

Variable	Mean	Std. Dev.	Min	Max
Headline Inflation, YoY (%)	16.45	9.07	7.50	54.10
Monetary Policy Rate (%)	18.92	4.97	12.50	29.50
91-Day Treasury Bill Discount Rate (%)	18.95	7.82	9.04	41.85
Broad Money M2 (GHS millions)	28,719.94	35,444.51	392.86	150,496.16
Private Sector Credit (GHS millions)	16,843.11	18,036.51	379.09	73,744.05
Brent Crude Oil (US$/barrel)	65.96	29.29	19.17	134.79
Cocoa (US$/tonne)	2,213.40	659.01	701.43	3,430.35
Gold (US$/fine ounce)	1,058.68	528.78	257.82	2,000.69

4.2 Why the Level-Target Learners Fail

We start with the baseline level-target learners. Their test RMSE ranges from 3.41 (Lasso) to 13.37 (SVR) against 2.41 for ARIMA (Table 4). Within the training sample they fit well: XGBoost's Train RMSE is 0.03 and Random Forest's is 0.83, which are signs of overfitting. We checked for target leakage and found none. The strongest correlation between any engineered feature and the contemporaneous target is the first lag of inflation, at 0.978, as expected for a persistent series.

Overfitting is only part of the explanation. The highest inflation rate in the Train and Validation windows is 41.95 percent. In four test months (November and December 2022, February and March 2023) inflation exceeded that value. Tree-based learners cannot follow it there (Table 5, Figure 2a): Random Forest never forecasts above 30.2 percent, XGBoost never above 26.5 and SVR never above 16.4, even though actual inflation reached 54.1. Ridge peaks at 38.6 and the MLP at 43.9. The Lasso is the exception, with a maximum of 48.6, because a linear model can extrapolate. In the four months above the training maximum, the RMSE of level-target Random Forest is 21.2 points and that of XGBoost is 24.7, compared with 6.0 and 5.7 for the anchored Random Forest and gradient boosting (Table 5).

Table 4. Test-window forecast accuracy (42 months, October 2019–April 2023; percentage points of inflation)

Group	Model	Val. RMSE	Test RMSE	Test MAE	MAPE (%)	Ratio to anchor	DM p vs anchor
Baseline	Random walk	0.53	2.65	1.66	7.8	1.23	0.018

Group	Model	Val. RMSE	Test RMSE	Test MAE	MAPE (%)	Ratio to anchor	DM p vs anchor
	Seasonal naive	2.40	15.08	9.32	35.5	7.00	0.002
Classical	ARIMA(3,1,0), baseline	0.51	2.41	1.34	6.8	1.12	0.114
Classical (anchor)	**AR(3) in differences (anchor)**	**0.54**	**2.15**	**1.22**	**6.4**	**1.00**	**–**
Classical	Markov-switching AR(3)	0.56	2.17	1.23	6.4	1.01	0.690
Level target	Ridge	0.59	7.01	4.64	19.2	3.25	0.001
	Lasso	0.96	3.41	2.12	9.9	1.58	0.059
	Random Forest	0.79	7.58	4.17	14.5	3.52	0.016
	XGBoost	0.73	9.04	4.97	16.7	4.20	0.011
	SVR	4.84	13.37	7.94	30.6	6.21	0.004
	MLP	7.61	5.35	3.85	23.0	2.49	0.005
Anchored change	Ridge	0.49	2.29	1.43	7.4	1.06	0.037
	Lasso (= drift)	0.50	2.67	1.70	8.1	1.24	0.013
	Random Forest	0.51	2.37	1.47	7.4	1.10	0.009
	Gradient boosting	0.51	2.17	1.33	7.1	1.01	0.887
	SVR	0.50	2.63	1.66	7.9	1.22	0.022
	MLP	0.73	1.90	1.44	9.0	0.88	0.635
	Mean of 7 learners	0.48	2.24	1.44	7.5	1.04	0.438
PARL	PARL-static (w = 1)	0.51	2.07	1.21	6.3	0.96	0.360
	PARL-fixed	0.52	2.08	1.22	6.3	0.97	0.405
	PARL-RMAS	0.52	2.08	1.22	6.3	0.97	0.408

Note: Val. = Validation window (24 months). DM p = two-sided p-value of the Diebold–Mariano test with the Harvey–Leybourne–Newbold correction on squared errors, against the anchor. Level-target models and the baseline ARIMA are the baseline-run forecasts. The anchored Lasso and Elastic Net coincide because both are reduced to an intercept.

Table 5. Extrapolation ceiling: highest forecast and errors in the four test months where inflation exceeded the Train and Validation maximum (41.95%)

Learner	Highest level-target forecast	Level-target forecast, Dec 2022	Level-target RMSE, 4 months above 41.95	Anchored forecast, Dec 2022	Anchored RMSE, same 4 months
RandomForest	30.2	29.6	21.2	53.4	6.0
XGBoost	26.5	25.7	24.7	55.3	5.7
SVR	16.4	16.4	34.3	50.3	6.6
Ridge	38.6	37.2	15.4	53.2	5.8
Lasso	48.6	45.7	8.2	50.3	6.6
MLP	43.9	39.6	12.1	59.1	4.0

Note: Actual inflation was 54.1% in December 2022. The XGBoost row is compared with gradient boosting in the anchored design.

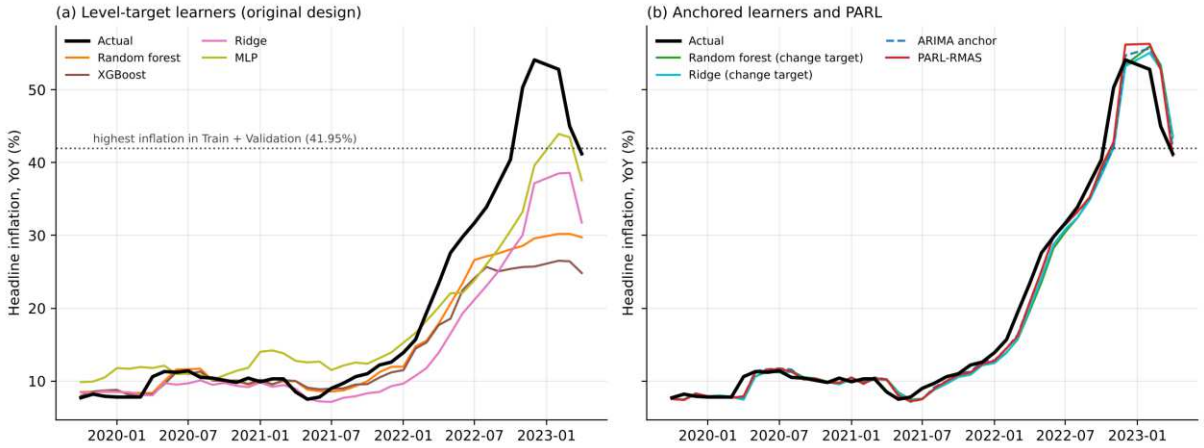


Figure 2. Actual and forecast headline inflation in the test window. (a) Baseline level-target learners; the dotted line marks the highest inflation rate in Train and Validation. (b) Anchored learners, the ARIMA anchor and PARL-RMAS.

4.3 The Effect of Anchoring the Target

Figure 3 and Table 4 show what happens when the same learners predict the monthly change. The test RMSE of the nonlinear learners falls by between 64 and 80 percent: Random Forest from 7.58 to 2.37, gradient boosting (against XGBoost) from 9.04 to 2.17, SVR from 13.37 to 2.63, the MLP from 5.35 to 1.90 and Ridge from 7.01 to 2.29. Most are now within 25 percent of the anchor's 2.15 (ratios in Table 4). The Clark–West test against the random walk, which these models nest, rejects at the 5 percent level for Ridge, Random Forest, gradient boosting, the MLP and the mean of the seven learners.

Three qualifications apply. First, anchoring brings most learners close to the anchor but not past it. The anchored Ridge and Random Forest are statistically worse than the anchor (Diebold–Mariano $p = 0.037$ and 0.009). Second, the Lasso and Elastic Net penalties chosen by cross-validation are large enough to set every coefficient to zero, so these two models forecast the training-sample average monthly change, which is a random walk with drift. Their test RMSE of 2.68 is therefore no better than the random walk's 2.65. Third, the anchored MLP has the lowest test RMSE of all models (1.90), but its Validation RMSE (0.73) is the worst among the anchored learners and its cross-validated error is not better than a zero-change forecast. We do not read this as evidence in its favor. A selection rule based on Validation would not have chosen it, and a 42-observation test window dominated by one episode can reward a lucky model.

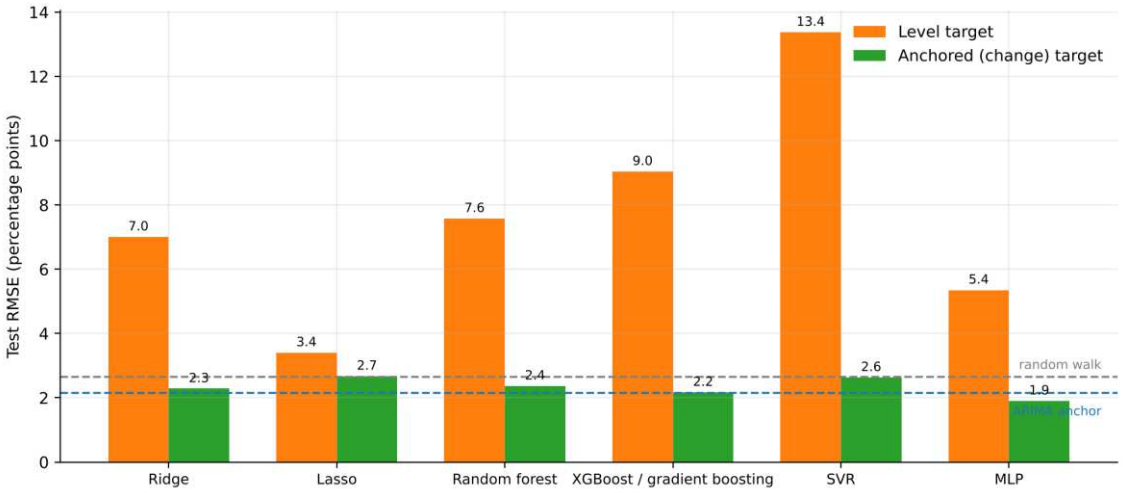


Figure 3. Test RMSE of each learner with a level target (baseline) and with the anchored change target. The dashed lines mark the random walk and the ARIMA anchor.

4.4 Residual Learning and the Adaptive Weight

The residual-correction hybrids give a small improvement over the anchor. PARL-static has a test RMSE of 2.075, PARL-fixed 2.083 and PARL-RMAS 2.083, against 2.153 for the anchor and 2.167 for the Markov-switching AR(3). The reduction is about 3 percent. It is not statistically significant against the anchor (Diebold–Mariano p = 0.36 to 0.41; Clark–West one-sided p = 0.09 to 0.11), but PARL is significantly better than the random walk (p = 0.03). Over all 135 origins from February 2012 to April 2023, PARL-RMAS has an RMSE of 1.27 against 1.31 for the anchor (p = 0.30), so the direction is the same in a longer record, and again not conclusive.

The regime-modulated weight did not improve on a single memory. The validation grid chose $\lambda_{calm} = 0.98$ and $\lambda_{turb} = 0.95$, so the two memories are nearly identical, and PARL-RMAS and PARL-fixed have the same test RMSE to three decimals. The Validation window (October 2017 to September 2019) contains no turbulent episode, so it carries almost no information about how fast memory should shorten in one. The regime model itself behaves sensibly. Its estimated standard deviation of monthly changes is 0.63 points in the calm state and 3.34 in the turbulent one, and the probability of the turbulent state is 0.4 at the March 2022 forecast origin and 1.0 by April (Figure 4). The weight on the correction falls to 0.13–0.29 between April and September 2022, as the learners' corrections fail in the early surge, and returns to 1 in November.

A note on how the corrections are built. Cross-validation chose the maximum shrinkage on the grid for Ridge, Lasso, Elastic Net and SVR in the residual design, so the corrections of these four models are almost constant (mean absolute value 0.04 to 0.05 points). Nearly all the variation in the ensemble correction comes from Random Forest (0.10), gradient boosting (0.20) and the MLP (0.78). The feature importances of the tree models are spread thinly. The most important feature (three-month growth in foreign-currency deposits) carries about 11 percent of the importance on average, followed by six-month volatility of changes (11 percent) and the anchor's own predicted change (7 percent) (Figure 5). Foreign-currency deposit growth may be picking up depreciation pressure, since the data set has no exchange-rate series, but we have not tested that reading.

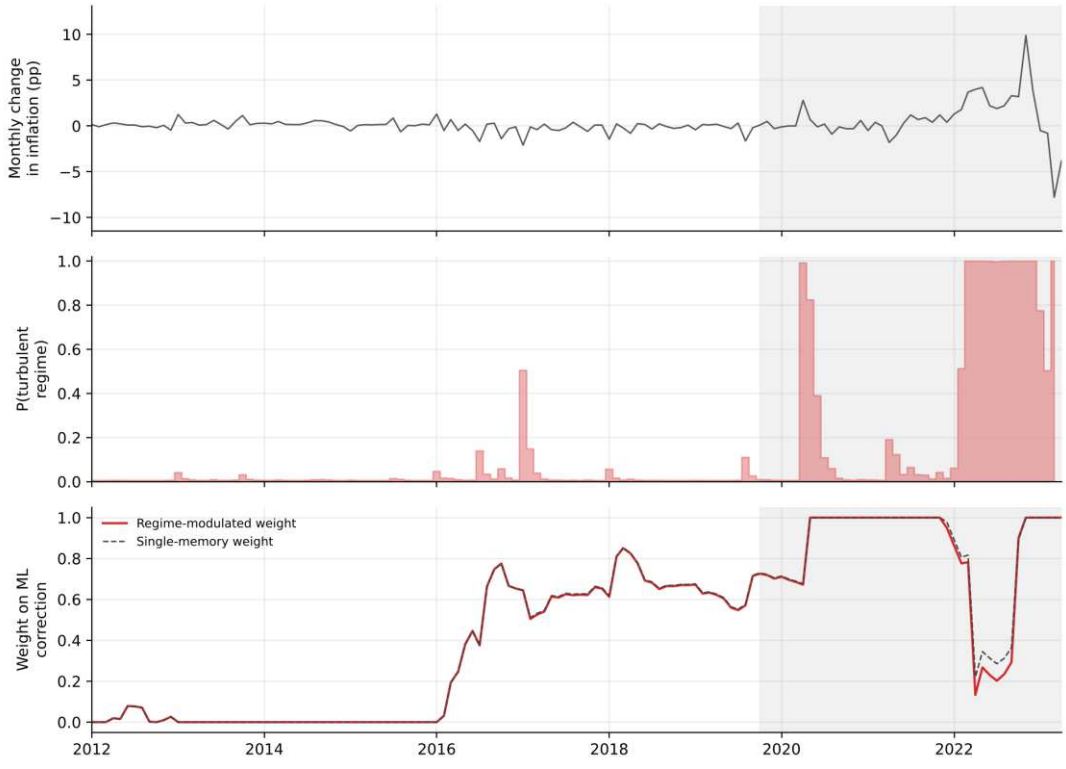


Figure 4. Regime model and weight. Top: monthly change in inflation. Middle: filtered probability of the turbulent state (final fit). Bottom: weight placed on the machine learning correction by PARL-RMAS and PARL-fixed, 2012–2023. The shaded area is the test window.

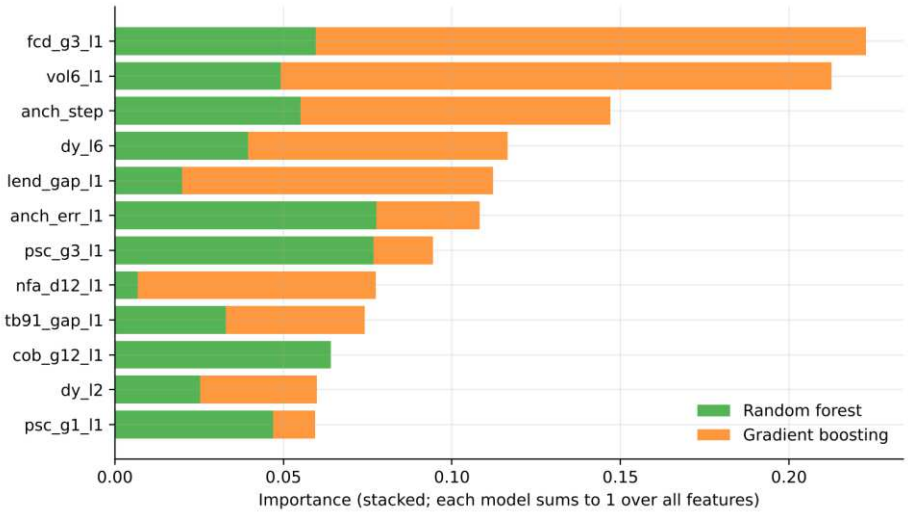


Figure 5. Twelve most important features of the residual-correction learners at the final origin (random forest and gradient boosting).

4.5 Robustness

The ranking is not an artifact of one test window. Table 6 splits the errors by period. In the pre-surge years (validation, and the pandemic months up to December 2021) all models except the MLP are within about 0.1 points of each other. The differences open up in the surge: PARL has an RMSE of 3.3 against 3.4 for

the anchor and 4.3 for the random walk, and the anchored MLP has 2.8. A rolling 12-month comparison (Figure 7) shows the same pattern, with PARL close to the anchor outside the surge and slightly below it within.

Table 6. RMSE by sub-period (percentage points). Pandemic and surge months use the 42 comparison dates

Model	Validation (n=24)	Pandemic, pre-surge (n=27)	Surge (n=15)	All origins 2012–23 (n=136)
Random walk	0.53	0.85	4.28	1.54
Anchor	0.54	0.84	3.42	1.31
Markov-switching AR(3)	0.56	0.85	3.44	1.33
Ridge, anchored	0.49	0.92	3.62	1.35
Random Forest, anchored	0.51	0.87	3.79	1.42
Gradient boosting, anchored	0.51	0.89	3.43	1.32
MLP, anchored	0.73	1.11	2.82	1.34
Mean of 7 anchored	0.48	0.88	3.55	1.32
PARL-static	0.51	0.81	3.30	1.26
PARL-RMAS	0.52	0.82	3.31	1.27

The Model Confidence Set at the 10 percent level (Table 7) excludes only three of the fifteen models considered: the level-target Ridge, MLP and XGBoost. The level-target Random Forest and the random walk remain inside the set. With 42 observations and a handful of large errors the set has little power, so we read it as a list of what cannot be ruled out. It does not rank the models.

Table 7. Model Confidence Set on squared errors, test window (T_{max} statistic, block bootstrap, 5,000 replications)

Model (in order of elimination)	MCS p-value	In 90% set	Test RMSE
Ridge, level target	0.046	no	7.01
MLP, level target	0.092	no	5.35
XGBoost, level target	0.092	no	9.04
Random Forest, level target	0.132	yes	7.58
Random walk	0.132	yes	2.65
Lasso, level target	0.325	yes	3.41
Random Forest, anchored	0.325	yes	2.37
Ridge, anchored	0.325	yes	2.29
Mean of 7 anchored	0.395	yes	2.24
Gradient boosting, anchored	0.412	yes	2.17
Markov-switching AR(3)	0.469	yes	2.17
Anchor	0.469	yes	2.15
PARL-RMAS	0.558	yes	2.08
PARL-static	0.585	yes	2.07
MLP, anchored	1.000	yes	1.90

The learning curves in Figure 6 test the sample-size explanation. Level-target Ridge improves from an RMSE of 18.6 when trained on the latest 48 months to 5.4 with 150 months, and then deteriorates to 7.0 with all 224 months. Level-target Random Forest and gradient boosting improve from about 15 to 7.6 and 6.1 as the window grows to 224 months. So more data helps, but even with the whole sample the level-target models remain two to three times worse than the anchored versions. The anchored learners stay between 2.4 and 2.9 for every window length from 48 to 224 months.

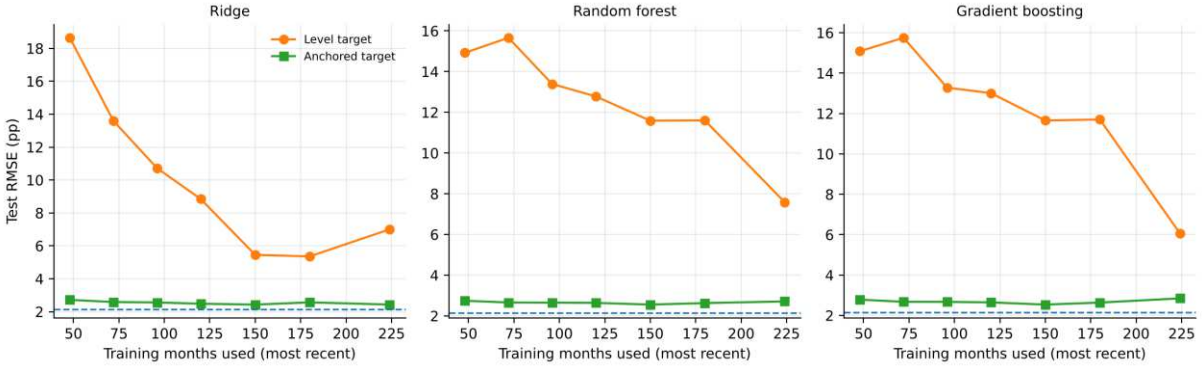


Figure 6. Learning curves. Test RMSE when learners are fitted to the most recent k months before the test window, for level-target and anchored designs (static fits). The dashed line is the ARIMA anchor.

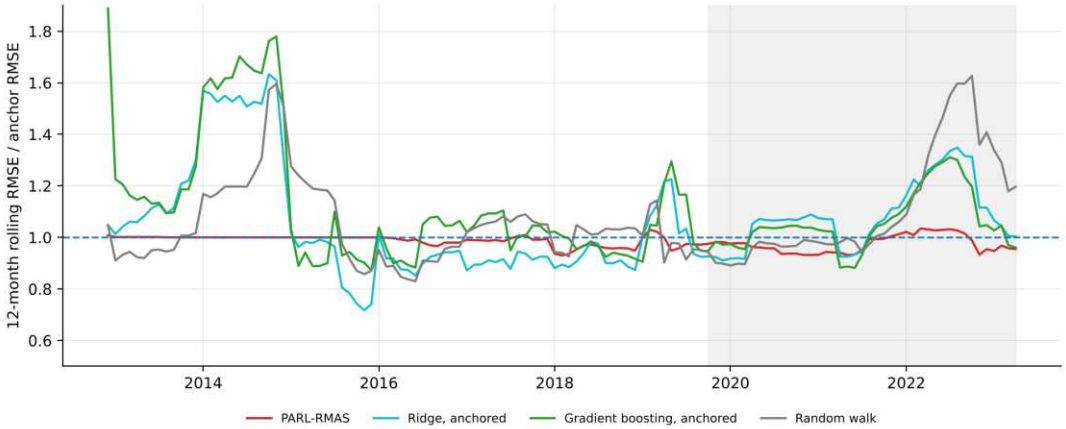


Figure 7. Twelve-month rolling RMSE relative to the anchor (values below 1 favor the model shown). The shaded area is the test window.

4.6 Prediction Intervals

Table 8 gives the interval results for the PARL-RMAS point forecast. In the test window the Gaussian intervals cover 59.5 percent of observations at the nominal 80 percent level and 66.7 percent at the 90 percent level. During the 2022–23 surge the 90 percent coverage is only 33 percent. Split-conformal intervals are no better, since a half-width calibrated on calm years is too narrow (64 percent at the 90 percent level). The locally scaled adaptive conformal intervals (LS) cover 73.8 and 85.7 percent, and their interval scores are lower than the Gaussian ones by about 15 percent at the 80 percent level (6.97 against 8.25) and by about 26 percent at the 90 percent level (9.94 against 13.41). Coverage is still below nominal in the surge (80 percent at the 90 percent level, 15 months), which is what one would expect after a structural break, but the shortfall is much smaller.

Table 8. Interval forecasts around the PARL-RMAS point forecast

Nominal level	Method	Coverage, test window	Coverage, surge (n=15)	Mean width	Interval score
80%	Gaussian	59.5%	26.7%	1.69	8.24
	Split conformal	42.9%	20.0%	1.11	9.25
	Adaptive conformal, local scale (LS)	73.8%	73.3%	3.41	6.97
	Adaptive conformal, regime scale (RS)	73.8%	73.3%	4.60	7.38
90%	Gaussian	66.7%	33.3%	2.17	13.41
	Split conformal	64.3%	26.7%	1.85	14.91
	Adaptive conformal, local scale (LS)	85.7%	80.0%	5.84	9.94
	Adaptive conformal, regime scale (RS)	85.7%	80.0%	7.70	11.94

Note: Intervals use information up to t−1 only. The interval score (Gneiting & Raftery, 2007) is the mean of width plus a penalty for misses; lower is better.

Making the scale regime-dependent (RS) widened the intervals without improving coverage, so it scored worse than LS (11.94 against 9.94 at the 90 percent level). As with the weight, regime information was not useful once adaptive scaling was in place. Intervals around the anchor give the same picture (supplementary table). With 42 test observations the standard error of an empirical coverage rate is about 5 to 7 points, so the differences between G and LS in the surge are clear but the exact coverage figures are imprecise.

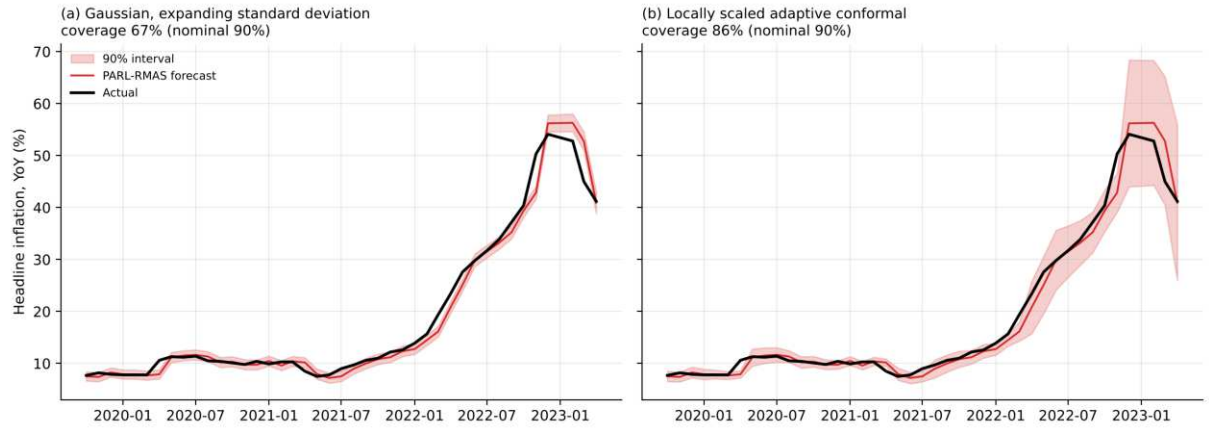


Figure 8. Ninety percent intervals around the PARL-RMAS forecast in the test window: Gaussian (a) and locally scaled adaptive conformal (b).

4.7 Residual Diagnostics

Figure 9 shows the PARL-RMAS residuals on the test window. Their mean is 0.33 points (the model is slightly too low on average), their standard deviation is 2.08 and their first-order autocorrelation is 0.26. The absolute error is positively related to the lagged level of inflation (correlation 0.59), so errors are larger when inflation is high. This supports the use of a local error scale in the intervals.

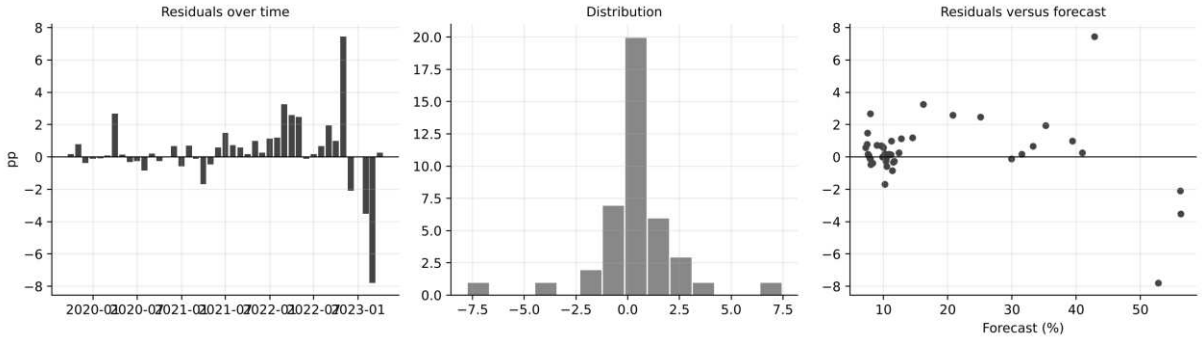


Figure 9. Residual diagnostics of PARL-RMAS in the test window: over time (left), distribution (center) and against the forecast (right).

5. Discussion

5.1 What the Results Say

Read on its own, the baseline comparison would suggest that simple models beat machine learning in Ghana. The full evidence supports a narrower conclusion. Level-target machine learning performs badly on a persistent series that breaks out of its training range, and part of that is mechanical: trees cannot extrapolate and every learner has to relearn persistence. Writing the problem in terms of changes removes most of the gap, and a hybrid that corrects an autoregressive anchor gains a few percent more. The simple anchor stays a very strong forecast for one month ahead. We would not claim from these data that machine learning adds reliable accuracy for Ghanaian inflation; we claim that it can be made competitive, and that the choice of target explains much more of its earlier failure than the choice of algorithm.

This agrees with the replication evidence of Naghi et al. (2024), who found that random forests lost their edge in the 2020–22 period, and with the argument of Goulet Coulombe et al. (2022) that target transformations determine whether nonlinearity helps. It also helps reconcile our baseline results with the favorable results in Adekoya et al. (2021), Akande et al. (2021) and Baidoo and Obeng (2024). Their designs differ from ours in target, validation and test period, so we do not claim a direct contradiction, but a model evaluated on a calm holdout would not meet the extrapolation ceiling at all.

5.2 Why Regime Modulation Did Not Help

We expected the regime model to make the weight and the interval scale more responsive when inflation turned turbulent. It did not, and we think there are two reasons. The practical one is calibration: a regime-specific memory can only be learned from turbulent observations, and Validation has none. The more basic one is that the filtered probability reacts after the first large changes, by which time the exponentially weighted statistics have already registered the error. The regime model is useful as a description (Figure 4) but, in this sample, adds nothing to forecasts that the adaptive weights and scales do not already give. We report this because the regime idea is natural and others may want to try it with a calibration window that includes a high-inflation episode.

5.3 Implications for Forecasting Practice

For a forecasting team the practical advice is modest. Keep an autoregressive anchor as the headline forecast. If machine learning models are used, give them the anchor's errors or the monthly change as the target, not the level, and check whether any tree-based component is hitting the largest value seen in training. Publish intervals that adapt to recent error size, because fixed-width or Gaussian intervals

understated uncertainty badly in 2022. Finally, treat any improvement from learners as provisional unless it survives a test period that includes a stress episode.

5.4 Limitations and Future Research

Several limitations remain. The test window has 42 observations and one dominant episode, with only 15 months in the surge proper, so statistical power is low and several of our tests cannot separate models that are within 5 percent of each other. The gate parameters were chosen on 24 calm validation months. The anchored ensemble includes learners whose penalties were chosen at the edge of the grid and therefore predict almost constant corrections; we report this and do not tune further. The anchor's estimator differs from the baseline ARIMA (2.15 against 2.41 RMSE), so the improvement of PARL over "ARIMA" depends on which version is meant; relative to the baseline ARIMA, PARL's RMSE is about 13 percent lower, but most of that difference comes from re-estimating the anchor, not from the learners.

The data have gaps in coverage that matter. There is no exchange-rate series, no food-price index and no survey of inflation expectations, all of which are plausible drivers in Ghana. The Bank of Ghana series are final-vintage, not real-time vintages, so some of what we call real-time forecasting benefits from later data revisions. The study covers one country, one target and one horizon. An LSTM (Hochreiter & Schmidhuber, 1997) was again not used because of the sample size, and conclusions about deep learning refer to the one-hidden-layer MLP only. XGBoost results come from the baseline run, and the anchored boosting results use scikit-learn's implementation.

Future work could add a calibration window that includes a high-inflation regime, for example by pooling several West African countries, to see whether regime-dependent memory then helps. Other natural steps are forecasts for food and core inflation, multi-step horizons, real-time data vintages, and richer predictors such as the exchange rate and news-based indicators.

6. Conclusions

This paper started from a negative result: nine standard models, including six machine learning and deep learning learners, failed to beat an ARIMA benchmark and a random walk for Ghanaian headline inflation through the 2022–23 surge. We have shown that much of that failure is due to the way the target is written. Level-target learners must relearn persistence from a short sample and, if they are tree based, cannot forecast above the highest inflation rate seen in training (41.95 percent, against a peak of 54.1). Predicting the monthly change instead lowers the test RMSE of the nonlinear learners by 64 to 80 percent, and the learners then perform close to the anchor regardless of training-window length. The proposed PARL framework, an autoregressive anchor with an ensemble of learners trained on its errors and an adaptive shrinkage weight, gives a test RMSE of 2.08 percentage points against 2.15 for the anchor. This 3 percent gain is not statistically significant against the anchor, though it is against the random walk. Making the weight or the interval scale depend on a Markov-switching regime probability did not help. Locally scaled adaptive conformal intervals were the most useful addition: they raised coverage of nominal 90 percent intervals from 67 to 86 percent and lowered the interval score by about a quarter. For the Bank of Ghana and similar institutions the lesson is that a persistence anchor should stay at the core of one-month-ahead inflation forecasts, that machine learning belongs in a supporting role if it is given a well-posed target, and that uncertainty ranges must adapt to recent error size. The limitations above, in particular the short test window and the missing exchange-rate and expectations data, mean that these conclusions should be tested on longer samples and on other countries.

References

Acquah-Sam, E. (2017). Influencers of Inflation in Ghana. *European Scientific Journal, ESJ*, *13*(7), 140. https://doi.org/10.19044/esj.2017.v13n7p140

Adekoya, A. F., Nti, I. K., & Weyori, B. A. (2021). Long short-term memory network for predicting exchange rate of the Ghanaian cedi. *FinTech*, *1*(1), 25–43. https://doi.org/10.3390/fintech1010002

Akande, E., Akanni, E., Taiwo, O. F., Joshua, J. D., & Anthony, A. (2021). *Predicting inflation component drivers in Nigeria: A stacked ensemble approach* [Preprint]. https://doi.org/10.21203/rs.3.rs-1080209/v1

Andrle, M., Berg, A., Morales, R. A., Portillo, R., & Vlcek, J. (2013). Forecasting and monetary policy analysis in low-income countries; food and non-food inflation in Kenya. *IMF Working Papers*, *13*(61), 1. https://doi.org/10.5089/9781475537444.001

Atkeson, A., & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review, 25*(1), 2–11.

Baidoo, T. G., & Obeng, A. (2024). Navigating Inflation in Ghana: How Can Machine Learning Enhance Economic Stability and Growth Strategies? arXiv preprint arXiv:2410.05630. https://doi.org/10.48550/arXiv.2410.05630

Box, G.E.P., & Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day.

Breiman, L. (2001). Random Forests. Machine Learning, 45. https://doi.org/10.1023/A:1010933404324

Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters. arXiv preprint arXiv:1909.13316. https://doi.org/10.48550/arXiv.1909.13316

Charemza, W., Kharin, Y., & Maevskiy, V. (2014). Bilinear forecast risk assessment for non-systematic inflation: Theory and evidence. In *Dynamic modeling and econometrics in economics and finance* (pp. 205–232). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-42039-9_6

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/2939672.2939785

Chugunov, I., Pasichnyi, M., & Nepytaliuk, A. (2019). Macroeconomic effects of inflation targeting in advanced and emerging market economies. *Banks and Bank Systems*, *14*(4), 153–165. https://doi.org/10.21511/bbs.14(4).2019.15

Chugunov, I., Pasichnyi, M., Kaneva, T., Nepytaliuk, A., & Koroviy, V. (2021). The influence of inflation targeting on economic growth in the OECD and developing countries. *Montenegrin Journal of Economics*, *17*(2), 173–186. https://doi.org/10.14254/1800-5845/2021.17-2.14

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*(1), 291–311. https://doi.org/10.1016/j.jeconom.2006.05.023

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in neural information processing systems*, *9*.

Filho, F. F., & Milan, M. (2018). Excess real interest rates and the inflation targeting regime in Brazil: Monetary policy ineffectiveness and rentiers' interests. *Applied Economics and Finance*, *5*(6), 84–100. https://doi.org/10.11114/aef.v5i6.3710

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

George-Anokwuru, C. C. (2023). Monetary policy and misery index in Nigeria. *European Journal of Economic and Financial Research*, *7*(2). https://doi.org/10.46827/ejefr.v7i2.1494

Gibbs, I., & Candès, E. (2021). Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems, 34*. https://doi.org/10.48550/arXiv.2106.00170

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association, 102*(477), 359–378. https://doi.org/10.1198/016214506000001437

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

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*(2), 357–384. https://doi.org/10.2307/1912559

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497. https://doi.org/10.3982/ECTA5771

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics, 135*(1–2), 399–426. https://doi.org/10.1016/j.jeconom.2005.07.033

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoerl, A.E., & Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. https://doi.org/10.1080/00401706.1970.10488634

Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

International Monetary Fund. (2013). *Uganda: First review under the policy support instrument* (IMF Staff Country Reports No. 13/375). https://doi.org/10.5089/9781475522754.002

International Monetary Fund. (2018). *Uganda: Technical assistance report—Monetary and foreign exchange operations, recapitalization, and act revision* (IMF Staff Country Reports No. 18/10). https://doi.org/10.5089/9781484337813.002

Islam, T. U., & Ahmed, D. (2023). Inflation targeting: A time-frequency causal investigation. *PLOS ONE*, *18*(12), e0295453. https://doi.org/10.1371/journal.pone.0295453

Jackson, E. A., Tamuke, E., Jabbie, M., & Ngombu, A. (2020). Adoption of inflation targeting in Sierra Leone: An empirical discourse. *Journal of Economic Policy Researches*, *7*(2), 21–50. https://doi.org/10.26650/jepr735604

Kunawotor, M. E., Bokpin, G. A., Asuming, P. O., & Amoateng, K. A. (2021). The impacts of extreme weather events on inflation and the implications for monetary policy in Africa. *Progress in Development Studies*, *22*(2), 130–148. https://doi.org/10.1177/14649934211063357

Madhou, A., Sewak, T., Moosa, I. A., Ramiah, V., & Gerth, F. (2021). Towards full-fledged inflation targeting monetary policy regime in Mauritius. *Journal of Risk and Financial Management*, *14*(3), 126. https://doi.org/10.3390/jrfm14030126

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics, 39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745

Muhammed-Gani, S. H., Nansir, A. M. M., & Ibrahim, A. (2022). Analysis of inflation convergence in the Economic Community of West African States (ECOWAS) zone (1987–2021). *International Journal of Research Publications*, *102*(1). https://doi.org/10.47119/ijrp1001021620223328

Naghi, A. A., O'Neill, E., & Danielova Zaharieva, M. (2024). The benefits of forecasting inflation with machine learning: New evidence. *Journal of Applied Econometrics, 39*(7), 1321–1331. https://doi.org/10.1002/jae.3088

Ncanywa, T., & Ralarala, O. (2020). The effects of expectation inflation on aggregate demand: Examining the South African inflation expectation channel of South African monetary policy. *Journal of Reviews on Global Economics*, *9*, 437–445. https://doi.org/10.6000/1929-7092.2020.09.41

Ngarambe, E., Shukla, & Hitayezu. (2016). Monetary policy and inflation control in Rwanda: 2006–2015. *International Journal of Advanced Research*, *4*(8), 1085–1096. https://doi.org/10.21474/ijar01/1308

Nyamao, N. R. (2014). A multiplicative autoregressive integrated moving average model for Kenya's inflation (2000:1–2013:12). *Science Journal of Applied Mathematics and Statistics*, *2*(6), 122. https://doi.org/10.11648/j.sjams.20140206.14

Ohemeng, W., Agyapong, E. K., & Ofori-Boateng, K. (2021). Exchange rate and inflation dynamics: Does the month or quarter of the year matter? *SN Business & Economics*, *1*(6), 1–24. https://doi.org/10.1007/s43546-021-00074-5

Oladipo, O. (2017). Inflation targeting and exchange rate pass-through to domestic prices: Evidence from South Africa. *Journal of Economic & Financial Studies*, *5*(5), 1–11. https://doi.org/10.18533/jefs.v5i05.294

Osei-Dwomoh, E., & Forkuo, G.O. (2026). Artificial intelligence in African economics: a systematic review of performance and ethical risks. *Future Business Journal,* 12(1), 84. https://doi.org/10.1186/s43093-026-00776-y

Otu, O. A., George, A. O., Jude, O., Ifeyinwa, M. H., & Andrew, I. I. (2014). Application of SARIMA models in modelling and forecasting Nigeria's inflation rates. *American Journal of Applied Mathematics and Statistics*, *2*(1), 16–28. https://doi.org/10.12691/ajams-2-1-4

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics, 137*(1), 134–161. https://doi.org/10.1016/j.jeconom.2006.03.010

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning Representations by Back-Propagating Errors. Nature. https://doi.org/10.1038/323533a0

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting, 36*(1), 75–85. https://doi.org/10.1016/j.ijforecast.2019.03.017

Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking, 39*(s1), 3–33. https://doi.org/10.1111/j.1538-4616.2007.00014.x

Svensson, L. E. O. (2010). *Inflation targeting* (NBER Working Paper No. 16654). National Bureau of Economic Research. https://doi.org/10.3386/w16654

Tassi, P. (2019). Introduction – The contributions of big data. *Économie et Statistique / Economics and Statistics*, 5–15. https://doi.org/10.24187/ecostat.2018.505d.1963

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society: Series B. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer.

World Bank. (2020a). *Nigeria development update, June 2020*. World Bank. https://doi.org/10.1596/34046

World Bank. (2020b). *Rwanda economic update, January 2020*. World Bank. https://doi.org/10.1596/33247

World Bank. (2012). *South Africa economic update, July 2012*. World Bank. https://doi.org/10.1596/26076

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B, 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x