Forecasting Inflation with Microdata: An Adaptive Machine Learning Approach*

Catherine Chen **Chen Gao** **Jonathon Hazell** **Lihua Lei** **Chen Lian**

March 26, 2026

Abstract

We develop a framework to forecast inflation using high dimensional microdata in a changing environment with structural breaks. We apply the method to microdata on price changes for the United Kingdom, over 1996-2024. Our contribution is threefold. First, we encode the distribution of price changes into quantiles and moments of the distribution, which combine with a machine learning algorithm to forecast inflation. Second, we develop a scan test for a composite null hypothesis that one forecast never outperforms a second—rejecting the null implies the first forecast outperforms at an ex ante unknown point in time, of unknown duration, for some unknown forecasting horizon. By applying the scan test, it is observed that micro forecast outperforms several benchmarks—a univariate forecast of inflation and a forecast with macro data—but only during the high and volatile inflation after 2020. This hypothesis is useful in a non-stationary environment where no single modality is expected to dominate. Third, we combine multiple modalities — micro, macro and univariate forecasts — using the fixed shares algorithm: an adaptive machine learning method with dynamic regret guarantees in adversarially non-stationary settings. The fixed shares algorithm outperforms forecasts that do not use microdata, and places significant weight on microdata after 2020. We discuss the implications of our results for models of inflation dynamics.

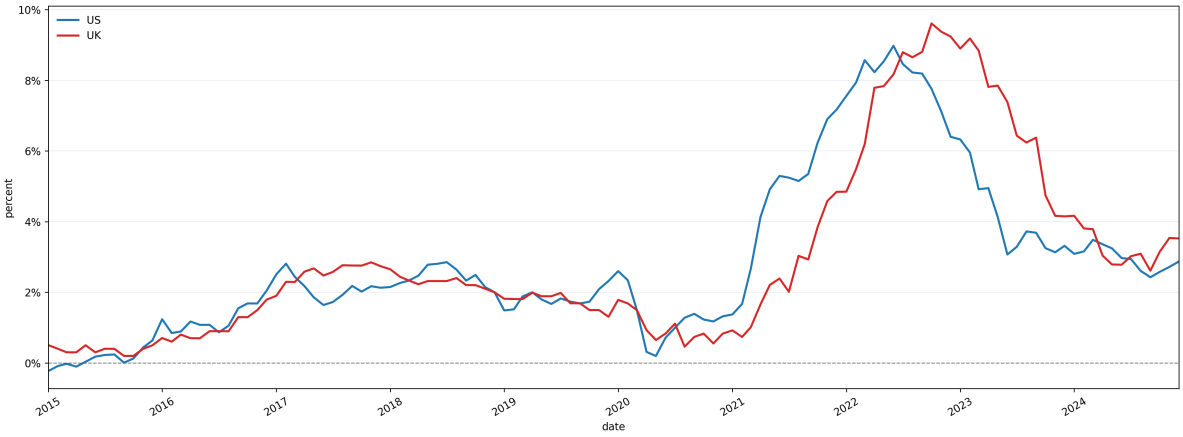
*Chen: Stanford University; cyc2152@stanford.edu. Gao: Peking University; chengao0716@gmail.com. Hazell: London School of Economics; j.hazell@lse.ac.uk. Lei: Stanford Graduate School of Business; lihualei@stanford.edu. Lian: UC Berkeley and NBER; chen_lian@berkeley.edu. We thank Nicholas Tokay and Sven van Holten for outstanding research assistance.

1 Introduction

How can we forecast the economy? Traditional forecasting approaches rely on macroeconomic data. Yet aggregate outcomes are generated by millions of microeconomic decisions about price setting, employment, consumption, and production. In recent decades, high-dimensional micro datasets tracking this behavior have become available. These data create an opportunity to use micro-level information to improve forecasts of aggregate variables.

This paper focuses on inflation, one of the most difficult macroeconomic variables to forecast. Even sophisticated models often struggle to outperform simple univariate benchmarks (Atkeson and Ohanian, 2001). A central challenge is non-stationarity: the process governing inflation dynamics changes over time. Figure 1 illustrates this clearly for the United Kingdom and the United States over 2015–2024, with markedly different inflation behavior before and after 2020.

Figure 1: Headline CPI inflation in the United States and United Kingdom, 2015–2024



Notes: Year-on-year headline CPI inflation for the United States (blue) and United Kingdom (red), January 2015 to December 2024. Data source: Federal Reserve Economic Data (FRED).

There is good reason to think that microdata—in our case, the millions of price changes observed each month—may improve inflation forecasts. Each price change reflects a forward-looking decision by an individual firm, and the cross section of such decisions may contain information about future aggregate inflation. Consistent with this idea, macroeconomic models imply that microdata should help forecast inflation in some environments but not others. When aggregate shocks are small, microdata do not add forecasting power conditional on macroeconomic aggregates in standard models of price setting (Auclert et al., 2024). When aggregate shocks are large, however, micro information can help predict aggregate inflation (Ascari and Sbordone, 2014; Blanco et al., 2024).

In this work, we introduce a method for inflation forecasting with high dimensional microdata. We adopt an adaptive machine learning approach, motivated by existing tools from the online learning literature (Herbster and Warmuth, 1998; Cesa-Bianchi and Lugosi, 2006; Shalev-Shwartz, 2025).

These algorithms are well suited for high dimensional prediction in a non-stationary environment, and produce easy-to-interpret statistics that connect to macroeconomic models.

Our primary dataset consists of official UK CPI micro-price quotes. The data contain tens of thousands of monthly observations on individual prices across the economy and underlie the construction of aggregate inflation. The sample runs from 1996 to 2024, covering both the low and stable inflation period before 2020 and the more volatile period thereafter (see Figure 1). Each observation is a price quote identified by an item code, shop, region, and shop type. We supplement these microdata with a large panel of macroeconomic predictors known to be useful for forecasting inflation (Goulet Coulombe, Marcellino and Stevanović, 2021).

Our method has three steps. The first step encodes the raw micro-price data into a stable, high-dimensional vector of distributional statistics. This step is necessary because individual products regularly enter and exit the sample, so the raw data do not admit a fixed predictor vector. We therefore summarize, for each month and each of 12 broad consumption categories, the cross-sectional distribution of price changes using statistics that are available in every period. These include the number of observations, the fraction of prices that do not change, the deciles and first five moments of the distribution of non-zero price changes, and corresponding summaries of the time since the last price change. Stacking these statistics across categories yields an interpretable but high-dimensional representation of the micro-price environment, which can then be used for forecasting.

Because this encoding produces a large predictor set relative to the available time series, conventional forecasting methods are prone to overfitting. We therefore use a machine learning method designed for high-dimensional prediction: gradient-boosted trees, implemented with XGBoost, which performs strongly in tabular forecasting settings (Grinsztajn, Oyallon and Varoquaux, 2022; Chen and Guestrin, 2016).[1] We estimate the model on rolling windows, tune hyperparameters using a nested validation procedure, and generate direct forecasts for horizons from 1 to 24 months ahead.

The second step of our method is a new scan test that evaluates whether the microdata forecast outperforms standard benchmarks over localized periods of time and forecasting horizons. Standard forecast comparison tests focus on average performance over the full sample (e.g. Diebold and Mariano, 1995). That is not the object of interest here. Inflation forecasting is non-stationary, and both theory and evidence suggest that microdata may help only in particular episodes and at particular horizons. We therefore ask whether there are sub-periods and forecasting distances for which the micro forecast systematically outperforms benchmark models.

We develop a scan test for the composite one-sided null hypothesis that one forecast is at least as accurate as another at every point in time and at every forecasting horizon, against the alternative

[1] Alternative ways to forecast with a large number of predictors, and prevent overfitting, include Bayesian shrinkage or factor analysis (e.g., Stock and Watson, 2002; Bańbura, Giannone and Reichlin, 2010).

that the second forecast outperforms over some contiguous interval. For each candidate interval and horizon, we compute a self-normalized t-statistic based on cumulative loss differentials, and define the scan statistic as the maximum of this object over all admissible intervals and horizons. Because this maximum has a non-standard distribution, we obtain critical values using a Gaussian multiplier bootstrap, after first testing the loss differentials for serial dependence to determine the appropriate bootstrap regime. The interval and horizon that maximize the scan statistic point to the period and forecasting distance at which the micro forecast outperforms by the most. The scan statistic is a maximum over a large collection of candidate intervals and horizons, meaning we require the recent high-dimensional Gaussian approximation results (Chernozhukov, Chetverikov and Kato, 2013, 2017; Chang, Chen and Wu, 2024) to derive our test. This scan test is of independent interest—there are many circumstances when one wishes to know whether one forecast outperforms another during some ex ante unknown length of time.

We compare the microdata forecast to two benchmarks from the literature. The first is a rolling-window seasonal ARIMA (SARIMA) model with nested hyperparameter selection, which nests several leading univariate inflation forecasts used in the literature, including those of Atkeson and Ohanian (2001) and Stock and Watson (2007). The second benchmark applies the same rolling-window XGBoost procedure used for microdata, but with a large set of macroeconomic predictors instead, allowing us to isolate the marginal contribution of micro-price information.

Our third step combines the micro, macro, and univariate forecasts in a non-stationary environment. We use the Fixed Share algorithm of Herbster and Warmuth (1998), which forms a weighted average of the three forecasts and updates those weights over time in response to realized forecast errors. This setting is well suited to inflation forecasting it has a regret guarantee in highly non-stationary environments against a dynamic comparator that may change across episodes and horizons for a finite number of times. Unlike standard forecast combination methods that effectively target a single best model, Fixed Share is designed to track the best *sequence* of experts over time. In our application, the experts are the Micro XGBoost forecast, the Macro XGBoost forecast, and the SARIMA forecast.

The Fixed Share algorithm also yields an interpretable time series of expert weights. In particular, the weight on the micro forecast measures how useful micro-price information is for forecasting inflation at a given date and horizon. This is valuable because macroeconomic models make predictions about when microdata should matter for aggregate inflation, for example after large inflation shocks. We can therefore use the evolution of the Fixed Share weights to assess whether those theoretical predictions line up with the data. This interpretability is an important advantage of the approach, especially relative to machine learning methods that are often criticized as opaque "black boxes" (Mullainathan and Ludwig, 2024).

Our first result is that the forecasting value of microdata is strongly state dependent. After en-

coding the microdata and combining them with XGBoost, we compare the resulting forecast with a rolling-window SARIMA benchmark over horizons from 1 to 24 months and across the pre-2020 (2015–2019) and post-2020 (2020–2024) windows. The micro forecast underperforms SARIMA uniformly before 2020. After 2020, however, the pattern reverses at medium and long horizons: the micro forecast improves on SARIMA at 18 of 24 horizons in mean absolute error and 19 of 24 horizons in root mean squared error, with especially large gains from about $h = 6$ onward. We formalize this localized outperformance using the scan test, which rejects the null that the micro forecast never outperforms SARIMA. The maximizing episode runs from February to December 2022, during the sharp inflation run-up and subsequent decline (see Figure 1). The same broad message also emerges relative to the macro XGBoost benchmark: microdata are most useful when inflation is high and volatile, rather than in stable periods.

Our second result is that adaptive forecast combination is valuable in this non-stationary environment, and that microdata add incremental value within that combination. The full three-expert Fixed Share forecast, which combines Micro XGBoost, Macro XGBoost, and SARIMA, delivers only modest gains before 2020 but broad-based improvements after 2020; with adaptive learning-rate selection, it improves on SARIMA at all 24 horizons in mean absolute error and at 22 of 24 horizons in root mean squared error in the post-2020 window. A powerful two-expert benchmark that combines only Macro XGBoost and SARIMA also performs well, but remains slightly weaker on average than the full three-expert combination. The expert weights clarify why: at short horizons and in the pre-2020 window, the algorithm places relatively little weight on the micro forecast, whereas after 2020 the weight on microdata rises, especially at medium and long horizons. The adaptive combination therefore uses microdata precisely when they become informative.

Our results also have implications for models of inflation. The scan-test evidence and the Fixed Share weights point to the same conclusion: microdata become especially valuable for aggregate forecasting when inflation is high and volatile. This pattern is consistent with benchmark state-dependent models of price setting. In these models, when aggregate shocks are small, aggregate inflation is largely forecastable from macro information alone (Auclert et al., 2024). When shocks are large, however, the cross section of price adjustment contains additional information about aggregate inflation dynamics (Blanco et al., 2024). The results are also consistent with models in which monetary policy normally offsets inflationary disturbances, limiting predictability, but periods of imperfect stabilization allow more predictability to emerge (McLeay and Tenreyro, 2020). In that sense, the post-2020 episode appears as a period in which both macro and micro information became unusually informative for forecasting inflation.

Related literature. Our paper contributes to the literature on forecasting inflation. A classic result is that inflation is hard to forecast, meaning simple univariate models such as random walks are

difficult to beat (Atkeson and Ohanian, 2001; Stock and Watson, 2007). Notwithstanding this result, models with various degrees of sophistication, using macroeconomic information, judgemental forecasts, and inflation expectations by households, forecasters and financial markets, have also been used to forecast inflation (see Faust and Wright (2013) for a review). There is a newer literature that seeks to forecast inflation using increasingly rich macroeconomic datasets combined with machine learning methods. For instance Medeiros et al. (2021) find that rich macroeconomic data and random forest algorithms lead to forecasting gains for US inflation, and Goulet Coulombe, Marcellino and Stevanović (2021) find similar for the UK. Likewise, recent work attempts to forecast inflation using aggregated statistics drawn from rich sources of microdata (e.g., Beck et al., 2024; Brandão-Marques et al., 2025). We develop a method to use the full richness of information from the microdata, via encoding, machine learning, and online learning. In doing so, we contribute more broadly to the literature on machine learning in time series forecasting (e.g., Giannone, Lenza and Primiceri, 2021; Chi et al., 2025).

There is previous work that forecasts macroeconomic variables using the distribution of microdata. A seminal, recent paper is Chang, Chen and Schorfheide (2024), which develops a "functional" vector autoregression (VAR) approach. The method involves summarizing the log density of a distribution of microdata by a low dimensional collection of objects: finite dimensional sieves with fixed basis functions and time varying coefficients. One can then estimate a vector autoregression using aggregate variables and the time varying sieve coefficients, allowing a forecast of macro variables with microdata. The main focus is on a stationary environment.[2] Meeks and Monti (2023) develop a related approach that summarizes cross-sectional distributions with functional principal components, and links them to inflation in a linear, Phillips Curve style model. We offer a different approach: summarizing microdata with a high dimensional series of statistics, which we combine with black-box forecasting algorithms and expert aggregation via the Fixed Share algorithm that suit a non-stationary environment. We also aim at a simpler goal—conducting unconditional forecasting, as opposed to the more structured analysis allowed by VARs and Phillips Curve models.

Our non-stationary environment leads us to develop a scan test, to test for sub-periods in which microdata forecasts well. Classic time series tests of forecasting performance test whether one forecast outperforms another on average, throughout a given time series (e.g. Diebold and Mariano, 1995). Instead, we test for whether there are occasional periods of time in which a chosen forecast out-performs a benchmark—without specifying ex ante when and for how long the forecast outperforms. As such, our test closely relates to Giacomini and Rossi (2010), which also develops a test for time-varying relative forecast performance. A key distinction from our framework is in the null hy-

[2] See Lenza and Savoia (2024) for an application of this approach to firm heterogeneity in the Euro Area.

pothesis. Giacomini and Rossi (2010) test equal predictive ability, whereas we study a one-sided null asserting that one method does not outperform the other on any relevant interval. Rejection of equal predictive ability is therefore not sufficient for our purposes, because it does not statistically identify the direction of the difference. [3] Another advantage of our test is that we do not require an ex ante specification of the length of time for which the chosen forecast outperforms the benchmark. Since we do not specify the length of time, the scan statistic takes a maximum over a large collection of candidate intervals. Lastly, to accommodate the low sample size in most macro-forecasting problems, we develop a full-fledged econometric theory for our test under a high-dimensional asymptotic regime that allows the minimal interval length to grow logarithmically in the sample size and the number of intervals under scrutiny to grow exponentially in the sample size.

Finally, we use an adaptive machine learning approach—the Fixed Share algorithm—to exploit that microdata helps to forecast inflation, but only occasionally (i.e. there is non-stationarity). A large literature studies how to combine forecasts (Bates and Granger, 1969; Timmermann, 2006), including with Bayesian model averaging approaches (e.g., Koop and Korobilis, 2012). Many methods require a full specification of the likelihood function, making it difficult to combine them with high dimensional covariates and black-box machine learning methods. The Fixed Share algorithm provides worst-case regret guarantees without distributional assumptions, in the presence of non-stationarity, and when some experts use high dimensional data and machine learning. An important previous paper also uses adaptive machine learning, namely Fouliard et al. (2023). Besides a different application, and the use of microdata, our application has several benefits. By combining the algorithm with a scan test, we include only forecasts that provably improve the regret bound and are likely to improve the overall performance of the final forecast.

The article proceeds as follows. Section 2 describes the data. Section 3 develops the methods: the encoding and machine learning algorithm, the scan test, and the online learning combination. Section 4 presents the main empirical results. Section 5 discusses implications for models of inflation. Section 6 concludes.

2 Data

This section describes the construction of the datasets used in the empirical analysis. We combine two main sources of information. First, we use official UK CPI micro-price quotes collected by the Office for National Statistics (ONS), which provide the micro-price panel underlying the official inflation

[3] While Giacomini and Rossi (2010) entertainsa one-sided alternative hypothesis, it is unclear whether the equal-predictive-ability case is least favorable for Type-I error over our one-sided null parameter space. The main obstacle is that, as in our test, the relevant t-statistics are self-normalized and hence need not be monotone in the data.

statistics. Second, we assemble a large set of macroeconomic and financial time series which we use to forecast inflation with macro data.

2.1 Microdata on Prices

Our primary dataset consists of official micro-level price quotes collected by the Office for National Statistics (ONS) and underlying the UK Consumer Price Index (CPI), spanning January 1996 to December 2024. For 1996–2016, we use the replication package of Adam and Weber (2023); for 2017 onward, we extend the dataset using price quotes and item index files from the ONS website.

Each observation is a price quote identified by four fields: a six-digit item code, a two-digit shop code, a two-digit region code, and a shop type (multiple vs. independent). For instance, an item might be a loaf of white bread, a shop code might be an individual Sainsbury's grocery store, based in a region such as London. There is also a flag for whether the price has a sale, as measured by the ONS price quote collector. For each observation, we calculate the monthly unit price change as

$$\Delta p_{jt} = 100 \times (\log p_{jt} - \log p_{j,t-1}),$$

the month-on-month log price change in percentage points.[4]

The raw ONS panel is first cleaned using standard validity, comparability, and price filters, which we list in Appendix D.1. We organise the data at the cell level, where a cell is defined by a monthly date and a product category. The product categories are broad groupings of items at the COICOP1 level (there are 12 COICOP1 categories, such as food and non-alcoholic beverages). We balance the data by retaining only COICOP1 codes present in more than 95% of months.

Table 1 reports the percentile distribution of observations per cell, separately for all price changes and for non-zero price changes only. The difference between the two panels reflects the well-documented fact that roughly 90% of prices are unchanged at monthly frequency. We include cell counts with and without items on sale.

We validate the data in two ways. First, we reconstruct the aggregate CPI from the underlying micro quotes and verify that the implied inflation rate closely tracks the official ONS series (see Appendix D.3). Second, we benchmark the moments of the price-change distribution—mean, standard deviation, and excess kurtosis of $|\Delta p|$—against the existing literature and find that our estimates replicate the well-known stylised facts closely (see Appendix D.4).

Our main dataset for the forecasting exercise will be all of the observations in the clean data, after dropping sales observations. To remove sales, we drop observations flagged by the ONS as temporary

[4] Two regions warrant note. Region 1 ("Catalogue collections") covers items collected centrally—such as shelter, university tuition fees, and rail fares—for which only item-level indices are published; following Blanco et al. (2024), we exclude these observations. Region 13 ("NI") denotes observations with no geographic attribution.

price reductions. In addition, we construct two alternative panels in which sale prices are adjusted rather than dropped, using the Nakamura–Steinsson filter (Nakamura and Steinsson, 2008) with two parameterisations. These filters identify likely sale episodes algorithmically and replace sale prices with the inferred regular price; Appendix D.2 describes the algorithm and its parameters in detail.

2.2 Macro Data

In addition to the micro-price data, we assemble a large set of macroeconomic and financial predictors. The macroeconomic dataset is built around the UK Monthly Database (UKMD) of Goulet Coulombe, Marcellino and Stevanović (2021), which provides a harmonised panel of monthly UK macroeconomic series designed for high-dimensional forecasting applications. We extend this database along three dimensions. First, we add a number of series not included in the original UKMD, notably commodity futures, measures of global supply-chain disruptions, housing indicators, uncertainty indices, and additional monetary aggregates. Second, whenever possible, we replace seasonally adjusted series with their non-seasonally adjusted counterparts. Using non-seasonally adjusted series avoids the look-ahead bias inherent in two-sided seasonal adjustment filters. Third, we update the sample through to November 2024.

The resulting dataset contains monthly non-seasonally adjusted series spanning the start of 1996 to November 2024. Appendix Table B.1 lists the 60 baseline predictors grouped into broad categories: labour market conditions, industrial production, retail and trade activity, exchange rates, commodities and global prices, monetary and credit variables, financial markets, and uncertainty.

3 Methods: Macro Forecasting with Micro Data

This section describes our three-step procedure for forecasting inflation with microdata. The first step encodes the high dimensional microdata into a stable vector of distributional statistics and feeds it into a machine learning algorithm. The second step develops a scan test that evaluates whether the microdata forecast outperforms standard benchmarks over localized time periods and forecasting horizons—a test designed for the non-stationary environment in which the power of microdata may come and go. To avoid mathematical complications, we present the full econometric theory of the scan test in Appendix A. The third step combines the microdata forecast with benchmark forecasts using an online learning algorithm that adapts to each model's recent performance.

3.1 Step 1: Encoding and a Machine Learning Algorithm

Encoding Micro Price Data. The aim is to forecast inflation with microdata. However, the raw microdata from the consumer price index is not in a form that we can combine with a prediction algorithm. In the raw microdata, products (i.e. an item in a particular store and region) regularly enter and exit the data. The typical product remains in the data for less than a year. This pattern is due to behavior by shops: for instance, a 6-pack of beer in a given grocery store would typically be substituted for an alternative at some point. Since there is no stable set of products in the database, there is no fixed vector that we can use for prediction.

Therefore the first step is to encode the raw microdata into a stable, but high dimensional vector of statistics. This vector summarizes the features of the microdata, but can be combined with a prediction algorithm. For each cell—an aggregated set of products at the COICOP1 level, and a month—we summarize the cross-sectional distribution of price changes. We use two sets of statistics, which account for the well-known fact that product-level prices tend to rarely change at monthly frequency. The first set captures the distribution of non-zero price changes. Specifically, for each cell we calculate:

1. The total number of price observations in the cell.
2. The fraction of observations with $\Delta p_{jt} = 0$.
3. The nine deciles (at probabilities $0.1, 0.2, \ldots, 0.9$) of the *non-zero* price-change distribution within the cell.
4. Five moments of the non-zero price-change distribution—mean (m_1), variance with Bessel's correction (m_2), and the standardised third, fourth and fifth central moments (m_3, m_4, m_5), where m_4 is expressed as excess kurtosis.

The second set captures the time elapsed since the last price change for items with zero price changes in the current period. For each cell, the dataset records:

1. The nine deciles of the distribution of time since the last price change, measured in months. The clock resets on each non-zero price-change period, on the first observation in a price spell, and whenever the gap between two consecutive observations of the same product exceeds 12 months.
2. The same five moments computed for the time-since-change distribution.

Taken together, the vector used for prediction stacks these moment and quantile summaries across all COICOP1 cells at time t, yielding a high-dimensional but interpretable representation of the micro-price environment. All series are constructed at the monthly frequency and are not seasonally adjusted (NSA). We additionally include a monthly indicator to allow for seasonal patterns.

Machine Learning Algorithm. The vector of statistics summarizing the microdata is high dimensional. In total, the vector contains 360 elements, and we also wish to use lags of this vector for forecasting. Therefore the vector of statistics is large relative to the length of the time series. Conventional time series methods risk overfitting, leading to poor forecasting performance. Instead, we use machine learning methods, which are well suited to prediction with high dimensional covariates. In particular, we use a version of the gradient boosted tree, which is one of the most successful machine learning algorithms for forecasting (Grinsztajn, Oyallon and Varoquaux, 2022). In particular, we use a variant called XGBoost (Chen and Guestrin, 2016).

We forecast with XGBoost estimated on a rolling window, with hyperparameter selection via a nested validation scheme. The procedure is described formally in pseudocode in Appendix E (Algorithm 1), and proceeds as follows. At each forecast origin τ, we select hyperparameters Γ^* from a grid $\mathscr{G}$ by minimizing out-of-sample root mean-squared-error (RMSE) over a 48-month validation window $[\tau-47-h, \tau-h]$. To avoid look-ahead bias, each candidate model ML(Γ) is trained on data up to four non-overlapping held-out periods ($l = 4,3,2,1$ years prior to $\tau-h$), and the resulting forecasts are aggregated to compute validation RMSE. The training window is capped at 120 months (10 years) to limit the influence of distant observations.

Having selected Γ^*, we re-train the model on the 120 months of data immediately preceding $\tau-h$, and produce the h-step-ahead forecast $\hat{\pi}^h_\tau$ using the feature vector X_τ. This procedure is repeated for every τ from 2000/01 to 2024/11, and for every horizon $h \in \{1,\dots,24\}$. The full procedure is described formally in Algorithm 1 in Appendix E, which also collects a table of notation (Table E.1).

The parameter grid $\mathscr{G}$ for XGBoost consists of all combinations of the number of trees $n_{\text{est}} \in \{100\}$ and the learning rate $\eta \in \{0.05, 0.1\}$. Forecasting is *direct*: we train a separate model for each horizon h, with the target variable being the h-step-ahead inflation realization π_τ.

3.2 Step 2: A Scan Test for Forecasting Performance

At this point, we have a forecast of inflation using microdata only. The second step of our method evaluates the forecast sequences produced in Step 1 against other forecasting models. As we have discussed in the introduction, the behavior of inflation seems to be non-stationary. Microdata may forecast better at some points in time than others. Therefore our objective is not only to compare average forecasting performance, but also to determine whether one forecast systematically outperforms

another over particular sub-periods and horizons. We therefore begin by defining the benchmark forecasts against which the micro and macro machine learning forecasts will be assessed, and then develop a scan-test procedure for detecting localized episodes of outperformance.

3.2.1 Benchmark I: SARIMA

Our first benchmark is a Seasonal Autoregressive Integrated Moving Average (SARIMA) model estimated on a rolling window with nested hyperparameter selection. The SARIMA order is denoted $\Upsilon = (p,d,q)(P,D,Q,M)$, with fitted parameters Θ. We search over $p,d,q,P,D,Q \in \{0,1\}$ and fix $M = 12$ to account for monthly seasonality. We also include a candidate that forecasts h-step ahead inflation as the trailing 12-month average, $\hat{\pi}_{t+h} = \frac{1}{12}\sum_{j=0}^{11} \pi_{t-j}$.

We choose this SARIMA forecast because it nests a wide range of popular, univariate specifications for inflation forecasting in the literature. Many of these specifications are known to forecast inflation as well as, or better than, richer alternatives. Our specification nests the random walk forecast of Atkeson and Ohanian (2001), the IMA(1,1) forecast of Stock and Watson (2007), the "airline model" of inflation forecasting used by statistics agencies (e.g. Eurostat, 2015), and the ARMA(1,1) model of Ang, Bekaert and Wei (2007).

At each forecast origin τ, hyperparameters are selected by minimizing out-of-sample RMSE over a 48-month validation window $[\tau-47-h, \tau-h]$. For each candidate order Υ, we estimate the model on four rolling 120-month windows ending at $\tau-h-12l$, $l \in \{4,3,2,1\}$, and use each fit to generate h-step-ahead forecasts for the subsequent 12 months. This produces 48 validation forecasts for each Υ; we select Υ^* as the order with the lowest validation RMSE. We then re-estimate SARIMA on the 120 months immediately preceding $\tau-h$ and generate the h-step forecast $\hat{\pi}^h_\tau$ by iterating the fitted model forward. The next two paragraphs define the training and forecasting steps used both in this validation exercise and in the final forecast at origin τ; the full procedure is summarized in Algorithm 2 in Appendix E.

SARIMA training. Training SARIMA$(\Upsilon)(\Theta^{\Upsilon}_{\tau-h-12l})$ on $\{\pi_t\}_{t=\tau-h-12l-119}^{\tau-h-12l}$ means fixing the order Υ and solving

$$\Theta^{\Upsilon}_{\tau-h-12l} = \arg\max_{\hat{\Theta}} \mathscr{L}\left(\hat{\Theta} \,\middle|\, \{\pi_t\}_{t=\tau-h-12l-119}^{\tau-h-12l}\right),$$

i.e. the maximum-likelihood parameters given the training window.

SARIMA forecasting. A generic SARIMA$(p,d,q)(P,D,Q,s)$ model takes the form

$$\Phi(B)\,\Phi_s(B^s)\,(1-B)^d\,(1-B^s)^D\,\pi_t = \Omega(B)\,\Omega_s(B^s)\,\epsilon_t, \qquad \epsilon_t \sim i.i.d.(0,\sigma^2),$$

where $\Phi, \Phi_s, \Omega, \Omega_s$ are lag polynomials and B is the backshift operator. For a forecast origin $\tau-h$, the iterated h-step ahead forecast $\hat{\pi}_{\tau|\tau-h}$ is obtained by feeding the observed history $\{\pi_t\}_{t\le\tau-h}$ into the lag

polynomial and setting future innovations $\epsilon_{t+j} = 0$ for $j > 0$.

3.2.2 Benchmark II: Macro XGBoost

Our second benchmark applies the rolling-window XGBoost procedure of Algorithm 1 to *macro* rather than micro features. Concretely, the feature vector X_t is replaced by lagged values of the macro time series described in Section 2, and all other aspects of the algorithm—rolling window length (120 months), hyperparameter grid, validation scheme, and direct h-step forecasting—are held fixed. This benchmark thus isolates the marginal contribution of micro-price distributional data over and above what standard macroeconomic predictors can achieve. Having defined the benchmark forecasts, we now formalize the pairwise comparison framework used to test whether one method outperforms another over particular sub-periods and horizons.

3.2.3 Null and Alternative Hypotheses

Having defined the benchmark forecasts, we now formalize the pairwise comparison framework used to test whether one method outperforms another over particular sub-periods and horizons.

Let n denote the number of time periods and H the maximal forecasting horizon. For a benchmark method A and a competing method B (e.g. SARIMA vs. micro), let $\varepsilon_{t,A}^{(h)}$ and $\varepsilon_{t,B}^{(h)}$ denote the h-step-ahead forecast residuals at time t, and define

$$D_t^{(h)} = \left|\varepsilon_{t,A}^{(h)}\right| - \left|\varepsilon_{t,B}^{(h)}\right|.$$

With this sign convention, positive values of $D_t^{(h)}$ indicate that method B outperforms method A at time t and horizon h.

The null hypothesis is

$$H_0 : \mathbb{E}[D_t^{(h)}] \le 0 \quad \text{for all } t \in [n] := \{1, \dots, n\},\ h \in [H] := \{1, \dots, H\},$$

so that under H_0, method A is at least as good as method B across all time periods and forecasting horizons. Note that it is possible for the null to be rejected in both directions, that is, for both A versus B and B versus A.

To make rejection substantively meaningful, we consider the restricted alternative

$$H_1 : \sum_{t \in [s,e]} \mathbb{E}[D_t^{(h)}] > 0 \quad \text{for some interval } [s, e] \text{ with } e - s + 1 \ge L_{\min},\ h \in [H].$$

That is, method B outperforms method A on average over some contiguous interval of length at least $L_{\min}$ across the horizons under study. In the empirical implementation below, we take $H = 24$ and $L_{\min} = 7$. Note that $L_{\min}$ is a user-specified parameter. The scan statistic introduced next is designed to

operationalize this alternative by searching over all admissible intervals and horizons for the strongest evidence of such localized outperformance.

3.2.4 Scan Statistic

To test the null hypothesis against the localized alternative, we construct a scan statistic that searches over all admissible time intervals and forecasting horizons for unusually large positive loss differentials.

Let $\mathscr{I}$ be the collection of all intervals $I = [s, e] \subset [n]$ with $|I| \geq L_{\min}$. For each $I = [s, e]$, $h \in [H]$, and $D_t = (D_t^{(1)}, \ldots, D_t^{(H)})^\top$, define

$$S_I^{(h)} = \sum_{t=s}^{e} D_t^{(h)}, \qquad \bar{D}_I^{(h)} = \frac{S_I^{(h)}}{|I|}, \qquad \tilde{Q}_I^{(h)} = \sum_{t=s}^{e} \big(D_t^{(h)} - \bar{D}_I^{(h)}\big)^2.$$

In words, $S_I^{(h)}$ is the cumulative loss differential over interval I, $\bar{D}_I^{(h)}$ is the corresponding within-interval mean, and $\tilde{Q}_I^{(h)}$ is the within-interval sum of squared deviations. We can then define the self-normalized t-statistic for interval I and horizon h as

$$T_I^{(h)} = \frac{S_I^{(h)}}{\sqrt{\tilde{Q}_I^{(h)}}}$$

and the scan statistic is defined as the double maximum

$$T^{\max} = \max_{I \in \mathscr{I}} \max_{1 \leq h \leq H} T_I^{(h)}.$$

This statistic searches jointly across time intervals and forecasting horizons for the most pronounced localized outperformance of method B relative to method A. To assess whether the observed value of $T^{\max}$ is unusually large under the null, we next approximate its sampling distribution with a Gaussian multiplier bootstrap.

3.2.5 Gaussian Multiplier Bootstrap Distribution

Because the null distribution of the scan statistic $T^{\max}$ is nonstandard, owing to the maximization over many intervals and forecasting horizons, we approximate it using a Gaussian multiplier bootstrap. This bootstrap delivers critical values and p-values for testing H_0 against the localized alternative described above.

Let

$$\xi = (\xi_1, \ldots, \xi_n)^\top \sim \mathcal{N}(0, \Theta)$$

be independent of the data, where Θ is a positive semidefinite $n \times n$ covariance matrix specified below. The role of Θ is to encode the temporal dependence structure imposed on the bootstrap multipliers,

so its choice should mirror the dependence properties of the loss differential process under the null. For each interval $I = [s, e]$ and horizon h, define

$$S_I^{*(h)} = \sum_{t=s}^{e} \xi_t (D_t^{(h)} - \bar{D}_{[s,e]}^{(h)}), \quad \text{where } \bar{D}_{[s,e]}^{(h)} = \frac{1}{e-s+1} \sum_{t=s}^{e} D_t^{(h)}$$

and

$$T_I^{*(h)} = \frac{S_I^{*(h)}}{\sqrt{\tilde{Q}_I^{(h)}}}.$$

Note that the bootstrap multipliers are applied only to the numerator of the self-normalized t-statistics. The bootstrap scan statistic is then defined as

$$T^{*,\max} = \max_{I \in \mathcal{I}} \max_{1 \le h \le H} T_I^{*(h)}.$$

Using B bootstrap draws of $T_{(1)}^{*,\max}, \ldots, T_{(B)}^{*,\max}$, the p-value is

$$\hat{p} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{T^{*,(b),\max} > T^{\max}\}}{B+1}.$$

The next subsection specifies Θ under alternative assumptions about temporal dependence in the loss differentials.

3.2.6 Choice of Θ

The remaining ingredient in the bootstrap construction is the covariance matrix Θ, which determines the dependence structure of the Gaussian multipliers and hence the null distribution used for inference. We consider two regimes for Θ, corresponding to whether the loss differentials are treated as independent or temporally dependent across time.

Regime I (Independent Noise). If $D_1, \ldots, D_n$ are independent across time, we take

$$\Theta = I_n.$$

Regime D (Dependent Noise). If $D_1, \ldots, D_n$ are temporally dependent, we take

$$\Theta(t, u) = K\left(\frac{t-u}{b_n}\right),$$

where K is a kernel and b_n is a bandwidth. Our empirical baseline uses the quadratic spectral kernel, introduced by Andrews (1991), because it satisfies the regularity condition; see Appendix A.3.6.

Quadratic Spectral Kernel. The quadratic spectral (QS) kernel is defined as

$$K(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin\left(\frac{6\pi x}{5}\right)}{\frac{6\pi x}{5}} - \cos\left(\frac{6\pi x}{5}\right) \right), \quad x \neq 0,$$

with $K(0) = 1$ by continuity. Unlike the standard Bartlett kernel in the Newey-West estimator (Newey and West, 1987), the QS kernel does not truncate at a finite lag and instead decays smoothly, placing nonzero weight on all lags. It is known to be optimal in a mean squared error sense for long-run variance estimation under certain conditions (Andrews, 1991).

Bandwidth Selection. Conditional on choosing a kernel, the remaining tuning parameter in Regime D is the bandwidth b_n, which controls how quickly dependence decays in the bootstrap covariance matrix. In practice, we adopt a grid-based approach:

$$b_n \in \mathscr{B} = \{b_1, \ldots, b_M\}.$$

For each b_n, we compute the bootstrap distribution and assess the stability of resulting p-values and critical values. We select a bandwidth that yields stable inference across nearby values, avoiding both under- and over-smoothing. In the empirical analysis, we report results for $b_n \in \mathscr{B} = \{0.5, 1.0, 1.5, 2.0\}$ and verify that inference is robust across this range.

3.2.7 Pre-Test for Serial Dependence and Regime Selection

Because the appropriate choice of Θ depends on whether the loss differentials are serially dependent, we conduct a pre-test before applying the regime-specific bootstrap algorithms. This step determines whether inference should proceed under the independent noise regime (Regime I) or the dependent noise regime (Regime D).

Visual Diagnostics. We begin with standard diagnostics for both temporal and cross-horizon dependence in the loss differentials. Specifically, we examine:

- Autocorrelation function (ACF) plots,
- Partial autocorrelation function (PACF) plots,
- Scree plots of the empirical covariance matrix of $D_t = (D_t^{(1)}, \ldots, D_t^{(H)})^\top$.

The ACF and PACF diagnose temporal dependence in the loss differential sequence: slowly decaying autocorrelations or significant lag spikes indicate persistence over time. The scree plot instead summarizes dependence across forecast horizons: a few dominant eigenvalues indicate that the residuals are strongly correlated across horizons, whereas a flatter spectrum suggests weaker cross-horizon dependence.

Ljung–Box Test. To formally test for serial correlation, we apply the Ljung–Box test to $\{D_t^{(h)}\}$ at multiple lags. The null hypothesis is that the sequence is independently distributed (no autocorrelation up to the tested lag). Rejection of the null indicates the presence of temporal dependence.

For each horizon $h \in \{1,\dots,12\}$, let $r_{t,h} = D_t^{(h)}$ denote the loss differential series. Fix a maximum lag m and compute the Ljung–Box statistic

$$Q_h \;=\; T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_h(k)^2}{T-k},$$

where $\hat{\rho}_h(k)$ is the sample autocorrelation of $\{r_{t,h}\}_{t=1}^{T}$ at lag k. We aggregate over horizons via $Q = \sum_{h=1}^{12} Q_h$.

To approximate the null distribution, we use a Gaussian multiplier bootstrap that preserves the cross-horizon dependence at each time t while breaking temporal dependence. Specifically, for each bootstrap draw $b = 1,\dots,B$, sample i.i.d. multipliers $\{\xi_t^{(b)}\}_{t=1}^{T}$ with $\xi_t^{(b)} \sim \mathcal{N}(0,1)$, and form the bootstrapped panel $r_{t,h}^{*(b)} = \xi_t^{(b)} r_{t,h}$ for all h. Compute $Q_h^{*(b)}$ and $Q^{*(b)} = \sum_{h=1}^{12} Q_h^{*(b)}$ from $\{r_{t,h}^{*(b)}\}_{t=1}^{T}$, and estimate the (one-sided) p-value as

$$\hat{p} \;=\; \frac{1+\sum_{b=1}^{B}\mathbf{1}\{Q^{*(b)} > Q\}}{B+1}.$$

Interpretation of the p-value. The null hypothesis of the Ljung–Box test corresponds to the absence of temporal dependence in the loss differential sequences. Under this null, any observed autocorrelation arises purely from random fluctuations. A small p-value indicates that the observed level of autocorrelation is unlikely under temporal independence, leading us to adopt the dependent noise regime (Regime D). A large p-value suggests insufficient evidence of dependence, and we proceed under the independent noise regime (Regime I).

In our application, we use $B = 5{,}000$ bootstrap draws at significance level $\delta = 0.05$. The full procedure is given in Algorithm 3 in Appendix E.

3.2.8 Implementing the Scan Test

In our experiments, we use $B = 5{,}000$ bootstrap draws and apply the test to $n = 239$ monthly observations. We consider the 2 ordered pairwise comparisons SARIMA–micro and SARIMA–macro. With $H = 24$, $L_{\min} = 7$, and $L_{\max} = 12$, the interval collection contains 1,383 candidate intervals, for a total of $24 \times 1{,}383 = 33{,}192$ interval-horizon pairs. The complete procedures for Regime I and Regime D are given in Algorithms 4 and 5 in Appendix E.

Overall, Step 2 delivers two objects. First, it provides formal evidence on whether the micro, macro, and/or SARIMA forecasts outperform one another over particular sub-periods and horizons. Second, it identifies when relative forecast performance shifts over time. These comparisons motivate the final step of the method, where we move from pairwise evaluation to forecast combination and allow the weights on the competing forecasts to adapt dynamically.

3.3 Step 3: Combining Data Modalities with Adaptive Machine Learning

Having produced separate forecasts from microdata, macro data, and univariate information, we combine them using the Fixed Share algorithm of Herbster and Warmuth (1998). The goal is to use all data modalities in order to maximize forecasting power—while acknowledging that because inflation dynamics are non-stationary, the identity of the best forecast may change over time. The algorithm maintains a weighted average of forecasts and updates the weights dynamically based on each model's realized forecasting performance, offering regret guarantees in a non-stationary environment where the identity of the best expert can change over time. This property is well-suited to inflation forecasting, where the relative informativeness of micro and macro data may shift.

Setup. Let K denote the number of experts, indexed by $k = 1, \ldots, K$. In our application, $K = 3$: expert $k = 1$ is the Micro XGBoost forecast, expert $k = 2$ is the Macro XGBoost forecast, and expert $k = 3$ is the SARIMA forecast. For each horizon $h \in \{1, \ldots, 24\}$ and time t, let $\widehat{\pi}_t^{h,k}$ denote the h-step-ahead forecast of expert k and π_t the realized inflation. The algorithm maintains a probability vector of weights $\{w_t^{h,k}\}_{k=1}^K$ over experts and outputs the weighted aggregate forecast

$$\widetilde{\pi}_t^h = \sum_{k=1}^K w_t^{h,k} \widehat{\pi}_t^{h,k}.$$

Fixed Share Update. At each t, after observing π_t, the weight of expert k is updated in two steps. First, a multiplicative weights (MW) step down-weights experts in proportion to their absolute forecast error:

$$\bar{w}_{t+1}^{h,(\eta),k} = \frac{w_t^{h,(\eta),k} \exp\left(-\eta \left|\pi_t - \widehat{\pi}_t^{h,k}\right|\right)}{\sum_{j=1}^K w_t^{h,(\eta),j} \exp\left(-\eta \left|\pi_t - \widehat{\pi}_t^{h,j}\right|\right)},$$

where $\eta > 0$ is the learning rate. Second, a fixed-share mixing step redistributes a fraction $\alpha \in [0,1]$ of weight uniformly across all experts, preventing any expert from being permanently abandoned:

$$w_{t+1}^{h,(\eta),k} = (1-\alpha)\, \bar{w}_{t+1}^{h,(\eta),k} + \frac{\alpha}{K}.$$

All weights are initialized uniformly, $w_{t_0}^{h,(\eta),k} = 1/K$. The full path of forecasts and weights for a given η is produced by Algorithm 6 in Appendix E.

Regret Guarantee. The Fixed Share algorithm provides a worst-case performance guarantee, clarifying precisely what the algorithm optimizes in a non-stationary environment. The relevant object is *regret*: the gap between the cumulative loss of the online learner and the cumulative loss of an ideal benchmark sequence of experts chosen in hindsight after seeing the full sample. Formally, let

$$L_n^{\mathrm{FS}} = \sum_{t=1}^n |\hat{y}_t - y_t|$$

denote the cumulative loss of the Fixed Share forecast over T periods, where $\hat{y}_t$ is the algorithm's prediction at date t, y_t is the realized outcome, and $|\hat{y}_t - y_t|$ is the one-period loss. Let $L_n^{(k_0,k_1,\ldots,k_S)}$ denote the cumulative loss of the best sequence of experts $k_0, k_1, \ldots, k_S$ in hindsight when we allow the identity of the expert to switch S times over the sample. Let $y_t, \hat{y}_t \in [0, y_{\max}]$ where $y_{\max}$ denotes the largest possible inflation rate. Herbster and Warmuth (1998) (Theorem 1 and 3) show that [5]

$$L_n^{\mathrm{FS}} \leq \frac{y_{\max}\eta}{1-e^{-y_{\max}\eta}} L_n^{(k_0,k_1,\ldots,k_S)} + \frac{y_{\max}}{1-e^{-y_{\max}\eta}} \left\{ (S+1)\log K + S\log\left(\frac{1}{\alpha}\right) + (n-S-1)\log\left(\frac{1}{1-\alpha}\right) \right\}.$$

where K is the number of experts, η is the learning rate, α is the fixed-share mixing parameter. Crucially, the guarantee holds for *any* sequence of outcomes. It does not require stationarity, ergodicity, or any specific model of the data generating process, which is exactly why the method is attractive for inflation forecasting after structural breaks.

Each term on the right-hand side has a simple interpretation. The first term is essentially the comparator loss, up to a multiplicative factor $1 + O(y_{\max}\eta)$ which is close to 1 for a small learning rate. The second term is proportional to the number of switches S for the comparator and the logarithm of the number of experts $\log K$. While it grows with K slowly, a large K can still result in a constant multiplicative increase in the loss. This motivates our scan test that can filter out experts that are uniformly dominated, which distinguishes us from Fouliard et al. (2023) that involves many experts. The last two terms implies α should be small when $S \ll n$. In particular, it is straightforward to show that the optimal α is given by $S/(n-1)$. In our empirical analysis, we set $\alpha \approx 4/n$.

To gain more insights about the impact of η, we set α to be the optimum and consider a local expansion around 0: [6]

$$L_n^{\mathrm{FS}} - L_n^{(k_0,k_1,\ldots,k_S)} = O\left(\eta n + \frac{\log(Kn)}{\eta}\right).$$

A small η yields stable weights but may adapt too slowly; a large η allows quick reallocation but can overreact to short-run noise.

Fixed Share can be understood as a generalization of the multiplicative weights (MW) algorithm from the online learning literature (Cesa-Bianchi and Lugosi, 2006). The MW algorithm provides a similar regret guarantee, but only relative to the single best expert in hindsight. That is too restrictive for our application: we do not expect the same forecast to be best both before and after large changes in inflation dynamics. The reason MW fails in this environment is that it updates weights multiplicatively based on losses, so an expert that performs poorly early on sees its weight decay exponentially toward zero. Even if that expert later becomes the best, it cannot recover meaningful weight. The key

[5] Herbster and Warmuth (1998) assume the $y_t, \hat{y}_t \in [0,1]$. If we replaced y_t by $y_t/y_{\max}$ and η by $y_{\max}\eta$, the sequence $\hat{y}_t$ would be $\hat{y}_t/y_{\max}$. This rescaling yields the stated regret bound.

[6] We used three facts: (a) for small η, $1-e^{-y_{\max}\eta} = y_{\max}\eta + O(\eta^2)$; (b) when $\alpha = S/(n-1)$, $S\log\left(\frac{1}{\alpha}\right) + (n-S-1)\log\left(\frac{1}{1-\alpha}\right) = S\log((n-1)/S) + (n-S-1)\log\left(\frac{n-1}{n-S-1}\right) \leq S\log((n-1)/S) + S = O(S\log n)$; (c) $L_n^{(k_0,k_1,\ldots,k_S)} = O(n)$.

innovation of Fixed Share is the mixing step governed by α: after each multiplicative update, a fraction of weight is redistributed across all experts. This keeps every expert alive, so the algorithm can track switches in the identity of the best forecast source over time.

Fixed Learning Rate. As a first implementation, we fix $\eta = 0.1$ and $\alpha = 0.02$ throughout, with all weights initialized uniformly at $1/K$ in January 2005. This specification serves as a transparent benchmark: the fixed learning rate is consistent with standard practice in the online learning literature (Herbster and Warmuth, 1998), and with η held constant the only free parameter is α. The combination reduces to a single pass of Algorithm 6 (Appendix E) with $\eta = 0.1$.

Adaptive Selection of the Learning Rate. While $\eta = 0.1$ provides a reasonable default, the optimal degree of responsiveness may itself vary over the forecast horizon and over time. We therefore also consider an adaptive variant that selects η at each t using a rolling-window MAE criterion, restricting the search to the immediate neighborhood of the previously selected value on a pre-specified grid $\mathscr{H} = \{\eta^{(1)} < \eta^{(2)} < \cdots < \eta^{(G)}\}$. In our implementation, $\mathscr{H}$ consists of $G = 20$ equally spaced values on $[0.01, 1.0]$, with rolling window $W = 60$ months and default $\eta_0 = 0.1$. Concretely, if $\eta^{\star}_{t-1} = \eta^{(g_{t-1})}$, the candidate set at time t is

$$\mathscr{I}_t = \{\max(1, g_{t-1} - 1),\ g_{t-1},\ \min(G, g_{t-1} + 1)\},$$

and the selected learning rate minimizes the rolling-window MAE over the past W periods among these neighbors:

$$\eta^{\star}_t = \eta^{(g_t)}, \qquad g_t = \arg\min_{g \in \mathscr{I}_t} \frac{1}{W} \sum_{s=t-W}^{t-1} \left| \pi_s - \widetilde{\pi}_s^{h,(\eta^{(g)})} \right|.$$

Restricting candidates to the neighborhood of $\eta^{\star}_{t-1}$ reduces computational cost relative to a full-grid search while preserving adaptivity. The complete procedure is described in Algorithms 6 and 7 in Appendix E.[7]

4 Results: Forecasting Inflation with Microdata

We present results in five steps. First, we establish a univariate SARIMA benchmark that exploits only the history of inflation itself. The benchmark performs well in the stable pre-2020 window (2015–2019), but deteriorates sharply in the post-2020 window (2020–2024), where errors rise to [0.32, 0.52]. Second, we evaluate the micro XGBoost forecast against this benchmark. Microdata adds little forecasting power during normal times, but in the post-2020 window the micro forecast improves on

[7] We have considered other algorithms for updating η, such as AdaHedge (de Rooij et al., 2014); however they performed poorly in our setting.

SARIMA at medium and long horizons. A scan test confirms that the micro model significantly outperforms SARIMA during the February–December 2022 episode of sharp inflation movements. Third, we carry out the same exercise for the macro XGBoost forecast, again finding that improvements relative to SARIMA are concentrated in the post-2020 window. Fourth, we combine all three forecasts—micro, macro, and univariate—using the Fixed Share algorithm. The combination improves on SARIMA at all 24 horizons in the post-2020 window, and the expert weights reveal that the importance of microdata rises in that period. Fifth, we study a two-expert online-learning benchmark that combines only macro and univariate forecasts. This benchmark performs strongly after 2020, but remains slightly weaker than the full three-expert combination, indicating that microdata retain incremental value in the combination stage.

4.1 Warmup: Univariate Forecasting of Inflation

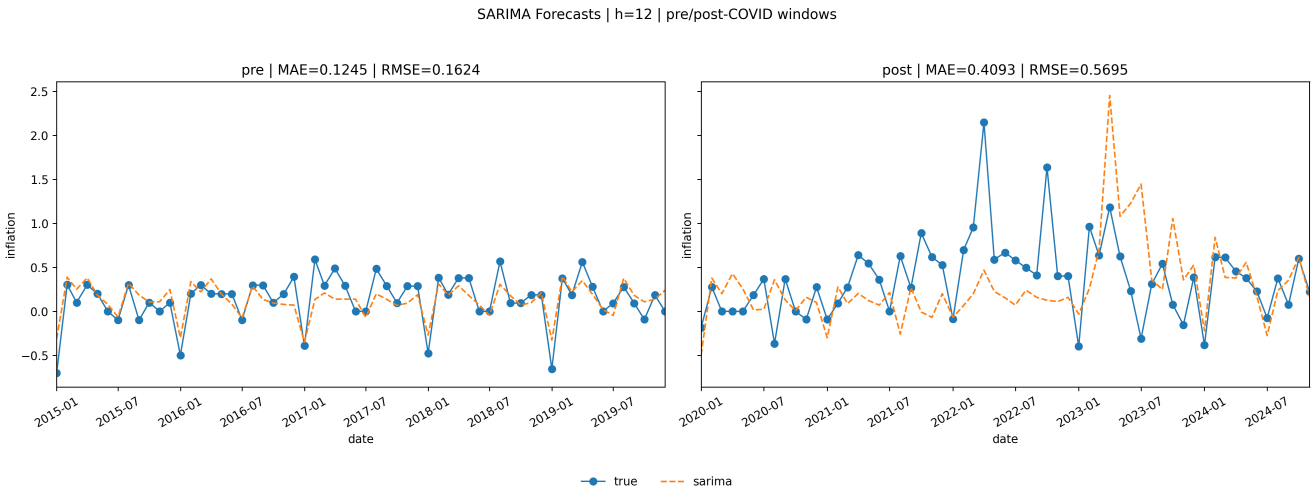
We begin with a univariate benchmark that exploits only the history of inflation itself. Specifically, we estimate a rolling-window SARIMA model with nested hyperparameter selection, as described in Section 3. This benchmark serves two purposes. First, it provides a transparent baseline for what can be achieved using only the time-series dynamics of inflation. Second, it helps isolate gains attributable to richer information sets from those that merely reflect predictable autoregressive or seasonal structure. For completeness, we report SARIMA performance for horizons $h = 1, \ldots, 24$ across two evaluation windows: pre-2020 (2015–2019) and post-2020 (2020–2024). The multivariate comparisons that follow focus on the common set of horizons used in the scan-test analysis.

Table 2 reveals two clear patterns. First, SARIMA performs substantially better over the pre-2020 window. Across all horizons, pre-2020 mean absolute error (MAE) remains tightly concentrated in the range [0.1065, 0.1260], while root mean squared error (RMSE) ranges from 0.1367 to 0.1654. The lowest errors are attained at $h = 1$, indicating that in a relatively stable macroeconomic environment a univariate model captures a substantial share of the predictable variation in inflation.

Second, the post-2020 window proves far more challenging for the univariate benchmark. Post-2020 mean absolute error rises to the [0.3236, 0.5232] range, and root mean squared error increases even more sharply, to [0.4420, 1.2738]. The deterioration is more pronounced at longer horizons: post-2020 mean absolute error increases from 0.3315 at $h = 1$ to 0.4093 at $h = 12$ and 0.5232 at $h = 24$, while the corresponding root mean squared error values rise from 0.4477 to 0.5695 and 1.2738, respectively. This pattern is consistent with the view that inflation dynamics became substantially less well described by stable autoregressive and seasonal structure after 2020.

Figure 2 illustrates these patterns for horizon $h = 12$. The solid line denotes realized inflation and the dashed line denotes the SARIMA forecast. The two panels correspond to 2015–2019 and 2020–

Figure 2: SARIMA forecasts vs. realized inflation: $h = 12$



Notes: The solid line plots realized inflation and the dashed line plots the rolling-window SARIMA 12-step-ahead forecast. The two panels correspond to the 2015–2019 and 2020–2024 evaluation periods. Each panel title reports the SARIMA mean absolute error (MAE) and root mean squared error (RMSE) for that sub-period.

2024, with each panel title reporting the sub-period-specific forecast errors. The model tracks inflation reasonably well during the pre-2020 window but fails to anticipate the sharp rise and subsequent decline in inflation after 2020. Appendix Figures C.5 and C.6 report the corresponding $h = 1$ and $h = 6$ forecasts.

Taken together, the SARIMA results establish a demanding benchmark at short horizons under normal conditions, while revealing the clear limits of purely univariate forecasting when inflation is subject to large structural shifts. These findings motivate the richer forecasting exercises that follow, which incorporate macroeconomic predictors and micro-level price-distribution features.

4.2 Forecasting Performance of Micro and Macro Data

We report the out-of-sample forecasting performance of the micro and macro XGBoost models, each benchmarked against SARIMA. A common pattern emerges: both models add little over the univariate baseline when inflation is stable, but deliver broad-based gains during the high-inflation episode after 2020.

4.2.1 Micro data

Aggregate evidence: pre-2020 versus post-2020 windows. Table 3 reports the mean absolute error and root mean squared error improvement of the micro XGBoost model relative to SARIMA across the two evaluation windows—pre-2020 (2015–2019) and post-2020 (2020–2024)—for all horizons $h = 1, \ldots, 24$, and two findings stand out.

Over the pre-2020 window, the micro model underperforms SARIMA at every horizon under both loss functions. Mean absolute error deterioration ranges from 13.28% to 42.52%, and root mean squared error deterioration from 6.72% to 55.99%. This uniform underperformance suggests that, in the relatively subdued inflation environment of 2015–2019, the cross-sectional distribution of price changes contributes little forecasting information beyond what a univariate time-series benchmark already captures.

The post-2020 window tells a materially different story. Although the micro model continues to underperform at several short horizons, it improves upon SARIMA at 18 of the 24 horizons in mean absolute error and at 19 horizons in root mean squared error. The gains become especially pronounced from roughly $h = 6$ onward, with the largest improvement reaching 23.45% in mean absolute error at $h = 14$ and 58.27% in root mean squared error at $h = 24$. This pattern suggests that micro-price distributional features are most informative precisely when inflation is elevated and volatile—when aggregate dynamics become less stable and a purely univariate benchmark is more likely to miss shifts in the underlying inflation process. The contrast between the two windows therefore indicates that the value of microdata is state dependent: limited during normal times but sharply elevated during periods of macroeconomic stress.

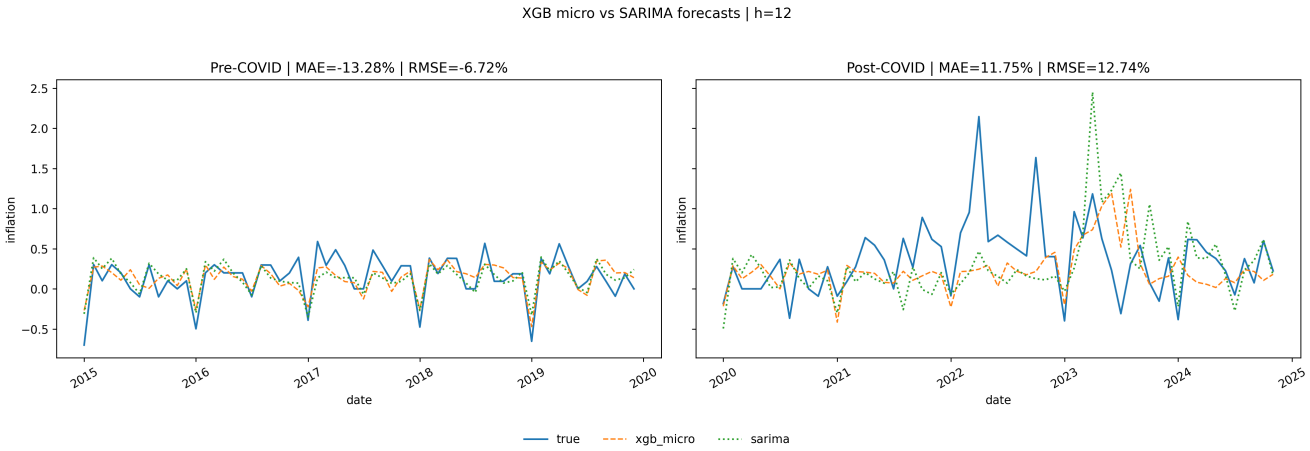
Visual evidence across forecast horizons. Figure 3 plots both sets of forecasts against realized inflation at horizon $h = 12$. In the pre-2020 panels, SARIMA and the micro XGBoost model track realized inflation similarly, with SARIMA generally remaining slightly closer to the realized series. In the post-2020 panels, by contrast, the micro model's relative advantage is clear: the micro-based forecast more closely captures the broad inflation run-up and subsequent decline, whereas SARIMA tends to lag or smooth through these changes. The visual patterns thus corroborate the table-based findings: forecasting gains from microdata are small or negative at short horizons but become meaningful at medium and long horizons during the high-inflation episode. Appendix Figures C.7, C.8, and C.9 report the $h = 1$, $h = 6$, and $h = 24$ forecasts.

4.2.2 Macro data

While the primary focus of the paper is on microdata, we also document the performance of the macro XGBoost model, which displays an analogous regime-dependent pattern.

Aggregate evidence: pre-2020 versus post-2020 windows. Table 4 reports the mean absolute error and root mean squared error improvement of the macro XGBoost model relative to SARIMA across the two evaluation windows—pre-2020 (2015–2019) and post-2020 (2020–2024)—for all horizons $h = 1, \ldots, 24$. Two findings stand out.

Figure 3: Micro XGBoost vs. SARIMA forecasts: $h = 12$



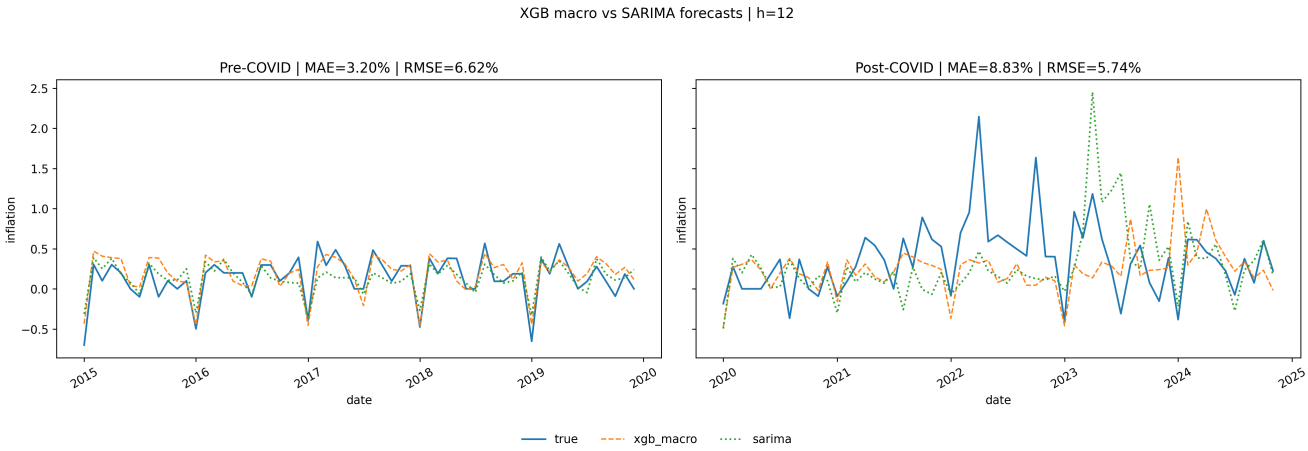
Notes: Each panel plots realized inflation together with the rolling-window XGBoost micro 12-step-ahead forecast and the rolling-window SARIMA forecast over one evaluation window. The two panels correspond to 2015–2019 and 2020–2024. Realized inflation is shown with a solid line, the XGBoost micro forecast with a dashed line, and the SARIMA forecast with a dotted line. Each panel title reports the percentage MAE and RMSE improvement of XGBoost micro relative to SARIMA in that window.

Over the pre-2020 window, the macro model mostly underperforms SARIMA, though the underperformance is not uniform. The macro model delivers modest positive gains at $h = 11$ and $h = 12$ (5.40% and 3.20% in mean absolute error; 7.32% and 6.62% in root mean squared error) and is close to parity at $h = 14$; at the remaining horizons it still deteriorates relative to SARIMA, with the largest losses reaching 35.57% in mean absolute error and 48.70% in root mean squared error. The overall picture is one in which macro covariates only occasionally add predictive value beyond the univariate benchmark in a stable-inflation environment.

The post-2020 window tells a materially different story. Although the macro model remains mixed at a few short horizons, it improves upon SARIMA at 21 of the 24 horizons in mean absolute error and at 19 horizons in root mean squared error. The gains become especially pronounced from roughly $h = 7$ onward, with the largest improvement reaching 30.89% in mean absolute error at $h = 17$ and 60.02% in root mean squared error at $h = 24$. This pattern suggests that macro covariates become most informative precisely when inflation is elevated and volatile—when aggregate dynamics become less stable and a purely univariate benchmark is more likely to miss shifts in the underlying inflation process. The contrast between the two windows therefore indicates that the value of macro data is also state dependent: limited during normal times but substantially higher during periods of macroeconomic stress.

Visual evidence across forecast horizons. Figure 4 plots both sets of forecasts against realized inflation at horizon $h = 12$. In the pre-2020 panels, SARIMA and the macro XGBoost model track realized inflation similarly, with SARIMA generally remaining slightly closer. In the post-2020 panels, by con-

Figure 4: Macro XGBoost vs. SARIMA forecasts: $h = 12$



Notes: Each panel plots realized inflation together with the rolling-window XGBoost macro forecast and the rolling-window SARIMA forecast over one evaluation window. The two panels correspond to 2015–2019 and 2020–2024. Realized inflation is shown with a solid line, the XGBoost macro forecast with a dashed line, and the SARIMA forecast with a dotted line. Each panel title reports the percentage MAE and RMSE improvement of XGBoost macro relative to SARIMA in that window.

trast, the macro model's relative advantage is clear: the macro-based forecast more closely captures the broad inflation run-up and subsequent decline, whereas SARIMA tends to lag these changes. The visual patterns thus corroborate the table-based findings: forecasting gains from macro data are limited and horizon-specific before 2020 but become meaningful at medium and long horizons during the high-inflation episode. Appendix Figures C.10, C.11, and C.12 report the $h = 1$, $h = 6$, and $h = 24$ forecasts.

4.3 Conducting the Scan Test

Having obtained the individual macro and micro forecasts, we now conduct the scan test for each set of predictions. This step serves two purposes. First, it determines which forecasts should enter the online-learning combination. Since the regret guarantees of the online-learning algorithm deteriorate as the number of experts expands, adding experts mechanically is not costless. We therefore want to include only forecasts that show evidence of improving on the univariate forecast baseline, rather than enlarging the expert set with models that do not add predictive power. Second, the scan test is of independent interest because in a non-stationary environment the relevant question is often not whether one forecast dominates on average over the full sample, but whether it outperforms during some ex ante unknown episode and horizon. We begin with the macro forecast relative to SARIMA, and then turn to the corresponding exercise for microdata.

Dependence structure and bootstrap regime of macrodata. We consider two dimensions of the macrodata forecast loss differential to inform the choice of bootstrap regime: cross-horizon dependence across the $H = 24$ horizons and serial persistence over time. Figure C.1 in Appendix C addresses the first through the scree plot of the sample covariance matrix of

$$r_{t,h} = |\pi_t - \hat{\pi}_t^{h,\text{SARIMA}}| - |\pi_t - \hat{\pi}_t^{h,\text{Macro}}|, \qquad h = 1, \ldots, 24,$$

for $T = 239$ observations. The eigenvalue decline is gradual: the first principal component explains 29.60% of the variance and the first five jointly account for 75.11%, so cross-horizon co-movement in the loss differential is real but spread over several latent dimensions rather than concentrated in a single common factor.

For serial dependence, the multiplier-bootstrap Ljung–Box test yields $Q = 1093.91$ ($m = 24$, $T = 239$, $B = 5{,}000$, $p = 0.0588$); we do not reject the null of no serial correlation at the 5% significance level. Figure C.2 in Appendix C corroborates this result through the ACF and PACF of the cross-horizon mean differential $\bar{r}_t = H^{-1} \sum_{h=1}^{H} r_{t,h}$, which at each date tracks the average advantage of the macro model over SARIMA. The 95% confidence band is $\pm 1.96/\sqrt{239} \approx \pm 0.127$; several positive spikes exceed these bounds, most notably at lags 1, 5, 6, 9, 15, and 20 in the ACF and at lags 1, 5, 6, 9, and 20 in the PACF, together with a small negative PACF spike at lag 18. Nevertheless, there is no slow-decay profile in the ACF and no clean low-order autoregressive structure in the PACF—the dependence is moderate and episodic rather than systematic. We accordingly adopt Regime I (i.i.d. Gaussian multipliers) for the baseline scan test.

Scan test results and robustness of macrodata. Under the baseline Regime I specification ($H = 24$, $T = 239$, $B = 5{,}000$), the maximum scan statistic is $T^{\max} = 10.289$ with a bootstrap p-value of 0.0044. The maximizing interval runs from September 2022 to March 2023, with the strongest signal at horizon $h = 16$ over a window of seven months. Because the scan statistic is constructed from

$$D_t^{(h)} = |\epsilon_{t,\text{SARIMA}}^{(h)}| - |\epsilon_{t,\text{Macro}}^{(h)}|,$$

this result implies that the macro XGBoost model significantly outperforms SARIMA during the localized episode covering the later phase of the inflation surge and the beginning of the subsequent disinflation. Economically, this is precisely the period in which macro covariates should be especially valuable: aggregate inflation dynamics were shifting rapidly, so a univariate benchmark was less able to track the persistence and turning point of the episode.

As a robustness check, we implement the Regime D version of the same multi-length scan test using a quadratic-spectral covariance kernel and the bandwidth grid $b_n \in \{0.5, 1.0, 1.5, 2.0\}$. The maximizing interval and horizon remain unchanged across all four specifications—September 2022 to March 2023 at horizon $h = 16$—with associated bootstrap p-values of 0.0072, 0.0112, 0.0334, and

0.0716, respectively. The null is therefore rejected across the first three bandwidth values but not at the largest bandwidth. The dependent-noise robustness check thus supports the baseline Regime I conclusion for moderate bandwidth choices, while also indicating that the macro-versus-SARIMA result is somewhat sensitive to the most aggressive covariance specification.

Dependence structure and bootstrap regime of microdata. As with the macrodata, we consider two dimensions of the microdata forecast loss differential to inform the choice of bootstrap regime: *cross-horizon dependence* across the $H = 24$ horizons and *serial persistence* over time. Figure C.3 in Appendix C addresses the first through the scree plot of the sample covariance matrix of

$$r_{t,h} = |\pi_t - \hat{\pi}_t^{h,\text{SARIMA}}| - |\pi_t - \hat{\pi}_t^{h,\text{Micro}}|, \qquad h = 1, \ldots, 24,$$

for $T = 239$ observations. The eigenvalue decline is gradual: the first principal component explains 31.15% of the variance and the first five jointly account for 71.46%, so cross-horizon co-movement in the loss differential is real but spread over several latent dimensions rather than concentrated in a single common factor.

For serial dependence, the multiplier-bootstrap Ljung–Box test yields $Q = 928.74$ ($m = 24$, $T = 239$, $B = 5{,}000$, $p = 0.0960$); we do not reject the null of no serial correlation at conventional significance levels. Figure C.4 in Appendix C corroborates this result through the ACF and PACF of the cross-horizon mean differential $\bar{r}_t = H^{-1} \sum_{h=1}^{H} r_{t,h}$, which at each date tracks the average advantage of the micro model over SARIMA. The 95% confidence band is $\pm 1.96/\sqrt{239} \approx \pm 0.127$; most spikes lie within these bounds, and the few isolated exceedances—at lags 2, 11, 12, 18, and 21 in the ACF and at lags 2 and 11 in the PACF—show no slow-decay profile or low-order autoregressive structure. We accordingly adopt Regime I (i.i.d. Gaussian multipliers) for the baseline scan test.

Scan test results and robustness of microdata. Under the baseline Regime I specification ($H = 24$, $T = 239$, $B = 5{,}000$), the maximum scan statistic is $T^{\max} = 12.665$ with a bootstrap p-value of 0.0028. The maximizing interval runs from February 2022 to December 2022, with the strongest signal at horizon $h = 13$ over a window of eleven months. Because the scan statistic is constructed from

$$D_t^{(h)} = |\epsilon_{t,\text{SARIMA}}^{(h)}| - |\epsilon_{t,\text{Micro}}^{(h)}|,$$

this result implies that the micro XGBoost model significantly outperforms SARIMA during the localized episode of sharp inflation acceleration and subsequent reversal. Economically, this is precisely the period in which cross-sectional price information should be most valuable. As we will discuss, benchmark macroeconomic models predict microdata forecasts inflation well, when shocks make inflation high and volatile.

As a robustness check, we implement the Regime D version of the same multi-length scan test

using a quadratic-spectral covariance kernel and the bandwidth grid $b_n \in \{0.5, 1.0, 1.5, 2.0\}$. The maximizing interval and horizon remain unchanged across all four specifications—February 2022 to December 2022 at horizon $h = 13$—with associated bootstrap p-values of 0.0022, 0.0062, 0.0194, and 0.0496, respectively. The null is therefore rejected across the full bandwidth range. The inference that microdata significantly outperform SARIMA over this localized high-inflation episode is therefore robust both to the baseline i.i.d. multiplier specification and to the dependent-noise alternative.

Taken together, the scan tests show that both the macro and micro forecasts improve on the SARIMA benchmark over localized episodes. This is the relevant criterion for expert selection in the next step. Because online learning should not include additional experts unless they add predictive content relative to the baseline, the scan test evidence justifies carrying both Macro XGBoost and Micro XGBoost, alongside SARIMA, into the online-learning combination.

4.4 Combining Data Modalities with Online Learning

We combine the three forecast sources studied above—Micro XGBoost, Macro XGBoost, and SARIMA—using fixed-share online learning. We consider two specifications. The first is a fixed-share benchmark that holds the learning rate constant at $\eta = 0.1$ throughout. The second is the adaptive rule from Algorithm 7, which selects the learning rate period by period via the neighbor-only rolling search. In both cases, we set $K = 3$, $\alpha = 0.02$, and $W = 60$ months, with all recursive updates beginning in January 2005. Results are reported for the pre-2020 (2015–2019) and post-2020 (2020–2024) evaluation windows.

Fixed-learning-rate benchmark. Table 5 shows that fixing $\eta = 0.1$ yields modest and uneven gains over SARIMA in the pre-2020 window but broad-based improvements in the post-2020 window. In the pre-2020 window, average gains are only 0.20% in mean absolute error and −1.78% in root mean squared error, indicating that a moderate fixed-share rule adds little over the univariate benchmark when the inflation environment is stable. In the post-2020 window, by contrast, the fixed-η rule improves on SARIMA with average gains of 13.87% in mean absolute error and 15.13% in root mean squared error, with positive MAE improvements at all 24 horizons. The expert weights under this specification, however, remain close to equal shares throughout. At $h = 12$, for example, the average weights on Micro XGBoost, Macro XGBoost, and SARIMA are $(0.318, 0.321, 0.361)$ in the pre-2020 window and $(0.341, 0.346, 0.313)$ in the post-2020 window, reflecting the limited reallocation capacity of a low fixed learning rate.

Adaptive learning-rate selection. Table 6 shows that the adaptive neighbor-only rule outperforms the fixed-η benchmark in both evaluation windows. In the pre-2020 window, the adaptive rule delivers

average improvements of 1.47% in mean absolute error and 0.29% in root mean squared error, with gains at 15 of 24 horizons in mean absolute error and 14 of 24 horizons in RMSE. In the post-2020 window, the adaptive rule improves on SARIMA at all 24 horizons in mean absolute error and at 22 of 24 horizons in root mean squared error, with average gains of 14.49% in MAE and 15.70% in RMSE. The largest post-2020 gains are 29.44% in mean absolute error at $h = 17$ and 53.94% in root mean squared error at $h = 24$. By allowing η_t to rise well above 0.1, the adaptive rule reallocates weight more aggressively toward the best-performing expert at each horizon—generating gains that the fixed-η rule cannot fully replicate, especially when the relative informativeness of the three forecast sources shifts rapidly.

Figure 5 displays the aggregate forecast path and the selected expert weights for the representative horizon $h = 12$. The algorithm is trained from 2005 onward, but diagnostics are shown only for the pre-2020 and post-2020 windows. The corresponding $h = 1$, $h = 6$, and $h = 24$ diagnostics, as well as the selected η_t paths for all four horizons, are reported in Appendix C.

Figure 5: Three-expert online-learning diagnostics: $h = 12$



Notes: The top panel plots realized inflation and the online-learning aggregate forecast for the pre-2020 (2015–2019) and post-2020 (2020–2024) windows. The bottom panel plots the time-varying weights on Micro XGBoost, Macro XGBoost, and SARIMA. The selected learning rate η_t path is reported in Appendix C.

At $h = 1$, the adaptive combination remains SARIMA-centered, especially in the pre-2020 window (see Appendix Figure C.13). The average SARIMA weight is 0.647 in the pre-2020 window and 0.459 in the post-2020 window. The average selected learning rate is already high in the pre-2020 window (0.873) and rises further to 0.991 in the post-2020 window, indicating that the selector frequently moves toward the upper end of the grid when short-horizon relative forecast performance changes quickly.

At $h = 6$, the combination becomes more balanced, especially in the post-2020 window (see Appendix Figure C.14). Micro XGBoost receives the largest average weight (0.381), followed by SARIMA (0.364) and Macro XGBoost (0.254). The post-2020 gains reach 4.39% in mean absolute error and 5.65% in root mean squared error, while the average selected learning rate drops to 0.386, consistent with smoother medium-horizon reallocation across experts than at $h = 1$.

At $h = 12$, the adaptive combination is most clearly diversified. In the post-2020 window, the average weights on Micro XGBoost and Macro XGBoost are 0.360 and 0.368, respectively, while the average SARIMA weight falls to 0.271. This is also the most balanced of the four highlighted horizons across both evaluation windows: the rule improves on SARIMA by 7.18% in mean absolute error and 7.98% in root mean squared error in the pre-2020 window, and by 14.46% in MAE and 14.30% in RMSE in the post-2020 window. The average post-2020 learning rate of only 0.243 suggests that these gains come from smoother diversification rather than rapid switching toward a single expert.

At $h = 24$, long-horizon performance is driven jointly by Macro XGBoost and SARIMA (see Appendix Figure C.15). In the post-2020 window, their average weights are 0.389 and 0.406, respectively, while Micro XGBoost retains a smaller but still meaningful average weight of 0.204. The corresponding gains rise to 20.90% in mean absolute error and 53.94% in root mean squared error, and the average selected learning rate reaches 0.998, indicating that the selector frequently moves to the top of the grid when long-horizon expert rankings shift sharply.

Overall, the online-learning results reinforce the main message of the paper. In the pre-2020 window, both specifications produce only modest average gains over SARIMA, with the fixed-η rule often close to neutral and the adaptive rule delivering small but more consistent improvements. In the post-2020 window, by contrast, both rules deliver broad-based gains across horizons, with the adaptive rule consistently outperforming the fixed-η benchmark. The combination is therefore valuable not because one expert always dominates, but because the identity of the most informative forecast source changes over time and the fixed-share rule can reallocate weight in real time.

4.5 SARIMA + Macro Online Learning as a Benchmark without Microdata

To isolate the incremental contribution of microdata to the online-learning step, we next consider a two-expert benchmark that combines only Macro XGBoost and SARIMA. As in Section 3, we study two implementations. The first is a fixed-share benchmark with constant learning rate $\eta = 0.1$. The second is the adaptive neighbor-only rule, with $\alpha = 0.02$, $W = 60$, and $\eta_0 = 0.1$. This exercise asks whether a flexible combination of macro and univariate forecasts can already replicate the gains that we attribute to adding microdata. More broadly, it helps motivate the full three-expert procedure. In a non-stationary environment, the objective is not to decide once and for all whether microdata should be used, but to employ an algorithm that is robust to changing conditions and can adaptively choose among experts over time. The scan test shows that microdata is useful only in some episodes, rather than uniformly over the sample. This is exactly the setting in which a three-expert online-learning rule is attractive: it can shift weight toward microdata when they become informative, while continuing to rely on the SARIMA + macro benchmark when they do not. The comparison in this section therefore asks whether that additional adaptive flexibility delivers empirical value.

Fixed-learning-rate benchmark. Table 7 shows that the fixed-η two-expert benchmark produces modest and uneven gains before 2020 but broad-based gains afterwards. In the pre-2020 window, average improvements are 1.35% in mean absolute error and −0.16% in root mean squared error, with positive gains at 15 of 24 horizons in mean absolute error and 13 of 24 horizons in root mean squared error. In the post-2020 window, by contrast, the fixed-η rule improves on SARIMA at all 24 horizons in mean absolute error and at 23 of 24 horizons in root mean squared error, with average gains of 13.52% and 14.10%, respectively. The fixed learning rate nevertheless keeps the weights close to an even split: at $h = 12$, for example, the average weights on Macro XGBoost and SARIMA are $(0.471, 0.529)$ in the pre-2020 window and $(0.526, 0.474)$ in the post-2020 window. As in the three-expert case, a low fixed learning rate limits the algorithm's ability to reallocate sharply when relative forecast performance changes.

Adaptive learning-rate selection. Table 8 shows that the adaptive neighbor-only rule modestly outperforms the fixed-η benchmark overall. In the pre-2020 window, the adaptive rule delivers average improvements of 1.97% in mean absolute error and 1.04% in root mean squared error, with positive gains at 17 of 24 horizons in mean absolute error and 14 of 24 horizons in root mean squared error. At $h = 12$, for example, the adaptive combination improves on SARIMA by 9.57% in mean absolute error and 9.53% in root mean squared error. This is materially better than the standalone micro forecast in the pre-2020 window and also improves upon the standalone macro forecast at the same horizon.

The post-2020 window is where the adaptive benchmark becomes most informative. The two-

expert rule improves on SARIMA at all 24 horizons in mean absolute error and at 23 of 24 horizons in root mean squared error, with average gains of 13.84% and 14.74%, respectively. The largest gains reach 29.05% in mean absolute error at $h = 17$ and 52.04% in root mean squared error at $h = 24$, both stronger than under the fixed-η benchmark. Compared with the standalone micro forecast, the two-expert benchmark is competitive at several horizons, indicating that adaptive combination of macro and univariate information already captures a substantial share of the post-2020 forecasting signal. At the same time, the full three-expert combination remains slightly stronger on average, suggesting that microdata still add incremental value in the combination stage.

Visual evidence across forecast horizons. Figure 6 displays the aggregate forecast path and expert weights for the representative horizon $h = 12$. In the pre-2020 window, the two-expert combination tracks realized inflation more closely than either standalone benchmark, reflecting modest but persistent diversification gains. In the post-2020 window, the aggregate forecast captures the broad inflation run-up and subsequent decline materially better than SARIMA alone, while also smoothing through some of the standalone macro forecast's volatility. Appendix Figures C.16, C.17, and C.18 report the corresponding diagnostics for $h = 1$, $h = 6$, and $h = 24$.

At $h = 1$, the adaptive combination remains heavily SARIMA-centered, especially before 2020. The average SARIMA weight is 0.746 in the pre-2020 window and 0.601 in the post-2020 window (see Appendix Figure C.16), and the pre-2020 gains are slightly negative. This is unsurprising: at very short horizons in stable times, the univariate benchmark remains difficult to beat.

At $h = 6$, the combination becomes more balanced, though it still stays close to an equal split in the post-2020 window, with average weights of 0.480 on Macro XGBoost and 0.520 on SARIMA (see Appendix Figure C.17). The corresponding post-2020 gains are modest, at 2.53% in mean absolute error and 3.56% in root mean squared error, which is weaker than the three-expert combination at the same horizon.

At $h = 12$, the macro model takes the lead in the post-2020 window: its average weight rises to 0.580, while the SARIMA weight falls to 0.420. The resulting gains are 10.76% in mean absolute error and 9.38% in root mean squared error. These improvements are close to the standalone micro forecast's post-2020 performance at the same horizon, but they do not dominate it, especially under root mean squared error.

At $h = 24$, long-horizon performance is driven by a sharp reallocation toward Macro XGBoost while still retaining substantial SARIMA weight. In the post-2020 window, the average weights are 0.531 on Macro XGBoost and 0.469 on SARIMA, and the average selected learning rate reaches 1.000 (see Appendix Figure C.18). The resulting gains rise to 20.04% in mean absolute error and 52.04% in root mean squared error. These are large improvements, but they remain slightly below the correspond-

Figure 6: SARIMA + macro online-learning diagnostics: $h = 12$



Notes: The top panel plots realized inflation and the online-learning aggregate forecast for the pre-2020 (2015–2019) and post-2020 (2020–2024) windows. The bottom panel plots the time-varying weights on Macro XGBoost and SARIMA. The corresponding diagnostics for $h = 1$, $h = 6$, and $h = 24$, including the selected learning rate η_t path, are reported in Appendix C.

ing gains from the full three-expert combination, again pointing to incremental value from keeping microdata in the expert pool.

Dependence structure and bootstrap regime. To compare the two-expert SARIMA + macro benchmark directly with the micro forecast, we examine the loss-differential panel constructed from

$$r_{t,h} = |\pi_t - \hat{\pi}_t^{h,\text{SARIMA+Macro}}| - |\pi_t - \hat{\pi}_t^{h,\text{Micro}}|, \qquad h = 1, \ldots, 24.$$

For serial dependence, the multiplier-bootstrap Ljung–Box test aggregated across the 24 horizons yields $Q = 762.79$ with $m = 24$, $T = 239$, and $B = 10{,}000$, corresponding to a bootstrap p-value of 0.1126. We therefore do not reject the null of no serial correlation at conventional significance levels. As in Section 4.3, this points to Regime I as the appropriate baseline specification for the scan test.

Scan test results and robustness. Under this baseline Regime I specification, the scan test rejects in both directions, implying that neither forecast uniformly dominates the other over the full sample. For the comparison

$$D_t^{(h)} = |\epsilon_{t,\text{SARIMA+Macro}}^{(h)}| - |\epsilon_{t,\text{Micro}}^{(h)}|,$$

the maximum scan statistic is $T^{\max} = 9.322$ at horizon $h = 22$ over a 7-month window running from March 2020 to September 2020. The associated critical value is 7.492 and the bootstrap p-value is 0.0090, so we reject the null that the micro forecast never outperforms the SARIMA + macro benchmark. Economically, this result indicates that once the inflation regime shift began in 2020, microdata contained information that was not fully spanned by the benchmark's combination of macro covariates and univariate inflation dynamics, especially at long horizons.

The reverse comparison also rejects. When the loss differential is constructed as

$$D_t^{(h)} = |\epsilon_{t,\text{Micro}}^{(h)}| - |\epsilon_{t,\text{SARIMA+Macro}}^{(h)}|,$$

the maximum scan statistic is $T^{\max} = 8.440$ at horizon $h = 24$ over a 7-month window from July 2016 to January 2017, with critical value 7.273 and bootstrap p-value 0.0200. We therefore also reject the null that the SARIMA + macro benchmark never outperforms the micro forecast. Taken together, the two directional rejections deliver a clear robustness message: relative forecast performance between microdata and the SARIMA + macro online-learning benchmark is state dependent. The micro forecast dominates during the early 2020 inflation break, while the SARIMA + macro benchmark dominates during a localized pre-2020 episode. This pattern is consistent with the broader theme of the paper that no single modality dominates uniformly in a non-stationary environment, which in turn motivates the full adaptive combination.

Overall, the SARIMA + macro online-learning benchmark is powerful: adaptive combination of macro and univariate information alone delivers broad-based post-2020 gains and, at some horizons,

comes close to the forecasting gains from microdata. But it does not fully subsume the role of microdata. The full three-expert combination remains the best-performing specification on average, which suggests that microdata contribute information that is not completely spanned by macro covariates and univariate dynamics.

4.6 Three-Expert Online Learning versus the SARIMA + Macro Benchmark

The previous subsection established that a two-expert online-learning benchmark combining Macro XGBoost and SARIMA is already strong. We now ask whether adding microdata to this benchmark through the full three-expert online-learning algorithm yields further gains. This comparison is useful for two reasons. First, it isolates the incremental value of microdata once the benchmark is already allowed to adapt optimally between macro and univariate information. Second, it sharpens the interpretation of the three-expert procedure as a robust adaptive algorithm. The scan-test results above showed that microdata are useful only in some episodes and horizons, rather than uniformly over the sample. In that setting, the value of the three-expert rule is precisely that it can choose when to rely on microdata and when to fall back on the SARIMA + macro benchmark. The question in this subsection is therefore whether that additional adaptive flexibility translates into statistically meaningful forecast improvements.

Dependence structure and bootstrap regime. To compare the full three-expert online-learning forecast directly with the two-expert SARIMA + macro benchmark, we examine the loss-differential panel constructed from

$$r_{t,h} = |\pi_t - \hat{\pi}_t^{h,\text{SARIMA+Macro+Micro}}| - |\pi_t - \hat{\pi}_t^{h,\text{SARIMA+Macro}}|, \qquad h = 1,\dots,24.$$

For serial dependence, the multiplier-bootstrap Ljung–Box test aggregated across the 24 horizons yields $Q = 701.68$ with $m = 24$, $T = 239$, and $B = 5{,}000$, corresponding to a bootstrap p-value of 0.7732. We therefore do not reject the null of no serial correlation at conventional significance levels. As in the previous scan-test exercises, this points to Regime I as the appropriate baseline specification for the scan test.

Scan test results and interpretation. Under this baseline Regime I specification, the directional scan tests deliver an asymmetric result. For the comparison

$$D_t^{(h)} = |\epsilon_{t,\text{SARIMA+Macro+Micro}}^{(h)}| - |\epsilon_{t,\text{SARIMA+Macro}}^{(h)}|,$$

the maximum scan statistic is $T^{\max} = 6.420$ at horizon $h = 7$ over a 7-month window running from August 2017 to February 2018. The associated critical value is 7.150 and the bootstrap p-value is 0.1072, so we do not reject the null that the SARIMA + macro benchmark never outperforms the full three-

expert combination. In other words, once microdata are added to the adaptive combination, we do not find evidence that the benchmark excluding microdata significantly beats it over any localized episode.

The reverse comparison yields stronger evidence. When the loss differential is constructed as

$$D_t^{(h)} = |\epsilon_{t,\text{SARIMA+Macro}}^{(h)}| - |\epsilon_{t,\text{SARIMA+Macro+Micro}}^{(h)}|,$$

the maximum scan statistic is $T^{\max} = 8.397$ at horizon $h = 11$ over a 7-month window from April 2014 to October 2014, with critical value 7.188 and bootstrap p-value 0.0140. We therefore reject the null that the full three-expert combination never outperforms the SARIMA + macro benchmark. This result implies that adding microdata to the adaptive expert pool yields a statistically significant forecasting gain over a localized episode at a medium horizon.

Taken together, these results support the view that microdata adds value inside the online-learning procedure. We find a statistically significant episode in which the three-expert rule outperforms the two-expert benchmark, but no statistically significant episode in which the two-expert benchmark outperforms the three-expert rule. This is exactly the pattern one would expect if microdata are occasionally useful and the online-learning algorithm is successfully deciding when to use them. The three-expert procedure is therefore valuable not simply because it contains one additional expert, but because it preserves robustness while allowing the algorithm to exploit microdata in the periods and horizons where they add predictive content beyond macro covariates and univariate inflation dynamics.

5 Implications for Models of Inflation

Our empirical results, that microdata improve inflation forecasts only during the high and volatile inflation after 2020, connect to benchmark models of price setting. We briefly develop this connection here.

In standard menu cost models, when aggregate shocks are small, a first-order approximation to the aggregate price level is valid (Auclert et al., 2024). Under this approximation, aggregate inflation admits a vector moving average (VMA) representation in the history of aggregate shocks:

$$\pi_t = \mu_\pi + \sum_{l=0}^{\infty} \Psi_l \epsilon_{t-l},$$

where $\mu_\pi \in \mathbb{R}$ is a scalar constant denoting steady state inflation, ϵ_t is a vector of aggregate shocks realized at date t that are unpredictable given period-$t-1$ information, $\mathbb{E}_{t-1}[\epsilon_t] = 0$, and $\{\Psi_l\}$ are impulse response coefficients that depend on the structural parameters of the model. The key implication for forecasting is that π_t is spanned by the history of aggregate shocks $\{\epsilon_{t-l}\}_{l\geq 0}$. If the macroeconomic

time series used in our macro forecast are sufficiently rich to span this shock history, that is, if past values of the macro covariates are invertible for $\{\epsilon_{t-l}\}_{l\geq 1}$, then conditional on the macro information set, microdata carry no additional predictive content for future inflation. Intuitively, when shocks are small, the mapping from aggregate shocks to aggregate inflation is linear, so all the information relevant for forecasting π_{t+h} is already contained in aggregate variables.

This conclusion, however, relies on the first-order approximation being accurate. When aggregate shocks are large, the first-order approximation breaks down (Ascari and Sbordone, 2014; Blanco et al., 2024). In this regime, the mapping from shocks to aggregate inflation becomes nonlinear, depending on features of the micro-level price distribution such as the cross-sectional distribution of price gaps. Microdata, which capture such features, may then contain forecasting information that is not spanned by aggregate variables alone.

This theoretical prediction is consistent with our empirical findings. During the stable period of 2015–2019, the micro forecast does not improve upon the SARIMA benchmark at any horizon, consistent with the first-order approximation being adequate and aggregate variables being sufficient for forecasting. After 2020, when the UK experienced large inflationary shocks (see Figure 1), the micro forecast outperforms SARIMA at the majority of horizons, with gains concentrated at medium and long horizons. The scan test rejects the null that the micro forecast never outperforms SARIMA, localizing the strongest outperformance to February–December 2022, the period of sharpest inflation movements. The Fixed Share algorithm also shifts weight toward the micro expert after 2020, particularly at medium horizons. These patterns are consistent with models in which microdata become more informative for forecasting when aggregate shocks are large.

6 Conclusion

This paper develops an adaptive machine learning framework to forecast inflation using high-dimensional microdata on prices. Applied to UK CPI micro-price quotes, our method proceeds in three steps. First, we encode the cross-sectional distribution of price changes into distributional statistics and use them as input for gradient-boosted trees to forecast inflation. Second, we introduce a scan test, supported by a full-fledged econometric theory, for whether one forecast outperforms another over some contiguous interval of unknown location and duration. The scan test reveals that the micro forecast significantly outperforms both a univariate SARIMA benchmark and a macro XGBoost benchmark, with the strongest gains during the February–December 2022 episode. Third, given this evidence that microdata improve forecasts in certain episodes, we combine micro, macro, and univariate forecasts using the Fixed Share algorithm, an online learning method with regret guarantees in non-stationary settings. The combined forecast improves on benchmarks that exclude microdata, and the expert

weights show that the importance of micro-price information rises after 2020.

Our findings connect to models of inflation dynamics. The finding that microdata improve forecasts only during high and volatile inflation is consistent with benchmark models of price setting, in which microdata carry additional predictive content when shocks are large. Several directions for future work emerge naturally. The same pipeline of distributional encoding, scan testing, and adaptive combination could be applied to forecast other macroeconomic aggregates, such as GDP growth or employment, using firm-level or household-level microdata. Extending the analysis to other countries would help assess whether the forecasting value of microdata during high and volatile inflation is a general feature of inflation dynamics.

Tables

Table 1: COICOP1 cell counts (observations per cell)

Treatment	p5	p10	p25	p50	p75	p90	p95
Panel A: All price changes							
Drops ONS sales	265	1,201	2,566	6,170	8,840	15,752	19,039
Includes sales	299	1,219	2,648	6,673	10,634	17,588	20,637
Panel B: Non-zero price changes only							
Drops ONS sales	26	58	222	559	1,036	2,273	3,303
Includes sales	49	94	271	908	2,063	3,789	4,831

Notes: This table reports percentiles of the distribution of cell counts across all (date, COICOP1 code) cells, separately for all price observations (Panel A) and for non-zero price changes only (Panel B). NS-filtered datasets have cell counts equal to the "includes sales" figures, since observations are adjusted rather than dropped.

Table 2: SARIMA MAE and RMSE across pre-2020 and post-2020 windows

h	Pre MAE	Pre RMSE	Post MAE	Post RMSE
1	0.1065	0.1367	0.3315	0.4477
2	0.1083	0.1408	0.3236	0.4420
3	0.1147	0.1491	0.3359	0.4586
4	0.1157	0.1505	0.3771	0.4995
5	0.1201	0.1525	0.3432	0.4629
6	0.1250	0.1579	0.3471	0.4732
7	0.1227	0.1565	0.3992	0.5439
8	0.1251	0.1624	0.4059	0.5580
9	0.1260	0.1654	0.4036	0.5674
10	0.1248	0.1617	0.3812	0.5165
11	0.1242	0.1613	0.4108	0.5685
12	0.1245	0.1624	0.4093	0.5695
13	0.1249	0.1593	0.4809	0.6984
14	0.1250	0.1585	0.4581	0.6246
15	0.1244	0.1562	0.5002	0.7977
16	0.1246	0.1582	0.4287	0.5860
17	0.1237	0.1573	0.4941	0.8132
18	0.1241	0.1576	0.4618	0.6599
19	0.1241	0.1570	0.4385	0.5981
20	0.1231	0.1558	0.4081	0.5482
21	0.1224	0.1542	0.4692	0.6385
22	0.1218	0.1529	0.4074	0.5497
23	0.1200	0.1512	0.4015	0.5554
24	0.1241	0.1554	0.5232	1.2738

Notes: This table reports the mean absolute error (MAE) and root mean squared error (RMSE) of the rolling-window SARIMA forecast for horizons $h = 1,\ldots,24$, across the pre-2020 (2015–2019) and post-2020 (2020–2024) evaluation windows. The SARIMA model is estimated with nested hyperparameter selection on a 120-month rolling window, as described in Section 3.

Table 3: XGB micro percent improvement relative to SARIMA across the pre- and post-COVID evaluation windows

h	Pre-COVID MAE	Pre-COVID RMSE	Post-COVID MAE	Post-COVID RMSE
1	-42.05	-39.87	-3.38	-7.53
2	-20.39	-23.36	-6.42	-9.68
3	-27.09	-28.69	-18.52	-23.06
4	-25.50	-22.44	-2.72	-3.67
5	-32.50	-26.27	-14.86	-10.69
6	-19.32	-16.23	6.68	4.46
7	-21.21	-17.02	4.80	4.83
8	-19.36	-14.55	5.13	2.26
9	-15.24	-15.04	5.35	6.05
10	-14.62	-12.15	12.55	9.09
11	-16.77	-13.35	13.31	14.14
12	-13.28	-6.72	11.75	12.74
13	-31.03	-27.78	21.96	25.06
14	-30.19	-31.59	23.45	22.51
15	-28.49	-31.12	17.31	29.51
16	-42.52	-37.80	6.38	9.73
17	-32.39	-26.79	16.90	33.51
18	-26.17	-23.10	6.19	9.93
19	-30.29	-36.15	-2.73	3.99
20	-25.27	-28.78	0.44	0.91
21	-31.38	-33.59	12.05	15.09
22	-26.20	-31.01	9.45	10.23
23	-37.51	-43.64	4.63	8.83
24	-42.23	-55.99	23.26	58.27

Table 4: XGB macro percent improvement relative to SARIMA across the pre- and post-COVID evaluation windows

h	Pre-COVID MAE	Pre-COVID RMSE	Post-COVID MAE	Post-COVID RMSE
1	-35.57	-48.70	10.10	10.98
2	-21.95	-33.89	-3.36	-10.27
3	-12.19	-27.84	4.41	-5.02
4	-12.97	-26.55	7.46	-0.83
5	-8.99	-14.73	-5.69	-12.39
6	-11.78	-22.78	-2.59	-9.28
7	-9.72	-15.23	8.40	3.39
8	-16.49	-14.05	5.88	2.30
9	-6.81	-8.24	11.71	5.16
10	-12.68	-11.18	9.43	2.37
11	5.40	7.32	9.20	1.79
12	3.20	6.62	8.83	5.74
13	-9.32	-8.25	21.11	22.57
14	-0.11	-0.17	19.45	16.82
15	-8.45	-7.40	26.58	38.59
16	-15.73	-12.52	18.31	18.61
17	-3.30	-4.45	30.89	41.87
18	-18.14	-16.55	28.42	30.39
19	-8.06	-7.61	23.99	22.05
20	-25.82	-26.46	23.12	17.00
21	-24.18	-27.25	29.14	27.17
22	-14.92	-23.30	18.13	16.62
23	-21.17	-32.12	8.84	6.58
24	-29.04	-42.50	29.57	60.02

Table 5: Fixed-share MAE and RMSE improvement vs SARIMA ($\eta = 0.1, \alpha = 0.02, W = 60$)

	2015-2019		2020-2024	
h	MAE Improvement	RMSE Improvement	MAE Improvement	RMSE Improvement
1	-10.83	-14.00	6.56	7.63
2	-1.74	-5.99	2.78	-1.61
3	-1.95	-1.52	2.69	-0.42
4	-0.87	-3.65	11.88	9.28
5	0.77	-0.58	4.12	3.15
6	2.80	-0.21	6.61	6.80
7	5.80	2.84	11.69	9.44
8	4.16	4.04	11.72	9.83
9	10.74	8.31	16.05	12.99
10	10.77	6.43	17.84	13.84
11	10.29	9.30	13.94	13.28
12	7.90	10.22	16.23	15.68
13	-1.53	-0.76	20.10	24.97
14	-0.59	1.82	19.25	17.92
15	-3.22	-2.39	21.89	30.76
16	-10.05	-5.08	13.69	15.11
17	-2.25	0.17	22.69	35.09
18	0.38	-1.71	18.64	21.46
19	3.24	-0.63	15.10	15.12
20	-0.80	-6.40	14.64	10.65
21	-3.77	-9.85	20.52	21.36
22	2.25	-5.04	15.17	12.57
23	-6.94	-9.79	10.16	8.63
24	-9.71	-18.19	19.00	49.51

Table 6: Neighbor-only MAE and RMSE improvement vs SARIMA ($\alpha = 0.02, \eta_0 = 0.1, W = 60$)

	2015-2019		2020-2024	
h	MAE Improvement	RMSE Improvement	MAE Improvement	RMSE Improvement
1	-4.70	-7.40	7.17	8.30
2	-2.00	-4.20	2.42	-1.75
3	-0.39	-1.22	4.15	-0.07
4	-2.14	-3.46	12.30	9.63
5	2.14	0.61	3.86	3.11
6	2.61	0.25	4.39	5.65
7	3.42	2.63	11.80	9.48
8	2.89	2.72	12.14	10.09
9	8.51	6.91	15.71	12.79
10	8.73	4.99	18.34	14.17
11	9.10	8.45	11.33	10.51
12	7.18	7.98	14.46	14.30
13	0.16	1.55	18.20	23.15
14	1.54	3.50	18.97	18.02
15	-0.86	0.50	21.41	31.70
16	-4.08	-0.33	15.22	15.88
17	1.33	2.43	29.44	40.44
18	2.78	1.30	22.62	25.12
19	4.46	0.06	20.25	18.94
20	1.33	-1.51	16.98	11.78
21	-1.21	-3.00	21.10	21.56
22	2.72	-1.99	15.15	12.93
23	-3.47	-4.76	9.32	7.19
24	-4.66	-9.01	20.90	53.94

Table 7: SARIMA + macro fixed-η MAE and RMSE improvement vs SARIMA ($\alpha = 0.02, \eta = 0.1, W = 60$)

	2015-2019		2020-2024	
h	MAE Improvement	RMSE Improvement	MAE Improvement	RMSE Improvement
1	-5.90	-14.33	9.38	10.73
2	-2.94	-7.77	3.85	-0.24
3	0.71	-1.65	8.50	3.28
4	-1.37	-4.46	10.48	8.05
5	2.33	1.02	1.30	1.97
6	2.96	-2.15	3.64	4.01
7	3.97	1.15	11.79	8.47
8	3.08	2.46	10.32	9.79
9	6.70	6.11	15.93	13.04
10	7.22	5.73	18.47	13.34
11	12.79	12.19	13.30	11.39
12	10.63	11.89	13.64	13.69
13	1.92	3.42	15.54	19.79
14	3.23	5.78	15.13	12.91
15	-0.65	2.54	20.92	28.14
16	-2.57	1.10	15.26	14.58
17	5.12	4.59	22.29	33.04
18	0.16	-0.82	18.13	21.61
19	6.64	5.60	17.18	16.30
20	-4.36	-5.04	17.89	12.46
21	-5.18	-7.09	22.30	21.86
22	1.24	-3.99	15.00	12.40
23	-6.33	-7.45	8.19	5.85
24	-7.07	-12.80	15.97	41.86

Table 8: SARIMA + macro neighbor-only MAE and RMSE improvement vs SARIMA ($\alpha = 0.02, \eta_0 = 0.1, W = 60$)

	2015-2019		2020-2024	
h	MAE Improvement	RMSE Improvement	MAE Improvement	RMSE Improvement
1	-1.21	-6.07	8.32	9.59
2	-1.87	-4.21	2.29	-1.25
3	0.58	-1.84	8.28	3.24
4	-1.60	-2.40	10.52	8.29
5	1.82	0.95	0.70	1.65
6	0.64	-2.94	2.53	3.56
7	2.81	1.75	11.84	8.51
8	2.94	2.76	9.81	9.57
9	5.87	5.41	14.99	11.90
10	7.71	5.06	18.62	13.46
11	10.45	10.01	10.36	7.84
12	9.57	9.53	10.76	9.38
13	1.21	3.24	14.20	17.93
14	2.45	4.64	15.55	13.41
15	0.32	2.58	20.18	30.75
16	-1.23	1.89	15.11	14.76
17	3.69	3.80	29.05	39.84
18	1.55	1.39	22.32	24.79
19	6.84	4.93	20.39	18.40
20	0.51	-0.60	19.93	13.90
21	-1.44	-3.76	24.24	23.77
22	2.50	-0.21	14.19	12.74
23	-3.51	-3.92	7.84	5.60
24	-3.44	-6.98	20.04	52.04

References

Adam, Klaus, and Henning Weber. 2023. "Estimating the optimal inflation target from trends in relative prices." *American Economic Journal: Macroeconomics*, 15(3): 1–42.

Andrews, Donald W. K. 1991. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59(3): 817–858.

Ang, Andrew, Geert Bekaert, and Min Wei. 2007. "Do macro variables, asset markets, or surveys forecast inflation better?" *Journal of Monetary Economics*, 54(4): 1163–1212.

Ascari, Guido, and Argia M. Sbordone. 2014. "The macroeconomics of trend inflation." *Journal of Economic Literature*, 52(3): 679–739.

Atkeson, Andrew, and Lee E. Ohanian. 2001. "Are Phillips curves useful for forecasting inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1): 2–11.

Auclert, Adrien, Rodolfo Rigato, Matthew Rognlie, and Ludwig Straub. 2024. "New pricing models, same old Phillips curves?" *The Quarterly Journal of Economics*, 139(1): 121–186.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin. 2010. "Large Bayesian vector auto regressions." *Journal of Applied Econometrics*, 25(1): 71–92.

Bates, J. M., and C. W. J. Granger. 1969. "The combination of forecasts." *Operational Research Quarterly*, 20(4): 451–468.

Beck, Günter W, Kai Carstensen, Jan-Oliver Menz, Richard Schnorrenberger, and Elisabeth Wieland. 2024. "Nowcasting consumer price inflation using high-frequency scanner data: Evidence from Germany."

Blanco, Andres, Corina Boar, Callum Jones, and Virgiliu Midrigan. 2024. *Non-linear inflation dynamics in menu cost economies.* National Bureau of Economic Research.

Brandão-Marques, Luis, Gaston Gelos, David Hofman, Julia Otten, Gurnain Kaur Pasricha, and Zoe Strauss. 2025. "Do household expectations help predict inflation?" *SSRN Working Paper*.

Cesa-Bianchi, Nicolò, and Gábor Lugosi. 2006. *Prediction, Learning, and Games.* Cambridge University Press.

Chang, Jinyuan, Xiaohui Chen, and Mingcong Wu. 2024. "Central limit theorems for high dimensional dependent data." *Bernoulli*, 30(1): 712–742.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide. 2024. "Heterogeneity and aggregate fluctuations." *Journal of Political Economy*, 132(12).

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A scalable tree boosting system." 785–794.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2013. "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors." *Annals of Statistics*, 41(6): 2786–2819.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2015. "Comparison and anti-concentration bounds for maxima of Gaussian random vectors." *Probability Theory and Related Fields*, 162(1): 47–70.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2017. "Central limit theorems and bootstrap in high dimensions." *Annals of Probability*, 45(4): 2309–2352.

Chi, Ta-Chung, Ting-Han Fan, Raffaele M. Ghigliazza, Domenico Giannone, and Zixuan (Kevin) Wang. 2025. "Macroeconomic forecasting and machine learning." *arXiv preprint arXiv:2510.11008*.

de Rooij, Steven, Tim van Erven, Peter D. Grünwald, and Wouter M. Koolen. 2014. "Follow the leader if you can, hedge if you must." *Journal of Machine Learning Research*, 15: 1281–1316.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing predictive accuracy." *Journal of Business & Economic Statistics*, 13(3): 253–263.

Eurostat. 2015. *ESS Guidelines on Seasonal Adjustment*. Luxembourg:Publications Office of the European Union.

Faust, Jon, and Jonathan H. Wright. 2013. "Forecasting inflation." In *Handbook of Economic Forecasting*. Vol. 2A, 2–56. Elsevier.

Fouliard, Jeremy, Michael Howell, Hélène Rey, and Vania Stavrakeva. 2023. "Answering the Queen: Machine learning and financial crises." *NBER Working Paper*, , (28302).

Giacomini, Raffaella, and Barbara Rossi. 2010. "Forecast comparisons in unstable environments." *Journal of Applied Econometrics*, 25(4): 595–620.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2021. "Economic predictions with big data: The illusion of sparsity." *Econometrica*, 89(5): 2409–2437.

Goulet Coulombe, Philippe, Massimiliano Marcellino, and Dalibor Stevanović. 2021. "Can machine learning catch the COVID-19 recession?" *National Institute Economic Review*, 256: 71–109.

Grinsztajn, Léo, Edouard Oyallon, and Gaël Varoquaux. 2022. "Why do tree-based models still outperform deep learning on typical tabular data?" *Advances in Neural Information Processing Systems*, 35: 507–520.

Herbster, Mark, and Manfred K. Warmuth. 1998. "Tracking the best expert." *Machine Learning*, 32(2): 151–178.

Karadi, Peter, and Adam Reiff. 2019. "Menu costs, aggregate fluctuations, and large shocks." *American Economic Journal: Macroeconomics*, 11(3): 111–146.

Klenow, Peter J, and Oleksiy Kryvtsov. 2008. "State-dependent or time-dependent pricing: Does it matter for recent US inflation?" *The Quarterly Journal of Economics*, 123(3): 863–904.

Koop, Gary, and Dimitris Korobilis. 2012. "Forecasting inflation using dynamic model averaging." *International Economic Review*, 53(3): 867–886.

Lenza, Michele, and Ettore Savoia. 2024. "Do we need firm data to understand macroeconomic dynamics?" *Sveriges Riksbank Working Paper Series*, , (438).

McLeay, Michael, and Silvana Tenreyro. 2020. "Optimal inflation and the identification of the Phillips curve." *NBER Macroeconomics Annual*, 34: 199–255.

Medeiros, Marcelo C., Gabriel F. R. Vasconcelos, Álvaro Veiga, and Eduardo Zilberman. 2021. "Forecasting inflation in a data-rich environment: the benefits of machine learning methods." *Journal of Business & Economic Statistics*, 39(1): 98–119.

Meeks, Roland, and Francesca Monti. 2023. "Heterogeneous beliefs and the Phillips curve." *Journal of Monetary Economics*, 139: 41–54.

Merlevède, Florence, Magda Peligrad, and Emmanuel Rio. 2011. "A Bernstein type inequality and moderate deviations for weakly dependent sequences." *Probability Theory and Related Fields*, 151(3): 435–474.

Mullainathan, Sendhil, and Jens Ludwig. 2024. "Machine Learning as a Tool for Hypothesis Generation." *Quarterly Journal of Economics*, 139(2): 751–827.

Nakamura, Emi, and Jón Steinsson. 2008. "Five facts about prices: A reevaluation of menu cost models." *The Quarterly Journal of Economics*, 123(4): 1415–1464.

Newey, Whitney K, and Kenneth D West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica: Journal of the Econometric Society*, 703–708.

Shalev-Shwartz, Shai. 2025. "Online learning and online convex optimization." *Foundations and Trends® in Machine Learning*, 4(2): 107–194.

Stock, James H., and Mark W. Watson. 2002. "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97(460): 1167–1179.

Stock, James H., and Mark W. Watson. 2007. "Why has U.S. inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39(s1): 3–33.

Timmermann, Allan. 2006. "Forecast combinations." In *Handbook of Economic Forecasting*. Vol. 1, , ed. Graham Elliott, Clive W. J. Granger and Allan Timmermann, 135–196. Elsevier.

Vershynin, Roman. 2018. *High-dimensional probability: An introduction with applications in data science*. Vol. 47, Cambridge university press.

A Econometric Theory of Scan Tests

A.1 Setup

A.1.1 Notation

Fix a collection of intervals $\mathscr{I}$ on $[n] := \{1, \dots, n\}$ and define

$$L_{\min} := \min_{I \in \mathscr{I}} |I|, \quad p := H|\mathscr{I}| \tag{A.1}$$

For notational convenience, define

$$A := \mathscr{I} \times [H]. \tag{A.2}$$

Quantities used in the procedure. We start by redefining all useful quantities based on A for completeness. For $a = (I, h) \in A$, write

$$S_a := \sum_{t \in I} D_t^{(h)}, \qquad \bar{D}_a := \frac{1}{|I|} \sum_{t \in I} D_t^{(h)}, \qquad \widetilde{Q}_a := \sum_{t \in I} \big(D_t^{(h)} - \bar{D}_a\big)^2.$$

The actual scan statistic is

$$T_{\max} := \max_{a \in A} \frac{S_a}{\sqrt{\widetilde{Q}_a}}$$

Write

$$\mu_t^{(h)} := \mathbb{E}[D_t^{(h)}], \qquad d_t^{(h)} := D_t^{(h)} - \mu_t^{(h)}, \qquad S_a^\circ := \sum_{t \in I} d_t^{(h)}.$$

Thus $S_a = S_a^\circ + M_a$ with $M_a := \sum_{t \in I} \mu_t^{(h)}$. Define the centered self-normalized comparison statistic

$$T_{\max}^\circ := \max_{a \in A} \frac{S_a^\circ}{\sqrt{\widetilde{Q}_a}}.$$

The Gaussian multiplier bootstrap is defined by

$$R_{t,a} := \mathbf{1}\{t \in I\}\big(D_t^{(h)} - \bar{D}_a\big), \qquad S_a^* := \sum_{t=1}^{n} \xi_t R_{t,a},$$

where $\xi_1, \dots, \xi_n \overset{\text{iid}}{\sim} N(0,1)$ are independent of the data. Set

$$T_{\max}^* := \max_{a \in A} \frac{S_a^*}{\sqrt{\widetilde{Q}_a}}$$

Let $c_{n,1-\alpha}^*$ denote the conditional $(1-\alpha)$-quantile of $T_{\max}^*$.

Additional quantities used in the assumptions and proof. For $a = (I, h)$, let

$$\bar{d}_a := \frac{1}{|I|} \sum_{t \in I} d_t^{(h)}, \qquad \widetilde{Q}_a^d := \sum_{t \in I} \big(d_t^{(h)} - \bar{d}_a\big)^2.$$

Also define

$$U_W^\circ := \max_{a\in A} \frac{S_a^\circ}{\sqrt{W_a}}, \qquad \bar{U}_W^\circ := \max_{a\in A} \frac{|S_a^\circ|}{\sqrt{W_a}}.$$

Define the centered oracle denominator

$$W_a := \sum_{t\in I} \mathbb{E}\big[d_t^{(h)2}\big].$$

This is the natural scale for the centered sums S_a°. We also define the interval-wise mean-flatness index

$$r_n := \max_{a=(I,h)\in A} \frac{1}{W_a} \sum_{t\in I} \big(\mu_t^{(h)} - \bar{\mu}_a\big)^2, \qquad \bar{\mu}_a := \frac{1}{|I|} \sum_{t\in I} \mu_t^{(h)}. \tag{A.3}$$

Observe that $r_n = 0$ whenever the mean is constant on each scanned interval, so arbitrary negative constant shifts are allowed.

Finally, define the deterministic Gaussian target $G_{Q,\circ} \sim N(0, \Sigma_{Q,\circ})$ in $\mathbb{R}^p$ by

$$(\Sigma_{Q,\circ})_{ab} := \frac{\sum_{t\in I_a\cap I_b} \mathbb{E}\big[d_t^{(h_a)} d_t^{(h_b)}\big]}{\sqrt{W_a W_b}}, \qquad a, b \in A.$$

Because $W_a = \mathrm{Var}(S_a^\circ)$ under independence across t, every diagonal entry satisfies $(\Sigma_{Q,\circ})_{aa} = 1$. Write

$$G_{Q,\circ,\max} := \max_{a\in A} G_{Q,\circ,a}, \qquad q_{Q,\circ,n,\beta} := \inf\{x \in \mathbb{R} : \ \mathbb{P}(G_{Q,\circ,\max} \le x) \ge \beta\}.$$

A.1.2 Common assumptions

We impose the following assumptions for both Regimes I and D.

Assumption A.1. *The centered coordinates are uniformly sub-Gaussian:*

$$\sup_{t\in[n],\, h\in[H]} \|d_t^{(h)}\|_{\psi_2} \le B_0$$

for some constant $B_0 < \infty$, where $\|\cdot\|_{\psi_2}$ denotes the Orclitz norm; see

Assumption A.2. *The centered variances are uniformly bounded away from 0 and ∞:*

$$0 < \sigma_-^2 \le \mathbb{E}\big[d_t^{(h)2}\big] \le \sigma_+^2 < \infty \qquad \forall\, t \in [n],\ h \in [H],$$

for some constants σ_-^2, σ_+^2.

A.2 Regime I Validity

A.2.1 Main theorem

We consider the following additional assumptions for Regime I.

Assumption A.3. *The vectors $D_1, \ldots, D_n$ are independent across t.*

Assumption A.4. *The interval-wise mean is asymptotically flat:*

$$r_n = o\big((\log p)^{-2}\big).$$

Assumption A.5. *With* $\ell_n := \log(epn)$,

$$\frac{\ell_n^7}{L_{\min}} \to 0.$$

Remark 1. Assumption A.4 constrains the *shape* of the mean on each scanned interval, not its level. If $\mu_t^{(h)} \equiv -c_h$ is constant in t, then $r_n = 0$ for every n regardless of the magnitude of c_h.

Theorem A.1. *Assume Assumptions A.1, A.2, and A.3–A.5. Let* $c^*_{n,1-\alpha}$ *be the conditional* $(1-\alpha)$*-quantile of* $T^*_{\max}$. *Then the Gaussian multiplier bootstrap with* $\Theta = I_n$ *is asymptotically conservative: under* H_0,

$$\limsup_{n\to\infty} \mathbb{P}\big(T_{\max} > c^*_{n,1-\alpha}\big) \le \alpha \qquad \forall \alpha \in (0,1).$$

We will split the proof into a few steps.

A.2.2 Preliminary bounds

Lemma A.2. *Under Assumptions A.1, A.2, and A.3,*

$$\bar{U}^\circ_W = O_p\big(\sqrt{\log p}\big).$$

Moreover, if

$$\mathcal{K} := \{I \cap J : \ I, J \in \mathcal{I},\ I \cap J \neq \varnothing\},$$

and for $K \in \mathcal{K}$, $h \in [H]$ *we set*

$$W_{K,h} := \sum_{t\in K} \mathbb{E}\big[d_t^{(h)2}\big], \qquad S^\circ_{K,h} := \sum_{t\in K} d_t^{(h)},$$

then

$$\max_{K\in\mathcal{K},\ h\in[H]} \frac{|S^\circ_{K,h}|}{\sqrt{W_{K,h}}} = O_p\big(\sqrt{\log p}\big).$$

Proof. Fix $a = (I,h)$. By Assumptions A.1, A.3, and Proposition A.26,

$$\big\|S^\circ_a\big\|_{\psi_2} \le CB_0\sqrt{|I|}$$

for a constant $C < \infty$ depending only on universal constants. Since

$$W_a \ge \sigma_-^2 |I|,$$

we obtain the uniform bound

$$\left\|\frac{S^\circ_a}{\sqrt{W_a}}\right\|_{\psi_2} \le \frac{CB_0}{\sigma_-}.$$

Hence, for some constants $c_1, c_2 > 0$,

$$\mathbb{P}\left(\left|\frac{S_a^\circ}{\sqrt{W_a}}\right| > u\right) \le c_1 e^{-c_2 u^2} \qquad (u \ge 0)$$

uniformly in $a \in A$. A union bound over the p coordinates gives

$$\mathbb{P}\big(\bar{U}_W^\circ > u\big) \le c_1 p e^{-c_2 u^2}.$$

Choosing $u = M\sqrt{\log p}$ with M large enough proves the first claim.

For the second claim, note that $|\mathscr{K}| \le |\mathscr{I}|^2$, hence

$$H|\mathscr{K}| \le H|\mathscr{I}|^2 = \frac{p^2}{H} \le p^2.$$

The same argument with p^2 in place of p yields the same $\sqrt{\log p}$ order. □

Lemma A.3 (Uniform denominator approximation)**.** *Define the rate*

$$\zeta_n := \sqrt{\frac{\log p}{L_{\min}}} + \frac{\log p}{L_{\min}} + \sqrt{\frac{r_n \log p}{L_{\min}}} + r_n. \tag{A.4}$$

Under Assumptions A.1, A.2, and A.3–A.5,

$$\Delta_Q := \max_{a \in A} \left|\frac{\widetilde{Q}_a}{W_a} - 1\right| = O_p(\zeta_n).$$

In particular, $\Delta_Q \xrightarrow{p} 0$ *and* $\mathbb{P}(\Delta_Q > 1/2) \to 0$.

Proof. Fix $a = (I, h)$. Write

$$\delta_{t,a} := \mu_t^{(h)} - \bar{\mu}_a.$$

Since $\sum_{t \in I} \delta_{t,a} = 0$, we have

$$D_t^{(h)} - \bar{D}_a = (d_t^{(h)} - \bar{d}_a) + \delta_{t,a},$$

and therefore

$$\widetilde{Q}_a = \widetilde{Q}_a^d + 2C_a + R_a, \tag{A.5}$$

$$C_a := \sum_{t \in I} (d_t^{(h)} - \bar{d}_a)\delta_{t,a} = \sum_{t \in I} d_t^{(h)} \delta_{t,a}, \tag{A.6}$$

$$R_a := \sum_{t \in I} \delta_{t,a}^2. \tag{A.7}$$

Also,

$$\widetilde{Q}_a^d - W_a = A_a^d - B_a^d,$$

where

$$A_a^d := \sum_{t \in I} \Big(d_t^{(h)2} - \mathbb{E}[d_t^{(h)2}] \Big), \qquad B_a^d := |I| \bar{d}_a^2 = \frac{(S_a^\circ)^2}{|I|}.$$

Combining these identities gives

$$\frac{\widetilde{Q}_a - W_a}{W_a} = \frac{A_a^d}{W_a} - \frac{B_a^d}{W_a} + 2\frac{C_a}{W_a} + \frac{R_a}{W_a}. \tag{A.8}$$

It therefore suffices to control the four terms on the right uniformly over $a \in A$.

Step 1: control of $\frac{A_a^d}{W_a}$. By Assumption A.1 and Proposition A.27,

$$d_t^{(h)2} - \mathbb{E}[d_t^{(h)2}]$$

is uniformly sub-exponential. Hence Bernstein's inequality yields, for every fixed $a = (I,h)$ and every $u \geq 1$,

$$\mathbb{P}\Big(|A_a^d| > C\big(\sqrt{|I|u} + u\big)\Big) \leq 2e^{-u}$$

for some constant $C < \infty$. Taking $u = x + \log p$ and a union bound over the p coordinates gives

$$\max_{a\in A} |A_a^d| = O_p\Big(\max_{a\in A}\{\sqrt{|I_a|\log p} + \log p\}\Big).$$

Since $W_a \geq \sigma_-^2 |I_a| \geq \sigma_-^2 L_{\min}$,

$$\max_{a\in A} \frac{|A_a^d|}{W_a} = O_p\left(\sqrt{\frac{\log p}{L_{\min}}} + \frac{\log p}{L_{\min}}\right).$$

Step 2: control of $\frac{B_a^d}{W_a}$. By Lemma A.2,

$$\max_{a\in A} \frac{B_a^d}{W_a} = \max_{a\in A} \frac{(S_a^\circ)^2}{|I_a| W_a} \leq \frac{(\bar{U}_W^\circ)^2}{L_{\min}} = O_p\left(\frac{\log p}{L_{\min}}\right).$$

Step 3: control of $\frac{C_a}{W_a}$. Conditionally on the deterministic coefficients $\delta_{t,a}$, the random variable C_a in (A.6) is a weighted sum of independent centered sub-Gaussian variables. By Proposition A.26,

$$\|C_a\|_{\psi_2} \leq C B_0 \left(\sum_{t\in I} \delta_{t,a}^2\right)^{1/2} = C B_0 R_a^{1/2}.$$

Therefore, for every fixed a and every $u \geq 0$,

$$\mathbb{P}\big(|C_a| > C R_a^{1/2}\sqrt{u}\big) \leq 2e^{-u}.$$

A union bound over $a \in A$ yields

$$\max_{a\in A} |C_a| = O_p\Big(\max_{a\in A} R_a^{1/2}\sqrt{\log p}\Big).$$

Since $R_a \leq r_n W_a$ and $W_a \geq \sigma_-^2 L_{\min}$,

$$\max_{a\in A} \frac{|C_a|}{W_a} = O_p\left(\sqrt{\frac{r_n \log p}{L_{\min}}}\right).$$

Step 4: control of $\frac{R_a}{W_a}$. By definition (A.3) of r_n,

$$\max_{a\in A}\frac{R_a}{W_a} = r_n.$$

Collecting the four bounds proves

$$\Delta_Q = O_p\left(\sqrt{\frac{\log p}{L_{\min}}} + \frac{\log p}{L_{\min}} + \sqrt{\frac{r_n \log p}{L_{\min}}} + r_n\right) = O_p(\zeta_n),$$

as claimed. □

A.2.3 Gaussian approximation for $T^\circ_{\max}$

Define the centered oracle-normalized vector

$$Y^\circ := \left(Y^\circ_a\right)_{a\in A}, \qquad Y^\circ_a := \frac{S^\circ_a}{\sqrt{W_a}}, \qquad U^\circ_W := \max_{a\in A} Y^\circ_a.$$

By construction, $\mathrm{Cov}(Y^\circ) = \Sigma_{Q,\circ}$.

Lemma A.4. *Under Assumptions A.1, A.2, and A.3–A.5,*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(T^\circ_{\max} \le x\right) - \mathbb{P}\left(G_{Q,\circ,\max} \le x\right)\right| \to 0.$$

Proof. We proceed in three steps.

Step 1. For $t \in [n]$ and $a = (I,h) \in A$, define

$$Z_{t,a} := \sqrt{n}\,\mathbf{1}\{t \in I\}\frac{d^{(h)}_t}{\sqrt{W_a}}. \tag{A.9}$$

Then

$$Y^\circ = \frac{1}{\sqrt{n}}\sum_{t=1}^n Z_t.$$

By Assumptions A.1, A.3, and Proposition A.26,

$$\|Z_{t,a}\|_{\psi_2} \le B_0\sqrt{\frac{n}{\sigma_-^2 L_{\min}}} \lesssim B_n \qquad \forall\, t,a.$$

Finally,

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{t,a}\right) = \mathrm{Var}\left(\frac{S^\circ_a}{\sqrt{W_a}}\right) = 1 \qquad \forall\, a \in A.$$

By Assumption A.5,

$$\frac{B_n^2 \ell_n^7}{n} \lesssim \frac{\ell_n^7}{L_{\min}} \to 0.$$

Together with Assumption A.2, the conditions of Proposition A.17 are satisfied. Thus,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(U^\circ_W \le x\right) - \mathbb{P}\left(G_{Q,\circ,\max} \le x\right)\right| \to 0.$$

Step 2. On the event $\{\Delta_Q \le 1/2\}$, we have

$$\frac{\widetilde{Q}_a}{W_a} \in [1/2, 3/2] \qquad \forall a \in A.$$

For $x \in [1/2, 3/2]$, the map $x \mapsto x^{-1/2}$ is Lipschitz, so

$$\left|x^{-1/2} - 1\right| \le 2|x-1|.$$

Therefore, on $\{\Delta_Q \le 1/2\}$,

$$\begin{aligned}\left|\frac{S_a^\circ}{\sqrt{\widetilde{Q}_a}} - \frac{S_a^\circ}{\sqrt{W_a}}\right| &= \frac{|S_a^\circ|}{\sqrt{W_a}}\left|\left(\frac{\widetilde{Q}_a}{W_a}\right)^{-1/2} - 1\right| \\ &\le 2\Delta_Q \frac{|S_a^\circ|}{\sqrt{W_a}} \le 2\Delta_Q \bar{U}_W^\circ.\end{aligned}$$

Taking maxima over $a \in A$ implies that

$$|T_{\max}^\circ - U_W^\circ| \le 2\Delta_Q \bar{U}_W^\circ \qquad \text{on } \{\Delta_Q \le 1/2\}.$$

By Lemma A.3 and Lemma A.2,

$$2\Delta_Q \bar{U}_W^\circ = O_p\big(\zeta_n \sqrt{\log p}\big).$$

Since Assumptions A.4 and A.5 imply

$$\zeta_n \sqrt{\log p} = \frac{\log p}{\sqrt{L_{\min}}} + \frac{(\log p)^{3/2}}{L_{\min}} + \log p \sqrt{\frac{r_n}{L_{\min}}} + r_n \sqrt{\log p} \to 0,$$

we may choose a deterministic sequence $\eta_n \downarrow 0$ such that

$$\mathbb{P}\big(2\Delta_Q \bar{U}_W^\circ > \eta_n\big) \to 0, \qquad \eta_n \sqrt{1 \vee \log(p/\eta_n)} \to 0.$$

Step 3. Since the diagonal entries of $\Sigma_{Q,\circ}$ equal 1, Proposition A.18 implies

$$\sup_{x \in \mathbb{R}} \mathbb{P}\big(|G_{Q,\circ,\max} - x| \le \eta_n\big) \to 0.$$

Now fix $x \in \mathbb{R}$. On the event $\{\Delta_Q \le 1/2,\ 2\Delta_Q \bar{U}_W^\circ \le \eta_n\}$,

$$T_{\max}^\circ \le x \Longrightarrow U_W^\circ \le x + \eta_n,$$

and similarly

$$U_W^\circ \le x - \eta_n \Longrightarrow T_{\max}^\circ \le x.$$

Hence

$$\begin{aligned}\mathbb{P}\big(U_W^\circ \le x - \eta_n\big) - o(1) &\le \mathbb{P}\big(T_{\max}^\circ \le x\big) \\ &\le \mathbb{P}\big(U_W^\circ \le x + \eta_n\big) + o(1),\end{aligned}$$

uniformly in x. Comparing U_W° to $G_{Q,\circ,\max}$ by Step 1 and then using Proposition A.18 yields

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\big(T_{\max}^\circ \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big)\right| \to 0.$$

The proof is then completed. □

A.2.4 Consistency of bootstrap covariance matrix

Throughout the section we define

$$\|\Sigma\|_{\max} := \max_{a,b}|\Sigma_{a,b}|$$

Lemma A.5. *Under Assumptions A.1, A.2, and A.3–A.5, the Gaussian multiplier bootstrap covariance matrix*

$$\widehat{\Sigma}_{ab} := \mathrm{Cov}^*\left(\frac{S_a^*}{\sqrt{\widetilde{Q}_a}}\frac{S_b^*}{\sqrt{\widetilde{Q}_b}}\right)$$

satisfies

$$\|\widehat{\Sigma} - \Sigma_{Q,\circ}\|_{\max} = O_p(\zeta_n) = o_p\big((\log p)^{-2}\big).$$

Proof. For $a=(I,h)$ and $b=(J,g)$, set $K := I\cap J$ and write

$$\nu_{t,a} := \mathbf{1}\{t\in I\}(d_t^{(h)} - \bar{d}_a), \qquad m_{t,a} := \mathbf{1}\{t\in I\}(\mu_t^{(h)} - \bar{\mu}_a).$$

Then

$$R_{t,a} = \nu_{t,a} + m_{t,a}.$$

By direct calculation,

$$\widehat{\Sigma}_{ab} = \frac{N_{ab}}{\sqrt{\widetilde{Q}_a\widetilde{Q}_b}} \qquad N_{ab} := \sum_{t=1}^n R_{t,a}R_{t,b} = \sum_{t\in K} R_{t,a}R_{t,b}.$$

The target covariance entry is

$$(\Sigma_{Q,\circ})_{ab} = \frac{\Omega_{ab}}{\sqrt{W_aW_b}}, \qquad \Omega_{ab} := \sum_{t\in K}\mathbb{E}\big[d_t^{(h)}d_t^{(g)}\big].$$

Hence

$$\widehat{\Sigma}_{ab} - (\Sigma_{Q,\circ})_{ab} = \frac{N_{ab}-\Omega_{ab}}{\sqrt{\widetilde{Q}_a\widetilde{Q}_b}} \tag{A.10}$$

$$+\,\Omega_{ab}\left(\frac{1}{\sqrt{\widetilde{Q}_a\widetilde{Q}_b}} - \frac{1}{\sqrt{W_aW_b}}\right). \tag{A.11}$$

We bound the two terms uniformly.

Step 1: bounding (A.11). On the event $\{\Delta_Q \le 1/2\}$, every $\widetilde{Q}_a \ge W_a/2 > 0$. Since $|\Omega_{ab}| \le \sqrt{W_aW_b}$ by

Cauchy–Schwarz,

$$\max_{a,b}\left|\Omega_{ab}\left(\frac{1}{\sqrt{\widetilde{Q}_a\widetilde{Q}_b}}-\frac{1}{\sqrt{W_aW_b}}\right)\right|\le C\Delta_Q=O_p(\zeta_n),$$

by Lemma A.3.

Step 2: bounding (A.10). Expand

$$N_{ab}=\sum_{t\in K}v_{t,a}v_{t,b}+\sum_{t\in K}v_{t,a}m_{t,b}+\sum_{t\in K}v_{t,b}m_{t,a}+\sum_{t\in K}m_{t,a}m_{t,b}. \tag{A.12}$$

We analyze the four pieces.

Bounding the first term of (A.12). A direct expansion gives

$$\sum_{t\in K}v_{t,a}v_{t,b}-\Omega_{ab}=M^d_{ab}-\bar{d}_bS^\circ_{K,h}-\bar{d}_aS^\circ_{K,g}+|K|\bar{d}_a\bar{d}_b,$$

where

$$M^d_{ab}:=\sum_{t\in K}\left(d^{(h)}_td^{(g)}_t-\mathbb{E}[d^{(h)}_td^{(g)}_t]\right),$$

and

$$S^\circ_{K,h}:=\sum_{t\in K}d^{(h)}_t,\quad S^\circ_{K,g}:=\sum_{t\in K}d^{(g)}_t.$$

By Assumption A.1, $d^{(h)}_t$ and $d^{(g)}_t$ are both sub-Gaussian. By Proposition A.27, $d^{(h)}_td^{(g)}_t$ is sub-exponential, Bernstein's inequality and a union bound over the p^2 pairs (a,b) yield

$$\max_{a,b}\frac{|M^d_{ab}|}{\sqrt{W_aW_b}}=O_p\left(\sqrt{\frac{\log p}{L_{\min}}}+\frac{\log p}{L_{\min}}\right).$$

Also, by Lemma A.2,

$$\max_b\frac{|\bar{d}_b|}{\sqrt{W_b}}=\max_b\frac{|S^\circ_b|}{|I_b|\sqrt{W_b}}\le\frac{\bar{U}^\circ_W}{L_{\min}}=O_p\left(\frac{\sqrt{\log p}}{L_{\min}}\right),$$

and

$$\max_{a,b}\frac{|S^\circ_{K,h}|}{\sqrt{W_a}}=O_p\big(\sqrt{\log p}\big),\quad\max_{a,b}\frac{|S^\circ_{K,g}|}{\sqrt{W_b}}=O_p\big(\sqrt{\log p}\big),$$

so the two sample-mean terms are both

$$O_p\left(\frac{\log p}{L_{\min}}\right).$$

Finally,

$$\max_{a,b}\frac{|K||\bar{d}_a\bar{d}_b|}{\sqrt{W_aW_b}}\le\max_{a,b}\frac{|K|}{|I_a||I_b|}\frac{|S^\circ_a|}{\sqrt{W_a}}\frac{|S^\circ_b|}{\sqrt{W_b}}\le\frac{(\bar{U}^\circ_W)^2}{L_{\min}}=O_p\left(\frac{\log p}{L_{\min}}\right),$$

where the last line follows from Lemma A.2. Therefore

$$\max_{a,b}\frac{\left|\sum_t v_{t,a}v_{t,b}-\Omega_{ab}\right|}{\sqrt{W_aW_b}}=O_p\left(\sqrt{\frac{\log p}{L_{\min}}}+\frac{\log p}{L_{\min}}\right). \tag{A.13}$$

Bounding the second and third terms of (A.12). Consider $\sum_{t\in K} v_{t,a} m_{t,b}$; the other term is symmetric. Write

$$d_{t,b} := \mathbf{1}\{t \in J\}(\mu_t^{(g)} - \bar{\mu}_b).$$

Then

$$\sum_t v_{t,a} m_{t,b} = \sum_{t\in I} (d_t^{(h_a)} - \bar{d}_a) d_{t,b} = \sum_{t\in I} d_t^{(h_a)} d_{t,b} - \bar{d}_a \sum_{t\in I} d_{t,b}.$$

The first term is a weighted sum of independent centered sub-Gaussian variables with coefficient vector $(d_{t,b})_{t\in I_a}$, hence is sub-Gaussian with scale at most $CB_0(\sum_{t\in I_a} d_{t,b}^2)^{1/2} \le CB_0 R_b^{1/2}$. The definition (A.3) of r_n and a union bound over (a,b) therefore gives

$$\max_{a,b} \frac{\left|\sum_{t\in I_a} d_t^{(h_a)} d_{t,b}\right|}{\sqrt{W_a W_b}} = O_p\left(\sqrt{\frac{r_n \log p}{L_{\min}}}\right).$$

For the second term, Cauchy–Schwarz yields

$$\left|\sum_{t\in I_a} d_{t,b}\right| \le \sqrt{|I_a| R_b}.$$

By definition of $\bar{U}_W^\circ$ and Lemma A.2,

$$\frac{|\bar{d}_a|\left|\sum_{t\in I_a} d_{t,b}\right|}{\sqrt{W_a W_b}} \le \frac{|S_a^\circ|}{\sqrt{W_a}} \frac{1}{\sqrt{|I_a|}} \sqrt{\frac{R_b}{W_b}} \le \bar{U}_W^\circ \sqrt{\frac{r_n}{L_{\min}}} = O_p\left(\sqrt{\frac{r_n \log p}{L_{\min}}}\right).$$

Hence

$$\max_{a,b} \frac{\left|\sum_t v_{t,a} m_{t,b}\right|}{\sqrt{W_a W_b}} = O_p\left(\sqrt{\frac{r_n \log p}{L_{\min}}}\right), \tag{A.14}$$

and the same bound holds for $\sum_t v_{t,b} m_{t,a}$.

Bounding the last term of (A.12). By Cauchy–Schwarz and the definition of r_n,

$$\max_{a,b} \frac{\left|\sum_t m_{t,a} m_{t,b}\right|}{\sqrt{W_a W_b}} \le \max_{a,b} \sqrt{\frac{\sum_t m_{t,a}^2}{W_a}} \sqrt{\frac{\sum_t m_{t,b}^2}{W_b}} \le r_n. \tag{A.15}$$

Combining (A.13), (A.14), and (A.15) yields

$$\max_{a,b} \frac{|N_{ab} - \Omega_{ab}|}{\sqrt{W_a W_b}} = O_p(\zeta_n).$$

On the event $\{\Delta_Q \le 1/2\}$, we have $\widetilde{Q}_a \widetilde{Q}_b \ge W_a W_b/4$, so

$$\max_{a,b} \frac{|N_{ab} - \Omega_{ab}|}{\sqrt{\widetilde{Q}_a \widetilde{Q}_b}} = O_p(\zeta_n).$$

Together with the denominator perturbation bound from Step 1 and the decomposition (A.10)–(A.11),

this proves

$$\|\widehat{\Sigma} - \Sigma_{Q,\circ}\|_{\max} = O_p(\zeta_n).$$

Finally, by Assumptions A.4 and A.5, $\zeta_n(\log p)^2 \to 0$. Thus,

$$\|\widehat{\Sigma} - \Sigma_{Q,\circ}\|_{\max} = o_p\big((\log p)^{-2}\big).$$

□

Lemma A.6. *Under Assumptions A.1, A.2, and A.3–A.5, there exists a deterministic sequence $\rho_n \downarrow 0$ such that*

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*\big(T^*_{\max} \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big)\right| > \rho_n\right) \to 0.$$

Proof. Conditionally on the data, the vector

$$T^* := \left(\frac{S^*_a}{\sqrt{\widetilde{Q}_a}}\right)_{a\in A}$$

is Gaussian, because it is a linear transformation of the Gaussian multiplier vector $(\zeta_1,\dots,\zeta_n)$. Its conditional covariance matrix is exactly $\widehat{\Sigma}$.

By Lemma A.5, there exists a deterministic sequence $r_n \downarrow 0$ such that

$$r_n(\log p)^2 \to 0, \qquad \mathbb{P}\big(\|\widehat{\Sigma} - \Sigma_{Q,\circ}\|_{\max} > r_n\big) \to 0.$$

Let

$$E_n := \big\{\|\widehat{\Sigma} - \Sigma_{Q,\circ}\|_{\max} \le r_n\big\}.$$

On E_n, Proposition A.19 implies

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*\big(T^*_{\max} \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big)\right| \le C\{r_n(\log p)^2\}^{1/3}$$

for a universal constant $C < \infty$. The right-hand side is deterministic and converges to 0. Define

$$\rho_n := C\{r_n(\log p)^2\}^{1/3} \downarrow 0.$$

Then

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*\big(T^*_{\max} \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big)\right| > \rho_n\right)$$
$$\le \mathbb{P}(E_n^c) \to 0.$$

This proves the lemma. □

A.2.5 Proof of Theorem A.1

Under H_0, we have $M_a \le 0$ for every $a \in A$, hence

$$S_a = S_a^\circ + M_a \le S_a^\circ \qquad \forall a \in A.$$

Because the same realized denominator $\widetilde{Q}_a$ appears in both coordinates, it follows pathwise that

$$T_{\max} \le T_{\max}^\circ.$$

Let

$$F_n^*(x) := \mathbb{P}^*\big(T_{\max}^* \le x\big), \qquad E_n := \left\{ \sup_{x\in\mathbb{R}} \big| F_n^*(x) - \mathbb{P}(G_{Q,\circ,\max} \le x) \big| \le \rho_n \right\},$$

where ρ_n is given by Lemma A.6. Then $\mathbb{P}(E_n^c) \to 0$.

On E_n, the definition of the conditional $(1-\alpha)$-quantile gives

$$F_n^*\big(c_{n,1-\alpha}^*\big) \ge 1-\alpha.$$

Therefore, still on E_n,

$$\mathbb{P}\big(G_{Q,\circ,\max} \le c_{n,1-\alpha}^*\big) \ge 1-\alpha-\rho_n.$$

By the definition of the Gaussian quantile $q_{Q,\circ,n,1-\alpha-\rho_n}$, this implies

$$c_{n,1-\alpha}^* \ge q_{Q,\circ,n,1-\alpha-\rho_n} \qquad \text{on } E_n.$$

Hence

$$\begin{aligned}
\mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*\big) &\le \mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*, E_n\big) + \mathbb{P}(E_n^c) \\
&\le \mathbb{P}\big(T_{\max}^\circ > c_{n,1-\alpha}^*, E_n\big) + \mathbb{P}(E_n^c) \\
&\le \mathbb{P}\big(T_{\max}^\circ > q_{Q,\circ,n,1-\alpha-\rho_n}\big) + \mathbb{P}(E_n^c).
\end{aligned}$$

Now let

$$d_n := \sup_{x\in\mathbb{R}} \Big| \mathbb{P}\big(T_{\max}^\circ \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big) \Big|.$$

By Lemma A.4, $d_n \to 0$. Therefore

$$\begin{aligned}
\mathbb{P}\big(T_{\max}^\circ > q_{Q,\circ,n,1-\alpha-\rho_n}\big) &\le \mathbb{P}\big(G_{Q,\circ,\max} > q_{Q,\circ,n,1-\alpha-\rho_n}\big) + d_n \\
&\le \alpha + \rho_n + d_n.
\end{aligned}$$

Combining the last two displays yields

$$\mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*\big) \le \alpha + \rho_n + d_n + \mathbb{P}(E_n^c).$$

Taking $\limsup_{n\to\infty}$ proves

$$\limsup_{n\to\infty} \mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*\big) \le \alpha.$$

A.3 Regime D Validity

A.3.1 Main theorem

We impose the following assumptions.

Assumption A.6. *The centered vector sequence $d_1, \dots, d_n$ is geometrically α-mixing: there exist constants $K_1 > 1$, $K_2 > 0$, and $\gamma_{\mathrm{mix}} > 0$ such that*

$$\alpha(\ell) := \sup_{1 \le k \le n-\ell} \sup_{A \in \sigma(d_1,\dots,d_k),\ B \in \sigma(d_{k+\ell},\dots,d_n)} \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| \le K_1 e^{-K_2 \ell^{\gamma_{\mathrm{mix}}}} \qquad \forall \ell \ge 1.$$

Assumption A.7. *The centered oracle coordinates are nondegenerate: there exist constants $0 < c_- \le c_+ < \infty$ such that*

$$c_- \le \mathrm{Var}\left(\frac{S_a^\circ}{\sqrt{W_a}} \right) \le c_+ \qquad \forall a \in A.$$

Assumption A.8. *The kernel is symmetric, continuously differentiable, satisfies $K(0) = 1$, and obeys*

$$|K(x)| \le C_K (1 + |x|)^{-\vartheta} \quad \forall x \in \mathbb{R}$$

for some constants $C_K < \infty$ and $\vartheta > 1$. The bandwidth satisfies

$$b_n \asymp n^\rho, \quad 0 < \rho < \frac{\vartheta - 1}{3\vartheta - 2}.$$

Remark 2. We prove in Section A.3.6 that the quadratic spectral kernel introduced by Andrews (1991) satisfies Assumption A.8.

Assumption A.9. *Let*

$$B_n := \sqrt{\frac{n}{L_{\min}}}.$$

Assume $p \ge n^{\kappa_0}$ eventually for some $\kappa_0 > 0$, and

$$(\log p)^{3-\gamma_{\mathrm{mix}}} = o\big(n^{\gamma_{\mathrm{mix}}/3}\big),$$

$$(B_n)^{2/3} \frac{(\log p)^{(1+2\gamma_{\mathrm{mix}})/(3\gamma_{\mathrm{mix}})}}{n^{1/9}} + B_n \frac{(\log p)^{7/6}}{n^{1/9}} \to 0,$$

and, for some constants $c_1, c_2 > 0$,

$$(B_n)^2 n^{-c_1} (\log p)^{c_2+2} \to 0, \qquad (B_n)^2 n^{-\rho} (\log p)^2 \to 0.$$

Assumption A.10. *The interval-wise mean is asymptotically flat:*

$$b_n r_n (\log p)^2 \to 0.$$

Assumption A.11. *The bandwidth is not too large relative to the minimum interval length:*

$$\frac{(\log p)^5}{L_{\min}} \to 0, \quad \frac{b_n (\log p)^3}{L_{\min}} \to 0.$$

Theorem A.7. *Assume Assumptions A.1, A.2, A.6–A.11. Let $c^*_{n,1-\alpha}$ be the conditional $(1-\alpha)$-quantile of $T^*_{\max}$. Then the feasible residual kernel bootstrap is asymptotically conservative: under H_0,*

$$\limsup_{n\to\infty} \mathbb{P}\big(T_{\max} > c^*_{n,1-\alpha}\big) \le \alpha \qquad \forall \alpha \in (0,1).$$

We will split the proof into a few steps.

A.3.2 Preliminary bounds

Lemma A.8. *Under Assumption A.8, there exists $C_\Theta < \infty$ such that*

$$\max_{1\le t\le n} \sum_{u=1}^{n} |\Theta_n(t,u)| \le C_\Theta b_n \qquad \text{for all } n.$$

Consequently,

$$|x^\top \Theta_n y| \le C_\Theta b_n \|x\|_2 \|y\|_2 \qquad \forall x, y \in \mathbb{R}^n.$$

Proof. Fix t. Since $|K(x)| \le C_K(1+|x|)^{-\vartheta}$ with $\vartheta > 1$,

$$\sum_{u=1}^{n} |\Theta_n(t,u)| \le \sum_{u=1}^{n} C_K \left(1 + \frac{|t-u|}{b_n}\right)^{-\vartheta} \le C_K \sum_{m\in\mathbb{Z}} \left(1+\frac{|m|}{b_n}\right)^{-\vartheta}.$$

Split the last sum into $|m| \le b_n$ and $|m| > b_n$. The first part contributes $O(b_n)$. For the second part,

$$\sum_{|m|>b_n} \left(1+\frac{|m|}{b_n}\right)^{-\vartheta} \le b_n \sum_{k\ge 1} (1+k)^{-\vartheta} = O(b_n)$$

because $\vartheta > 1$. This proves the row-sum bound. The quadratic-form bound follows from

$$|x^\top \Theta_n y| \le \|x\|_2 \|\Theta_n y\|_2 \le \|x\|_2 \left(\max_t \sum_u |\Theta_n(t,u)|\right) \|y\|_2.$$

□

Lemma A.9. *Lemma A.2 continuous to hold under Assumptions A.1, A.2, and A.6–A.9,*

Proof. As in the proof of Lemma A.2, define

$$Z^\circ_{t,a} := \sqrt{n}\, \mathbf{1}\{t \in I\} \frac{d_t^{(h)}}{\sqrt{W_a}}, \qquad a = (I,h) \in A.$$

Then

$$Y^\circ := (Y^\circ_a)_{a\in A} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z^\circ_t.$$

By Assumptions A.1, A.2, and Proposition A.26,

$$\|Z^\circ_{t,a}\|_{\psi_2} \le B_0 \sqrt{\frac{n}{\sigma_-^2 L_{\min}}} \lesssim B_n \qquad \forall t, a.$$

By Assumption A.6, the $Z_{t,a}$ sequence inherits the same geometric α-mixing rate, and by Assumption A.7 every coordinate variance of $n^{-1/2}\sum_t Z^\circ_{t,a}$ lies in $[c_-, c_+]$. Hence Proposition A.20 applies under

Assumption A.9 and gives

$$\sup_{A\in\mathscr{R}^p}\left|\mathbb{P}(Y^\circ\in A)-\mathbb{P}(G_{Q,\circ}\in A)\right|\to 0.$$

Applying this with $A=[-x,x]^p$ yields

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(\bar{U}_W^\circ\le x)-\mathbb{P}\left(\max_{a\in A}|G_{Q,\circ,a}|\le x\right)\right|\to 0.$$

Since $\max_a \mathrm{Var}(G_{Q,\circ,a})\le c_+$ and $\|G_{Q,\circ,a}\|_{\psi_2}\lesssim\sqrt{\mathrm{Var}(G_{Q,\circ,a})}$, Proposition A.25 implies

$$\max_{a\in A}|G_{Q,\circ,a}|=O_p\big(\sqrt{\log p}\big),$$

and therefore $\bar{U}_W^\circ=O_p(\sqrt{\log p})$.

For the second display, note that $|\mathscr{K}|\le|\mathscr{I}|^2$, so $H|\mathscr{K}|\le p^2$. Repeating the same argument with $\mathscr{K}\times[H]$ in place of A gives the same $\sqrt{\log p}$ order. □

Lemma A.10. *Define*

$$X_t^{(h)}:=\big(d_t^{(h)}\big)^2-\mathbb{E}\Big[\big(d_t^{(h)}\big)^2\Big].$$

Assume Assumptions A.1, A.2, and A.6–A.11. Then there exist constants $C_A,c_A,\eta_0\in(0,\infty)$, depending only on $(B_0,K_1,K_2,\gamma_{\mathrm{mix}})$, such that for every interval $I\subset[n]$ with $m:=|I|$ and every $x\ge 1$ satisfying

$$x\le c_A m^{\gamma_{\mathrm{mix}}/(2+\gamma_{\mathrm{mix}})},$$

one has

$$\mathbb{P}\left(\left|\sum_{t\in I}X_t^{(h)}\right|>C_A\sqrt{mx}\right)\le 2e^{-x}.$$

Proof. Fix an interval $I=[s,s+m-1]$ and define $Z_r:=X_{s+r-1}^{(h)}$ for $r=1,\dots,m$. Since $X_t^{(h)}$ is a measurable function of D_t, the sequence $(Z_r)_{r=1}^m$ inherits geometric strong mixing condition (Assumption A.6). After possibly decreasing the exponent constant, the prefactor K_1 can be absorbed into the exponential rate, so there exists $\widetilde{K}_2>0$, depending only on $(K_1,K_2,\gamma_{\mathrm{mix}})$, such that

$$\alpha_Z(\ell)\le\exp\big(-\widetilde{K}_2\ell^{\gamma_{\mathrm{mix}}}\big)\qquad\text{for all }\ell\ge 1.$$

By Assumption A.1 and Proposition A.27,

$$\|(d_t^{(h)})^2\|_{\psi_1}\le B_0^2\Longrightarrow\|X_t^{(h)}\|_{\psi_1}\le B_0^2.$$

Therefore there exists $b_0<\infty$, depending only on B_0, such that

$$\sup_{r\in[m]}\mathbb{P}(|Z_r|>u)\le\exp\big(1-u/b_0\big)\qquad\text{for all }u>0.$$

Thus Proposition A.23 applies with

$$\gamma_1=\gamma_{\mathrm{mix}},\qquad\gamma_2=1,\qquad\gamma=\left(\frac{1}{\gamma_{\mathrm{mix}}}+1\right)^{-1}=\frac{\gamma_{\mathrm{mix}}}{1+\gamma_{\mathrm{mix}}}.$$

Consequently,

$$\frac{\gamma}{2-\gamma} = \frac{\gamma_{\mathrm{mix}}}{2+\gamma_{\mathrm{mix}}}.$$

Hence there exist constants $C_0, C_1, C_2, \eta_0 \in (0,\infty)$, depending only on $(B_0, K_1, K_2, \gamma_{\mathrm{mix}})$, such that for every $y \geq C_0(\log m)^{\eta_0}$,

$$\mathbb{P}\left(\max_{1\leq j\leq m}\left|\sum_{r=1}^{j} Z_r\right| \geq y\right) \leq (m+1)\exp\left(-\frac{y^{\gamma}}{C_1}\right) + \exp\left(-\frac{y^2}{C_2 + C_2 mV}\right).$$

By Proposition A.24, $V = O(1)$. Since the full interval sum is one of the partial sums, the same bound applies to $\left|\sum_{t\in I} X_t^{(h)}\right|$.

Now set $y = A\sqrt{mx}$ with $A > 0$ to be chosen. Because $x \geq 1$, one has $y \geq A\sqrt{m}$, which dominates $C_0(\log m)^{\eta_0}$ for all sufficiently large m; the finitely many small values of m are absorbed into the final constant C_A. Also, from

$$x \leq c_A m^{\gamma/(2-\gamma)} = c_A m^{\gamma_{\mathrm{mix}}/(2+\gamma_{\mathrm{mix}})},$$

it follows that $x \lesssim m^{\gamma/2}x^{\gamma/2}$, so by choosing A large enough and then c_A small enough,

$$(m+1)\exp\left(-\frac{A^{\gamma} m^{\gamma/2} x^{\gamma/2}}{C_1}\right) \leq e^{-x}.$$

For the second term, $1 + mV \leq Cm$ for a suitable constant C, hence

$$\exp\left(-\frac{A^2 mx}{C_2 + C_2 mV}\right) \leq \exp\left(-\frac{A^2 x}{C}\right) \leq e^{-x}$$

for A chosen large enough. Combining the two bounds gives the claimed inequality after renaming A as C_A. □

Lemma A.11. *Lemma A.3 continues to hold under Assumptions A.1, A.2, A.6–A.11. Moreover,*

$$\Delta_d := \max_{a\in A}\left|\frac{\widetilde{Q}_a^d}{W_a} - 1\right| = O_p\left(\sqrt{\frac{\log p}{L_{\min}}} + \frac{\log p}{L_{\min}}\right),$$

and

$$\max_{a\in A}\left|\frac{\widetilde{Q}_a - \widetilde{Q}_a^d}{W_a}\right| = O_{P^*}\left(\frac{\log p}{\sqrt{L_{\min}}} + \frac{(\log p)^{3/2}}{L_{\min}}\right).$$

In particular, $\Delta_d \xrightarrow{p} 0$ *and* $\mathbb{P}(\Delta_d > 1/2) \to 0$.

Proof. Recall the decomposition (A.8) in the proof of Lemma A.3:

$$\frac{\widetilde{Q}_a - W_a}{W_a} = \frac{A_a^d}{W_a} - \frac{B_a^d}{W_a} + 2\frac{C_a}{W_a} + \frac{R_a}{W_a}.$$

By definition,

$$\Delta_d = \frac{A_a^d}{W_a} - \frac{B_a^d}{W_a}, \quad \frac{\widetilde{Q}_a - \widetilde{Q}_a^{(d)}}{W_a} = 2\frac{C_a}{W_a} + \frac{R_a}{W_a}.$$

We control the four terms uniformly.

Step 1: control of $\frac{A_a^d}{W_a}$. By Lemma A.10, there exist constants $C_A, c_A > 0$ such that, for every fixed $a = (I,h)$ and every $x \geq 1$ with

$$x + \log p \leq c_A |I|^{\gamma_{\mathrm{mix}}/(2+\gamma_{\mathrm{mix}})},$$

one has

$$\mathbb{P}\Big(|A_a^d| > C_A \sqrt{|I|(x+\log p)}\Big) \leq 2e^{-(x+\log p)}.$$

Assumption A.9 implies in particular

$$\log p = o\big(L_{\min}^{\gamma_{\mathrm{mix}}/(2+\gamma_{\mathrm{mix}})}\big),$$

so the restriction is eventually satisfied uniformly over all intervals. A union bound over the p coordinates then gives

$$\max_{a\in A} |A_a^d| = O_p\Big(\max_{a\in A} \sqrt{|I|\log p}\Big).$$

Since $W_a \geq \sigma_-^2 |I| \geq \sigma_-^2 L_{\min}$,

$$\max_{a\in A} \frac{|A_a^d|}{W_a} = O_p\left(\sqrt{\frac{\log p}{L_{\min}}}\right).$$

Step 2: control of $\frac{B_a^d}{W_a}$. By Lemma A.9,

$$\max_{a\in A} \frac{B_a^d}{W_a} = \max_{a\in A} \frac{(S_a^\circ)^2}{|I| W_a} \leq \frac{(\bar{U}_W^\circ)^2}{L_{\min}} = O_p\left(\frac{\log p}{L_{\min}}\right).$$

Step 3: control of $\frac{C_a}{W_a}$. Fix $a = (I,h)$ and define $X_{t,a} := \delta_{t,a} d_t^{(h)}$ for $t \in I$. If $R_a = 0$, then $\delta_{t,a} = 0$ for every $t \in I$, hence $C_a = 0$, and there is nothing to prove. Assume therefore that $R_a > 0$, write $I = [s, s+m-1]$ with $m := |I|$, and introduce the rescaled sequence

$$Z_{r,a} := \frac{\delta_{s+r-1,a}}{R_a^{1/2}} d_{s+r-1}^{(h)}, \qquad r = 1, \ldots, m.$$

Then

$$\frac{C_a}{R_a^{1/2}} = \sum_{r=1}^{m} Z_{r,a}.$$

Each $Z_{r,a}$ is centered because $\delta_{s+r-1,a}$ is deterministic and $\mathbb{E}[d_{s+r-1}^{(h)}] = 0$. Because $Z_{r,a}$ is a measurable function of $d_{s+r-1}^{(h)}$, the sequence $(Z_{r,a})_{r=1}^m$ inherits the same geometric α-mixing rate as $(d_t^{(h)})_{t\in I}$. After possibly decreasing the exponential constant to absorb the prefactor K_1, there exists $\widetilde{K}_2 > 0$, depending only on $(K_1, K_2, \gamma_{\mathrm{mix}})$, such that

$$\alpha_{Z_a}(\ell) \leq \exp\big(-\widetilde{K}_2 \ell^{\gamma_{\mathrm{mix}}}\big) \qquad \forall \ell \geq 1.$$

Also, since $|\delta_{t,a}| \leq R_a^{1/2}$,

$$\|Z_{r,a}\|_{\psi_2} \leq B_0.$$

Hence there exists $b_0 < \infty$, depending only on B_0, such that

$$\sup_{1\le r\le m} \mathbb{P}(|Z_{r,a}| > u) \le \exp\bigl(1 - u/b_0\bigr) \qquad \forall\, u > 0.$$

Therefore Proposition A.23 applies to $(Z_{r,a})_{r=1}^m$ with

$$\gamma_1 = \gamma_{\mathrm{mix}}, \qquad \gamma_2 = 1, \qquad \gamma = \left(\frac{1}{\gamma_{\mathrm{mix}}} + 1\right)^{-1} = \frac{\gamma_{\mathrm{mix}}}{1+\gamma_{\mathrm{mix}}}, \qquad \frac{\gamma}{2-\gamma} = \frac{\gamma_{\mathrm{mix}}}{2+\gamma_{\mathrm{mix}}}.$$

By Proposition A.24, $V = O(1)$. Now

$$\sup_{1\le r\le m} \mathbb{E}[Z_{r,a}^2] \le \sigma_+^2 \sup_{t\in I} \frac{\delta_{t,a}^2}{R_a} \le \sigma_+^2,$$

while $\|Z_{r,a}\|_{\psi_2} \le B_0$ implies

$$Q_a(u) \lesssim B_0 \sqrt{\log(e/u)} \qquad \forall\, u \in (0,1],$$

so $\int_0^1 Q_a(u)^2\, du < \infty$. Geometric mixing gives $\sum_{k\ge 1} \alpha_{Z_a}(k) < \infty$. Hence $V \le C$ for a constant depending only on $(B_0, K_1, K_2, \gamma_{\mathrm{mix}}, \sigma_+^2)$, uniformly in a.

Applying Proposition A.23 to $(Z_{r,a})_{r=1}^m$ in this rescaled weighted-sum form, exactly as recorded in the dependent studentization note, therefore gives constants $C_C, c_C > 0$ such that

$$\mathbb{P}\left(\left|\frac{C_a}{R_a^{1/2}}\right| > C_C\sqrt{x+\log p}\right) \le 2e^{-(x+\log p)}$$

whenever

$$x + \log p \le c_C m^{\gamma_{\mathrm{mix}}/(2+\gamma_{\mathrm{mix}})}.$$

Equivalently,

$$\mathbb{P}\Bigl(|C_a| > C_C\sqrt{R_a(x+\log p)}\Bigr) \le 2e^{-(x+\log p)}$$

for all fixed a. Assumption A.9 implies, exactly as in Step 1, that this admissibility restriction is eventually satisfied uniformly over all scanned intervals. A union bound over $a \in A$ yields

$$\max_{a\in A} |C_a| = O_p\left(\max_{a\in A} \sqrt{R_a \log p}\right).$$

Since $R_a \le r_n W_a$ and $W_a \ge \sigma_-^2 L_{\min}$,

$$\max_{a\in A} \frac{|C_a|}{W_a} = O_p\left(\sqrt{\frac{r_n \log p}{L_{\min}}}\right).$$

Step 4: control of $\frac{R_a}{W_a}$. By definition (A.3) of r_n,

$$\max_{a\in A} \frac{R_a}{W_a} = r_n.$$

Collecting the bounds from Steps 1–4 proves that

$$\Delta_Q = O_p(\zeta_n), \quad \Delta_d = O_p\left(\sqrt{\frac{\log p}{L_{\min}}} + \frac{\log p}{L_{\min}}\right) = O_p(\zeta_n),$$

and

$$\max_{a\in A}\left|\frac{\widetilde{Q}_a - \widetilde{Q}_a^d}{W_a}\right| = O_{P^*}\left(\frac{\log p}{\sqrt{L_{\min}}} + \frac{(\log p)^{3/2}}{L_{\min}}\right).$$

Since $\zeta_n \to 0$ under Assumption A.10, we also have $\Delta_Q, \Delta_d \xrightarrow{p} 0$ and hence $\mathbb{P}(\Delta_Q > 1/2) \to 0, \mathbb{P}(\Delta_d > 1/2) \to 0$. □

A.3.3 Gaussian approximation for $T^\circ_{\max}$

Define the centered oracle-normalized vector

$$Y^\circ := \left(Y_a^\circ\right)_{a\in A}, \qquad Y_a^\circ := \frac{S_a^\circ}{\sqrt{W_a}}, \qquad U_W^\circ := \max_{a\in A} Y_a^\circ.$$

By construction, $\mathrm{Cov}(Y^\circ) = \Sigma_{Q,\circ}$.

Lemma A.12. *Under Assumptions A.1, A.2, and A.6–A.9,*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(T^\circ_{\max} \le x\right) - \mathbb{P}\left(G_{Q,\circ,\max} \le x\right)\right| \to 0.$$

Proof. We proceed in three steps.

Step 1. For $t \in [n]$ and $a = (I,h) \in A$, define

$$Z_{t,a} := \sqrt{n}\,\mathbf{1}\{t \in I\}\frac{d_t^{(h)}}{\sqrt{W_a}}.$$

Then

$$Y^\circ = \frac{1}{\sqrt{n}}\sum_{t=1}^n Z_t.$$

The $Z_{t,a}$ sequence inherits the same geometric α-mixing rate, and Assumption A.7 yields

$$c_- \le \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{t,a}\right) \le c_+ \qquad \forall\, a \in A.$$

By Assumptions A.2, A.6, Proposition A.23 with $\gamma = 2$, and Proposition A.24,

$$\|Z_{t,a}\|_{\psi_2} \lesssim B_n, \qquad \forall\, t, a.$$

Therefore Proposition A.20 applies under Assumption A.9 and implies

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(U_W^\circ \le x\right) - \mathbb{P}\left(G_{Q,\circ,\max} \le x\right)\right| \to 0.$$

Step 2. Using exactly the same argument as in Step 2 of the proof of Lemma A.4, we can prove that

there exists a deterministic sequence $\eta_n \downarrow 0$ such that

$$\mathbb{P}\big(2\Delta_Q \bar{U}_W^\circ > \eta_n\big) \to 0, \qquad \eta_n \sqrt{1 \vee \log(p/\eta_n)} \to 0.$$

Step 3. Since the diagonal entries of $\Sigma_{Q,\circ}$ lie in $[c_-, c_+]$, Proposition A.18 implies

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(|G_{Q,\circ,\max} - x| \le \eta_n\big) \to 0.$$

Using exactly the same argument as in Step 3 of the proof of Lemma A.4, we can prove that

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\big(T_{\max}^\circ \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big)\Big| \to 0.$$

This completes the proof. □

A.3.4 Bootstrap validity

Consider an oracle Gaussian multiplier bootstrap based on $T_{\max}^{*,\circ}$ where

$$M_a^* := \sum_{t\in I} \xi_t d_t^{(h)}, \qquad T_{\max}^{*,\circ} := \max_{a\in A} \frac{M_a^*}{\sqrt{\widetilde{Q}_a^d}}.$$

Lemma A.13. *Under Assumptions A.1, A.2, A.6–A.9, there exists a deterministic sequence $\rho_n^\circ \downarrow 0$ such that*

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} \Big|\mathbb{P}^*\big(T_{\max}^{*,\circ} \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big)\Big| > \rho_n^\circ\right) \to 0.$$

Moreover, if

$$\bar{U}_W^* := \max_{a\in A} \frac{|M_a^*|}{\sqrt{W_a}},$$

then

$$\bar{U}_W^* = O_{P^*}\big(\sqrt{\log p}\big) \quad \textit{in probability.}$$

Proof. Define the centered oracle bootstrap vector

$$Y^{*,\circ} := \left(\frac{M_a^*}{\sqrt{W_a}}\right)_{a\in A}, \quad U_W^* := \max_{a\in A} \frac{M_a^*}{\sqrt{W_a}}.$$

Conditionally on the data, $Y^{*,\circ}$ is Gaussian with covariance matrix

$$\widehat{\Gamma}^\circ := \mathrm{Cov}^*(Y^{*,\circ}) = \frac{1}{n}\sum_{t=1}^n \sum_{u=1}^n \Theta_n(t,u) Z_t^\circ Z_u^{\circ\top},$$

where $Z_t^\circ = (Z_{t,a})_{a\in A}$ and $Z_{t,a}$ is defined in (A.9). By Proposition A.21,

$$\big\|\widehat{\Gamma}^\circ - \Sigma_{Q,\circ}\big\|_{\max} = o_p\big((\log p)^{-2}\big).$$

In particular, because the diagonal of $\Sigma_{Q,\circ}$ lies in $[c_-, c_+]$ by Assumption A.7,

$$\max_{a\in A} \widehat{\Gamma}_{aa}^\circ = O_p(1) \quad \text{and} \quad \min_{a\in A} \widehat{\Gamma}_{aa}^\circ \ge c_-/2$$

with probability tending to one. Applying Proposition A.25 conditional on $(d_t^{(h)})$ with probability tending to one, we have

$$\bar{U}_W^* = O_{P^*}\big(\sqrt{\log p}\big) \quad \text{in probability.}$$

By Proposition A.22,

$$\sup_{A\in\mathscr{R}^p} \Big|\mathbb{P}^*(Y^{*,\circ}\in A) - \mathbb{P}(G_{Q,\circ}\in A)\Big| \xrightarrow{p} 0.$$

In particular,

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}^*(U_W^* \le x) - \mathbb{P}(G_{Q,\circ,\max}\le x)\Big| \xrightarrow{p} 0.$$

Finally, on the event $\{\Delta_d \le 1/2\}$,

$$|T_{\max}^{*,\circ} - U_W^*| \le 2\Delta_d \bar{U}_W^*.$$

By Lemma A.11 and the displayed bound for $\bar{U}_W^*$,

$$2\Delta_d \bar{U}_W^* = O_{P^*}\left(\frac{\log p}{\sqrt{L_{\min}}} + \frac{(\log p)^{3/2}}{L_{\min}}\right), \quad \text{in probability.}$$

Under Assumption A.11, the right-hand side is $o(1)$. Choose any deterministic sequence $\eta_n^\circ \downarrow 0$ so slowly that

$$\mathbb{P}\Big(\mathbb{P}^*\big(2\Delta_d\bar{U}_W^* > \eta_n^\circ\big) > \eta_n^\circ\Big) \to 0, \qquad \eta_n^\circ\sqrt{1\vee\log(p/\eta_n^\circ)} \to 0.$$

Since U_W^* is conditionally Gaussian and its conditional diagonal variances are bounded away from 0 and ∞ in probability, Proposition A.18 gives

$$\sup_{x\in\mathbb{R}} \mathbb{P}^*\big(|U_W^* - x| \le \eta_n^\circ\big) \xrightarrow{p} 0.$$

The usual inclusion argument then yields

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}^*\big(T_{\max}^{*,\circ}\le x\big) - \mathbb{P}\big(G_{Q,\circ,\max}\le x\big)\Big| \xrightarrow{p} 0.$$

A standard diagonal argument therefore yields a deterministic sequence $\rho_n^\circ \downarrow 0$ with the stated probability bound. □

Lemma A.14. *Under Assumptions A.1, A.2, and A.6–A.11,*

$$T_{\max}^* - T_{\max}^{*,\circ} = o_{P^*}(1) \quad \textit{in probability.}$$

More precisely, with

$$\beta_n := (\log p)\sqrt{\frac{b_n}{L_{\min}}} + \sqrt{b_n r_n \log p},$$

$$\gamma_n := \beta_n + \left(r_n + \sqrt{\frac{r_n\log p}{L_{\min}}}\right)\sqrt{\log p},$$

there exist deterministic sequences $\eta_n^* \downarrow 0$ *and* $\delta_n^* \downarrow 0$ *such that*

$$\mathbb{P}\Big(\mathbb{P}^*\big(|T_{\max}^* - T_{\max}^{*,\circ}| > \eta_n^*\big) > \delta_n^*\Big) \to 0, \qquad \eta_n^* \sqrt{1 \vee \log(p/\eta_n^*)} \to 0.$$

Consequently,

$$\sup_{x \in \mathbb{R}} \Big| \mathbb{P}^*\big(T_{\max}^* \le x\big) - \mathbb{P}^*\big(T_{\max}^{*,\circ} \le x\big) \Big| \xrightarrow{p} 0.$$

Proof. We compare the feasible statistic to the centered oracle bootstrap in two steps. Define the intermediate statistic

$$U_{\max}^{*,\circ} := \max_{a \in A} \frac{M_a^*}{\sqrt{\widetilde{Q}_a}}.$$

Then

$$T_{\max}^* - T_{\max}^{*,\circ} = \big(T_{\max}^* - U_{\max}^{*,\circ}\big) + \big(U_{\max}^{*,\circ} - T_{\max}^{*,\circ}\big).$$

Note that

$$T_{\max}^* := \max_{a \in A} \frac{N_a^*}{\sqrt{\widetilde{Q}_a}} \mathbf{1}\{\widetilde{Q}_a > 0\},$$

where

$$N_a^* := \sum_{t=1}^n \xi_t R_{t,a}.$$

Step 1. For $a = (I, h)$,

$$N_a^* - M_a^* = -\bar{d}_a \xi_a + \Lambda_a^*,$$

where

$$\xi_a := \sum_{t \in I} \xi_t, \qquad \Lambda_a^* := \sum_{t \in I} \xi_t \big(\mu_t^{(h)} - \bar{\mu}_a\big).$$

On the event $\{\Delta_Q \le 1/2\}$, one has $\widetilde{Q}_a \ge W_a/2$ for every a, so

$$|T_{\max}^* - U_{\max}^{*,\circ}| \le \sqrt{2} \max_{a \in A} \frac{|N_a^* - M_a^*|}{\sqrt{W_a}}.$$

We bound the two pieces on the right.

First, by Lemma A.8,

$$\mathrm{Var}^*(\xi_a) = \sum_{t,u \in I} \Theta_n(t,u) \le \sum_{t,u \in I} |\Theta_n(t,u)| \le C_\Theta b_n |I|.$$

Therefore

$$\mathrm{Var}^*\left(\frac{\bar{d}_a \xi_a}{\sqrt{W_a}}\right) \le C_\Theta b_n \frac{|I| \bar{d}_a^2}{W_a} = C_\Theta b_n \frac{(S_a^\circ)^2}{|I| W_a} \le C_\Theta b_n \frac{(\bar{U}_W^\circ)^2}{L_{\min}}.$$

Conditionally on the data, the family $\{\bar{d}_a \xi_a / \sqrt{W_a}\}_{a \in A}$ is Gaussian. Hence Proposition A.25 gives

$$\max_{a \in A} \frac{|\bar{d}_a \xi_a|}{\sqrt{W_a}} = O_{P^*}\left(\sqrt{\frac{b_n}{L_{\min}}}\, \bar{U}_W^\circ \sqrt{\log p}\right) \quad \text{in probability.}$$

By Lemma A.9, $\bar{U}_W^\circ = O_p(\sqrt{\log p})$, so unconditionally

$$\max_{a\in A} \frac{|\bar{d}_a \xi_a|}{\sqrt{W_a}} = O_p\left((\log p)\sqrt{\frac{b_n}{L_{\min}}}\right).$$

Second, again by Lemma A.8,

$$\mathrm{Var}^*(\Lambda_a^*) \le C_\Theta b_n R_a \le C_\Theta b_n r_n W_a.$$

Therefore, conditionally on the data,

$$\max_{a\in A} \frac{|\Lambda_a^*|}{\sqrt{W_a}} = O_{P^*}\left(\sqrt{b_n r_n \log p}\right) \quad \text{in probability.}$$

Combining the last three displays yields

$$|T_{\max}^* - U_{\max}^{*,\circ}| = O_{P^*}(\beta_n) \quad \text{in probability.}$$

Step 2: denominator perturbation. On the event

$$E_n := \{\Delta_Q \le 1/2\} \cap \{\Delta_d \le 1/2\},$$

both denominators are bounded below by one half of W_a. For every $a \in A$,

$$\left|\frac{1}{\sqrt{\widetilde{Q}_a}} - \frac{1}{\sqrt{\widetilde{Q}_a^d}}\right| \le \frac{2}{\sqrt{W_a}} \left|\frac{\widetilde{Q}_a - \widetilde{Q}_a^d}{W_a}\right| \qquad \text{on } E_n.$$

Therefore

$$|U_{\max}^{*,\circ} - T_{\max}^{*,\circ}| \le 2\bar{U}_W^* \max_{a\in A} \left|\frac{\widetilde{Q}_a - \widetilde{Q}_a^d}{W_a}\right| \qquad \text{on } E_n.$$

By Lemma A.11,

$$\max_{a\in A} \left|\frac{\widetilde{Q}_a - \widetilde{Q}_a^d}{W_a}\right| = O_p\left(r_n + \sqrt{\frac{r_n \log p}{L_{\min}}}\right).$$

By Lemma A.13,

$$\bar{U}_W^* = O_{P^*}\left(\sqrt{\log p}\right) \quad \text{in probability.}$$

Hence

$$|U_{\max}^{*,\circ} - T_{\max}^{*,\circ}| = O_{P^*}\left(\left(r_n + \sqrt{\frac{r_n \log p}{L_{\min}}}\right)\sqrt{\log p}\right)$$

in probability.

Combining Steps 1 and 2,

$$|T_{\max}^* - T_{\max}^{*,\circ}| = O_{P^*}(\gamma_n) \quad \text{in probability.}$$

Under Assumptions A.10 and A.5,

$$\gamma_n \to 0, \quad \gamma_n \sqrt{\log p} \to 0.$$

Choose any deterministic sequence $M_n \uparrow \infty$ so slowly that

$$\eta_n^* := M_n \gamma_n \downarrow 0, \quad \eta_n^* \sqrt{1 \vee \log(p/\eta_n^*)} \to 0.$$

Since $|T_{\max}^* - T_{\max}^{*,\circ}| = O_{P^*}(\gamma_n)$ in probability, a standard diagonal argument yields a deterministic sequence $\delta_n^* \downarrow 0$ such that

$$\mathbb{P}\Big(\mathbb{P}^*\big(|T_{\max}^* - T_{\max}^{*,\circ}| > \eta_n^*\big) > \delta_n^*\Big) \to 0.$$

To pass from statistic closeness to CDF closeness, note that conditionally on the data, $T_{\max}^{*,\circ}$ is the maximum of a centered Gaussian vector. By Lemma A.13, its conditional diagonal variances are bounded away from 0 and ∞ in probability. Therefore Proposition A.18 implies

$$\sup_{x \in \mathbb{R}} \mathbb{P}^*\big(|T_{\max}^{*,\circ} - x| \le \eta_n^*\big) \xrightarrow{p} 0.$$

The standard inclusion argument now yields

$$\sup_{x \in \mathbb{R}} \Big| \mathbb{P}^*\big(T_{\max}^* \le x\big) - \mathbb{P}^*\big(T_{\max}^{*,\circ} \le x\big) \Big| \xrightarrow{p} 0.$$

□

Lemma A.15. *Under Assumptions A.1, A.2, and A.6–A.11, there exists a deterministic sequence $\rho_n \downarrow 0$ such that*

$$\mathbb{P}\left(\sup_{x \in \mathbb{R}} \Big| \mathbb{P}^*\big(T_{\max}^* \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big) \Big| > \rho_n \right) \to 0.$$

Proof. By the triangle inequality,

$$\begin{aligned} \sup_{x \in \mathbb{R}} \Big| \mathbb{P}^*\big(T_{\max}^* \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big) \Big| \le & \sup_{x \in \mathbb{R}} \Big| \mathbb{P}^*\big(T_{\max}^* \le x\big) - \mathbb{P}^*\big(T_{\max}^{*,\circ} \le x\big) \Big| \\ & + \sup_{x \in \mathbb{R}} \Big| \mathbb{P}^*\big(T_{\max}^{*,\circ} \le x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \le x\big) \Big|. \end{aligned}$$

The first term converges to 0 in probability by Lemma A.14, and the second by Lemma A.13. A standard diagonal argument therefore yields the deterministic sequence $\rho_n \downarrow 0$ in the statement. □

A.3.5 Proof of Theorem A.7

Proof. Under H_0, we have $M_a \le 0$ for every $a \in A$, hence

$$S_a = S_a^\circ + M_a \le S_a^\circ \qquad \forall a \in A.$$

Because the same realized denominator $\widetilde{Q}_a$ appears in both coordinates, it follows pathwise that

$$T_{\max} \le T_{\max}^\circ.$$

Let

$$F_n^*(x) := \mathbb{P}^*\big(T_{\max}^* \le x\big), \qquad E_n := \left\{ \sup_{x \in \mathbb{R}} \big| F_n^*(x) - \mathbb{P}(G_{Q,\circ,\max} \le x) \big| \le \rho_n \right\},$$

where ρ_n is given by Lemma A.15. Then $\mathbb{P}(E_n^c) \to 0$.

On E_n, the definition of the conditional $(1-\alpha)$-quantile gives

$$F_n^*\big(c_{n,1-\alpha}^*\big) \geq 1-\alpha.$$

Therefore, still on E_n,

$$\mathbb{P}\big(G_{Q,\circ,\max} \leq c_{n,1-\alpha}^*\big) \geq 1-\alpha-\rho_n.$$

By the definition of the Gaussian quantile $q_{Q,\circ,n,1-\alpha-\rho_n}$, this implies

$$c_{n,1-\alpha}^* \geq q_{Q,\circ,n,1-\alpha-\rho_n} \qquad \text{on } E_n.$$

Hence

$$\begin{aligned}
\mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*\big) &\leq \mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*, E_n\big) + \mathbb{P}(E_n^c) \\
&\leq \mathbb{P}\big(T_{\max}^\circ > c_{n,1-\alpha}^*, E_n\big) + \mathbb{P}(E_n^c) \\
&\leq \mathbb{P}\big(T_{\max}^\circ > q_{Q,\circ,n,1-\alpha-\rho_n}\big) + \mathbb{P}(E_n^c).
\end{aligned}$$

Now let

$$\delta_n := \sup_{x\in\mathbb{R}} \Big| \mathbb{P}\big(T_{\max}^\circ \leq x\big) - \mathbb{P}\big(G_{Q,\circ,\max} \leq x\big) \Big|.$$

By Lemma A.12, $\delta_n \to 0$. Therefore

$$\begin{aligned}
\mathbb{P}\big(T_{\max}^\circ > q_{Q,\circ,n,1-\alpha-\rho_n}\big) &\leq \mathbb{P}\big(G_{Q,\circ,\max} > q_{Q,\circ,n,1-\alpha-\rho_n}\big) + \delta_n \\
&\leq \alpha + \rho_n + \delta_n.
\end{aligned}$$

Combining the last two displays yields

$$\mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*\big) \leq \alpha + \rho_n + \delta_n + \mathbb{P}(E_n^c).$$

Taking $\limsup_{n\to\infty}$ proves

$$\limsup_{n\to\infty} \mathbb{P}\big(T_{\max} > c_{n,1-\alpha}^*\big) \leq \alpha.$$

□

A.3.6 Quadratic spectral kernel satisfies Assumption A.8

Lemma A.16. *Let*

$$K_{\mathrm{QS}}(x) := \begin{cases} \dfrac{25}{12\pi^2 x^2}\left\{\dfrac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right\}, & x \neq 0, \\ 1, & x = 0. \end{cases}$$

Then K_{QS} is symmetric and continuously differentiable on $\mathbb{R}$, satisfies $K_{\mathrm{QS}}(0) = 1$, and obeys

$$|K_{\mathrm{QS}}(x)| \leq C_{\mathrm{QS}}(1+|x|)^{-2} \qquad \forall x \in \mathbb{R}$$

for some constant $C_{\mathrm{QS}} < \infty$. Consequently, Assumption A.8 holds for the quadratic spectral kernel with $\vartheta = 2$ whenever

$$b_n \asymp n^{\rho}, \qquad 0 < \rho < \frac{1}{4}.$$

Proof. Symmetry is immediate because K_{QS} depends on x only through x^2 and the even functions $\sin(z)/z$ and $\cos(z)$. For $x \neq 0$, the formula is smooth. Set $c := 6\pi/5$. Using the Taylor expansions

$$\frac{\sin(cx)}{cx} = 1 - \frac{c^2x^2}{6} + \frac{c^4x^4}{120} + O(x^6), \qquad \cos(cx) = 1 - \frac{c^2x^2}{2} + \frac{c^4x^4}{24} + O(x^6),$$

we obtain

$$\frac{\sin(cx)}{cx} - \cos(cx) = \frac{c^2x^2}{3} - \frac{c^4x^4}{30} + O(x^6).$$

Therefore,

$$K_{\mathrm{QS}}(x) = 1 - \frac{18\pi^2}{125}x^2 + O(x^4) \qquad (x \to 0).$$

Hence K_{QS} extends continuously at 0 with value 1, and the expansion also gives $K'_{\mathrm{QS}}(0) = 0$; thus $K_{\mathrm{QS}} \in C^1(\mathbb{R})$.

For the tail bound, again with $c = 6\pi/5$, for $x \neq 0$ we have

$$|K_{\mathrm{QS}}(x)| \le \frac{25}{12\pi^2x^2}\left(\left|\frac{\sin(cx)}{cx}\right| + |\cos(cx)|\right) \le \frac{25}{12\pi^2x^2}\left(1 + \frac{1}{c|x|}\right).$$

In particular, for $|x| \ge 1$,

$$|K_{\mathrm{QS}}(x)| \le \frac{25}{12\pi^2}\left(1 + \frac{1}{c}\right)|x|^{-2} \le 4\frac{25}{12\pi^2}\left(1 + \frac{1}{c}\right)(1 + |x|)^{-2}.$$

On the compact set $[-1, 1]$, continuity implies $\sup_{|x|\le 1} |K_{\mathrm{QS}}(x)| < \infty$. Enlarging the constant if necessary yields

$$|K_{\mathrm{QS}}(x)| \le C_{\mathrm{QS}}(1 + |x|)^{-2} \qquad \forall x \in \mathbb{R}.$$

Thus Assumption A.8 is satisfied with $\vartheta = 2$. The bandwidth restriction in that assumption then becomes

$$0 < \rho < \frac{\vartheta - 1}{3\vartheta - 2} = \frac{1}{4}.$$

□

A.4 Miscellany

This section states the key technical results our proof relies on for completeness.

Proposition A.17 (Chernozhukov, Chetverikov and Kato (2017), Theorem 2.1 and Proposition 2.1)**.** *Let $X_1, \ldots, X_n$ be independent centered random vectors in $\mathbb{R}^p$, and let*

$$S_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i, \qquad \Sigma := \mathrm{Cov}(S_n), \qquad G \sim N(0, \Sigma).$$

Assume that, for some constants $b > 0$ and $B_n \geq 1$,

$$\min_{1\leq j\leq p} \Sigma_{jj} \geq b, \qquad \max_{1\leq i\leq n, 1\leq j\leq p} \|X_{ij}\|_{\psi_1} \leq B_n.$$

Then there exists a constant $C = C(b) < \infty$ such that

$$\sup_{A\in\mathscr{R}^p} \left|\mathbb{P}(S_n \in A) - \mathbb{P}(G \in A)\right| \leq C\left(\frac{B_n^2 \log^7(pn)}{n}\right)^{1/6},$$

where $\mathscr{R}^p$ denotes the class of coordinatewise hyperrectangles in $\mathbb{R}^p$. In particular, if $B_n^2 \log^7(pn) = o(n)$, then the left-hand side converges to 0.

Proposition A.18 (Chernozhukov, Chetverikov and Kato (2015), Theorem 3 and Corollary 1)**.** *Let $G = (G_1, \ldots, G_p)^\top$ be a centered Gaussian vector such that*

$$0 < \underline{\sigma} \leq \sqrt{\mathrm{Var}(G_j)} \leq \bar{\sigma} < \infty \qquad \forall j \in [p].$$

Then there exists a constant $C < \infty$, depending only on $\underline{\sigma}$ and $\bar{\sigma}$, such that for every $\epsilon > 0$,

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(|\max_{1\leq j\leq p} G_j - x| \leq \epsilon\right) \leq C\epsilon\left\{1 \vee \sqrt{\log(p/\epsilon)}\right\}.$$

In particular, if $\epsilon_n \downarrow 0$ and $\epsilon_n\sqrt{1 \vee \log(p/\epsilon_n)} \to 0$, then

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(|\max_{1\leq j\leq p} G_j - x| \leq \epsilon_n\right) \to 0.$$

The same conclusion holds with $\max_{1\leq j\leq p} |G_j|$ in place of $\max_{1\leq j\leq p} G_j$ after replacing p by $2p$.

Proposition A.19 (Chernozhukov, Chetverikov and Kato (2015), Theorem 2)**.** *Let $X = (X_1, \ldots, X_p)^\top$ and $Y = (Y_1, \ldots, Y_p)^\top$ be centered Gaussian vectors with covariance matrices Σ^X and Σ^Y. Assume that*

$$0 < \underline{\sigma} \leq \sqrt{\mathrm{Var}(X_j)} \vee \sqrt{\mathrm{Var}(Y_j)} \leq \bar{\sigma} < \infty \qquad \forall j \in [p],$$

and define

$$\Delta := \|\Sigma^X - \Sigma^Y\|_{\max}.$$

Then there exists a constant $C < \infty$, depending only on $\underline{\sigma}$ and $\bar{\sigma}$, such that

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(\max_{1\leq j\leq p} X_j \leq x\right) - \mathbb{P}\left(\max_{1\leq j\leq p} Y_j \leq x\right)\right| \leq C\Delta^{1/3}\{1 \vee \log(p/\Delta)\}^{2/3},$$

with the convention that the right-hand side equals 0 when $\Delta = 0$. In particular, if $\Delta_n = o((\log p)^{-2})$, then the Kolmogorov distance between the two Gaussian maxima converges to 0.

Proposition A.20 (Chang, Chen and Wu (2024), Theorem 1)**.** *Let $X_1, \ldots, X_n$ be centered random vectors in $\mathbb{R}^p$, and let*

$$S_n := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t, \qquad \Xi := \mathrm{Cov}(S_n), \qquad G \sim N(0, \Xi).$$

Assume that $\{X_t\}$ is geometrically α-mixing with

$$\alpha(\ell) \le K_1 e^{-K_2 \ell^{\gamma_{\mathrm{mix}}}} \qquad (\ell \ge 1)$$

for some constants $K_1 > 1$, $K_2 > 0$, and $\gamma_{\mathrm{mix}} > 0$, that $p \ge n^{\kappa_0}$ eventually for some $\kappa_0 > 0$, and that

$$\max_{t\in[n],\, j\in[p]} \|X_{t,j}\|_{\psi_2} \le B_n, \qquad \min_{1\le j\le p} \zeta_{jj} \ge b > 0.$$

If, in addition,

$$(\log p)^{3-\gamma_{\mathrm{mix}}} = o\big(n^{\gamma_{\mathrm{mix}}/3}\big)$$

and

$$B_n^{2/3} \frac{(\log p)^{(1+2\gamma_{\mathrm{mix}})/(3\gamma_{\mathrm{mix}})}}{n^{1/9}} + B_n \frac{(\log p)^{7/6}}{n^{1/9}} \to 0,$$

then

$$\sup_{u\in\mathbb{R}^p} \big|\mathbb{P}(S_n \le u) - \mathbb{P}(G \le u)\big| \to 0.$$

In particular,

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(\max_{1\le j\le p} S_{n,j} \le x\right) - \mathbb{P}\left(\max_{1\le j\le p} G_j \le x\right)\right| \to 0.$$

Proposition A.21 (Chang, Chen and Wu (2024), Theorem 11(i) with $\gamma_1 = 2$)**.** *Let $(X_j)_{j\ge1}$ be centered real-valued random variables, and define their strong-mixing coefficients by*

$$\alpha_X(\ell) := \sup_{k\ge1} \sup_{\substack{A\in\sigma(X_1,\dots,X_k),\\ B\in\sigma(X_{k+\ell},X_{k+\ell+1},\dots)}} \big|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\big|, \qquad \ell \ge 1.$$

Assume that, for some constant $\gamma_1 \ge 1$ and deterministic sequence $B_n \ge 0$,

$$\|X_{t,j}\|_{\psi_{\gamma_1}} \le B_n, \quad j = 1,\dots,\dim(X_j),$$

and, for some constant $K_1, K_2, \gamma_2 > 0$

$$\alpha_X(\ell) \le K_1 \exp\{-K_2 k^{\gamma_2}\}, \quad \ell \ge 1.$$

Let K be a kernel function that satisfies Assumption A.8 and

$$\bar{X} := \frac{1}{n}\sum_{t=1}^{n} X_t, \quad \widehat{\Xi}_n := \sum_{j=-n+1}^{n-1} K\left(\frac{j}{b_n}\right)\widehat{H}_j,$$

where

$$\widehat{H}_j := \frac{1}{n}\sum_{t=j+1}^{n} (X_t - \bar{X})(X_{t-j} - \bar{X})^\top \quad \text{if } j \ge 0,$$

$$\widehat{H}_j := \frac{1}{n}\sum_{t=-j+1}^{n} (X_{t+j} - \bar{X})(X_t - \bar{X})^\top \quad \text{if } j < 0,$$

and $\Xi := \mathrm{Cov}\big(n^{-1/2}\sum_{t=1}^{n} X_t\big)$. Assume $p > n^\kappa$ for some universal constant $\kappa > 0$ and $b_n \asymp n^\rho$. Then there

exist constants $c_1 > 0$ depending only on (ρ, ϑ) and $c_2 > 0$ depending only on $(\gamma_1, \gamma_2, \vartheta)$ such that

$$\|\widehat{\Xi}_n - \Xi\|_{\max} = O_p\{B_n^2 n^{-c_1} (\log p)^{c_2}\} + O\big(B_n^2 n^{-\rho}\big).$$

Proposition A.22 (Chang, Chen and Wu (2024), Theorem 10(i))**.** *Let $X_1, \ldots, X_n$ be centered random vectors in $\mathbb{R}^p$, and define*

$$S_{n,x} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t, \qquad \Xi := \mathrm{Cov}(S_{n,x}).$$

Let $G \sim N(0, \Xi)$. Given a random covariance estimator $\widehat{\Xi}_n$, let $\widehat{G}$ be a conditionally Gaussian random vector satisfying

$$\widehat{G} \mid X_1, \ldots, X_n \sim N(0, \widehat{\Xi}_n).$$

For a class $\mathscr{A}$ of Borel subsets of $\mathbb{R}^p$, define

$$\rho_n(\mathscr{A}) := \sup_{A \in \mathscr{A}} \big|\mathbb{P}(S_{n,x} \in A) - \mathbb{P}(G \in A)\big|, \quad \widehat{\rho}_n(\mathscr{A}) := \sup_{A \in \mathscr{A}} \big|\mathbb{P}(S_{n,x} \in A) - \mathbb{P}(\widehat{G} \in A \mid X_1, \ldots, X_n)\big|.$$

Let $\mathscr{A}^{\mathrm{re}}$ denote the class of all hyper-rectangles in $\mathbb{R}^p$, namely all sets of the form

$$A = \prod_{j=1}^{p} [a_j, b_j], \qquad -\infty \le a_j \le b_j \le \infty,$$

Assume that, for each $j \in [p]$, let

$$V_{n,j} := \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t,j}\right).$$

There exists a universal constant $K_3 > 0$ such that $\min_{j \in [p]} V_{n,j} \ge K_3$. Further, assume that $p \ge n^{\kappa}$ for some universal constant $\kappa > 0$. Then

$$\widehat{\rho}_n(\mathscr{A}^{\mathrm{re}}) \lesssim \rho_n(\mathscr{A}^{\mathrm{re}}) + \big\|\widehat{\Xi}_n - \Xi\big\|_{\max}^{1/3} (\log p)^{2/3},$$

where $\lesssim$ hides a universal positive constant.

Proposition A.23 (Merlevède, Peligrad and Rio (2011), Theorem 1 and Remark 1)**.** *Let $(X_j)_{j \ge 1}$ be centered real-valued random variables, and define their strong-mixing coefficients by*

$$\alpha_X(\ell) := \sup_{k \ge 1} \sup_{\substack{A \in \sigma(X_1, \ldots, X_k), \\ B \in \sigma(X_{k+\ell}, X_{k+\ell+1}, \ldots)}} \big|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)\big|, \qquad \ell \ge 1.$$

Assume that there exist constants $c > 0$, $b > 0$, $\gamma_1 > 0$, and $\gamma_2 \in (0, \infty]$ such that

$$\alpha_X(n) \le \exp\big(-c n^{\gamma_1}\big) \qquad \textit{for all } n \ge 1,$$

and

$$\sup_{j \ge 1} \mathbb{P}(|X_j| > t) \le \exp\Big(1 - (t/b)^{\gamma_2}\Big) \qquad \textit{for all } t > 0.$$

Let

$$\gamma := \left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2}\right)^{-1}, \qquad \varphi_M(x) := (x \wedge M) \vee (-M),$$

and assume $\gamma < 1$. *Define*

$$V := \sup_{M>0} \sup_{i\geq 1} \left(\mathrm{Var}\big(\varphi_M(X_i)\big) + 2 \sum_{j>i} \big|\mathrm{Cov}\big(\varphi_M(X_i), \varphi_M(X_j)\big)\big| \right). \tag{A.16}$$

Then there exist constants $C_0, C_1, C_2, \eta_0 \in (0,\infty)$, *depending only on* $(b, c, \gamma_1, \gamma_2)$, *such that for every* $n \geq 4$ *and every* $y \geq C_0(\log n)^{\eta_0}$,

$$\mathbb{P}\left(\max_{1\leq j\leq n} \left| \sum_{r=1}^{j} X_r \right| \geq y \right) \leq (n+1)\exp\left(-\frac{y^\gamma}{C_1}\right) + \exp\left(-\frac{y^2}{C_2 + C_2 n V}\right). \tag{A.17}$$

Proposition A.24 (Merlevède, Peligrad and Rio (2011), Remark 3)**.** *Consider the setting of Proposition* **??**. *For* $u \in (0,1]$, *define the marginal quantiles*

$$Q_{|X_i|}(u) := \inf\{t > 0 : \mathbb{P}(|X_i| > t) \leq u\}, \qquad Q(u) := \sup_{i\geq 1} Q_{|X_i|}(u),$$

and let G *be the inverse function of*

$$x \longmapsto \int_0^x Q(v)\, dv.$$

Then

$$V \leq \sup_{i\geq 1} \mathbb{E}[X_i^2] + 8\left(\sum_{k>0} \alpha_X(k)\right) \int_0^1 Q(u)^2\, du. \tag{A.18}$$

Orlicz ψ_2 and ψ_1 norms. Following Vershynin (2018), Definition 2.5.6 and Definition 2.7.5, define for a real-valued random variable X

$$\|X\|_{\psi_2} := \inf\big\{t > 0 :\ \mathbb{E}\exp(X^2/t^2) \leq 2\big\}, \qquad \|X\|_{\psi_1} := \inf\big\{t > 0 :\ \mathbb{E}\exp(|X|/t) \leq 2\big\}.$$

We call X *sub-Gaussian* if $\|X\|_{\psi_2} < \infty$ and *sub-exponential* if $\|X\|_{\psi_1} < \infty$. By Proposition 2.5.2 of Vershynin (2018), the ψ_2 norm is equivalent to the standard sub-Gaussian tail and moment conditions; by Proposition 2.7.1, the ψ_1 norm is equivalent to the standard sub-exponential tail and moment conditions.

Proposition A.25 (Vershynin (2018), Proposition 2.5.2, and Exercise 2.5.10))**.** *Let* $X_1, \ldots, X_m$ *be real-valued random variables satisfying*

$$\max_{1\leq j\leq m} \|X_j\|_{\psi_2} \leq K.$$

Then there exist universal constants $C, c \in (0,\infty)$ *such that*

$$\mathbb{E}\Big[\max_{1\leq j\leq m} |X_j|\Big] \leq CK\sqrt{1 \vee \log m}$$

and, for every $u \geq 0$,

$$\mathbb{P}\Big(\max_{1\leq j\leq m} |X_j| > CK\big(\sqrt{\log(2m)} + u\big)\Big) \leq 2e^{-cu^2}.$$

In particular,

$$\max_{1\le j\le m} |X_j| = O_p\big(K\sqrt{\log m}\big).$$

No independence is required.

Proposition A.26 (Vershynin (2018), Proposition 2.6.1)**.** *Let $X_1,\dots,X_m$ be independent centered real-valued random variables, and set*

$$K_j := \|X_j\|_{\psi_2}, \qquad V_2 := \left(\sum_{j=1}^m K_j^2\right)^{1/2}.$$

Then there exists a universal constant $C<\infty$ such that

$$\Big\|\sum_{j=1}^m X_j\Big\|_{\psi_2} \le C V_2.$$

More generally, for deterministic weights $a_1,\dots,a_m\in\mathbb{R}$,

$$\Big\|\sum_{j=1}^m a_j X_j\Big\|_{\psi_2} \le C\left(\sum_{j=1}^m a_j^2 K_j^2\right)^{1/2}.$$

Proposition A.27 (Vershynin (2018), Lemma 2.7.7)**.** *Let X and Y be sub-gaussian random variables. Then XY is sub-exponential. Moreover,*

$$\|XY\|_{\psi_1} \le \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

Proposition A.28 (Vershynin (2018), Proposition 2.7.1))**.** *Let $Y_1,\dots,Y_m$ be real-valued random variables satisfying*

$$\max_{1\le j\le m} \|Y_j\|_{\psi_1} \le K.$$

Then there exist universal constants $C,c\in(0,\infty)$ such that

$$\mathbb{E}\Big[\max_{1\le j\le m}|Y_j|\Big] \le CK(1\vee\log m)$$

and, for every $u\ge 0$,

$$\mathbb{P}\Big(\max_{1\le j\le m}|Y_j| > CK\big(\log(2m)+u\big)\Big) \le 2e^{-cu}.$$

In particular,

$$\max_{1\le j\le m}|Y_j| = O_p\big(K\log m\big).$$

Again, no independence is required.

B Appendix Tables

Table B.1: Main macro series by category

Variable	Description	Source	Tr.
Labour market			
EMP	Total employment, aged 16+	ONS	5
EMP_PART	Part-time employment	ONS	5
EMP_RATE	Employment rate, aged 16–64 (%)	ONS	2
EMP_ACT	Economically active, aged 16–64	ONS	5
EMP_ACT_RATE	Economic activity rate, aged 16–64 (%)	ONS	2
UNEMP_RATE	Unemployment rate, aged 16+ (%)	ONS	2
UNEMP_DURA_6mth	Unemployed up to 6 months	ONS	5
UNEMP_DURA_6-12mth	Unemployed 6–12 months	ONS	5
UNEMP_DURA_12mth+	Unemployed over 12 months	ONS	5
CLAIMS	Claimant count (thousands)	ONS	5
CLAIMS_RATE	Claimant count rate (%)	ONS	2
AVG_WEEK_HRS	Avg. actual weekly hours, all workers	ONS	5
AVG_WEEK_HRS_FULL	Avg. actual weekly hours, full-time workers	ONS	5
WAGE	Hourly earnings, manufacturing	FRED	5
Industrial output (Index of Production)			
IOP_PROD	Total production	ONS	5
IOP_CAP_GOOD	Capital goods	ONS	5
IOP_DUR	Consumer durables	ONS	5
IOP_ENER	Energy	ONS	5
IOP_INT_GOOD	Intermediate goods	ONS	5
IOP_MACH	Machinery & equipment	ONS	5
IOP_MANU	Manufacturing	ONS	5
IOP_MINE	Mining & quarrying	ONS	5
IOP_NON_DUR	Consumer non-durables	ONS	5
IOP_PETRO	Coke & refined petroleum	ONS	5
IOP_OIL_EXTRACT	Crude oil & natural gas extraction	ONS	5
Retail & consumption			
RSI	Retail sales index, volume (incl. fuel)	ONS	5
AVG_WEEK_RETAIL_SALE	Avg. weekly retail sales value, all	ONS	5
AVG_WEEK_RETAIL_SALE_NON_FOOD	Avg. weekly retail sales, non-food stores	ONS	5
Trade			
EXP_GOOD	Exports of goods, BOP (CVM)	ONS	5

Continued on next page

(continued)

Variable	Description	Source	Tr.
IMP_GOOD	Imports of goods, BOP (CVM)	ONS	5
Exchange rates			
GBP_BROAD	Broad effective exchange rate (Jan 2005=100)	BOE	5
GBP_CAN	GBP/CAD spot rate	BOE	5
GBP_JAP	GBP/JPY spot rate	BOE	5
GBP_US	GBP/USD spot rate	BOE	5
Commodities, energy & global prices			
OIL_PRICE	Brent crude oil price ($/barrel)	FRED	5
CRUDE_OIL_FUT	Brent crude oil futures (CO1)	Bloomberg	1
CRUDE_GAS_FUT	Natural gas futures (FN1)	Bloomberg	1
FOODFUT	Wheat futures index (INFAWHP)	Bloomberg	1
FUEL	UK fuel/gas cost measure	DESNZ	1
COMP_P	World Bank primary commodity price index	World Bank	1
GSCPI	Global supply chain pressure index	Iacoviello	1
TOTINF	OECD total CPI inflation	OECD	1
UKHPI	UK house price index	ONS	1
Monetary & credit			
M1	Money supply M1	BOE	5
M2	Money supply M2	BOE	5
M3	Money supply M3	BOE	5
M4	Money supply M4 (sterling, private sector)	BOE	5
BANK_RATE	Official Bank Rate	BOE	2
SONIA	Sterling overnight index average rate	FRED	2
BGS_5yrs_yld	5-year gilt nominal par yield	BOE	2
BGS_10yrs_yld	10-year gilt nominal par yield	BOE	2
MORT_FIXED_RATE_2YRS	2-year fixed mortgage rate, 75% LTV	BOE	2
MORT_FIXED_RATE_5YRS	5-year fixed mortgage rate, 75% LTV	BOE	2
TOT_HOUSE_APP	Total sterling house purchase approvals	BOE	5
Financial markets			
FTSE_ALL	FTSE All Share index	Yahoo	5
FTSE250	FTSE 250 index	Yahoo	5
VIX	CBOE Volatility Index	Yahoo	1
SP500	S&P 500	Yahoo	5
UK_focused_equity	iShares MSCI UK ETF (EWU)	Yahoo	5
Uncertainty			

Continued on next page

(continued)

Variable	Description	Source	Tr.
EUR_UNC_INDEX	Economic policy uncertainty index, Europe	FRED	2

Notes: This table lists the 60 baseline macroeconomic and financial predictors used in the forecasting exercises. The "Tr." column indicates the transformation applied to render each series stationary before estimation: 1 = levels; 2 = first differences; 4 = log levels; 5 = first differences of logs. All series are non-seasonally adjusted where available.

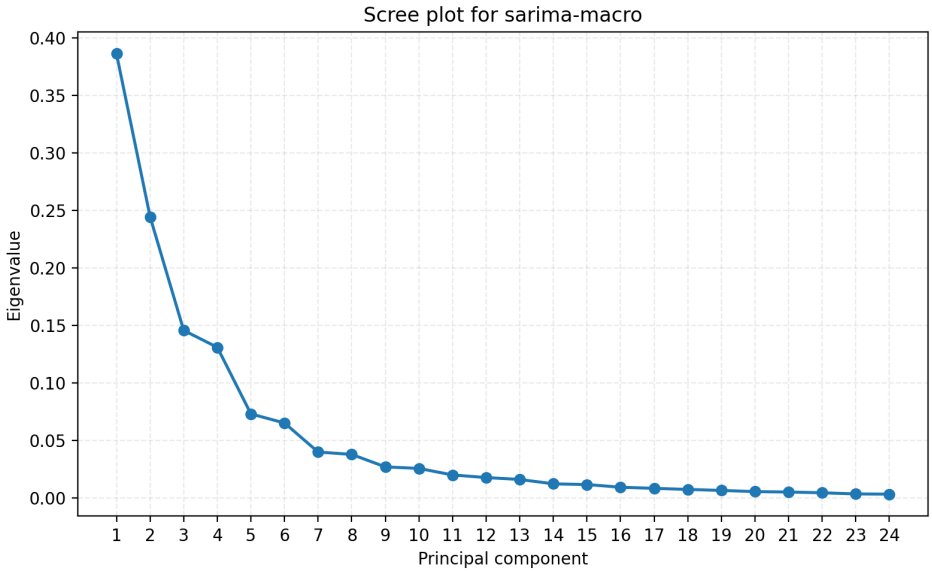
Table B.2: Moments of $|\Delta p|$ across sales treatments and literature (2% trim)

	Mean $\lvert\Delta p\rvert$	Std dev $\lvert\Delta p\rvert$	Excess kurtosis
This paper			
All prices, no trim	0.213	0.202	70.22
All prices (2% trim)	0.202	0.171	3.97
ONS sales dropped	0.139	0.126	5.01
NS Filter A	0.202	0.171	3.97
NS Filter B	0.202	0.171	3.97
Literature			
Klenow and Kryvtsov (2008)	0.113	–	–
Nakamura and Steinsson (2008)	0.085	–	–
Blanco et al. (2024)	0.110	–	4.02
Karadi and Reiff (2019)	0.099	–	3.98
Blanco et al. (2024) (UK)	0.153	0.205	3.81

Notes: Nakamura and Steinsson (2008) report the median $|\Delta p|$. Dropping ONS-flagged sales compresses the distribution markedly, consistent with sale price changes being larger than regular price changes. NS-filtered moments are identical to the all-prices baseline because imputation preserves the number of observations and replaces sales with the pre-sale price, leaving the tail-trimmed distribution nearly unchanged.

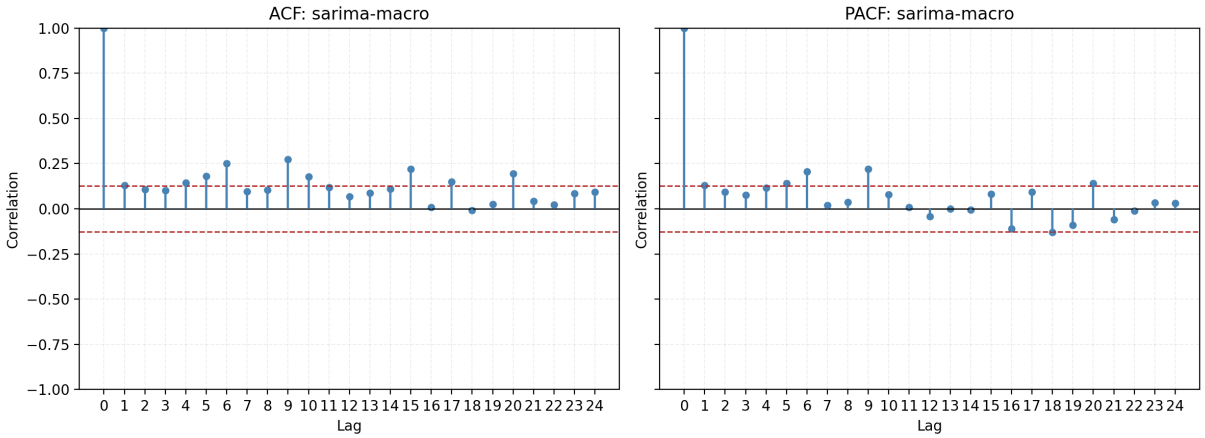
C Appendix Figures

Figure C.1: Scree plot: SARIMA–macro residual-difference panel



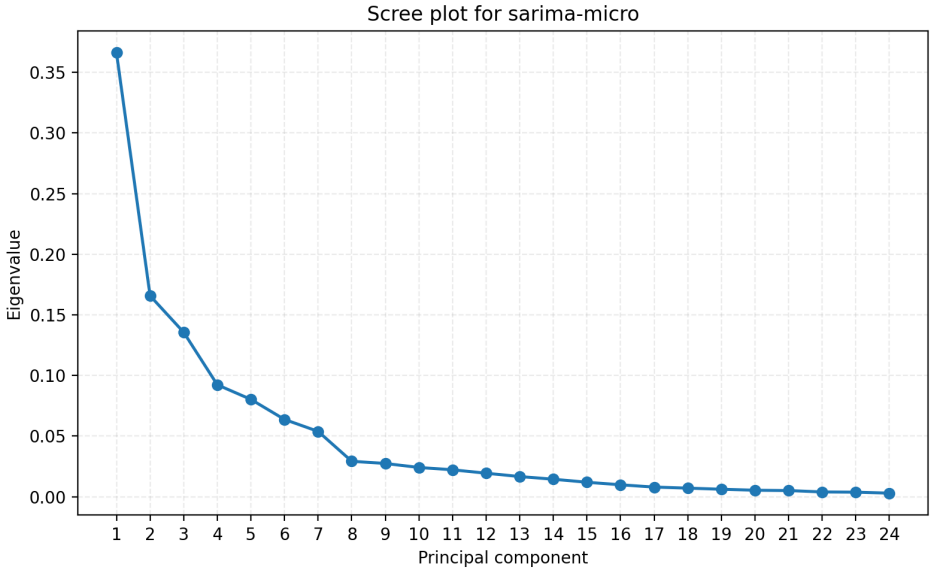
Notes: The line plots the eigenvalues of the sample covariance matrix of the absolute-error differential panel $r_{t,h} = |\pi_t - \hat{\pi}_t^{h,\text{SARIMA}}| - |\pi_t - \hat{\pi}_t^{h,\text{Macro}}|$, for $h = 1, \ldots, 24$.

Figure C.2: ACF and PACF: cross-horizon mean SARIMA–macro loss differential



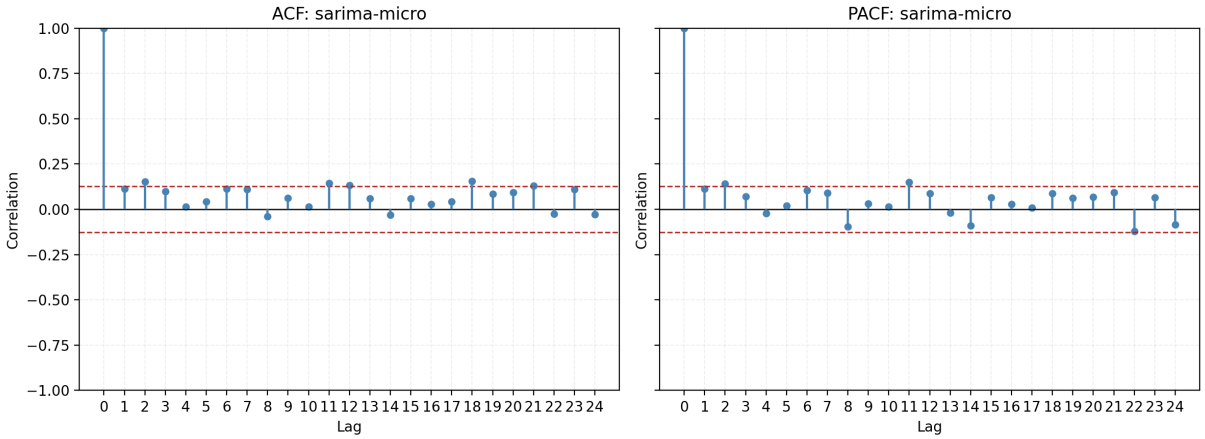
Notes: The figure plots the autocorrelation and partial autocorrelation functions of $\bar{r}_t = 24^{-1}\sum_{h=1}^{24}\left(|\pi_t - \hat{\pi}_t^{h,\text{SARIMA}}| - |\pi_t - \hat{\pi}_t^{h,\text{Macro}}|\right)$. Dashed lines show the approximate 95% confidence band $\pm 1.96/\sqrt{T}$ with $T = 239$.

Figure C.3: Scree plot: SARIMA–micro residual-difference panel



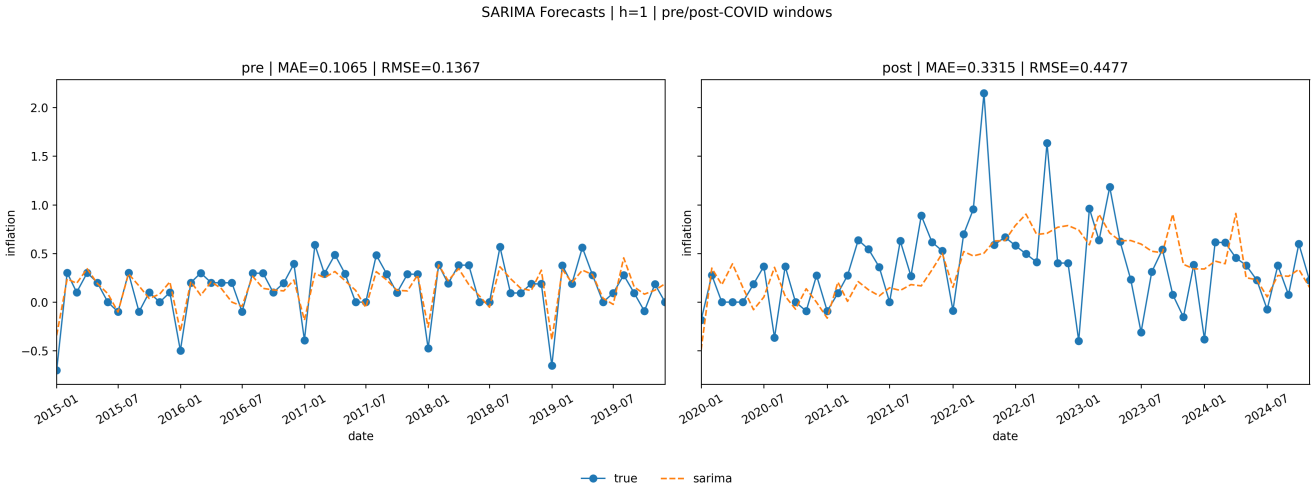
Notes: The line plots the eigenvalues of the sample covariance matrix of the absolute-error differential panel $r_{t,h} = |\pi_t - \hat{\pi}_t^{h,\text{SARIMA}}| - |\pi_t - \hat{\pi}_t^{h,\text{Micro}}|$, for $h = 1,\ldots,24$.

Figure C.4: ACF and PACF: cross-horizon mean SARIMA–micro loss differential



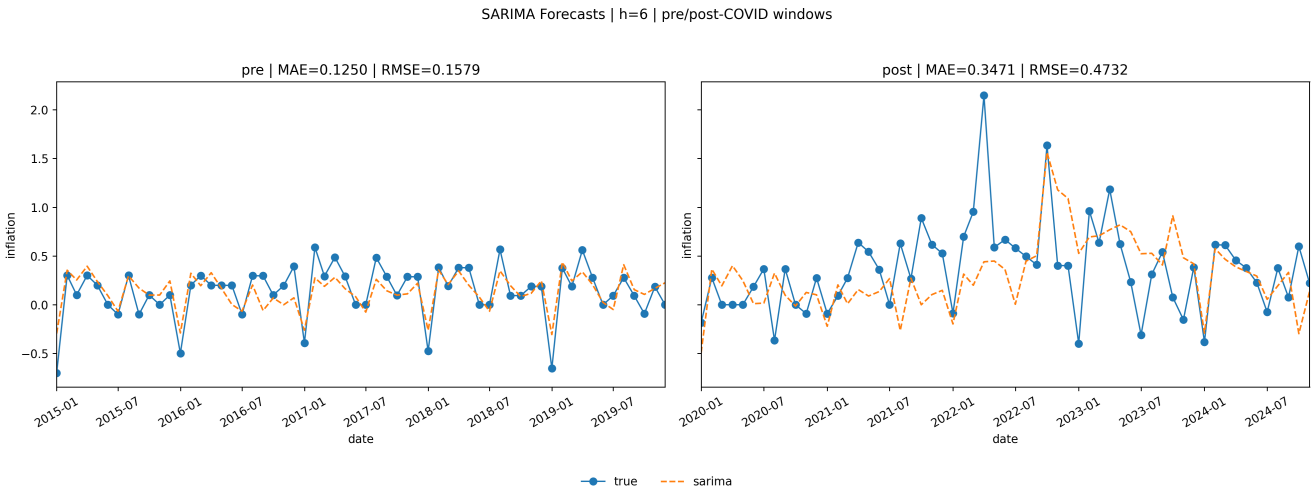
Notes: The figure plots the autocorrelation and partial autocorrelation functions of $\bar{r}_t = 24^{-1}\sum_{h=1}^{24}\left(|\pi_t - \hat{\pi}_t^{h,\text{SARIMA}}| - |\pi_t - \hat{\pi}_t^{h,\text{Micro}}|\right)$. Dashed lines show the approximate 95% confidence band $\pm 1.96/\sqrt{T}$ with $T = 239$.

Figure C.5: SARIMA forecasts vs. realized inflation: $h = 1$



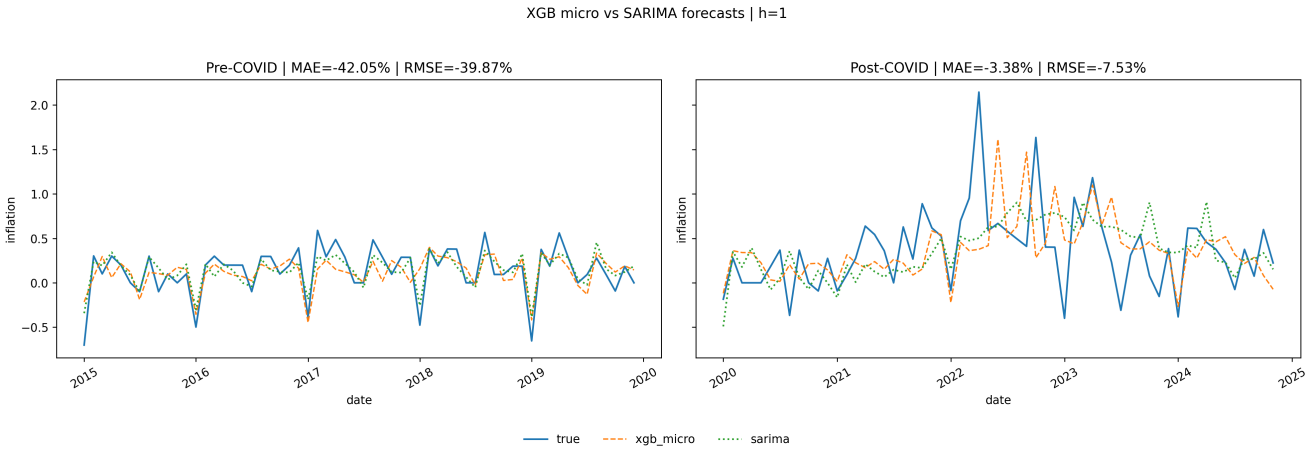
Notes: The solid line plots realized inflation and the dashed line plots the rolling-window SARIMA forecast. The two panels correspond to the 2015–2019 and 2020–2024 evaluation periods. Each panel title reports the SARIMA mean absolute error (MAE) and root mean squared error (RMSE) for that sub-period.

Figure C.6: SARIMA forecasts vs. realized inflation: $h = 6$



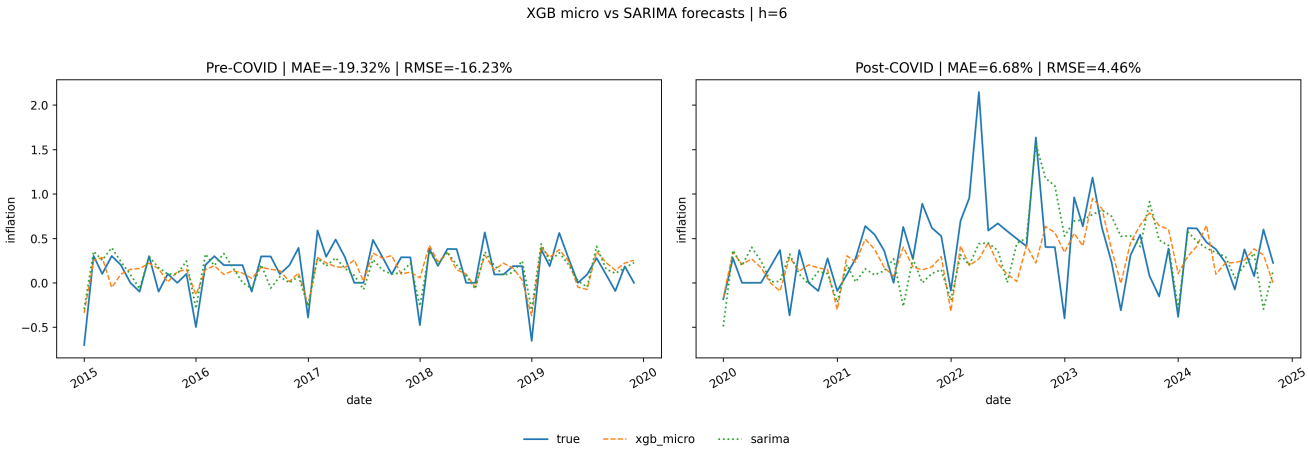
Notes: As in Figure C.5, but for the 6-step-ahead forecast.

Figure C.7: Micro XGBoost vs. SARIMA forecasts: $h = 1$



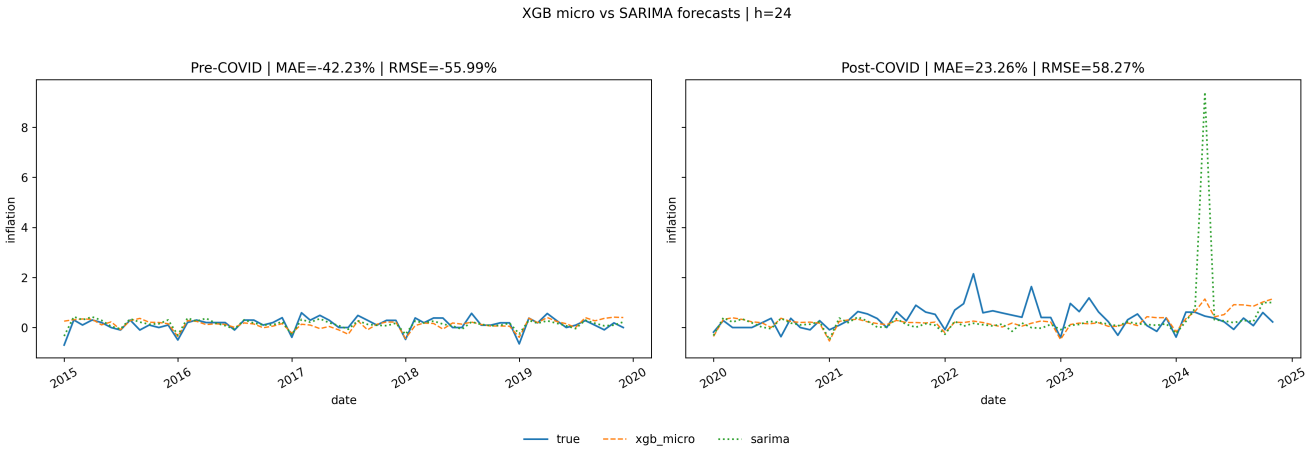
Notes: Each panel plots realized inflation together with the rolling-window XGBoost micro forecast and the rolling-window SARIMA forecast over one evaluation window. The two panels correspond to 2015–2019 and 2020–2024. Realized inflation is shown with a solid line, the XGBoost micro forecast with a dashed line, and the SARIMA forecast with a dotted line. Each panel title reports the percentage MAE and RMSE improvement of XGBoost micro relative to SARIMA in that window.

Figure C.8: Micro XGBoost vs. SARIMA forecasts: $h = 6$



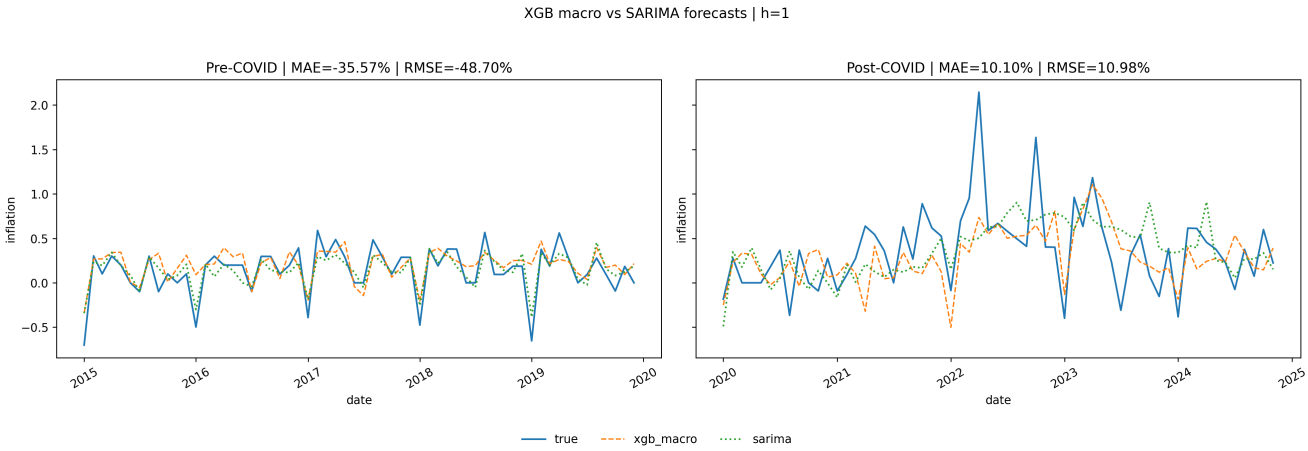
Notes: As in Figure C.7, but for the 6-step-ahead forecast.

Figure C.9: Micro XGBoost vs. SARIMA forecasts: $h = 24$



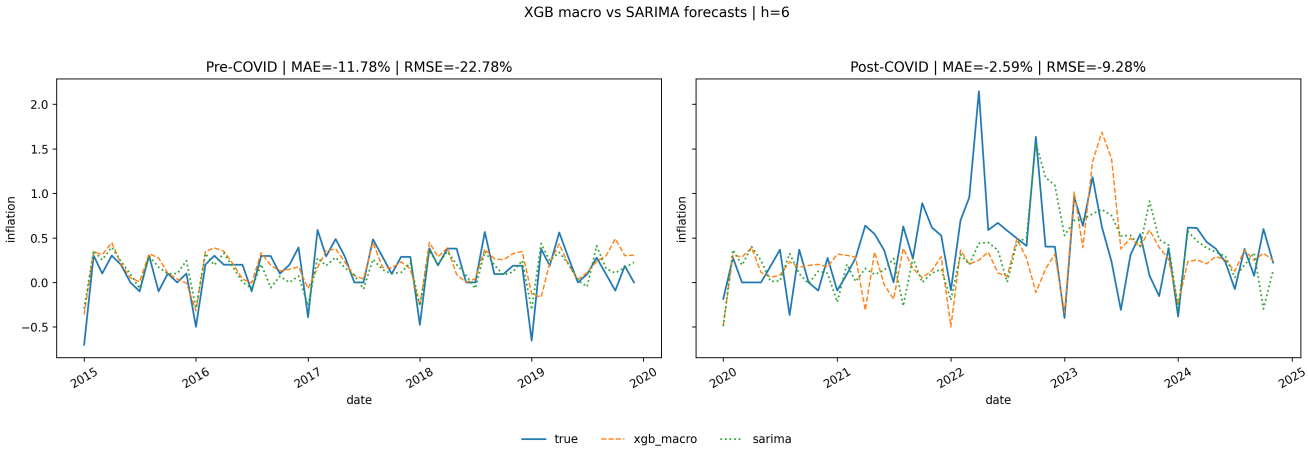
Notes: As in Figure C.7, but for the 24-step-ahead forecast.

Figure C.10: Macro XGBoost vs. SARIMA forecasts: $h = 1$



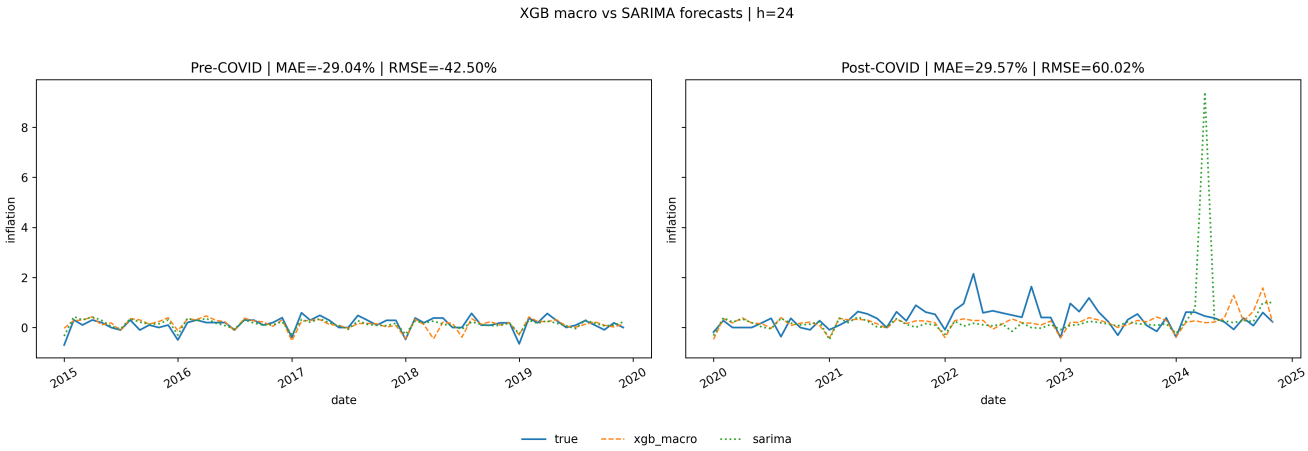
Notes: Each panel plots realized inflation together with the rolling-window XGBoost macro forecast and the rolling-window SARIMA forecast over one evaluation window. The two panels correspond to 2015–2019 and 2020–2024. Realized inflation is shown with a solid line, the XGBoost macro forecast with a dashed line, and the SARIMA forecast with a dotted line. Each panel title reports the percentage MAE and RMSE improvement of XGBoost macro relative to SARIMA in that window.

Figure C.11: Macro XGBoost vs. SARIMA forecasts: $h = 6$



Notes: As in Figure C.10, but for the 6-step-ahead forecast.

Figure C.12: Macro XGBoost vs. SARIMA forecasts: $h = 24$



Notes: As in Figure C.10, but for the 24-step-ahead forecast.

Figure C.13: Three-expert online-learning diagnostics: $h = 1$



Notes: The top row plots realized inflation and the online-learning aggregate forecast for the pre-2020 (2015–2019) and post-2020 (2020–2024) windows. The middle row plots the time-varying weights on Micro XGBoost, Macro XGBoost, and SARIMA. The bottom row plots the selected learning rate η_t under the neighbor-only search rule.

Figure C.14: Three-expert online-learning diagnostics: $h = 6$



Notes: As in Figure C.13, but for the 6-step-ahead horizon.

Figure C.15: Three-expert online-learning diagnostics: $h = 24$



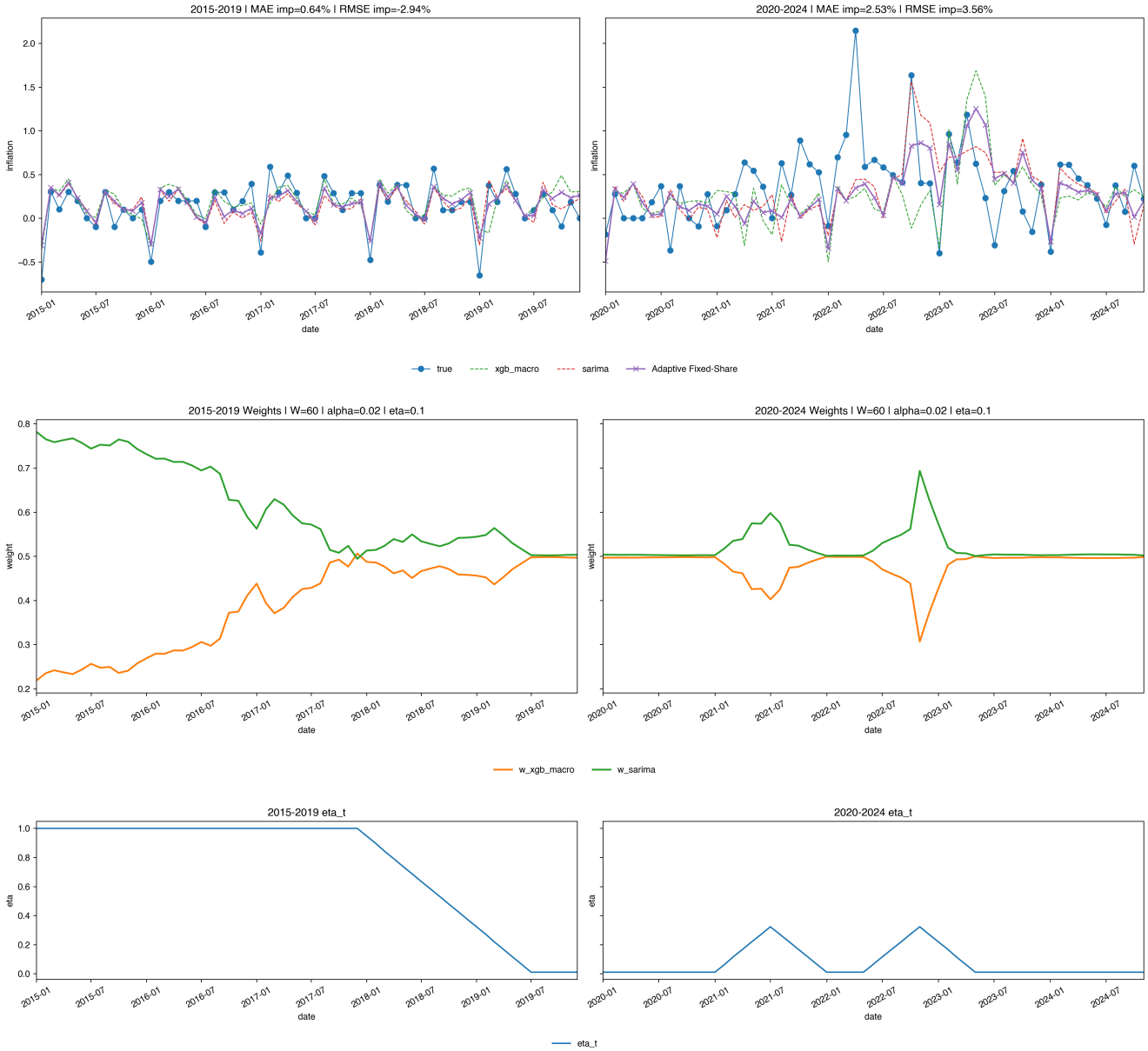
Notes: As in Figure C.13, but for the 24-step-ahead horizon.

Figure C.16: SARIMA + macro online-learning diagnostics: $h = 1$



Notes: The top row plots realized inflation and the online-learning aggregate forecast for the pre-2020 (2015–2019) and post-2020 (2020–2024) windows. The middle row plots the time-varying weights on Macro XGBoost, and SARIMA. The bottom row plots the selected learning rate η_t under the neighbor-only search rule.

Figure C.17: SARIMA + macro online-learning diagnostics: $h = 6$



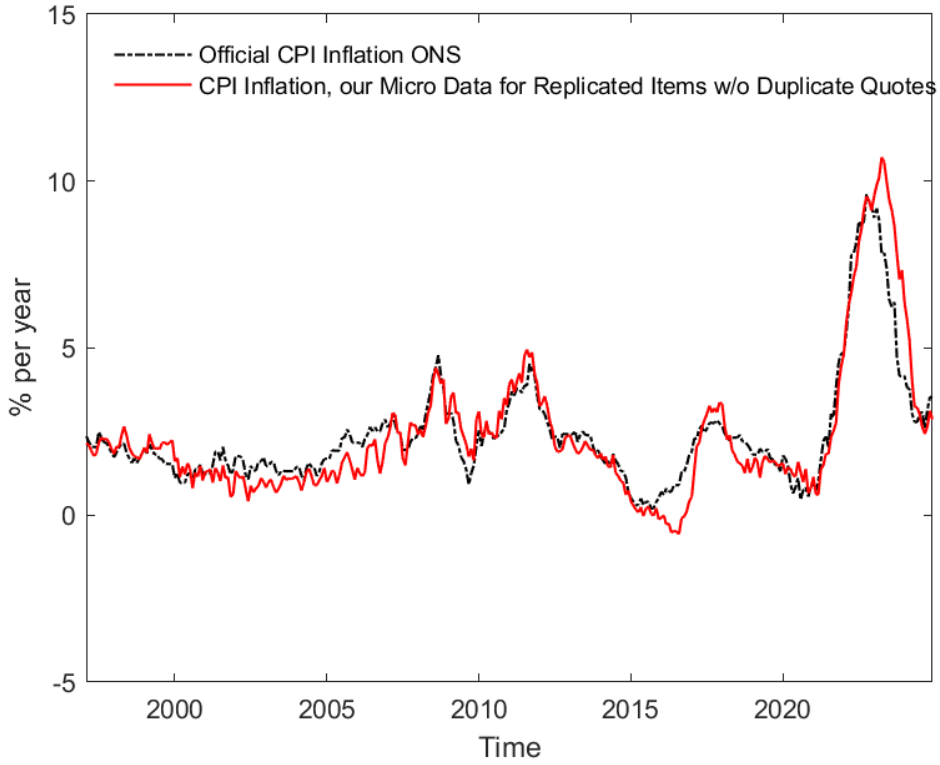
Notes: As in Figure C.16, but for the 6-step-ahead horizon.

Figure C.18: SARIMA + macro online-learning diagnostics: $h = 24$



Notes: As in Figure C.16, but for the 24-step-ahead horizon.

Figure C.19: Official ONS inflation and replicated CPI inflation: all posted prices



Notes: The figure compares year-on-year CPI inflation computed from the official ONS aggregate series with the inflation rate implied by our reconstructed chained index using the baseline micro-price panel with all posted prices.

D Data Appendix

This appendix gathers implementation details on data construction that are useful for replication but are not central to the main exposition. We organise the material into four subsections: cleaning of the ONS micro-price panel (Appendix D.1), treatment of temporary sales (Appendix D.2), aggregation and CPI replication (Appendix D.3), and benchmarking of the price-change distribution (Appendix D.4).

D.1 Price Quotes Data Cleaning

D.1.1 Cleaning

We follow Adam and Weber (2023) and apply the steps below in order.

1. *Validity filter.* Retain only observations whose current-price and base-price are valid according to the Office of National Statistics codes.

2. *Duplicate removal.* If a product identifier appears more than once in any given period, the entire series is dropped.

3. *Price rounding.* All prices are rounded to two decimal places.

4. *Implausible prices.* Zero prices, zero base prices, and prices below £0.001 are removed.

5. *Flag-based exclusions.* Non-comparable observations are dropped. The observation immediately preceding a substitution is treated as the end of the original price spell; the replacement begins a new series.

6. *Short-spell removal.* Rather than splitting product series at gaps, we store prices in a time × product matrix where missing months are represented as `NaN`; gaps therefore propagate naturally as missing observations. Single isolated observations (with `NaN` on both sides) are set to missing, and any product column with fewer than two non-missing observations is dropped.[8]

In total, these steps remove 0.24% of raw observations.

D.2 Treatment of Sales

The ONS S flag (sale) incompletely captures temporary price reductions. We therefore identify additional sales observations using algorithms from Nakamura and Steinsson (2008). The filter operates

[8] Adam and Weber (2023) instead split each product series into new sub-series at any gap exceeding one month. Their stricter rule removes 3.3% of price-quote observations and 23% of product series.

product by product on the price level series. It is governed by three integer parameters: a short rebound window JJ, a minimum number of distinct price points KK, and a longer look-ahead window LL. At each period t with observed price p_t, a *regular price* r_t is assigned via the following ordered rules:

1. **Continuity.** $p_t = r_{t-1}$: set $r_t = r_{t-1}$.
2. **Price increase.** $p_t > r_{t-1}$: set $r_t = p_t$.
3. **Drop with rebound.** $p_t < r_{t-1}$ and $p_{t+s} = r_{t-1}$ for some $s \le JJ$: classify period t as a sale, set $r_t = r_{t-1}$.
4. **Drop, no rebound, sufficient variation.** If the price does not rebound within JJ periods but the following LL observations contain at least KK distinct price values, treat the drop as permanent and set $r_t = p_t$.
5. **Drop, no rebound, modal-price rule.** If the maximum price in the LL-period window reappears later in the series, adopt it as the regular price.
6. **Default.** Set $r_t = p_t$.

The output is the adjusted series $\{r_t\}$ and a binary indicator equal to one when $r_t \neq p_t$. We identify sales using the following two parameterisations:

Table D.1: Nakamura–Steinsson filter parameters

	Rebound window JJ	Min. distinct prices KK	Look-ahead LL
Filter A (baseline)	3	3	3
Filter B (conservative)	1	1	3

D.3 Aggregation and CPI Replication

Aggregation proceeds in four steps.

D.3.1 Step 1 – Stratum index

Within each item i and stratum cell k, the price quotes are aggregated into a stratum-level price index using a shop-weighted geometric mean of log price relatives to the January base:

$$\text{stratum index}_{t,k} = 100 \times \exp\left(\frac{\sum_{j \in k} w_j^{\text{shop}} \cdot \log(p_{jt}/p_{j0})}{\sum_{j \in k} w_j^{\text{shop}}}\right), \tag{D.1}$$

where j runs over all products in stratum k, p_{j0} is the January base price, and w_j^{shop} is the shop weight. Shop weights are set to `NaN` wherever the corresponding price is missing, so the denominator always sums only over observed products. If a stratum cell code is zero (no sub-item stratification), a uniform stratum weight of 100 is applied.

D.3.2 Step 2 – Item index

Stratum indices are combined into the item index by a weighted arithmetic mean using stratum weights w_k^{stratum}:

$$\text{item index}_{t,i} = \frac{\sum_k w_{k,t}^{\text{stratum}} \cdot \text{stratum index}_{t,k}}{\sum_k w_{k,t}^{\text{stratum}}}. \tag{D.2}$$

Each price quote j receives an implicit product weight

$$\omega_j^t = \frac{w_j^{\text{shop}}}{\sum_{j' \in k} w_{j'}^{\text{shop}}} \times \frac{w_{k,t}^{\text{stratum}}}{\sum_{k'} w_{k',t}^{\text{stratum}}}, \tag{D.3}$$

which gives its proportional contribution to the item index.

D.3.3 Step 3 – Quality screen

Each reconstructed item index is compared to the corresponding official ONS item series over their joint sample. An item is deemed *suitable* if the RMSE of the log-level difference between the two series is below 2% and there are no gaps in the reconstructed series. Over the 1996–2016 sample, 1,091 items are suitable; only 2 fail the RMSE screen.

D.3.4 Step 4 – CPI aggregation and chaining

The reconstructed headline index is a weighted average of suitable item indices, with item weights normalised to sum to one in each period:

$$\text{CPI}_t^{\text{unchained}} = \sum_i \tilde{w}_{i,t}^{\text{item}} \cdot \text{item index}_{t,i}, \qquad \tilde{w}_{i,t}^{\text{item}} = \frac{w_{i,t}^{\text{item}}}{\sum_{i'} w_{i',t}^{\text{item}}}. \tag{D.4}$$

Because each annual index is expressed relative to the January base of that year, the unchained series restarts at 100 each January. The resulting chained index tracks the official ONS CPI closely over 1996–2016. The post-2016 extension is somewhat noisier, likely reflecting major ONS weight revisions introduced in 2017.

To assess the quality of the aggregation more transparently, we compare the inflation rate implied by our reconstructed CPI with the corresponding official ONS inflation series, for the baseline panel

with all posted prices.

Figure C.19 plots the corresponding year-on-year inflation series. The replicated series closely tracks the official ONS benchmark, indicating that the sequence of cleaning, weighting, item-index construction, and chaining reproduces the aggregate inflation dynamics well. This figure therefore serves as an important validation exercise for the data construction pipeline: it shows that the resulting micro-price panel remains consistent with the official CPI at the aggregate level.

D.4 Benchmarking the Price-Change Distribution

To validate the pipeline we compute moments of $|\Delta p_t|$, trimming the top and bottom 2% of observations and excluding zeros, following Blanco et al. (2024). These aggregate statistics are not used in estimation; they serve as a check that the cleaning and aggregation are correctly implemented. Table B.2 reports the mean, standard deviation, and excess kurtosis of $|\Delta p|$ across sales treatments, alongside estimates from the literature. Our estimates replicate the well-known stylised facts closely: price changes are infrequent and large when they occur, with moments broadly in line with previous studies of both US and UK micro-price data.

E Notation and Pseudocode

Table E.1 summarizes the main notation used in the forecasting, scan-test, and online-learning procedures described in Section 3.

Table E.1: Main notation used in the methods section

Symbol	Description
t	Calendar month index.
τ	Forecast origin.
h	Forecast horizon in months ahead.
H	Maximum forecast horizon considered in the scan test.
π_t	Realized inflation at time t.
X_t	Feature vector observed at time t; depending on the application, this contains either micro-distribution summaries or macro predictors.
Γ	XGBoost hyperparameter vector.
$\mathcal{G}$	Grid of candidate XGBoost hyperparameters.
$\hat{\pi}_\tau^h$	Forecast of inflation made at origin τ for horizon h.
$\tilde{\pi}_t^h$	Validation-period forecast used in rolling hyperparameter selection.
Υ	Candidate SARIMA order $(p,d,q)(P,D,Q,M)$.
Θ_τ^Υ	Estimated SARIMA parameter vector for order Υ using data available through time τ.
$\mathcal{L}(\cdot)$	SARIMA log-likelihood function.
ϵ_t	SARIMA innovation term.
$r_{t,h}$	Absolute-error differential between two competing forecasts at time t and horizon h.
$\bar{r}_h$	Within-horizon sample mean of the loss differential series, $\bar{r}_h = T^{-1}\sum_{t=1}^{T} r_{t,h}$; used in the autocorrelation pre-test.
$D_t^{(h)}$	Loss differential at time t and horizon h, defined as the difference in absolute forecast errors across two methods.
n	Number of evaluation periods used in the scan test.
$I = [s,e]$	Candidate contiguous time interval with start date s and end date e.

Continued on next page

Table E.1 continued

Symbol	Description
$\mathscr{I}$	Collection of candidate intervals considered by the scan test.
$L_{\min}$	Minimum admissible interval length in the scan test.
p	Total number of (interval, horizon) pairs over which the scan statistic searches, $p = H\lvert\mathscr{I}\rvert$.
$S_I^{(h)}$	Sum of loss differentials over interval I at horizon h.
$\bar{D}_I^{(h)}$	Mean loss differential over interval I at horizon h.
$\tilde{Q}_I^{(h)}$	Sum of squared deviations of $D_t^{(h)}$ within interval I at horizon h.
$T_I^{(h)}$	Self-normalized scan statistic for interval I and horizon h.
$T^{\max}$	Maximum scan statistic across all intervals and horizons.
$\xi_t^{(b)}$	Gaussian multiplier in bootstrap draw b for time t.
Θ	Covariance matrix of the Gaussian multiplier bootstrap. In Regime I, $\Theta = I_n$; in Regime D, it is kernel-based.
$K(\cdot)$	Kernel function used to construct the dependent-bootstrap covariance matrix in Regime D.
b_n	Bandwidth parameter used in the kernel-based covariance estimator.
B	Number of bootstrap draws.
K	Number of forecasting experts in the fixed-share combination step.
$\widehat{\pi}_t^{h,k}$	Forecast for horizon h at time t produced by expert k.
$\widetilde{\pi}_t^h$	Final combined forecast at time t and horizon h; in the adaptive variant, $\widetilde{\pi}_t^h = \widetilde{\pi}_t^{h,(\eta_t^\star)}$ (see Algorithm 7).
$w_t^{h,k}$	Weight placed on expert k at time t for horizon h; written $w_t^{h,(\eta),k}$ in the algorithms when the dependence on η is made explicit.
$\bar{w}_{t+1}^{h,(\eta),k}$	Post-multiplicative-weights, pre-sharing weight on expert k before the fixed-share mixing step.
η	Learning rate in the multiplicative-weights update.
α	Fixed-share mixing parameter.
$\mathscr{H}$	Grid of candidate learning-rate values for the adaptive fixed-share procedure.
W	Rolling-window length used to adaptively select the learning rate.
$\eta_t^\star$	Learning rate selected at time t by the adaptive fixed-share rule.

Continued on next page

Table E.1 continued

Symbol	Description
t_0	Start period for the fixed-share combination; in the baseline implementation, January 2005.
$\ell_t^{h,k}$	Absolute loss of expert k at time t for horizon h.

This appendix also collects the pseudocode for all algorithms used in Section 3.

Algorithm 1 Machine Learning (Rolling Window)

Input: Inflation $\{\pi_t\}_{t=1996/01}^{2025/01}$, features $\{X_t\}_{t=1997/01}^{2025/01}$, hyperparameter grid $\mathcal{G}$, horizon h.
Output: Forecasts $\{\hat{\pi}_\tau^h\}_{\tau=2000/01}^{2024/11}$ and forecast errors $\{\hat{\pi}_\tau^h - \pi_\tau\}_{\tau=2000/01}^{2024/11}$.
for $\tau = 2000/01, \ldots, 2024/11$ **do**
    for $\Gamma \in \mathcal{G}$ **do**
        for $l = 4, \ldots, 1$ **do**
            Train ML(Γ) on $\left\{(X_t, \pi_t)\right\}_{t=t_0}^{\tau-h-12l}$
                ▷ $t_0 = \max\{\tau - h - 12l - 119, \underline{t}\}$, where $\underline{t}$ is the earliest date available in $\{X_t\}$
            for $m = 1, \ldots, 12$ **do**
                Set $t' = \tau - h - 12l + m$; forecast $\tilde{\pi}_{t'}^h$ using fitted ML(Γ) and $X_{t'}$
            end for
        end for
        Compute RMSE$\left(\{\tilde{\pi}_t^h\}_{t=\tau-47-h}^{\tau-h}, \{\pi_t\}_{t=\tau-47-h}^{\tau-h}\right)$
    end for
    Select $\Gamma^* \in \mathcal{G}$ achieving the lowest validation RMSE
    Re-train ML(Γ^*) on $\left\{(X_t, \pi_t)\right\}_{t=\tau-h-119}^{\tau-h}$; forecast $\hat{\pi}_\tau^h$ using fitted ML(Γ^*) and X_τ
end for

Algorithm 2 SARIMA (Rolling Window)

Input: Inflation $\{\pi_t\}_{t=1985/01}^{2025/01}$, horizon h, parameter grid $\mathscr{T}$.
Output: Forecasts $\{\hat{\pi}_\tau^h\}_{\tau=1996/03}^{2024/11}$ and forecast errors $\{\hat{\pi}_\tau^h - \pi_\tau\}_{\tau=1996/03}^{2024/11}$.
for $\tau = 1996/03, \ldots, 2024/11$ **do**
    for $\Upsilon \in \mathscr{T}$ **do**
        for $l = 4, \ldots, 1$ **do**
            Train SARIMA(Υ)($\Theta^{\Upsilon}_{\tau-h-12l}$) on $\{\pi_t\}_{t=\tau-h-12l-119}^{\tau-h-12l}$
            for $m = 1, \ldots, 12$ **do**
                Forecast $\tilde{\pi}^h_{\tau-h-12l+m}$ via SARIMA(Υ)($\Theta^{\Upsilon}_{\tau-h-12l}$)
            end for
        end for
        Compute RMSE$\big(\{\tilde{\pi}_t^h\}_{t=\tau-47-h}^{\tau-h}, \{\pi_t\}_{t=\tau-47-h}^{\tau-h}\big)$
    end for
    Select $\Upsilon^* \in \mathscr{T}$ with lowest RMSE
    Train SARIMA(Υ^*)($\Theta^{\Upsilon^*}_{\tau-h}$) on $\{\pi_t\}_{t=\tau-h-119}^{\tau-h}$; forecast $\hat{\pi}_\tau^h$ via SARIMA(Υ^*)($\Theta^{\Upsilon^*}_{\tau-h}$)
end for

Algorithm 3 Multiplier-Bootstrap Ljung–Box Test (Aggregated over Horizons)

Input: Residual panel $\{r_{t,h}\}$, $t = 1,\dots,T$, $h = 1,\dots,H$, where $r_{t,h} = |\pi_t - \hat{\pi}_t^{h,A}| - |\pi_t - \hat{\pi}_t^{h,B}|$; maximum lag m; bootstrap draws B.

Output: p-value $\hat{p}$.

for $h = 1,\dots,H$, $k = 1,\dots,m$ **do**

  $\hat{\rho}_h(k) \leftarrow \dfrac{\sum_{t=k+1}^{T}(r_{t,h} - \bar{r}_h)(r_{t-k,h} - \bar{r}_h)}{\sum_{t=1}^{T}(r_{t,h} - \bar{r}_h)^2}$; $Q_h \leftarrow T(T+2)\sum_{k=1}^{m} \dfrac{\hat{\rho}_h(k)^2}{T-k}$

end for

$Q \leftarrow \sum_{h=1}^{H} Q_h$

for $b = 1,\dots,B$ **do**

  Draw $\xi_1^{(b)},\dots,\xi_T^{(b)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$; form $r_{t,h}^{*(b)} \leftarrow \xi_t^{(b)}\, r_{t,h}$ for all t, h

  for $h = 1,\dots,H$ **do**

    Compute $\hat{\rho}_h^{*(b)}(k)$ from $\{r_{t,h}^{*(b)}\}_{t=1}^{T}$; $Q_h^{*(b)} \leftarrow T(T+2)\sum_{k=1}^{m} \dfrac{\hat{\rho}_h^{*(b)}(k)^2}{T-k}$

  end for

  $Q^{*(b)} \leftarrow \sum_{h=1}^{H} Q_h^{*(b)}$

end for

$\hat{p} = \left(1 + \sum_{b=1}^{B} \mathbf{1}\{Q^{*(b)} > Q\}\right)/(B+1)$

Algorithm 4 Scan Test with Gaussian Multiplier Bootstrap (Regime I: $\Theta = I_n$)

Input: Residuals $\{\varepsilon_{t,A}^{(h)}, \varepsilon_{t,B}^{(h)}\}$, $t = 1, \dots, n$, $h = 1, \dots, H$; $L_{\min}$; bootstrap draws B.

Output: p-value $\hat{p}$.

for $t = 1, \dots, n$, $h = 1, \dots, H$ **do**

  $D_t^{(h)} \leftarrow |\varepsilon_{t,A}^{(h)}| - |\varepsilon_{t,B}^{(h)}|$

end for

$\mathscr{I} \leftarrow \{[s,e] \subset [n] : e - s + 1 \geq L_{\min}\}$; $T^{\max} \leftarrow -\infty$

for each $I = [s,e] \in \mathscr{I}$, $h = 1, \dots, H$ **do**

  $S_I^{(h)} \leftarrow \sum_{t=s}^{e} D_t^{(h)}$; $\bar{D}_I^{(h)} \leftarrow S_I^{(h)}/|I|$; $\tilde{Q}_I^{(h)} \leftarrow \sum_{t=s}^{e} (D_t^{(h)} - \bar{D}_I^{(h)})^2$

  $T_I^{(h)} \leftarrow S_I^{(h)} / \sqrt{\tilde{Q}_I^{(h)}}$ if $\tilde{Q}_I^{(h)} > 0$, else 0; $T^{\max} \leftarrow \max(T^{\max}, T_I^{(h)})$

end for

for $b = 1, \dots, B$ **do**

  Draw $\xi_1^{(b)}, \dots, \xi_n^{(b)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$; $T^{*,(b),\max} \leftarrow -\infty$

  for each $I = [s,e] \in \mathscr{I}$, $h = 1, \dots, H$ **do**

    $\bar{D}_{[s,e]}^{(h)} \leftarrow \frac{1}{e-s+1} \sum_{t=s}^{e} D_t^{(h)}$; $S_I^{(h)*} \leftarrow \sum_{t=s}^{e} \xi_t^{(b)} D_t^{(h)}$

    $T_I^{*,(h)} \leftarrow S_I^{(h)*} / \sqrt{\tilde{Q}_I^{(h)}}$ if $\tilde{Q}_I^{(h)} > 0$, else 0

    $T^{*,(b),\max} \leftarrow \max(T^{*,(b),\max}, T_I^{*,(h)})$

  end for

end for

$\hat{p} = \left(1 + \sum_{b=1}^{B} \mathbf{1}\{T^{*,(b),\max} > T^{\max}\}\right) / (B+1)$

Algorithm 5 Scan Test with Gaussian Multiplier Bootstrap (Regime D: kernel Θ)

Input: Residuals $\{\varepsilon_{t,A}^{(h)}, \varepsilon_{t,B}^{(h)}\}$, $t = 1, \dots, n$, $h = 1, \dots, H$; $L_{\min}$; bootstrap draws B; kernel K; bandwidth b_n.

Output: p-value $\hat{p}$.

for $t = 1, \dots, n$, $h = 1, \dots, H$ **do**

  $D_t^{(h)} \leftarrow |\varepsilon_{t,A}^{(h)}| - |\varepsilon_{t,B}^{(h)}|$

end for

Compute $\Theta \in \mathbb{R}^{n \times n}$ with $\Theta(t,u) = K\left(\dfrac{t-u}{b_n}\right)$ for all t, u

$\mathscr{I} \leftarrow \{[s,e] \subset [n] : e - s + 1 \geq L_{\min}\}$; $\quad T^{\max} \leftarrow -\infty$

for each $I = [s,e] \in \mathscr{I}$, $h = 1, \dots, H$ **do**

  $S_I^{(h)} \leftarrow \sum_{t=s}^{e} D_t^{(h)}$; $\quad \bar{D}_I^{(h)} \leftarrow S_I^{(h)}/|I|$; $\quad \tilde{Q}_I^{(h)} \leftarrow \sum_{t=s}^{e} (D_t^{(h)} - \bar{D}_I^{(h)})^2$

  $T_I^{(h)} \leftarrow S_I^{(h)} / \sqrt{\tilde{Q}_I^{(h)}}$ if $\tilde{Q}_I^{(h)} > 0$, else 0; $\quad T^{\max} \leftarrow \max(T^{\max}, T_I^{(h)})$

end for

for $b = 1, \dots, B$ **do**

  Draw $\xi^{(b)} = (\xi_1^{(b)}, \dots, \xi_n^{(b)})^\top \sim \mathcal{N}(0, \Theta)$; $\quad T^{*,(b),\max} \leftarrow -\infty$

  for each $I = [s,e] \in \mathscr{I}$, $h = 1, \dots, H$ **do**

    $\bar{D}_{[s,e]}^{(h)} \leftarrow \frac{1}{e-s+1} \sum_{t=s}^{e} D_t^{(h)}$; $\quad S_I^{(h)*} \leftarrow \sum_{t=s}^{e} \xi_t^{(b)} D_t^{(h)}$

    $T_I^{*,(h)} \leftarrow S_I^{(h)*} / \sqrt{\tilde{Q}_I^{(h)}}$ if $\tilde{Q}_I^{(h)} > 0$, else 0

    $T^{*,(b),\max} \leftarrow \max(T^{*,(b),\max}, T_I^{*,(h)})$

  end for

end for

$\hat{p} = \left(1 + \sum_{b=1}^{B} \mathbf{1}\{T^{*,(b),\max} > T^{\max}\}\right)/(B+1)$

Algorithm 6 FIXEDSHAREPATH(η, α, h)

Input: $\eta > 0$, $\alpha \in [0,1]$, horizon h; realized $\{\pi_t\}_{t=t_0}^{T}$; expert forecasts $\{\widehat{\pi}_t^{h,k}\}$ for $k = 1, \dots, K$.

Output: Aggregate forecasts $\{\widetilde{\pi}_t^{h,(\eta)}\}_{t=t_0}^{T}$ and weights $\{w_t^{h,(\eta),k}\}_{t=t_0}^{T+1}$.

Initialize $w_{t_0}^{h,(\eta),k} \leftarrow 1/K$ for all k.

for $t = t_0, \dots, T$ **do**

  $\widetilde{\pi}_t^{h,(\eta)} \leftarrow \sum_{k=1}^{K} w_t^{h,(\eta),k}\, \widehat{\pi}_t^{h,k}$

  Observe π_t; compute $\ell_t^{h,k} \leftarrow |\pi_t - \widehat{\pi}_t^{h,k}|$ for all k

  MW update: $\bar{w}_{t+1}^{h,(\eta),k} \leftarrow w_t^{h,(\eta),k} \exp(-\eta\, \ell_t^{h,k}) \,/\, \sum_j w_t^{h,(\eta),j} \exp(-\eta\, \ell_t^{h,j})$

  Fixed-share mix: $w_{t+1}^{h,(\eta),k} \leftarrow (1-\alpha)\, \bar{w}_{t+1}^{h,(\eta),k} + \alpha/K$ for all k

end for

Algorithm 7 Fixed Share with Rolling-Window Neighbor Selection of η

Input: $\alpha \in [0,1]$, $K \geq 1$, horizon h; $\{\pi_t\}_{t=t_0}^{T}$; $\{\widehat{\pi}_t^{h,k}\}$ for $k = 1,\ldots,K$; grid $\mathscr{H} = \{\eta^{(1)} < \cdots < \eta^{(G)}\}$; window length W; default η_0.
Output: Combined forecasts $\{\widetilde{\pi}_t^h\}_{t=t_0}^{T}$ and MAE over sub-periods.
for each $\eta \in \mathscr{H}$ **do**
    Run FIXEDSHAREPATH(η, α, h) to obtain $\{\widetilde{\pi}_t^{h,(\eta)}\}_{t=t_0}^{T}$
end for
for $t = t_0, \ldots, T$ **do**
    if $t < t_0 + W$ **then**
        $\eta_t^\star \leftarrow \eta_0$
    else
        Let g_{t-1} be the index s.t. $\eta_{t-1}^\star = \eta^{(g_{t-1})}$; $\quad \mathscr{I}_t \leftarrow \{\max(1, g_{t-1}-1),\, g_{t-1},\, \min(G, g_{t-1}+1)\}$
        $g_t \leftarrow \arg\min_{g \in \mathscr{I}_t} \frac{1}{W} \sum_{s=t-W}^{t-1} |\pi_s - \widetilde{\pi}_s^{h,(\eta^{(g)})}|$; $\quad \eta_t^\star \leftarrow \eta^{(g_t)}$
    end if
    $\widetilde{\pi}_t^h \leftarrow \widetilde{\pi}_t^{h,(\eta_t^\star)}$
end for