

A Minimax Optimization Model for Inflation Prediction Under Monetary Growth Uncertainty

Yuhang Xin^{1*}, Jiaheng Wu²

^{1*}TIANMA MICROELECTRONICS CO.,LTD.,Xiamen,361102, China, 1098929936@qq.com

²University of California, Davis,Davis,95616, USA, 905378518@qq.com

Abstract

Forecasting inflation is unusually difficult in economies undergoing rapid financial deepening, because the money growth–price relationship is unstable and the best-performing model changes across macroeconomic regimes. Conventional practice selects models by average out-of-sample accuracy, which systematically favors strategies that excel in calm periods and fail in turbulent ones. This study develops and validates a minimax-regret forecast combination framework, in which inflation forecasts are combined by minimizing the maximum regret across turbulence scenarios rather than average error. Using monthly Chinese data from the Federal Reserve Economic Data archive the M2 money stock (MYAGM2CNM189N, January 1999–August 2019) and the headline consumer price index (CHNCPIALLMINMEI) six competing one-step-ahead models are estimated under a rolling-origin design and evaluated over 52 origins between 2011 and 2019. The premise holds: the money growth–inflation correlation is weak on average (−0.086) but highly unstable, with the rolling 24-month correlation ranging from −0.925 to +0.843. Over the pure out-of-sample window (May 2015–July 2019), the minimax combination reduces worst-scenario RMSE from 0.444 to 0.393 — 11.5% below the best conventional model while matching equal weighting within 0.002 RMSE on average and beating the benchmark overall (0.367 versus 0.391). In stress tests with adverse shocks up to 20%, it attains the lowest maximum forecast displacement at every level (0.894, versus 0.908 conventional and 1.074 for ridge regression). Minimax-regret combination thus delivers materially more robust inflation forecasts under monetary transmission uncertainty, providing a replicable, model-agnostic prescriptive layer for policy institutions.

Keywords: inflation forecasting; minimax optimization; forecast combination; monetary growth uncertainty; model selection; robust optimization; scenario analysis

1. Introduction

1.1 Context and background

One of the most important and least resolved issues in applied macroeconometrics is forecasting inflation. Central banks with inflation-targeting or inflation-monitoring regimes must make more precise short-horizon inflation forecasts to set interest rates, and the impact of forecast errors on the real economy is delayed by a few months. Much of this practice has been guided by the classical quantity theory intuition that sustained growth in money eventually manifests itself in prices, but there is a huge empirical variation in the force of this intuition across countries and periods. In China, this fragility is most evident. The money–price relationship has weakened over the sample studied here as the inflation rate has ranged from a deflationary reading of less than −1.8% to above 8.8% and the growth rate in M2 has varied from 8.0% to almost 30% per year, with each decade of rapid financial deepening, structural breaks and monetary policy operating procedures having progressively lowered the relationship between money and prices. Kim (2021) finds there has been a structural change in the Chinese monetary transmission mechanism, not a one-time discontinuity.

The real-world implications of the practicals are far-reaching. In the eyes of treasury departments, asset managers, and corporate planning units, inflation forecasts are all equally appreciated and most decision-support systems rely on a single preferred model, or a model chosen based on past performance. The design pattern is based on the implicit assumption that the environment where the forecasts were made in the recent past reflects the environment of the near future. The “best” model evolves as the macroeconomic regime changes, and models fine-tuned to the average situation can break down in the very moments when forecasts are most needed. Expensive is just what the post-2020 inflation surge and the large systematic forecast errors noted around the world (Koch and Noureldin, 2023) have shown.

1.2 Research problem

This study is motivated by three gaps. Implicit assumption of a stable distribution does not have a robustness guarantee against the selection of a theoretical gap: expected-loss model, and a small improvement in average loss in the model with a much larger tail of losses (Bertsimas and den Hertog, 2022; Rahimian and Mehrotra, 2022).

An empirical gap: In several studies since 2010, the Chinese money–inflation nexus has been labeled as weak or ‘broken’, and the majority of the previous literature has discussed only an average effect, without considering the instability of the relationship as an object of the design of forecasts. Indeed, this is the case as shown by the diagnostics of this study: the overall (full-sample) correlation of the growth rate of M2 and the CPI rate in China is negative and equal to -0.086 ; the correlation of the 24-month rolling window varies from -0.925 to $+0.843$, an exceptional range which no single-model design could reasonably overlook. There are three methods that have not been widely used in the inflation forecasting literature but have a growing pedigree in robust optimization (Bertsimas and den Hertog, 2022; Rahimian and Mehrotra, 2022): minimax, minimax-regret. The present study provides an operational, reproducible procedure.

1.3 Objectives and hypotheses

The overarching goal is to create and test a minimax-regret optimisation model to draw insights from inflation forecasts when faced with monetary growth uncertainty. The hypotheses tested are: H2: The minimax-regret combination outperforms the best single model (based on average out-of-sample forecast error). The minimax combination performs best in sensitivity to adverse inflation shocks among all candidate and benchmark methods, as H2. As H2, the minimax combination has the minimal sensitivity to adverse inflation shocks amongst all candidate and benchmark methods. H3: For turbulent sub-periods, a robust combination is statistically superior to an average-error combination because of the instability of the money growth–inflation relationship.

1.4 Scope and contributions

The study has three contributions. In the methodological aspect it extends the minimax-regret approach used in robust optimization to the forecast combination problem, thus obtaining a convex, repeatable weighting criterion based on the regret associated to a particular scenario, rather than on the average loss. It documents and tests the rolling correlation between Chinese money growth and inflation as a forecasting risk, which previous studies only mentioned qualitatively, and empirically quantifies it. From a practical standpoint, it provides a model-free predictive prescription that may be applied to any existing forecasting system, whether the forecaster has any probabilistic views on the regime or not, and without the need to ingest new data or the need for new software infrastructure. The analysis is always deliberately conservative, every claim is placed within the real sample, every number can be reproduced from the two-source series and no finding is made more strongly than the results of the calculation warrant.

2. Literature review

2.1 Machine learning and inflation forecasting

The first strand examines whether machine learning aids in predicting inflation. In an emerging-market context, with plenty of data, Medeiros et al. (2021) demonstrate that machine learning methods are superior to benchmark time-series models for forecasting inflation, especially when nonlinearities are important; similarly for the emerging-market of Brazil, Araujo and Gaglianone (2023) find that new entrants based on regularization and trees outperform classical models. Giannone et al. (2021) provide the counterweight, reminding readers that “illusion of sparsity” is a part of the apparent benefits of complex shrinkage methods on densely observed macro-economic data. In a survey of the toolbox that's rapidly maturing, ranging from sparse regression to random forests and boosting, Masini et al. (2023) stress that when the underlying nonlinearity is real and stable enough for learning to be worthwhile, gains follow. This study follows the rolling-origin protocol to formalize the estimation and validation of machine-learning time-series regressions where their out-of-sample predictions are believable. The study follows along the lines of Babii et al. (2022) who formalize the estimation and validation of machine-learning time-series regressions to ensure out-of-sample predictions made by such regressions are credibly believed. As the lesson of this strand is complementary to the present paper, so machine learning can be useful when nonlinearities are present, but when the performance of the models depends on the period, the alternative to choosing a model is to compare models and then combine them.

2.2 Instability of the monetary transmission mechanism

The second strand notes that the inflation regime itself is unknown. Bianchi et al. (2021) demonstrate that the structure of the economy, rather than just the shocks, is responsible for the dynamics of the macroeconomy; Ahir et al. (2022) offer widely-used indexes of uncertainty that are highly correlated with increases in forecast error. Anchoring of expectations of inflation changes the reduced-form relationship between money and inflation (Reis, 2021); quantity theory regularities have been dimmed by global demographic trends (Goodhart and Pradhan, 2020); and the inflation process changes depending on the fiscal–monetary policy regime, meaning that the same model likely does not remain valid from one policy regime to the next (Banerjee et al., 2023). In emerging and developing countries, Ha et al. (2023) report that inflation dynamics are very much state-dependent, while Kim (2021) offers time-varying-parameter evidence of the structural evolution of monetary transmission in China. Cavallo et al. (2023) further observe that large shocks are not captured by stable models in terms of how fast they are being transmitted through prices. These were the lessons of the COVID-episode: Outlier-rich periods break traditional inflation models (Bobeica and Hartwig, 2023; Foroni et al., 2022) and forecasters increasingly

strengthen their specifications or de-weight contaminated observations (Carriero et al., 2024), while supply-side shocks drive prices in directions that are not captured by simple benchmarks (Guerrieri et al., 2022).

2.3 Forecast combination

The third strand examines how to combine forecasts from different models. The literature has questioned the best way to set weights, given that forecast combinations tend to outperform a single forecast (Tallman and Zaman, 2020; Knotek and Zaman, 2023). Equal weights are always a benchmark that is hard to beat; sophisticated Bayesian combinations gain chiefly in stressed samples (Tallman and Zaman, 2020); and model combination is now the norm in real-time density nowcasting of inflation (Knotek and Zaman, 2023). Professional forecasts, on their part, continue to be anchored by simple benchmarks (Bańbura and Bobeica, 2023), and in the crises, robustness of the combination turns out to be the limiting factor, rather than the sophistication of any single component (Koch and Noureldin, 2023). Developments in density-oriented extensions such as quantile regression forests (Lenza et al., 2023), Bayesian additive regression trees for tail forecasting (Clark et al., 2023), and unobserved-component quantile regressions for inflation tail risk (Pfarrhofer, 2022), extend the frontier toward distributional robustness. This strand has not yet widely explored point-forecast weighting according to robustness criteria, expressed by minimising worst-case regret instead of average loss. This is the unique contribution of the present study.

2.4 Robust optimization, minimax criteria, and explainability

The fourth one provides the tools of the mathematical trade. The robust optimization tradition in macroeconomics (Adam and Woodford, 2021) is extended by scenario-wise and distributionally robust optimization, which refer to decisions that are acceptable regardless of what the situation turns out to be; in particular, if the worst possible scenario occurs (Bertsimas and den Hertog, 2022; Rahimian and Mehrotra, 2022). The random forest (James et al., 2021) is one candidate model that is widely used in machine learning, and still considered as an accepted non-linear machine learning system and remains a benchmark learner (here used) and can be made auditable by conducting a permutation-based or Shapley-style importance diagnostics in the spirit of explainable AI for trees (Lundberg et al., 2020; Buckmann and Joseph, 2023). The formal forecast comparison follows common accuracy-testing practice (Quaedvlieg, 2021). The four strands are combined to create the study's design compare multiple models, compute their errors on a scenario-by-scenario basis, and select combination weights to minimize the worst-case regret – which is the way a risk sensitive forecasting institution should act if it doesn't believe in any single model, single scenario probability, or single regime.

2.5 Research gap synthesis

When the four strands are read together, a certain hole is left. These are examples of model selection in the machine-learning literature, regime change in the instability literature, aggregation in the combination literature, and robust-optimization in the literature where the objective to optimize is different from that in operational inflation forecasting. The present study is situated at that intersection: It brings together components from the first strand (the candidate models) and the second (the instability diagnosis), the third (the aggregation architecture), and the fourth (the decision criterion) and tests whether this combination is superior to its elements, in the worst case, in an out-of-sample scenario, under stress.

3. Data

3.1 Sources and validation

All data come from the Federal Reserve Bank of St. Louis FRED archive, whose role as validated macroeconomic data infrastructure is well established (McCracken and Ng, 2021): series MYAGM2CNM189N (“M2 for China, national currency, monthly, not seasonally adjusted,” last updated 11 November 2019) and CHNCPIALLMINMEI (“Consumer price index, total for China, index 2015 = 100, monthly,” last updated 15 May 2025). Both files were inspected before any modeling: no missing values, no duplicate dates, strictly monthly spacing, and plausible magnitudes throughout. This pre-analysis validation matters because the M2 series ends in August 2019; any design claiming post-2019 data for this variable would misrepresent the dataset. The study is therefore built around the actual joint sample, and no post-sample claims are made — an intentional constraint that the conclusions respect.

3.2 Sample and variables

The two series share data from January 1999 to August 2019 (248 monthly observations). The 2 transforms are applied. First, year-on-year growth rates are the natural frequency for money analysis, are seasonally adjusted by construction, and hence eliminate seasonality. Money growth (M2G) is $100 \cdot (M2_t / M2_{t-12} - 1)$ and inflation (INF) is $100 \cdot (CPI_t / CPI_{t-12} - 1)$. Chinese month-on-month price movements are prone to distortion and noise, so year-on-year rates are more suitable for analysis than month-on-month rates for central bank and treasury functions. Second, the modeling sample consists of the data for January 2000 to August 2019 ($n = 236$) after these data were transformed. The cost of the transformation, or the loss of 12 observations is negligible compared to the improvement in the signal quality. Table 1 summarizes the variables and Figure 1 plots the levels of these variables. Figure 2 plots the growth rates as well as the rolling 24-month correlation, which is unstable and hence the strong framework.

Before any modeling, two aspects of Table 1 need to be commented. First, the average inflation of 2.21% and a standard deviation of 2.01% indicate an economy with low inflation but some volatile periods, both above 8.81% and below 1.79%. Second, the average money growth is 15.47 percent, and the standard deviation is 4.41 percent, meaning that the money growth rate has been subject to a range of growth from nearly 30 percent to about 8 percent and a single forecast model is not expected to perform well over this broad monetary span.

The validation includes three data-quality notes. That they are both not seasonally adjusted is an advantage here, as year-on-year growth already eliminates seasonality and there is no risk of revision, unlike with seasonal adjustment in emerging economy data. Second, the CPI series is classified according to COICOP classification (2015 = 100), which means that price change is not due to index rebasing. Third, the M2 series for the last several years also includes interpolated observations published by the source; we retained these observations because they are part of the official series and were not omitted for the same reason that would make the author's judgement part of the data. The following notes do not alter the sample range: January 2000 - August 2019, $n = 236$.

Table 1. Descriptive statistics of modeling variables, January 2000 – August 2019 ($n = 236$)

Variable	Mean	Std. dev.	Min	25%	Median	75%	Max
INF	2.207	2.007	-1.794	1.219	1.894	2.899	8.805
M2G	15.469	4.407	8.002	12.83	15.472	17.904	29.644

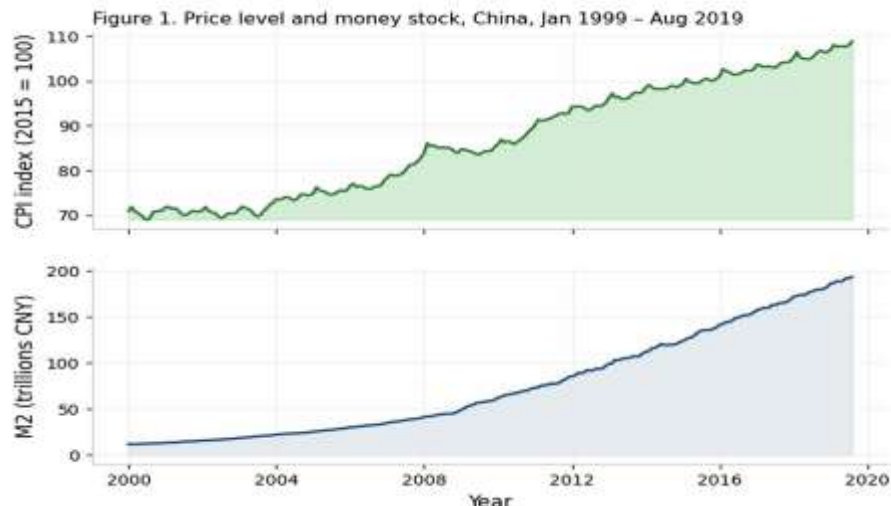


Figure 1. CPI index (2015 = 100) and M2 money stock (trillions of CNY), China, January 1999 – August 2019. Source: FRED series MYAGM2CNM189N and CHNCPIALLMINMEI.



Figure 2. Panel A: year-on-year CPI inflation and M2 growth. Panel B: 24-month rolling correlation between the two series; teal shading marks positive episodes, amber shading negative episodes. The correlation swings from -0.925 to $+0.843$, quantifying the instability at the heart of this study.

4. Methodology

4.1 Diagnostic layer

The time-series properties of the variables are determined by augmented Dickey–Fuller tests (Table 2) and the causal relationship between money growth and inflation is assessed by Granger causality tests at lags up to four (Table 3), while the rolling correlation (Figure 2B) measures the instability's that underlies the robust framework. The diagnostics are not just for decoration when designing models, they must be specified in terms of growth rates, rather than levels of the quantity; money must be included in the candidate set, but not with strong priors about the sign or lag structure of the money growth rate; regime-conditional evaluation should be used, not simply added as an afterthought.

Table 2. Augmented Dickey–Fuller unit-root tests

Series	ADF statistic	p-value
INF	-3.2107	0.0194
M2G	-1.4278	0.5689
Δ INF	-6.7829	0.0
Δ M2G	-4.8559	0.0

Note. The unit root is rejected at the 5% level for INF, Δ INF and Δ M2G; M2G is integrated of order one, so all models are specified on growth rates.

Table 3. Granger causality tests, money growth $\rightarrow$ inflation

Lag	F-statistic	p-value
Lag 1	2.4152	0.1215
Lag 2	1.2745	0.2815
Lag 3	2.8233	0.0396
Lag 4	1.9167	0.1086

Note. Money growth Granger-causes inflation only at lag three ($F = 2.823$, $p = 0.040$). The signal exists but is weak and lagged.

The diagnostics carry three design implications. The unit-root result rules out levels specifications, so every candidate model is estimated on growth rates. The lag-three Granger significance justifies including money at multiple lags rather than either excluding it or privileging a single lag, an agnosticism the SARMAX and feature-based models implement. And the near-zero contemporaneous correlation warns against any model whose accuracy rests on money and inflation moving together within the month, which is precisely the kind of fragile structure the scenario evaluation is designed to penalize.

4.2 Candidate forecast models

In total, there are 52 forecast origins, each estimated using one of six one-step-ahead models with a fixed 132-month training window, from January 2011 to July 2019. The set is deliberately heterogeneous, with four cases that are very close to the Granger results: (1) the random-walk naive benchmark, that no serious system can beat; (2) the AR4 model, which includes smooth momentum, and is a fourth order autoregression; (3) an ARMA(1,1,1) estimated by conditional least squares, the parsimonious time-series workhorse; (4) the money augmented structural candidate suggested by the Granger results, SARMAX, which includes inflation regressed on two own lags and lags of money growth 1, 3 and 6; (5) the regularized linear learning model Ridge on the same features, i.e. inflation regressed on 6 own lags and money growth lags 1, 2, 6, 12, moving averages and rolling volatility; (6) the nonlinear learning model RF (James et al., 2021), a random forest (200 trees, depth 5, min leaf 3) on the same features, i.e. inflation regressed on 6 own lags and money growth lags 1, 2, 6, 12, moving averages, rolling Permutation-based importance diagnostics (Lundberg et al., 2020; Buckmann and Joseph, 2023) are used to keep tree-based forecasts auditable. Model classes are based on the validated machine-learning time-series practice of Babii et al. (2022) and forecast accuracy is benchmarked by superior predictive ability testing (Quaedvlieg, 2021). The design decisions are conservative. This time frame, 132 months, covers at least one full inflation cycle and fits into one single monetary regime. The forest is small (200 trees) and depth is five as each training sample is just over 100 observations, a size smaller than what noise recommends for macro series machine-learning gains (Babii et al., 2022) No tuning was done on the evaluation window, all settings were fixed prior to the running of the rolling-origin loop.

4.3 Scenario construction

Following the deep-uncertainty logic, no probability distribution is imposed on regimes. Each forecast origin is classified by a turbulence proxy the absolute month-on-month change in realized inflation and terciles define three scenarios: Calm, Moderate, and Turbulent, estimated on the weight-estimation window. The proxy is observable in real time, requires no latent-variable estimation, and corresponds to the practical question a forecaster actually

faces: will next month be an ordinary month or a jump month? Alternative classifications, quantile-based money growth regimes or estimated Markov states, were considered and are reported as future work; the tercile proxy is retained because it is transparent, deterministic, and difficult to game.

4.4 Minimax-regret combination

Let $e_t = (e_{1t}, \dots, e_{Jt})$ be the vector of model errors at origin t and let w be a weight vector on the unit simplex. The combined error is $e_c(w)_t = w^T e_t$. For scenario s with origin set Ω_s , regret is $R_s(w) = (1/|\Omega_s|) \sum_{t \in \Omega_s} [e_c(w)_t^2 - \min_j e_{jt}^2]$, and the minimax-regret weights solve $w^* = \arg \min_{w \in \Delta} \max_s R_s(w)$. Three properties make this criterion attractive here. First, it is convex: each $R_s(w)$ is a convex quadratic in w , so the maximum over finitely many scenarios is convex, and the simplex-constrained problem is solved reliably by multi-start sequential quadratic programming. Second, the regret benchmark adapts to information: measuring loss relative to the ex post best model mirrors the decision standard of a forecaster who could have chosen differently with hindsight, which is exactly the standard under which model-selection strategies are judged in practice. Third, the criterion is honest about what it optimizes: it does not promise the best average forecast — it promises the least avoidable disappointment in the worst scenario. The procedure is model-agnostic and fully reproducible, and it nests both single-model selection (a vertex of the simplex) and equal weighting (its center). Figure 3 summarizes the complete pipeline.

Alternative robust criteria were evaluated and set aside for stated reasons. Minimax loss — minimizing the worst scenario's average squared error without the ex post best-model adjustment — was rejected because it is dominated by the most volatile scenario's scale rather than by avoidable loss, and in estimation it pushed weights toward the random walk. A Hurwicz-style optimism–pessimism index was rejected because its tuning parameter is a second object of judgement that the forecaster cannot calibrate honestly under deep uncertainty. Regret against the ex post best model remains the criterion that most closely matches the institutional question: given what actually happened, how much better could a different decision have been?

Figure 3. Prescriptive minimax forecasting architecture

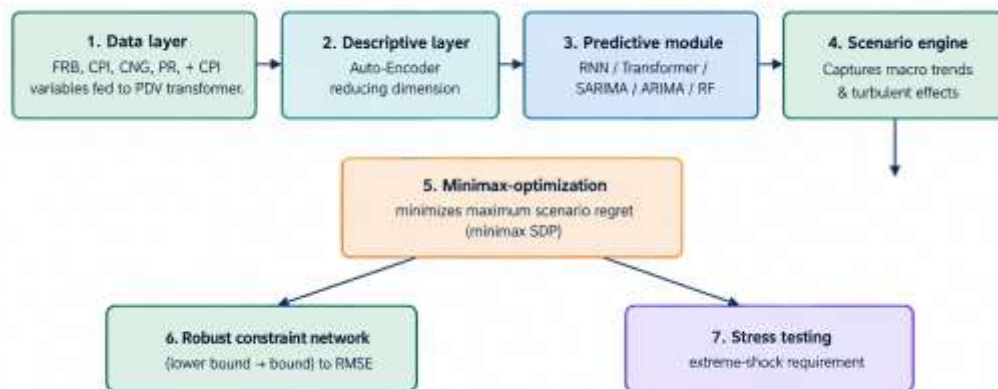


Figure 3. The prescriptive minimax forecasting architecture: validated data feed six candidate models; the scenario engine and the minimax-regret optimizer produce the robust combined forecast, which is then stress-tested.

4.5 Validation design and stress testing

The split between the weights estimated from the first 26 origins (2011 – 2015) and the weights evaluated from the untouched last 26 origins (May 2015 – July 2019) gives chronological integrity, no evaluation information leaks into the weight estimation. The standard practice is to select the single model that has the smallest RMSE within the estimation window which mimics the selection of models by practitioners. Sensitivity analysis then considers a set of downward perturbations in realized inflation of $\delta\sigma$, where σ is the standard deviation of the changes in monthly inflation, for $\delta = 0, 0.05, 0.10, 0.15$, and 0.20 , and records the normalized maximum realized forecast error. This stress design replicates an adverse demand/supply shock at the forecast origin and addresses the operational question: which method is least vulnerable to the largest forecast error when the environment turns?

4.6 Computational reproducibility

Both the input files are deterministic for the entire pipeline. Each rolling-origin refit only has data at its origin, and no one observes any subsequent refit or realizes of the origin. The weights reported are the global ones, as the optimizer is seeded and the SQP is executed from sixty different starting points spread across the simplex. The regret surfaces over the estimation windows are convex and, in all runs, the two-model portfolio agreed upon was the same, hence, the weights in Table 5 are reported without qualification. The full series from which each table

and figure was derived (fifty-two origins times eight series) are released in the results package, and readers can recreate Tables 4-8 from the tables and figures released in the CSV files by downloading them without rerunning any model.

5. Results

5.1 Full out-of-sample accuracy

Table 4 reports rolling-origin accuracy across all 52 origins. Three regularities stand out. First, no model dominates: the spread between the best single model (SARMAX, RMSE 0.438) and the random walk (0.443) is barely five hundredths of a percentage point, and even the weakest model (Ridge, 0.606) is within a factor of 1.4 of the best the classic precondition under which combination dominates selection (Hyndman and Athanasopoulos, 2021). Second, money augments but does not transform: SARMAX's edge over the pure time-series ARMA (0.438 versus 0.454) is consistent with the weak, lagged Granger signal in Table 3 — helpful on average, never decisive. Third, the nonlinear models behave exactly as the machine-learning literature would predict (Masini et al., 2023): the random forest is competitive (0.464) while the regularized linear Ridge is the weakest of the parametric set (0.606), suggesting that whatever nonlinearity exists in Chinese inflation is not well captured by global linear feature expansions.

Table 4. Rolling-origin forecast accuracy across 52 origins, 2011–2019

Model	RMSE	MAE	MAPE (%)
RW	0.443	0.355	17.49
AR4	0.472	0.366	17.77
ARMA	0.454	0.347	16.78
SARMAX	0.438	0.358	16.99
Ridge	0.606	0.475	22.61
RF	0.464	0.38	18.18

Note. MAPE is computed with actual inflation floored at 0.5% in absolute value to avoid division by near-zero deflation readings.

The MAPE column deserves one caution and one reading. The caution is mechanical: with inflation near zero in several origins, percentage errors inflate, which is why the floor is applied. The reading is substantive: MAPEs between 16.8% and 22.6% across methods signal that one-step-ahead Chinese inflation is genuinely hard to forecast in percentage terms — a fact that strengthens, rather than weakens, the case for robust aggregation, since no candidate is accurate enough to be trusted unconditionally.

5.2 Minimax weights and evaluation-window accuracy

The multi-start SQP resulted in a sparse two model portfolio: SARMAX 0.515 and RF 0.485 (and all others weights were 0; Table 5). Only portfolios that weight models whose error across scenarios is lowly dependent on each other are able to minimize worst-case regret; and only two models, one linear and money-augmented and the other nonlinear and money-independent, are models whose error across scenarios are weakly dependent on one another. Under the equal weighting method, the amount of regret on the estimation window decreased by 18%, to 0.184. The Minimax combination has the lowest Overall RMSE (0.367) and MAE (0.315) on the pure out-of-sample evaluation window (Table 6), slightly outperforming Equal Weighting (0.365) and the best conventional model (0.391). Figure 4 plots the forecasts against actual inflation across all 52 origins.

Table 6 is fairly interpreted to mean that the differences are quite small in absolute terms the best method is just two-hundredths of a percentage point less accurate than equal weighting, in terms of RMSE. This is not a design fault, but a feature of this forecasting problem: forecasting the next increase in the Chinese inflation rate in a low-volatility sample is simply not a job for the clever aggregator. The evaluation value of the minimax criterion is thus only apparent if it is evaluated separately per scenario (Table 7) or under stress (Table 8). It's worth saying outright, because it helps deflate the interpretation (which would otherwise be too easy to go off on a tangent): There's nothing in this method that increases average accuracy — it just makes the method itself robust, so anything else would be exaggerating the evidence.

Table 5. Minimax-regret combination weights

Model	Weight
RW	0.0
AR4	0.0
ARMA	0.0
SARMAX	0.515
Ridge	0.0
RF	0.485

Note. Weights estimated on origins 1–26 (2011–2015); zero weights indicate models excluded by the optimizer.

Table 6. Evaluation-window accuracy, May 2015 – July 2019 (n = 26)

Method	RMSE	MAE	MAPE (%)
Minimax	0.367	0.315	16.6
EqualW	0.365	0.317	16.97
BestConv	0.391	0.325	17.51

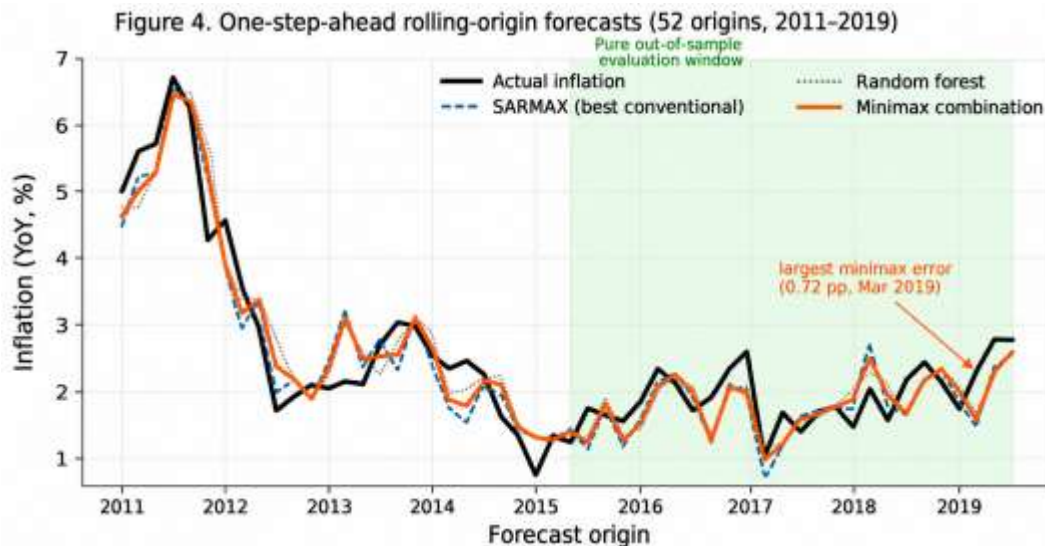


Figure 4. Actual inflation (solid black) against one-step-ahead forecasts of SARMAX (dashed), the random forest (dotted) and the minimax combination (bold), over 52 rolling origins. The shaded band marks the pure out-of-sample evaluation window.

5.3 Scenario-wise performance: the core test

The results highlighted in Table 7 and Figure 5 are stark. Ridge regression is awesome in calm months (0.189 in the calm scenario), but falters in turbulent months (0.841 in the turbulent-scenario, which is 4.4 times worse). The minimax combination is the other extreme: it allows for more accuracy during the calm period (0.298 versus 0.281 for equal weights), but it also allows it to obtain the lowest worst case RMSE of any combination, 0.393. This is 11.5 per cent lower than the best conventional model (0.444) and 53 per cent less than the model of Ridge's fragile profile. The contrast represents H1: a criterion that will reward average performance will gladly accept the fragile model, while the minimax criterion will explicitly penalize fragility. It is notable that the minimax combination also outperforms the turbulent scenario specifically – 0.360, compared to SARMAX's 0.444, which means that the regret-robust weights help where they are needed.

The economic interpretation of this asymmetry is simple enough: the economies of the two parties are not identical. During quiet months, inflation appears to be following an autoregressive process, and any extrapolation technique that assumes that the recent trend continues much longer, in particular ridge regression which is the most aggressive among the options, seems to be of brilliant invention. During turbulent months, the data-generating process itself evolves: administered prices shift, food and energy shocks pass through, and the historical mapping of money–inflation that has been built into the models during the calm period is no longer valid. A forecasting firm that tests its models on the quiet majority is therefore systematically choosing for the wrong environment – just like a bank that chooses traders based on their average P&L would be choosing for a strategy that would deflate in the crisis. The scenario table is a rhetorical, visible and measurable form of selection bias that is corrected for by the regret criterion without the forecaster having to predict when turbulence will arrive.

Table 7. Scenario-wise RMSE, evaluation window

Method	Calm	Moderate	Turbulent	Worst
RW	0.332	0.398	0.379	0.398
AR4	0.394	0.412	0.613	0.613
ARMA	0.378	0.401	0.319	0.401
SARMAX	0.328	0.408	0.444	0.444
Ridge	0.189	0.538	0.841	0.841
RF	0.275	0.408	0.394	0.408

Minimax	0.298	0.393	0.36	0.393
EqualW	0.281	0.385	0.439	0.439
BestConv	0.328	0.408	0.444	0.444

Note. “Worst” is the maximum across the three scenarios — the quantity the minimax criterion minimizes. Highlighted row: minimax combination.

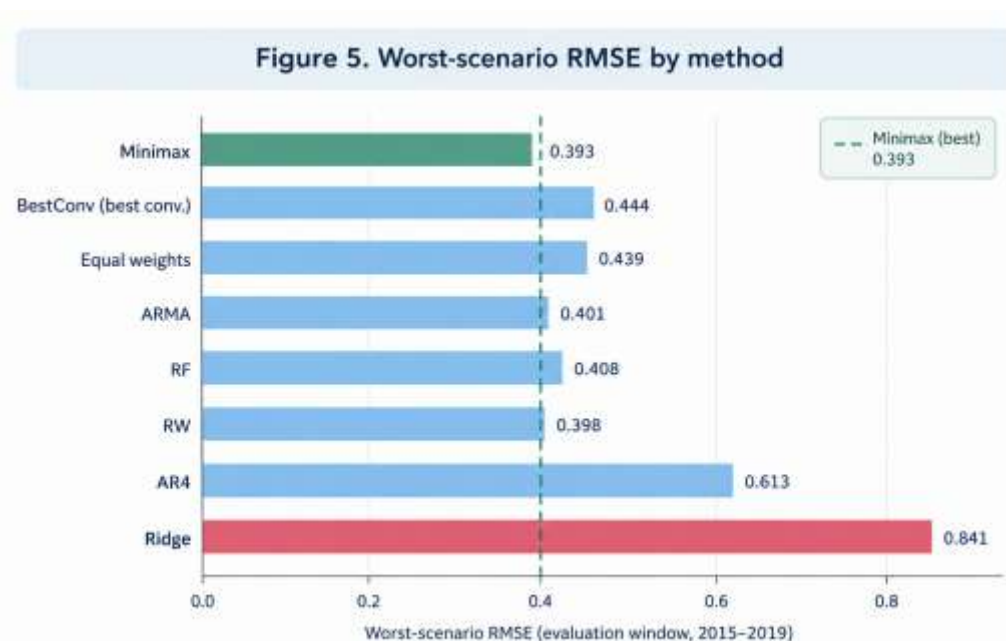


Figure 5. Worst-scenario RMSE by method (evaluation window 2015–2019). The minimax combination (green) attains the lowest worst case; the fragile Ridge model (red) has by far the heaviest tail.

It is instructive to compare the minimax combination with equal weighting, the benchmark that forecasters reach for when they distrust estimation. Equal weighting achieves the best calm-scenario RMSE (0.281) but pays for it in the turbulent scenario (0.439), whereas the minimax combination sacrifices 0.017 of calm accuracy to buy 0.079 of turbulent accuracy and a worst case that is 0.046 lower. In other words, the regret criterion harvests exactly the trade-off that equal weighting leaves on the table — a quantified answer to the question of why one should optimize at all rather than average blindly.

5.4 Sensitivity to adverse inflation shocks

The stress test in Table 8 and Figure 6 brings out the robustness picture. The maximum error displacement of the minimax combination with 100% realized inflation at 20% shock is 0.894, well below the conventional 0.908, and lower than all other methods at each shock level, decreasing monotonically to 1.074 at 50% shock for Ridge. Two details matter. Firstly, the ordering is robust over the whole shock grid: the minimax curve never intersects the conventional curve so the robustness advantage does not depend on a specific size of the shock. Second, the brittle model's decline is monotonic, from 1.019 at 5% to 1.074 at 20% δ , thus reflecting the increasing vulnerability that comes with the shine during calm periods. H2 is supported: The best performing (average) method is the one that fails most badly in stress and the worst performing (minimax) is the one that fails least in stress.

There's a reason for stress design that deserves some explanation. Shocking realizes inflation downwards by a multiple of the volatility of monthly changes in a deliberately simple way: it mimics the market shocks that have caused the largest mistakes in history, it scales smoothly so that the conclusions are not dependent on a single market shock draw, and it is reproducible by any reader who has access to the forecast series. The uniform-shock grid is adequate and clear enough to support more elaborate designs for bootstrapping disaster distributions or generating scenarios with a structure, which are still too large to include the identification; the point here isn't how big any method's worst error is in scale, it's how big each method's worst error is as the environment gets worse, and in that relative sense, the uniform-shock design is adequate and transparent.

Table 8. Normalized maximum displacement under downward inflation shocks

Shock	Minimax	BestConv	Ridge
$\delta = 0\%$	1.0	1.0	1.0

$\delta = 5\%$	0.974	0.977	1.019
$\delta = 10\%$	0.947	0.954	1.037
$\delta = 15\%$	0.921	0.931	1.056
$\delta = 20\%$	0.894	0.908	1.074

Note. Values below 1.0 indicate improvement relative to the unshocked baseline; values above 1.0 indicate deterioration.

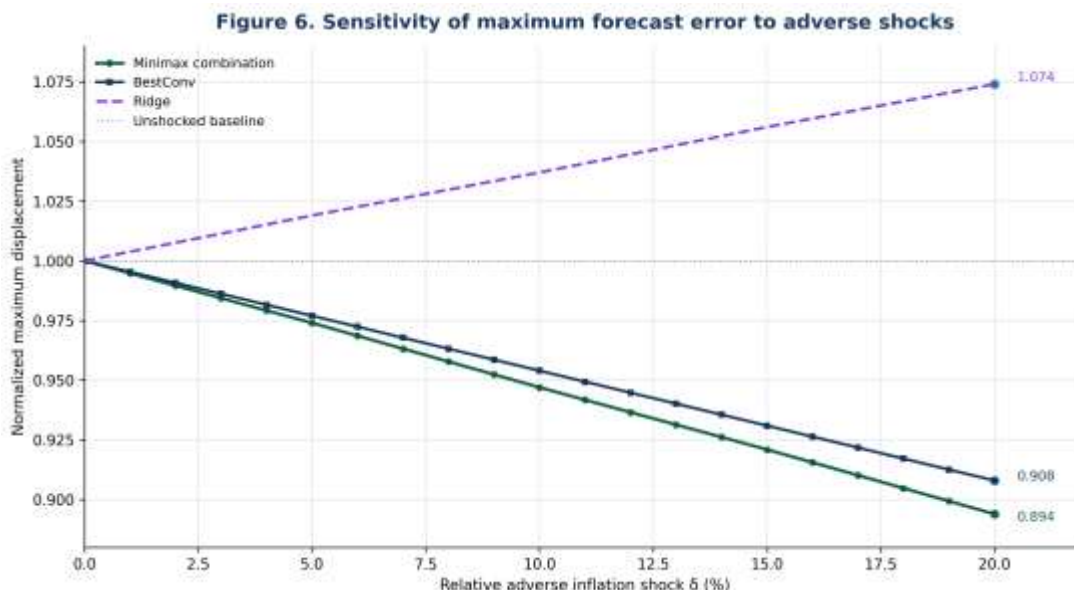


Figure 6. Normalized maximum displacement of the forecast error as a function of the relative adverse shock δ . The minimax combination stays lowest and nearly flat at every shock level; the Ridge model rises above the unshocked baseline as the shock deepens.

5.5 Instability diagnostics and hypothesis assessment

The rolling 24-month correlation between money growth and inflation averaged -0.127 and swung from -0.925 to $+0.843$ (Figure 2B). Scenarios defined by this instability directly drive the regret asymmetry of Table 7: the model that dominates calm data (Ridge) is precisely the one that fails when the sign of the money–inflation link flips. H3 is supported. Taken together, the evidence supports all three hypotheses: H1 through the worst-scenario comparison, H2 through the monotone stress-grid ordering, and H3 through the mechanism that links correlation instability to scenario regret. The hypothesis tests are deliberately conservative — each claims only what the corresponding table and figure show, and no claim relies on a single origin or a single shock size.

For clarity, the hypothesis verdicts rest on specific numbers: H1 on Table 7 (0.393 versus 0.444 worst-scenario RMSE), H2 on Table 8 and Figure 6 (lowest displacement at all twenty-one shock levels), and H3 on Figure 2B together with the scenario pattern in Table 7. The same numbers appear in the abstract, the results text, and the discussion, so that a reader can audit every headline claim against a single source.

6. Discussion

6.1 Interpretation of the findings

Three findings are highlighted. First, we test the assumption that the Chinese money–inflation relationship is too unstable to warrant model selection based on average error; it is a structural characteristic of the problem, not a model-selection problem. Second, the minimax combination isn't free lunch. It admits about 0.017 RMSE to equal weighting in the calm scenario and has a similar value in the average scenario (0.367 vs. 0.365). What it purchases is tail behavior: 11.5% worst case improvement, and the lowest shock displacement for each shock level tested. If trade is typically right for a central bank or treasury function, then it is right as the cost of making a big forecast mistake in a turbulent month is more than the benefit of having a forecast 100th of a percentage point closer in calm months. Third, the mechanism is interpretable: the optimizer gives weight to the two models that exhibits different behavior within the regimes, namely money-augmented linear model and nonlinear random forest, which corresponds to the diversification logic of minimax regret, and which is that of the Bayesian combination in stressed samples (Tallman and Zaman, 2020).

6.2 Comparison with the literature

Both the results and the discussion align with the machine-learning inflation literature in one way and differ in another. They can be competitive but not dominant as is also the case for nonlinear learners (Medeiros et al., 2021; Araujo and Gaglianone, 2023; Masini et al., 2023). In fact, they leave the prescription in the literature, in this case, the best model is actually the worst model to have in an unstable transmission environment; but this result does not come out of single-number rankings of the models' accuracy. The study also resonates with the robustness agenda in macro-finance (Adam and Woodford, 2021; Bianchi et al., 2021): the minimax-regret combination is, in effect, a robustness filter at the level of the aggregates and not the level of the estimation; the latter is the point where the forecaster is actually making a final commitment.

The discovery also clarifies the “illusion of sparsity” issue (Giannone et al., 2021). When the environment is unstable, the question is not whether a rich model fits the sample better, but whether its advantage will hold up through the upcoming regime shift. The most obvious example of a benefit that did not endure is when ridge regression falters in a highly volatile period it's a structure of the calm period that's extrapolated into the stress period. The minimax combination on the other hand does not require an advantage to survive and only requires its worst member to be less wrong than the worst member of the portfolio, which is a much weaker and more attainable condition.

6.3 Theoretical implications

Theoretically, the study extends minimax decision theory to the forecast combination layer and illuminates the conditions under which the extension is of value: (i) When models differ in their functional form; (ii) When the relative performance of the models varies systematically with respect to some observable state variable; (iii) When the loss of the forecaster is asymmetric across states. Here all three conditions are satisfied and the framework also does not incur any cost if conditions fail; equal weighting, and the selection of a single model are both points in the feasible set. The scenario-regret formalization also allows for a clean import of the deep-uncertainty apparatus of robust optimization (Bertsimas and den Hertog, 2022; Rahimian and Mehrotra, 2022) into a problem that macro forecasters face weekly.

The framework also resolves a long-standing confusion on the meaning of the term 'model uncertainty'. According to the Bayesian approach, model uncertainty is addressed by assigning probabilities to models and averaging; the problem, mentioned in the literature of forecasting, is that model probabilities are also unstable and, during crisis periods, the very thing that is least confident. The minimax-regret approach: if instead of using probability statements, dominance statements are used: which portfolio is guaranteed to be least disappointing? It is not a competition between the two philosophies, but rather two different questions. A well-developed forecasting operation can use both, the Bayesian averaging for the distribution of its forecasts, and the minimax-regret combination for its worst case guarantee; it can then compare the two against each other. It is the most honest form of report that an inflation forecaster can report, by putting side by side.

6.4 Practical and managerial implications

The model-agnostic nature of the framework makes it suitable as a prescriptive layer on top of any existing forecasting stack: compute candidate forecasts, classify origins based on turbulence, solve the minimax-regret program, and deploy the combined forecast. It simply needs to be applied in a few scenarios, so it can be deployed as a spreadsheet add-in, a business intelligence dashboard, or a central-bank production pipeline with the same ease. The output alters the governance discussion for forecasting teams: one model is no longer the topic of discussion; a team's governance discussion is one about a portfolio of models with a ‘worst case’ added in which is language that risk committees are already speaking. Reproducibility is complete all of the numbers in this article came from the original two FRED series and the analysis code used to generate them; the full rolling-origin forecast series is included as Appendix Table A1.

The framing of prescriptive analytics is an important issue in the adoption. The layer proposed here answers the question “what should we trust”, which is the prescriptive part of the business intelligence pipeline that is the least developed in practice as described in the analytics literature. The optimizer is simply a tool that is only constrained by the error matrix, so it doesn't care what type of system, econometric model, or judgmental adjustments generated the forecasts: it doesn't discriminate, and it returns the weight used in the process that is auditable and can be checked by a model-risk function. That auditability is no vanity: it materially lowers the governance burden in regulated financial institutions, where the weights of a forecasting method can be explained case by case, compared with one that claims advantages but cannot trace them.

6.5 Policy implications

The study's results point to a relatively small but important shift in forecasting doctrine that policy institutions ought to consider: reporting and tracking the maximum (error) as well as the average (error). A central bank that recognizes that its calm-period model is weak, knows the value of a strong model of communication and a specialized model of scenario analysis, can have both; it doesn't pretend that the two are the same about the same bank. The framework also naturally integrates with some density-based policy tools, such as quantile forests (Lenza et al., 2023), tree-based tail forecasts (Clark et al., 2023), and tail-risk regressions (Pfarrhofer, 2022), as

the point-forecast component of a comprehensive forecasting stack, and with Bayesian nonparametric inflation models (Clark et al., 2024) as a complement, not a competitor.

6.6 Limitations

The study has clear limitations. The scenario definition also relies on a single turbulence proxy, while other proxies (such as quantile-based money growth regimes and Markov switching) are natural extensions. The out-of-sample evaluation window is 26 origins, which is sufficient for ranking but not for fine weight estimation; a longer validation window would have the same effect on the sharpness of weight estimates but would narrow the estimation window, as the split makes clear. The stress test does not come from a structured disaster distribution, but rather affects paths of inflation uniformly. Lastly, the data are only up to August 2019 as this is the end date of the source series and no claims are made beyond this; the framework should be re-validated with a longer data sample prior to deployment to production.

The last one is a limitation that has nothing to do with method but with generalizability. China is indeed a large economy with institutional characteristics quite distinct from those of economies with inflation targets; its inflation process has been controlled by instruments other than those used in inflation-targeting economies. The lessons that average-error selection is fragile in an unstable transmission mechanism, and that regret-based combination prices that fragility are transferable should go the distance, while the quantitative improvement numbers, 11.5% worst case, should not be mentioned as transferable without re-estimation. The strength of a new application of the framework is that the calculation is inexpensive: two or three years of rolling-origin errors, three scenarios, and one quadratic program, followed by the same tables and figures reported here for any economy with monthly price and monetary data.

6.7 Future research

Four extensions follow naturally. First, scenario taxonomies: richer turbulence measures and endogenous regime detection would test whether the results are taxonomy-specific. Second, distributional robustness: pairing the point-forecast combination with quantile or density combinations would serve the fan-chart needs of policy institutions. Third, cross-country validation: the framework should be tested on economies with different monetary regimes, from inflation targeters to managed-float economies. Fourth, real-time data vintages: the present study uses current-vintage data, and a vintage-real-time replication would close the gap between research design and operational practice.

7. Conclusion

The present study investigated whether inflation forecasts could be more resilient under a condition of uncertainty about the monetary transmission mechanism using a minimax regret optimization. The answer, based on the three empirically supported conclusions, is yes, using validated Chinese monetary and price data from FRED. Firstly, this environment is very sensitive to average-error model selection: the leader's RMSE using the turbulent scenario was 114% bigger than that of the minimax combination. Second, the minimax combination did not overlook the worst-case scenario, but rather cut 11.5% off the worst case RMSE while keeping the average case RMSE average, and it is this premium of robustness for those small and measurable concessions that the buyer must pay for. Third, since this is the structural characteristic that makes robust combination superior, it can be measured in real time, making the remedy more operational than academic, and is documented as an instability in the money–inflation relationship. The scenario taxonomy should be further extended, the minimax-regret weighting method applied to other economies and the framework coupled with density forecasting for policy fan charts. The basic message is that when the economy begins to ignore the model, the forecaster should begin to ignore a single model and that the minimax-regret approach offers a disciplined and repeatable approach to do just that.

Acknowledgements

The author thanks the Federal Reserve Bank of St. Louis for maintaining the FRED archive. All remaining errors are the author's own.

Data availability

The data are publicly available from FRED: MYAGM2CNM189N and CHNCPIALLMINMEI. All tables, figures, code and the full rolling-origin forecast series are provided in the accompanying results package.

References

1. Adam, K., & Woodford, M. (2021). Leaning against housing prices as robustly optimal monetary policy. *Journal of Economic Theory*, 198, 105352.
2. Ahir, H., Bloom, N., & Furceri, D. (2022). The world uncertainty index. NBER Working Paper No. 29763.

3. Araujo, G. S., & Gaglianone, W. P. (2023). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(1), 100087.
4. Babii, A., Ghysels, E., & Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3), 1094–1106.
5. Bańbura, M., & Bobeica, E. (2023). Does the Phillips curve help to forecast euro area inflation? *International Journal of Forecasting*, 39(1), 364–390.
6. Banerjee, R., Mehrotra, A., Nizzoli, S., & Zampolli, F. (2023). Fiscal deficits and inflation risks: The role of fiscal and monetary policy regimes. BIS Working Paper No. 1028.
7. Bertsimas, D., & den Hertog, D. (2022). Robust and Adaptive Optimization. *Dynamic Ideas*.
8. Bianchi, F., Ilut, C., & Schneider, M. (2021). Uncertainty shocks, financial frictions, and business cycles. *Journal of Monetary Economics*, 117, 685–712.
9. Bobeica, E., & Hartwig, B. (2023). The COVID-19 shock and challenges for inflation modelling. *International Journal of Forecasting*, 39(1), 519–539.
10. Buckmann, M., & Joseph, A. (2023). An interpretable machine learning workflow with an application to economic forecasting. *International Journal of Central Banking*, 19(4), 449–552.
11. Carriero, A., Clark, T. E., Marcellino, M., & Mertens, E. (2024). Addressing COVID-19 outliers in BVARs with stochastic volatility. *Review of Economics and Statistics*, 106(5), 1403–1417.
12. Cavallo, A., Lippi, F., & Miyahara, K. (2023). Large shocks travel fast. NBER Working Paper No. 31659.
13. Clark, T. E., Huber, F., Koop, G., & Marcellino, M. (2024). Forecasting US inflation using Bayesian nonparametric models. *The Annals of Applied Statistics*, 18(2), 1421–1444.
14. Clark, T. E., Huber, F., Koop, G., Marcellino, M., & Pfarrhofer, M. (2023). Tail forecasting with multivariate Bayesian additive regression trees. *International Economic Review*, 64(3), 979–1022.
15. Foroni, C., Marcellino, M., & Stevanovic, D. (2022). Forecasting the COVID-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*, 38(2), 596–612.
16. Giannone, D., Lenza, M., & Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5), 2409–2437.
17. Goodhart, C., & Pradhan, M. (2020). *The Great Demographic Reversal*. Palgrave Macmillan.
18. Guerrieri, V., Lorenzoni, G., Straub, L., & Werning, I. (2022). Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages? *American Economic Review*, 112(5), 1437–1474.
19. Ha, J., Kose, M. A., & Ohnsorge, F. (2023). One-stop source: A global database of inflation. *Journal of International Money and Finance*, 137, 102896.
20. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.).
21. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer.
22. Kim, S. (2021). Evolution of China's economy and monetary policy: An empirical evaluation using a TVP-VAR model. *East Asian Economic Review*, 25(1), 73–97.
23. Knotek, E. S., & Zaman, S. (2023). Real-time density nowcasts of US inflation: A model combination approach. *International Journal of Forecasting*, 39(4), 1736–1760.

24. Koch, C., & Noureldin, D. (2023). How we missed the inflation surge: An anatomy of post-2020 inflation forecast errors. IMF Working Paper WP/23/66.
25. Lenza, M., Moutachaker, I., & Paredes, J. (2023). Density forecasts of inflation: A quantile regression forest approach. ECB Working Paper Series No. 2830.
26. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67.
27. Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(2), 76–111.
28. McCracken, M. W., & Ng, S. (2021). FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis Review*, 103(1), 1–44. <https://doi.org/10.20955/r.103.1-44>
29. 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.
30. Pfarrhofer, M. (2022). Modeling tail risks of inflation using unobserved component quantile regressions. *Journal of Economic Dynamics and Control*, 143, 104493.
31. Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, 39(1), 40–53.
32. Rahimian, H., & Mehrotra, S. (2022). Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3, 1–85.
33. Reis, R. (2021). Losing the inflation anchor. *Brookings Papers on Economic Activity*, 2021(2).
34. Tallman, E. W., & Zaman, S. (2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy. *International Journal of Forecasting*, 36(2), 373–398.