Empirical Economics (2026) 70:23
https://doi.org/10.1007/s00181-025-02879-x

Forecasting inflation with the hedged random forest

Elliot Beck[1] · **Michael Wolf**[2,3]

Received: 3 July 2025 / Accepted: 15 December 2025

Abstract
Accurate inflation forecasting is essential for economic policy, financial markets, and broader societal stability. In recent years, machine learning methods, particularly the random forest, have shown strong potential for improving forecasting accuracy, often outperforming traditional benchmarks. Building on this success, our paper adapts the hedged random forest (HRF) of Beck et al. (The hedged random forest, 2024) to the task of forecasting inflation. Unlike the standard random forest, the HRF assigns non-equal, and potentially negative, weights to individual trees to enhance forecasting performance. We develop customized estimators for the HRF's two key inputs: the mean and covariance matrix of the vector of forecast errors corresponding to individual trees. An extensive empirical analysis demonstrates that our approach consistently outperforms the standard random forest.

Keywords Exponentially weighted moving average · Linear shrinkage · Machine learning

JEL Classification C21 · C53 · E31 · E37 · E47

1 Introduction

Forecasting inflation is a crucial challenge for academics, policymakers, and industry actors, with profound implications for economic and social stability, investment decisions, and monetary policy. Over the past few decades, researchers have introduced a wide range of econometric models aimed at improving the accuracy of inflation forecasts; see Faust and Wright (2013) for a comprehensive review of the literature. Despite these efforts, and the clear benefits of more accurate inflation forecasts, achieving consistent improvements over simple univariate models, such as the naïve random-walk

✉ Michael Wolf
michael.wolf@econ.uzh.ch

1 Swiss National Bank, Zurich, Switzerland

2 Department of Economics, University of Zurich, Zurich, Switzerland

3 ADIA Lab, Abu Dhabi, United Arab Emirates

Published online: 22 January 2026

model of Atkeson and Ohanian (2001) and the unobserved-components model with stochastic volatility of Stock and Watson (2007), has proven notoriously difficult.

Recent advancements, however, have provided a new perspective on these challenges. In a series of influential papers, Medeiros and Mendes (2016), Garcia et al. (2017), and Medeiros et al. (2021) have demonstrated that machine learning (ML) methods, when applied to a large number of features consistently outperform traditional benchmark inflation-forecasting models in terms of forecasting accuracy.

Building on this insight, we focus on enhancing one specific machine learning method that has proven particularly effective for inflation forecasting: the random forest. This choice is motivated by the extensive and comprehensive empirical analysis of Medeiros et al. (2021) which evaluates a large number of competing proposals, including (i) traditional benchmark models, such as the random walk; (ii) shrinkage methods applied to linear regression, such as LASSO, Ridge regression, and elastic net; (iii) factor models; (iv) ensemble methods, such as bagging, complete subset regression, and jackknife model averaging; (v) random forests; and (vi) hybrid linear random forests. A backtest analysis applied to U.S. data establishes the random forest as the clear winner, "as it robustly delivers the smallest [forecast] errors".

The enhancement of the random forest we suggest is the hedged random forest (HRF) of Beck et al. (2024), adapted for the purpose of inflation forecasting. The random forest is an equal-weighted ensemble of tree-based forecasts. The hedged random forest starts with the same ensemble but uses non-equal weights instead of equal weights. The weights are designed with the aim of minimizing the mean-squared error of forecasts and are derived from two (estimated) inputs: the vector of means and the covariance matrix of the forecast errors corresponding to the individual trees.

Importantly, unlike the standard random forest and other variations that apply non-equal weights, such as those proposed by Winham et al. (2013), Pham and Olafsson (2020), and Chen et al. (2024), the method introduced by Beck et al. (2024) allows for negative weights. This feature is beneficial because, as noted by Goulet Coulombe et al. (2024), incorporating negative weights can improve the forecasting accuracy of random forests, a finding also confirmed by the empirical analysis of Beck et al. (2024).

This paper makes two major contributions to the literature. First, we propose specific estimators of the two inputs required by the hedged random forest that are customized for the task of forecasting inflation, where one is faced with time-series data. Second, we demonstrate that our approach consistently outperforms not only the standard random forest, as used in Medeiros et al. (2021), but also the hedged random forest using off-the-shell estimators of the two inputs, which works well with independent and identically distributed (i.i.d.) data.

The remainder of the paper is organized as follows. Section 2 presents the generics of the hedged random forest and discusses how to estimate relevant inputs for the purpose of forecasting inflation. Section 3 introduces the data used for the evaluation of our proposal. Section 4 presents the details of our forecast and backtest procedure. Section 5 presents the corresponding results. Section 6 presents concluding remarks. To streamline the presentation of the paper, certain mathematical results are relegated to an appendix that also contains robustness checks and data descriptions.

2 Methodology

This section describes how to adapt the hedged random forest for the purpose of forecasting inflation. The reader is referred to Beck et al. (2024) for a detailed motivation and description of the hedged random forest. Nevertheless, we aim to make this paper self-contained, at least for the initial reading.

2.1 The generics of the hedged random forest

The random forest (Breiman 2001) is one of the most popular tree-based methods for supervised machine learning. Depending on the nature of the response variable, categorical or numerical, the random forest can be used for either *classification* or *regression*. Since inflation is a numerical variable, the context of this paper is regression.

In the standard implementation of the random forest, one first grows an ensemble of (decorrelated) trees and then use the simple average of the trees, that is, the equal-weighted ensemble. Consequently, the individual forecasts of the trees are simply averaged to arrive at a final, combined forecast. For a textbook treatment on the random forest the reader is referred to Hastie et al. (2017, Chapter 15), for an example. The idea of the hedged random forest is to deviate from equal weighting with the aim of improving the forecasting accuracy.

In a general forecasting problem, the goal is to forecast (or predict) a variable $y \in \mathbb{R}$ on the basis of a set of variables $x \in \mathbb{R}^d$, which are also called regressors, attributes or features, and may contain lagged values of y. A generic forecast is denoted by $\hat{f}$. Then its mean-squared error (MSE) is given by $\mathrm{MSE}(\hat{f}) := \mathbb{E}\big(y - \hat{f}(x)\big)^2$. Letting

$$\mathrm{Bias}(\hat{f}) := \mathbb{E}\big(y - \hat{f}(x)\big) \quad \text{and} \quad \mathrm{Var}(\hat{f}) := \mathbb{V}\mathrm{ar}\big(y - \hat{f}(x)\big) = \mathbb{E}\big((y - \hat{f}(x)\big)^2 - \mathrm{Bias}^2(\hat{f}),$$

there exists the well-known decomposition

$$\mathrm{MSE}(\hat{f}) = \mathrm{Bias}^2(\hat{f}) + \mathrm{Var}(\hat{f}). \tag{2.1}$$

Random forests are based on a set (or ensemble) of p tree-based forecasting methods, denoted by $\{\mathcal{M}_j\}_{j=1}^p$. The standard random forest (RF) uses equal weighting, resulting in $\hat{f}_{\mathrm{EW}}(x) := \frac{1}{p}\sum_{j=1}^p \mathcal{M}_j(x)$, whereas a weighted random forest uses a general linear combination of the kind

$$\hat{f}_w(x) := \sum_{j=1}^p w_j \mathcal{M}_j(x) \quad \text{with} \quad w := (w_1, \ldots, w_p)' \quad \text{and} \quad \sum_{j=1}^p w_j = 1. \tag{2.2}$$

Denote by $e_j := y - \mathcal{M}_j(x)$ the forecast error corresponding to tree $\mathcal{M}_j$; these errors are collected into the vector $e := (e_1, \ldots, e_p)'$ with expectation (vector) $\mu :=$

$\mathbb{E}(e)$ and covariance matrix $\Sigma := \mathbb{V}\text{ar}(e)$. The MSE of the forecast (2.2) is then given by

$$\text{MSE}(\hat{f}_w) = (w'\mu)^2 + w'\Sigma w \ .$$

Therefore, the optimal forecast in terms of the MSE in the class (2.2) is the solution to the following optimization problem:

$$\min_w \ (w'\mu)^2 + w'\Sigma w \tag{2.3}$$

$$\text{s.t.} \quad w'\mathbb{1} = 1 \ , \tag{2.4}$$

where $\mathbb{1}$ denotes a conformable vector of ones.

The problem in practice is that the inputs μ and Σ are unknown. A feasible solution is to replace them with sample-based estimates $\hat{\mu}$ and $\hat{\Sigma}$.

Being agnostic, for the time being, about the nature of the estimators $\hat{\mu}$ and $\hat{\Sigma}$, we then solve the feasible optimization problem

$$\min_w \ (w'\hat{\mu})^2 + w'\hat{\Sigma} w \tag{2.5}$$

$$\text{s.t.} \quad w'\mathbb{1} = 1 \quad \text{and} \tag{2.6}$$

$$||w||_1 \leq \kappa \ , \tag{2.7}$$

where $||w||_1 := \sum_{j=1}^{p} |w_j|$ denotes the L_1 norm of w and $\kappa \in [1, \infty]$ is a constant chosen by the user. We shall denote the solution to this optimization problem by $\hat{w}$.

The addition of the constraint (2.7) is motivated by the related problem of portfolio selection in finance, in which context the constraint is called a "gross-exposure constraint". This constraint protects (or hedges) the user against extreme "positions", that is, against weights $\hat{w}_j$ that are unduly large in absolute value. Therefore, Beck et al. (2024) have coined this method the "hedged random forest" (HRF). In the extreme case $\kappa = 1$, all weights $\hat{w}_j$ must be nonnegative. However, allowing for negative weights turns out to be beneficial and Beck et al. (2024) recommend the default choice $\kappa = 2$.

2.2 Estimation of inputs

The two key inputs of the HRF are $\hat{\mu}$ and $\hat{\Sigma}$, estimates of the mean μ and the covariance matrix Σ of the vector of forecast errors e. One trains a random forest consisting of p trees on a training dataset of size n. After training, one extracts the forecasts of each tree $\mathcal{M}_j$ on the entire training set and thus obtains an in-sample error (or residual) matrix R of size $n \times p$.

Under the assumption of a strictly stationary time series, it might seem appealing to use the (column-wise) sample mean of R as $\hat{\mu}$ and the sample covariance matrix computed from R as $\hat{\Sigma}$. A feature of these sample estimators is that they give equal weight $1/n$ to all observations (that is, rows) of the residual matrix R, since they are invariant to row-wise permutations of R. Instead, we feel that for the task of forecasting inflation it is more useful to give higher weights to recent observations relative to more distant (in the past) observations. Arguably, real-life economic time series are often not

strictly stationary due to time-varying economic conditions, such as business cycles, recessions, and structural breaks. However, even if time series are strictly stationary, it can still be beneficial to give higher weights to recent observations in the case of time-varying *conditional* (co)variances, such as ARCH-GARCH effects. For example, for the task of estimating the covariance matrix of a vector of financial returns on the basis of daily (or at least weekly) data, many people use multivariate GARCH models, which give higher weights to recent observations.

Another estimation scheme that gives higher weights to recent observations is exponentially weighted moving average (EWMA) estimation, alternatively called exponential smoothing; for example, see Longerstaey and Zangari (1996) and Chatfield and Xing (2019, Chapter 5). The dimension p of μ and Σ is large in our context, given that one typically uses $p = 500$ or even $p = 1000$ trees in random forests. Therefore, we combine EWMA estimation with linear shrinkage to arrive at well-conditioned estimators; the corresponding details are described in Appendix D. EWMA estimation requires the choice of a model parameter $\lambda \in (0, 1)$ which determines the speed at which the weights on the observations decrease as one moves from the recent past to the more distant past; the larger λ, the faster the weights decrease. As the default choice for daily data, Longerstaey and Zangari (1996) recommend $\lambda = 0.06$. Since our data are monthly rather than daily, we opt for the larger default choice $\lambda = 0.15$; robustness checks regarding this choice can be found in Appendix B.

Notably, unlike when one enjoys the "comfort" of i.i.d. data where "one recipe fits all", when one is faced with time-series data, the estimation of μ and Σ generally requires a "case-by-case approach". For i.i.d. data, Beck et al. (2024) recommend using the sample mean of R as the estimator of μ and to use the QIS nonlinear shrinkage estimator of Ledoit and Wolf (2022) computed from R as the estimator of Σ; no further thinking is required. However, when one is faced with time-series data, some choices must be made, depending on the nature, frequency, and sample size of the data.

3 Data

We follow Medeiros et al. (2021) and use data from the FRED-MD database, which is designed for the empirical analysis of "big data". This dataset is updated in real time through the FRED database and can be accessed via Michael McCracken's webpage.[1] For detailed information, we refer the reader to McCracken and Ng (2016).

We utilize the vintage as of March 2025. Our sample period ranges from January 1960 through March 2025, comprising 783 monthly observations. From a total of 127 available variables, we include only the 111 variables that have a complete history for the entire period. In addition, the feature set includes four autoregressive lags of the target variable, the first four principal components extracted from the selected variables, and four lags of each selected variable. As a result, the dataset used in this study contains 464 features. Appendix A provides details on the data transformations that were used for any of the variables.

[1] See https://research.stlouisfed.org/econ/mccracken/fred-databases/.

It should be pointed out that we use a single vintage of the FRED-MD database, just as Medeiros et al. (2021) do in their main empirical analysis. As an alternative, one can carry out a "real-time analysis" with monthly vintages. Medeiros et al. (2021, Section 5.4) provide such an analysis as a robustness check and find that their main results do not qualitatively change. Therefore, in the interests of simplicity and easier reproducibility, we opt for the single-vintage analysis.

In contrast to Medeiros et al. (2021), who define inflation as log-differences of corresponding price indices, we define inflation as year-over-year (YoY) percentage changes. Specifically, inflation, π_t, is computed as $\pi_t := P_t/P_{t-12} - 1$, where P_t denotes the price index at time t. This approach aligns with the standard practice of central banks and policymakers, who typically use year-over-year inflation forecasts to guide monetary policy decisions.[2]

We consider six price indices, which are then converted to inflation measures, all sourced from the FRED Economic Data platform.[3] The corresponding FRED IDs and descriptions are as follows:

- **CPIAUCSL** (Consumer Price Index (CPI) for All Urban Consumers: All Items, USA): The headline CPI for urban consumers in the USA as reported by the U.S. Bureau of Labor Statistics.
- **CPILFESL** (Consumer Price Index for All Urban Consumers: All Items Less Food and Energy, USA): Commonly referred to as "core CPI", this measure excludes volatile food and energy components.
- **PCEPI** (Personal Consumption Expenditures, USA): Headline PCE is the Federal Reserve's preferred measure of inflation and is produced by the U.S. Bureau of Economic Analysis.
- **PCEPILFE** (Personal Consumption Expenditures Excluding Food and Energy, USA): Similar to the core CPI, this series excludes food and energy prices, and is often referred to the "core PCE".
- **CHECPIALLMINMEI** (Consumer Price Index: All Items, Switzerland): The headline CPI for Switzerland as produced by the Swiss Federal Statistical Office (SFSO). For simplicity, we refer to this ID as CHECPIALL throughout the remainder of this study.
- **CHECPICORMINMEI** (Consumer Price Index: All Items Less Food and Energy, Switzerland): Commonly referred to as Switzerland's "core CPI", it excludes volatile food and energy prices. For simplicity, we refer to this ID as CHECPICOR throughout the remainder of this study.

Figure 1 displays the evolution of the six inflation measures over the sample period. For either country, headline inflation is notably more volatile and reacts more strongly during periods of financial turmoil than does core inflation. Whereas the US headline and core measures exhibit broadly similar trends, the Swiss inflation measures show

[2] For definitions of inflation, we refer the reader to the explanation by the Federal Reserve at https://www.federalreserve.gov/economy-at-a-glance-inflation-pce.htm and the overview of the monetary policy strategy of the Swiss National Bank at https://www.snb.ch/en/the-snb/mandates-goals/monetary-policy/strategy.

[3] See https://fred.stlouisfed.org.

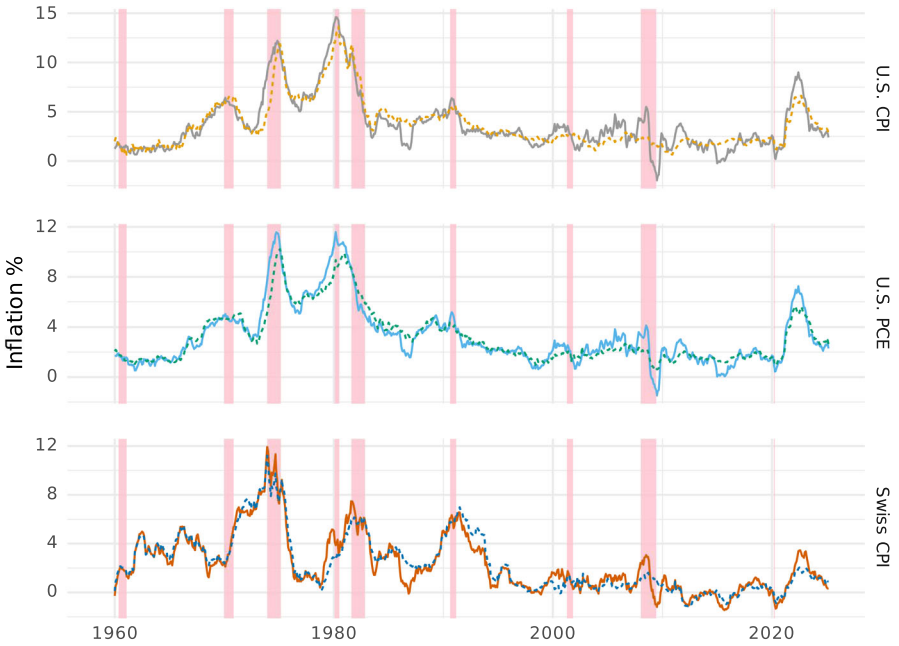

Fig. 1 The figure displays the U.S. CPI, U.S. PCE, and Swiss CPI inflation measures from January 1960 through March 2025 (783 observations). Headline inflation measures are displayed with solid lines, whereas core inflation measures are displayed with dashed lines. Inflation is calculated as $\pi_t := P_t / P_{t-12} - 1$, where P_t denotes a given price measure. The shaded areas indicate periods of economic recession, as defined by the NBER (https://www.nber.org/research/business-cycle-dating)

distinct patterns, influenced by Switzerland's openness to international markets and its sensitivity to exchange-rate movements.

4 Backtesting procedure

This section describes the backtesting procedure, evaluates two approaches to forecasting year-over-year inflation, and addresses the selection of hyperparameters along with other training considerations.

4.1 Framework

To evaluate the performance of the forecasting methods, we employ a backtesting framework based on the methodology of Medeiros et al. (2021). Consequently, our approach also employs the "direct" forecasting strategy, meaning that we build a separate model for each forecast horizon h. The forecasts are computed using a rolling-window approach with a fixed length of 360 observations. However, owing to the direct forecasting approach and the inclusion of lagged variables, the effective number of

training observations is somewhat less than 360. Specifically, the training sample size for horizon h is $R_h := 360 - h - l$, where l denotes the number of lags included.

Following Medeiros et al. (2021), we start our out-of-sample evaluation in January 1990. However, with access to more recent data, we extend the evaluation through March 2025, thus covering a longer period. We consider forecasts for horizons ranging from $h = 1$ to $h = 12$ months. For each combination of month t, forecasting method m, and horizon h, we compute inflation forecasts denoted by $\hat{\pi}_{t,m,h}$. The corresponding forecast errors are $\hat{e}_{t,m,h} := \hat{\pi}_{t,m,h} - \pi_t$. Section 4.3 provides relevant implementation details for the hedged random forest.

We evaluate the accuracy of the forecasts using two standard performance criteria, root-mean-squared error (RMSE) and mean absolute error (MAE), defined as

$$\text{RMSE}_{m,h} := \sqrt{\frac{1}{N}\sum_{t=1}^{N} \hat{e}_{t,m,h}^2} \quad \text{and} \quad \text{MAE}_{m,h} := \frac{1}{N}\sum_{t=1}^{N} |\hat{e}_{t,m,h}| \,,$$

where N denotes the number of out-of-sample observations.

4.2 Forecasting inflation: two alternative approaches

Following Goulet Coulombe et al. (2021), we consider two approaches to forecasting year-over-year inflation: (i) a "one-shot" forecast of year-over-year inflation and (ii) forecasting individual month-over-month (MoM) price changes, which are then aggregated into a forecast of year-over-year inflation.

4.2.1 A single forecast: the one-shot forecasting approach

A straightforward approach to forecasting year-over-year inflation is to use a single model with year-over-year inflation as the target variable for each forecast horizon; we call this the "one-shot approach". To this end, we consider the following model:

$$\pi_{t+h} := f_h(x_t) + u_{t+h} \,, \tag{4.1}$$

where π_{t+h} denotes inflation at month $t + h$ and x_t is an d-dimensional vector of features, including lags of π_t and other predictors. The mapping $f_h(\cdot)$ relates the features to future inflation, and u_{t+h} is a zero-mean random error term. As discussed in Sect. 4.1, we employ a direct forecasting strategy, meaning a separate mapping $f_h(x_t)$ is estimated for each forecast horizon h. The forecasting equation for each time t and horizon h is:

$$\hat{\pi}_{t+h,YoY|t} = \hat{f}_{h,t-R_h:t}(x_t) \,, \tag{4.2}$$

where $\hat{f}_{h,t-R_h:t}$ denotes the estimated mapping using data from $t - R_h$ through t, with $R_h := 360 - h - l$.

4.2.2 Forecast aggregation: the path-average approach

Alternatively, one can forecast individual month-over-month price changes and then aggregate these forecasts into a forecast of year-over-year inflation; Goulet Coulombe et al. (2021) call this the "path-average approach".[4]

A generic month-over-month percentage change in prices is defined as

$$\Delta P_t = \frac{P_t}{P_{t-1}} - 1 \, . \tag{4.3}$$

To the end of forecasting such a change, we consider the following model:

$$\Delta P_{t+h} := g_h(x_t) + v_{t+h} \, , \tag{4.4}$$

where $g_h(\cdot)$ is the mapping between features x_t and future MoM price changes and v_t is a zero-mean random error term. The forecasting equation for this approach is as follows:

$$\widehat{\Delta P}_{t+h|t} = \hat{g}_{h,t-R_h:t}(x_t) \, . \tag{4.5}$$

The path-average approach first forecasts the price level after h months as

$$\hat{P}_{t+h} = P_t \prod_{j=1}^{h} \left(1 + \widehat{\Delta P}_{t+j|t}\right) \tag{4.6}$$

and then aggregates these forecasts into

$$\hat{\pi}_{t+h,MoM|t} = \frac{\hat{P}_{t+h} - P_{t+h-12}}{P_{t+h-12}} \, . \tag{4.7}$$

4.2.3 Comparison of the two forecasting approaches

We assess the relative performance of the two forecasting approaches using the backtesting framework introduced in Sect. 4.1. For this comparison, we employ the hedged random forest outlined in Sect. 2 as the estimator for forecasting Eqs. (4.2) and (4.5).

Table 1 summarizes the backtest results. Panel (a) presents the RMSE values, whereas Panel (b) presents the MAE values for the six inflation measures across forecast horizons of 1, 3, 6, 9, and 12 months, along with the mean across all twelve horizons (1–12). It can be seen that, on balance, the path-average approach clearly outperforms the one-shot approach in terms of both the RMSE and the MAE, confirming the findings of Goulet Coulombe et al. (2021). Consequently, we will adopt the path-average forecasting approach for the remainder of the analysis.

[4] We take the liberty of using a hyphen in "path-average", which differs from the original syntax.

Table 1 Comparison of one-shot and path-average forecasting approaches for six inflation measures in the U.S. and Switzerland, evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL						
one-shot	0.008	0.011	0.013	0.015	0.017	0.013
path-average	0.003	0.006	0.01	0.014	0.018	0.011
CPILFESL						
one-shot	0.004	0.006	0.008	0.010	0.011	0.008
path-average	0.001	0.003	0.005	0.008	0.011	0.006
PCEPI						
one-shot	0.006	0.009	0.011	0.013	0.014	0.011
path-average	0.002	0.005	0.008	0.011	0.014	0.008
PCEPILFE						
one-shot	0.003	0.005	0.007	0.008	0.009	0.007
path-average	0.001	0.003	0.005	0.007	0.010	0.005
CHECPIALL						
one-shot	0.005	0.008	0.012	0.014	0.014	0.011
path-average	0.003	0.006	0.009	0.012	0.014	0.009

4.3 Implementation details

Estimating forecasting Eq. (4.5) using the hedged random forest involves choosing hyperparameters for the standard random forest, as well as choosing certain model parameters for estimating μ and Σ described in Sects. 2.2 and Appendix D.

The random forest is trained using the "ranger" package, a computationally efficient implementation of the random forest; see Wright and Ziegler (2017). A key hyperparameter in training a random forest is the number of features that are (randomly) selected from the universe of all d features at any given split. In "ranger" this hyperparameter is called `mtry` and the default is $\sqrt{d}$ for both tasks, classification and regression. However, in line with most of the literature, such as Liaw and Wiener (2002) and Hastie et al. (2017, Chapter 15), we set `mtry` equal to $d/3$ for our task of regression.[5] This choice aligns with Breiman (2001), who finds that higher `mtry` values are beneficial for regression tasks. All other hyperparameters are set to their default values, as recommended by Wright and Ziegler (2017) for the "ranger" package. We refrain from hyperparameter tuning for the random forest, as Medeiros et al. (2021) report that their results remain stable across a range of hyperparameter settings.

The EWMA parameter λ used in the estimation of μ and Σ is set to $\lambda = 0.15$, as detailed in Sect. 2.2. The bandwidth parameter H required for the estimator in Eq. (D.8) is chosen as $H = 6$.

[5] If need be, values of `mtry` are rounded down to the nearest integer by common convention.

Table 1 continued

Target	Forecast horizon 1	3	6	9	12	Mean
CHECPICOR						
one-shot	0.004	0.005	0.008	0.010	0.012	0.008
path-average	0.003	0.006	0.008	0.012	0.013	0.009
Panel (b): MAE ratios						
CPIAUCSL						
one-shot	0.005	0.008	0.009	0.011	0.013	0.009
path-average	0.002	0.004	0.007	0.010	0.012	0.007
CPILFESL						
one-shot	0.003	0.004	0.006	0.007	0.008	0.006
path-average	0.001	0.002	0.004	0.006	0.008	0.004
PCEPI						
one-shot	0.004	0.006	0.008	0.009	0.010	0.008
path-average	0.001	0.003	0.005	0.007	0.010	0.006
PCEPILFE						
one-shot	0.002	0.003	0.005	0.006	0.007	0.005
path-average	0.001	0.002	0.003	0.005	0.007	0.004
CHECPIALL						
one-shot	0.004	0.006	0.009	0.010	0.011	0.008
path-average	0.002	0.005	0.007	0.009	0.010	0.007
CHECPICOR						
one-shot	0.003	0.004	0.006	0.008	0.009	0.006
path-average	0.002	0.005	0.007	0.009	0.010	0.007

Many inflation series contain outliers, such as November 2008, which might adversely affect the training of a random forest. To mitigate such effects, we winsorize inflation values at their respective 1st and 99th percentiles during the corresponding (rolling) window used for training. Of course, forecast errors are based on actual inflation values, not on winsorized values.

5 Results

We want to assess the performance of the hedged random forest relative to the standard random forest employed by Medeiros et al. (2021). Both methods are evaluated using the rolling-window backtesting framework described in Sect. 4 across all six inflation measures considered in this study. Specifically, for each combination of out-of-sample month, forecast horizon, and inflation measure, we generate forecasts using both the standard random forest and the hedged random forest, applying the path-average forecasting approach outlined in Sect. 4.2.2.

Forecasting accuracy is measured using the RMSE and MAE criteria across all forecast horizons, methods, and inflation measures. To facilitate a direct comparison between the two methods, we calculate the ratio of these criteria with the HRF criterion in the numerator and the RF criterion in the denominator. Hence, the RMSE ratio is defined as:

$$\text{RMSE ratio}_{h,i} := \frac{\text{RMSE}_{h,i}^{\text{HRF}}}{\text{RMSE}_{h,i}^{\text{RF}}}, \tag{5.1}$$

where $\text{RMSE}_{h,i}^{\text{HRF}}$ and $\text{RMSE}_{h,i}^{\text{RF}}$ denote the RMSE values for forecast horizon h and inflation measure i generated by the hedged random forest and the standard random forest, respectively. Similarly, the MAE ratio is defined as:

$$\text{MAE ratio}_{h,i} := \frac{\text{MAE}_{h,i}^{\text{HRF}}}{\text{MAE}_{h,i}^{\text{RF}}}. \tag{5.2}$$

In either case, a ratio below one indicates that the hedged random forest outperforms the standard random forest, and vice versa.

Table 2 is the key table in this paper and presents performance ratios for the six inflation measures across forecast horizons of 1, 3, 6, 9, and 12 months, as well as mean ratios over all twelve horizons (1–12).

Panel (a) presents the RMSE ratios. The HRF consistently achieves lower RMSEs across all forecast horizons and target variables. Core inflation measures, such as CPILFESL and CHECPICOR, show the most significant gains, with reductions exceeding 9% for some horizons. For headline inflation measures, many improvements are nearly 5%, and one is above 5%. The mean RMSE ratios (across all twelve horizons) are less than one for all target variables, indicating that the HRF consistently outperforms the RF, with mean reductions ranging from 3.2 to 8.2%.

Panel (b) presents the MAE ratios. The HRF consistently achieves lower RMSEs across all forecast horizons and target variables. Core inflation measures, such as CPILFESL and CHECPICOR, show the most significant gains, with reductions exceeding 9% for some horizons. For headline inflation measures, many improvements are nearly 6%, and two are above 6%. The mean MAE ratios (across all twelve horizons) are less than one for all target variables, indicating that the HRF consistently outperforms the RF, with mean reductions ranging from 4.2 to 8.6%.

As a robustness check, Appendix B re-runs the results of Table 2 using alternative choices of the EWMA parameter λ. The findings are robust in the sense that a wide range of other values of $\lambda \in [0.02, 0.3]$ also result in clear outperformance over the standard random forest.

To evaluate whether the observed gains in forecast accuracy of the hedged random forest over the (standard) random forest are statistically significant, we apply a modified version of the Diebold and Mariano (1995) test. The modification consists of using heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; see Appendix C for a detailed description. The results displayed in Fig. 4 show that for the squared-error loss 61 out of 72 ($\approx$ 85%) p-values are below 0.1, whereas for

Table 2 RMSE and MAE ratios of the hedged random forest relative to the standard random forest for headline and core inflation measures in the U.S. and Switzerland evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL	0.993	0.971	0.955	0.963	0.963	0.966
CPILFESL	0.983	0.955	0.934	0.926	0.924	0.940
PCEPI	0.995	0.964	0.946	0.952	0.959	0.958
PCEPILFE	0.972	0.958	0.953	0.938	0.937	0.950
CHECPIALL	0.990	0.981	0.966	0.961	0.950	0.968
CHECPICOR	0.958	0.932	0.913	0.906	0.901	0.918
Panel (b): MAE ratios						
CPIAUCSL	0.972	0.953	0.946	0.954	0.960	0.955
CPILFESL	0.969	0.920	0.906	0.903	0.904	0.914
PCEPI	0.979	0.949	0.933	0.941	0.952	0.946
PCEPILFE	0.974	0.947	0.935	0.922	0.927	0.937
CHECPIALL	0.990	0.971	0.964	0.941	0.933	0.958
CHECPICOR	0.953	0.939	0.902	0.901	0.900	0.916

Ratios below one indicate that the hedged random forest outperforms the standard random forest

the absolute-error loss 70 out of 72 ($\approx 97\%$) p-values are below 0.1. In the related literature it is common to use the original Diebold–Mariano (DM) test or the more conservative HNL small-sample modification of Harvey et al. (1997). If we adopted such a convention, we would obtain even more significant results (in our favor). For example, switching to the HNL modification of the DM test would yield 97% of the p-values below 0.1 for the squared-error loss and 99% of the p values below 0.1 for absolute-error loss.

5.1 Impact of estimation scheme for inputs

The two inputs of the hedged random forest are the estimated mean vector $\hat{\mu}$ and the estimated covariance matrix $\hat{\Sigma}$ of the vector of the forecast errors corresponding to the individual trees. For i.i.d. data, Beck et al. (2024) recommend using the sample mean of the residual matrix R to estimate μ and the QIS nonlinear shrinkage estimator of Ledoit and Wolf (2022) computed from R to estimate Σ.

In the time-series setting of this paper, on the other hand, we employ EWMA estimation, combined with linear shrinkage, to estimate μ and Σ. To evaluate the benefit of this customized approach, we compare three versions of the HRF:

- HRF-EWMA: Our proposed approach uses EWMA estimation, combined with linear shrinkage, for $\hat{\mu}$ and $\hat{\Sigma}$.
- HRF-NL: Using the sample mean for $\hat{\mu}$ and the QIS nonlinear shrinkage estimator for $\hat{\Sigma}$; this is the generic approach proposed by Beck et al. (2024) for i.i.d. data.

- HRF-Sample: Using the sample mean for $\hat{\mu}$ and the sample covariance matrix for $\hat{\Sigma}$.

We then compute forecasting-performance ratios, as defined in Eqs. (5.1) and (5.2), for each of the three HRF versions. Table 3 presents the results.

These results demonstrate that the proposed HRF-EWMA version consistently outperforms the other two versions across all inflation measures and all forecast horizons, as indicated by lower RMSE and MAE ratios relative to the standard random forest. The performance gains are particularly pronounced for core inflation measures such as CHECPICOR, where the mean RMSE and MAE ratios (across all twelve horizons) of HRF-EWMA are more than seven percentage points lower than those of both HRF-NL and HRF-Sample.

Consequently, it is clearly beneficial to customize the estimators of μ and Σ for the task of forecasting inflation, where one is faced with time-series data.

5.2 Performance over time

To assess the stability of the hedged random forest's forecasting performance over time, we follow the approach of Borup and Schütte (2022). Specifically, we compute the Cumulative Sum of Squared Error Differences (CSSED) and the Cumulative Sum of Absolute Error Differences (CSAED) to compare the hedged random forest against the standard random forest.

The CSSED at time s is defined as

$$\text{CSSED}(s) := \sum_{t=F}^{s} \left(\hat{e}_{t,\text{HRF}}^2 - \hat{e}_{t,\text{RF}}^2 \right)$$

whereas the CSAED at time s is defined as

$$\text{CSAED}(s) := \sum_{t=F}^{s} \left(|\hat{e}_{t,\text{HRF}}| - |\hat{e}_{t,\text{RF}}| \right) ,$$

where F denotes the start of the out-of-sample evaluation period, s denotes the current month, and $\hat{e}_{t,\text{HRF}}$, and $\hat{e}_{t,\text{RF}}$ denote the respective forecast errors of the hedged random forest and the random forest at time t.

A negative value of the CSSED indicates that the hedged random forest has achieved lower cumulative squared errors than the standard random forest through month s, and vice versa. Analogously, a negative value of CSAED indicates that the hedged random forest has achieved lower cumulative absolute errors than the standard random forest through month s, and vice versa.

The CSSED and CSAED trajectories over time show how the relative cumulative performance of the two random forests, HRF and RF, develops over time. Moreover, one can also judge "local relative performance" in addition to "cumulative relative performance" from these trajectories: At any given point, a downward (upward) slope of a trajectory indicates better (worse) performance of the HRF relative to the RF

Table 3 Comparison of RMSE ratios (left) and MAE ratios (right) for inflation forecasts using three different HRF versions evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

Target	Panel (a): RMSE ratios						Panel (b): MAE ratios					
	Forecast horizon						Forecast horizon					
	1	3	6	9	12	Mean	1	3	6	9	12	Mean
CPIAUCSL												
HRF EWMA	0.993	0.971	0.955	0.963	0.963	0.966	0.972	0.953	0.946	0.954	0.960	0.955
HRF NL	1.025	0.999	0.976	0.977	0.973	0.986	1.016	0.990	0.980	0.979	0.977	0.986
HRF Sample	1.030	0.986	0.977	0.981	0.975	0.986	1.026	0.993	0.985	0.989	0.982	0.991
CPILFESL												
HRF EWMA	0.983	0.955	0.934	0.926	0.924	0.940	0.969	0.920	0.906	0.903	0.904	0.914
HRF NL	1.027	1.013	1.011	0.997	0.993	1.007	1.027	1.009	1.007	0.997	0.992	1.005
HRF Sample	1.038	1.021	1.012	1.002	0.999	1.013	1.035	1.009	1.008	1.000	0.998	1.010
PCEPI												
HRF EWMA	0.995	0.964	0.946	0.952	0.959	0.958	0.979	0.949	0.933	0.941	0.952	0.946
HRF NL	1.024	0.990	0.979	0.977	0.982	0.986	1.004	0.981	0.974	0.981	0.995	0.984
HRF Sample	1.012	0.993	0.978	0.978	0.984	0.985	1.003	0.991	0.982	0.987	1.001	0.990
PCEPILFE												
HRF EWMA	0.972	0.958	0.953	0.938	0.937	0.950	0.974	0.947	0.935	0.922	0.927	0.937
HRF NL	1.021	1.020	1.009	0.989	0.985	1.004	1.021	1.024	1.002	0.987	0.992	1.004
HRF Sample	1.008	1.013	1.003	0.989	0.983	0.999	1.003	1.009	0.993	0.992	0.995	0.999
CHECPIALL												
HRF EWMA	0.990	0.981	0.966	0.961	0.950	0.968	0.990	0.971	0.964	0.941	0.933	0.958
HRF NL	1.017	1.007	1.033	1.038	1.032	1.027	1.026	1.007	1.031	1.028	1.036	1.025
HRF Sample	1.002	1.003	1.024	1.032	1.029	1.021	1.007	0.990	1.024	1.023	1.026	1.015
CHECPICOR												
HRF EWMA	0.958	0.932	0.913	0.906	0.901	0.918	0.953	0.939	0.902	0.901	0.900	0.916
HRF NL	0.994	0.986	0.981	0.995	0.998	0.990	0.987	0.996	0.979	0.999	1.015	0.995
HRF Sample	0.990	1.006	0.997	1.005	0.998	0.999	0.988	1.006	0.983	0.997	1.008	0.994

Ratios below one indicate outperformance over RF

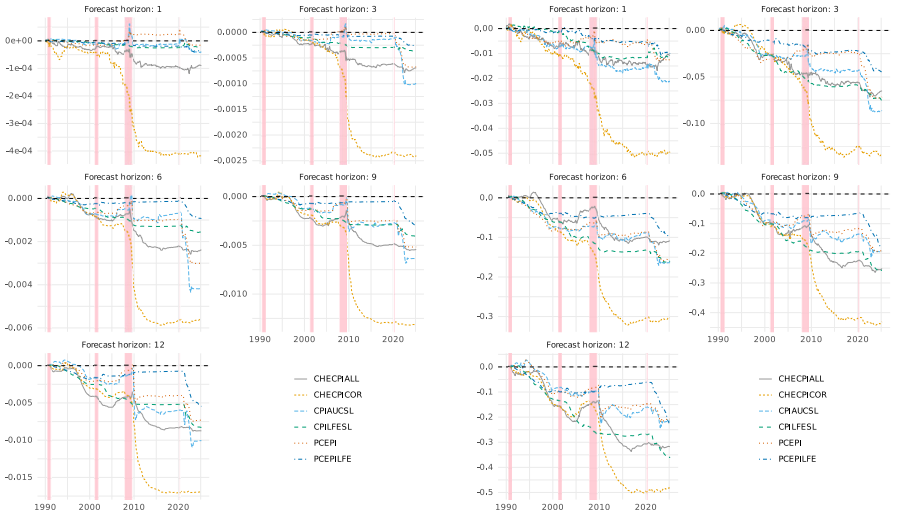

Fig. 2 CSSED (left panel) and CSAED (right panel) trajectories of relative cumulative performance of the HRF and the RF over time, across different inflation-forecast horizons and inflation measures. Negative values indicate that the HRF has outperformed the RF in a cumulative sense until this point in time, whereas downward slopes identify periods where the HRF outperforms the RF in a local sense

according to the corresponding criterion, squared errors or absolute errors. Finally, one can also judge "relative performance during a sub-period": If the trajectory is lower (higher) at the end of the period than at the beginning of the period, this indicates better (worse) performance of the HRF relative to the RF according to the corresponding criterion, squared errors or absolute errors.

Figure 2 displays the CSSED trajectories in the left panel and the CSAED trajectories in the right panel. Both types of trajectories exhibit a general downward trend over time, and rarely cross above the zero line, indicating that the hedged random forest consistently outperforms the standard random forest according to both criteria. With respect to the CSSED trajectories, the most pronounced improvements (that is, the steepest slopes) are observed during post-recession periods, highlighting the ability of the HRF to adapt to and perform in volatile economic environments. The CSAED trajectories exhibit similar, though somewhat less steep, patterns which can be attributed to absolute errors reacting less strongly than squared errors during volatile periods.

Notably, the CSAED trajectories reveal improvements of up to 50 percentage points for certain forecast horizons and targets, a substantial and economically meaningful gain. Together, these results emphasize the ability of the HRF to reduce cumulative errors relative to the RF across varying economic conditions, ranging from turbulence to stability.

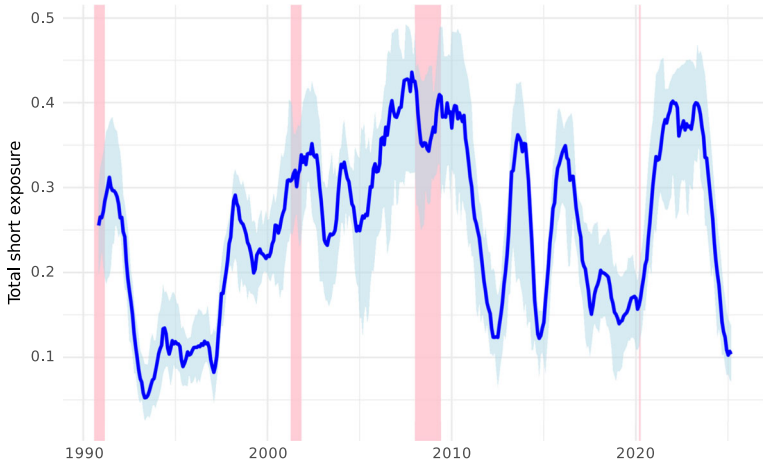

Fig. 3 The figure displays the evolution of three total-short-exposure (TSE) summary statistics over time for forecasting CPIAUCSL inflation. The lower end of the light-blue band is minimum TSE; the dark blue line is average TSE; and the upper end of the light-blue band is maximum TSE. Here, minimum, average, and maximum are with respect to the 12 TSE values corresponding to forecast horizons $h = 1, \ldots, 12$

5.3 Total short exposure over time

Our choice of $\kappa = 2$ for the gross-exposure constraint allows for a total short exposure of up to 0.5 in the hedged random forest, where total short exposure is defined as

$$\text{TSE} := \sum_{w_j < 0} |w_j| \, .$$

It is of interest to see how TSE evolves over time. We study this for the specific case of forecasting CPIAUCSL inflation. At any point in time, we compute three values: minimum TSE, average TSE, and maximum TSE over the 12 forecast horizons $h = 1, \ldots, 12$. Wen then plot these values over time. The results are displayed in Fig. 3 and lead to three insights (at least). First, negative weights do always occur; apart from very few exceptions, TSE is always above 0.1 and a typical value is 0.25. Second, the constraint $\kappa = 2$ is actually never binding; there are a few periods where maximum TSE gets close to 0.5 but it never quite reaches it. Third, and as to be expected, TSE evolves rather smoothly over time and is generally higher around periods of economic recessions compared to calm(er) periods.

6 Conclusion

This paper contributes to the growing literature on machine learning methods for inflation forecasting. We build on the influential work of Medeiros and Mendes (2016), Garcia et al. (2017), and Medeiros et al. (2021) who have demonstrated that the random

forest is the leading machine learning method for this task and delivers consistent and pronounced gains over traditional benchmarks.

We demonstrate that a further improvement can be achieved by the hedged random forest of Beck et al. (2024) which, in contrast to the standard random forest, employs non-equal (and even negative) weights of the individual trees. To obtain these weights, the hedged random forest requires two inputs that need to be estimated in practice: the vector of means and the covariance matrix of the forecast errors corresponding to the individual trees. To this end, we propose estimators that are customized for the task of forecasting inflation, where one is faced with time-series data; these estimators combine exponentially weighted moving average estimation with linear shrinkage.

An extensive backtest analysis, covering twelve forecast horizons and six inflation measures, has demonstrated clear and consistent outperformance of the hedged random forest over the standard random forest. A typical improvement in forecasting accuracy is on the order of five percent in terms of the root mean-squared error and on the order of six percent in terms of the mean absolute error.

There are several promising avenues for future research, such as (i) using the HRF with inflation data from other countries; (ii) employing the HRF to forecast other macroeconomic indicators, such as (changes in) gross domestic product or unemployment; and (iii) evaluating the performance of the HRF in nowcasting macroeconomic indicators, a practice which has gained increasing attention in recent years.

A Variable descriptions

Tables 4, 5, 6, 7, 8, 9, 10, 11 describe the variables used in the analysis, sorted by groups. In each table, the column `tcode` indicates the data transformation applied to a given series $\{x_t\}$ according to the following list:

1. x_t, that is, no transformation
2. Δx_t
3. $\Delta^2 x_t$
4. $\log(x_t)$
5. $\Delta \log(x_t)$
6. $\Delta^2 \log(x_t)$
7. $\Delta(x_t/x_{t-1} - 1)$

In all tables, the column `fred` provides mnemonics in FRED (followed by a short description), whereas the column `gsi` provides corresponding mnemonics in Global Insight (followed by a short description).

B Results for other choices of λ

This appendix re-runs the results of Table 2 using alternative choices of the EWMA parameter λ. The findings turn out to be rather robust in the sense that a wide range of other values of λ also result in clear outperformance over the standard random forest (Tables 12, 13, 14, 15, 16).

Table 4 Output and income variables

Group 1: Output and income variables						
	id	tcode	fred	fred.description	gsi	gsi.description
1	1	5	RPI	Real Personal Income	M_14386177	PI
2	2	5	W875RX1	Real Personal Income Ex Transfer Receipts	M_145256755	PI less transfers
3	6	5	INDPRO	IP Index	M_116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	M_116460981	IP: products
5	8	5	IPFINAL	IP: Final Products (Market Group)	M_116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M_116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M_116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M_116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M_116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M_116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M_116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M_116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing (SIC)	M_116461013	IP: mfg
14	17	5	IPB51222s	IP: Residential Utilities	M_116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M_116461275	IP: fuels
16	20	2	CUMFNS	Capacity Utilization: Manufacturing	M_116461602	Cap util

Table 5 Labor market variables

Group 2: Labor market variables						
	id	tcode	fred	fred,description	gsi	gsi.description
1	21	2	HWI	Help-Wanted Index for United States		Help wanted indx
2	22	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	M_110156531	Help wanted/une
3	23	5	CLF16OV	Civilian Labor Force	M_110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M_110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M_110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	M_110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	M_110156527	U < 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed for 5–14 Weeks	M_110156523	U 5–14 wks
9	29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	M_110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed for 15–26 Weeks	M_110156525	U 15–26 wks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M_110156526	U 27+ wks
12	32	5	CLAIMSx	Initial Claims	M_15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M_123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing Industries	M_123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining	M_123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M_123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M_123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable Goods	M_123109573	Emp: dble gds
19	39	5	NDMANEMP	All Employees: Nondurable Goods	M_123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service-Providing Industries	M_123109193	Emp: services
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities	M_123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M_123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M_123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M_123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M_123114411	Emp: Govt
26	46	1	CES0600000007	Avg Weekly Hours: Goods-Producing	M_140687274	Avg hrs
27	47	2	AWOTMAN	Avg Weekly Overtime Hours: Manufacturing	M_123109554	Overtime: mfg
28	48	1	AWHMAN	Avg Weekly Hours: Manufacturing	M_14386098	Avg hrs: mfg
29	127	6	CES0600000008	Avg Hourly Earnings: Goods-Producing	M_123109182	AHE: goods
30	128	6	CES2000000008	Avg Hourly Earnings: Construction	M_123109341	AHE: const
31	129	6	CES3000000008	Avg Hourly Earnings: Manufacturing	M_123109552	AHE: mfg

Table 6 Housing variables

Group 3: Housing variables						
	id	tcode	fred	fred.description	gsi	gsi.description
1	50	4	HOUST	Housing Starts: Total New Privately Owned	M_110155536	Starts: nonfarm
2	51	4	HOUSTNE	Housing Starts, Northeast	M_110155538	Starts: NE
3	52	4	HOUSTMW	Housing Starts, Midwest	M_110155537	Starts: MW
4	53	4	HOUSTS	Housing Starts, South	M_110155543	Starts: South
5	54	4	HOUSTW	Housing Starts, West	M_110155544	Starts: West
6	55	4	PERMIT	New Private Housing Permits (SAAR)	M_110155532	BP: total
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	M_110155531	BP: NE
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	M_110155530	BP: MW
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)	M_110155533	BP: South
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)	M_110155534	BP: West

Table 7 Consumption, orders, and inventory variables

Group 4: Consumption, orders, and inventory variables						
	id	tcode	fred	fred,description	gsi	gsi.description
1	3	5	DPCERA3M086SBEA	Real Personal Consumption Expenditures	M_123008274	Real Consumption
2	5	5	RETAILx	Retail and Food Services Sales	M_130439509	Retail sales
3	65	5	AMDMNOx	New Orders for Durable Goods	M_14386110	Orders: dble gds
4	67	5	AMDMUOx	Unfilled Orders for Durable Goods	M_14385946	Unf orders: dble

Table 8 Money and credit variables

Group 5: Money and credit variables						
	id	tcode	fred	fred,description	gsi	gsi.description
1	70	6	M1SL	M1 Money Stock	M_110154984	M1
2	71	6	M2SL	M2 Money Stock	M_110154985	M2
3	72	5	M2REAL	Real M2 Money Stock	M_110154985	M2 (real)
4	73	6	BOGMBASE	Monetary Base	M_110154995	MB
5	74	6	TOTRESNS	Total Reserves of Depository Institutions	M_110155011	Reserves tot
6	75	7	NONBORRES	Reserves of Depository Institutions	M_110155009	Reserves nonbor
7	76	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
8	77	6	REALLN	Real Estate Loans at All Commercial Banks	BUSLOANS	DC&I loans
9	134	6	INVEST	Securities in Bank Credit at All Commercial Banks	N.A	N.A

Table 9 Interest and exchange rate variables

Group 6: Interest and exchange rate variables						
	id	tcode	fred	dfred.description	gsi	gsi.description
1	84	2	FEDFUNDS	Effective Federal Funds Rate	M_110155157	Fed Funds
2	86	2	TB3MS	3-Month Treasury Bill	M_110155165	3 mo T-bill
3	87	2	TB6MS	6-Month Treasury Bill	M_110155166	6 mo T-bill
4	88	2	GS1	1-Year Treasury Rate	M_110155168	1 yr T-bond
5	89	2	GS5	5-Year Treasury Rate	M_110155174	5 yr T-bond
6	90	2	GS10	10-Year Treasury Rate	M_110155169	10 yr T-bond
7	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield		Aaa bond
8	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield		Baa bond
9	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS		3 mo-FF spread
10	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS		6 mo-FF spread
11	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS		1 yr-FF spread
12	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS		5 yr-FF spread
13	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS		10 yr-FF spread
14	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS		Aaa-FF spread
15	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS		Baa-FF spread
16	102	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	M_110154768	Ex rate: Switz
17	103	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	M_110154755	Ex rate: Japan
18	104	5	EXUSUKx	USA / U.K. Foreign Exchange Rate	M_110154772	Ex rate: UK
19	105	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	M_110154744	EX rate: Canada

Table 10 Price variables

Group 7: Price variables						
	id	tcode	fred	fred.description	gsi	gsi.description
1	106	6	WPSFD49207	PPI: Finished Goods	M110157517	PPI: fin gds
2	107	6	WPSFD49502	PPI: Finished Consumer Goods	M110157508	PPI: cons gds
3	108	6	WPSID61	PPI: Intermediate Materials	M_110157527	PPI: int matls
4	109	6	WPSID62	PPI: Crude Materials	M_110157500	PPI: crude matls
5	110	6	OILPRICEx	Crude Oil, Spliced WTI and Cushing	M_110157273	Spot market price
6	111	6	PPICMM	PPI: Metals and Metal Products	M_110157335	PPI: nonferrous
7	113	6	CPIAUCSL	CPI: All Items	M_110157323	CPI-U: all
8	114	6	CPIAPPSL	CPI: Apparel	M_110157299	CPI-U: apparel
9	115	6	CPITRNSL	CPI: Transportation	M_110157302	CPI-U: transp
10	116	6	CPIMEDSL	CPI: Medical Care	M_110157304	CPI-U: medical
11	117	6	CUSR0000SAC	CPI: Commodities	M_110157314	CPI-U: comm
12	118	6	CUSR0000SAD	CPI: Durables	M_110157315	CPI-U: dbles
13	119	6	CUSR0000SAS	CPI: Services	M_110157325	CPI-U: services
14	120	6	CPIULFSL	CPI: All Items Less Food	M_110157328	CPI-U: ex food
15	121	6	CUSR0000SA0L2	CPI: All Items Less Shelter	M_110157329	CPI-U: ex shelter
16	122	6	CUSR0000SA0L5	CPI: All items Less Medical Care	M_110157330	CPI-U: ex med
17	123	6	PCEPI	Personal Cons. Expend.: Chain Index	gmdc	PCE defl
18	124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable Goods	gmdcd	PCE defl: dlbes
19	125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable Goods	gmdcn	PCE defl: nondble
20	126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	gmdcs	PCE defl: service

Table 11 Stock market variables

Group 8: Stock market variables						
	id	tcode	fred	fred.description	gsi	gsi.description
1	80	5	S&P 500	S&P's Common Stock Price Index: Composite	M_110155044	S&P 500
2	81	5	S&P: indust	S&P's Common Stock Price Index: Industrials	M_110155047	S&P: indust

Table 12 $\lambda = 0.02$: RMSE and MAE ratios of the random forest relative to the standard random forest for headline and core inflation measures in the U.S. and Switzerland evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL	1.006	0.979	0.961	0.970	0.963	0.973
CPILFESL	1.004	0.987	0.983	0.979	0.974	0.984
PCEPI	0.994	0.982	0.960	0.966	0.969	0.972
PCEPILFE	0.966	0.975	0.976	0.968	0.967	0.971
CHECPIALL	1.001	0.985	0.986	0.989	0.983	0.988
CHECPICOR	0.959	0.942	0.935	0.931	0.929	0.938
Panel (b): MAE ratios						
CPIAUCSL	0.994	0.980	0.963	0.975	0.965	0.973
CPILFESL	1.004	0.977	0.979	0.976	0.972	0.980
PCEPI	0.986	0.966	0.954	0.968	0.978	0.968
PCEPILFE	0.968	0.967	0.965	0.958	0.963	0.964
CHECPIALL	0.998	0.969	0.984	0.962	0.966	0.974
CHECPICOR	0.957	0.944	0.921	0.920	0.919	0.931

Ratios below one indicate that the hedged random forest outperforms the standard random forest

Table 13 $\lambda = 0.06$: RMSE and MAE ratios of the hedged random forest relative to the standard random forest for headline and core inflation measures in the US and Switzerland evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL	1.011	0.981	0.953	0.958	0.958	0.967
CPILFESL	0.992	0.969	0.956	0.951	0.947	0.959
PCEPI	0.999	0.970	0.951	0.953	0.958	0.961
PCEPILFE	0.977	0.964	0.962	0.946	0.945	0.957
CHECPIALL	0.963	0.964	0.966	0.957	0.953	0.961
CHECPICOR	0.952	0.927	0.916	0.906	0.902	0.918
Panel (b): MAE ratios						
CPIAUCSL	0.992	0.964	0.950	0.951	0.955	0.959
CPILFESL	0.984	0.948	0.940	0.936	0.933	0.943
PCEPI	0.977	0.956	0.935	0.949	0.960	0.950
PCEPILFE	0.977	0.954	0.947	0.932	0.936	0.947
CHECPIALL	0.969	0.955	0.962	0.932	0.929	0.949
CHECPICOR	0.944	0.930	0.896	0.896	0.893	0.908

Ratios below one indicate that the hedged random forest outperforms the standard random forest

Table 14 $\lambda = 0.1$: RMSE and MAE ratios of the hedged random forest relative to the standard random forest for headline and core inflation measures in the US and Switzerland evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL	1.008	0.967	0.950	0.961	0.961	0.965
CPILFESL	0.979	0.961	0.942	0.935	0.930	0.946
PCEPI	0.985	0.959	0.942	0.948	0.955	0.954
PCEPILFE	0.974	0.959	0.955	0.939	0.943	0.952
CHECPIALL	0.983	0.970	0.954	0.949	0.939	0.958
CHECPICOR	0.951	0.935	0.918	0.910	0.906	0.922
Panel (b): MAE ratios						
CPIAUCSL	0.990	0.958	0.943	0.952	0.956	0.956
CPILFESL	0.969	0.927	0.917	0.914	0.912	0.924
PCEPI	0.962	0.942	0.932	0.942	0.953	0.944
PCEPILFE	0.972	0.951	0.938	0.926	0.932	0.940
CHECPIALL	0.981	0.954	0.951	0.925	0.917	0.945
CHECPICOR	0.948	0.938	0.905	0.899	0.895	0.914

Ratios below one indicate that the hedged random forest outperforms the standard random forest

Table 15 $\lambda = 0.2$: RMSE and MAE ratios of the hedged random forest relative to the standard random forest for headline and core inflation measures in the US and Switzerland evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL	0.997	0.977	0.960	0.969	0.968	0.971
CPILFESL	0.988	0.954	0.928	0.922	0.922	0.938
PCEPI	0.990	0.957	0.944	0.950	0.957	0.955
PCEPILFE	0.977	0.964	0.954	0.936	0.937	0.951
CHECPIALL	0.982	0.98	0.965	0.962	0.953	0.967
CHECPICOR	0.964	0.939	0.930	0.918	0.909	0.930
Panel (b): MAE ratios						
CPIAUCSL	0.976	0.967	0.953	0.961	0.964	0.962
CPILFESL	0.970	0.922	0.906	0.901	0.903	0.915
PCEPI	0.969	0.940	0.931	0.936	0.948	0.941
PCEPILFE	0.970	0.952	0.939	0.921	0.925	0.938
CHECPIALL	0.987	0.974	0.961	0.948	0.935	0.960
CHECPICOR	0.963	0.935	0.918	0.910	0.906	0.924

Ratios below one indicate that the hedged random forest outperforms the standard random forest

Table 16 $\lambda = 0.3$: RMSE and MAE ratios of the random forest relative to the standard random forest for headline and core inflation measures in the US and Switzerland evaluated across forecast horizons of 1, 3, 6, 9, and 12 months, as well as the mean over all twelve horizons (1–12)

	Forecast horizon					
Target	1	3	6	9	12	Mean
Panel (a): RMSE ratios						
CPIAUCSL	0.999	0.971	0.955	0.965	0.968	0.968
CPILFESL	0.990	0.958	0.935	0.928	0.927	0.942
PCEPI	0.986	0.959	0.948	0.951	0.957	0.957
PCEPILFE	0.972	0.966	0.960	0.947	0.947	0.958
CHECPIALL	0.989	0.987	0.976	0.976	0.963	0.978
CHECPICOR	0.966	0.951	0.937	0.934	0.926	0.941
Panel (b): MAE ratios						
CPIAUCSL	0.976	0.962	0.951	0.956	0.963	0.960
CPILFESL	0.974	0.928	0.915	0.911	0.911	0.922
PCEPI	0.967	0.946	0.940	0.941	0.952	0.947
PCEPILFE	0.966	0.960	0.946	0.939	0.939	0.948
CHECPIALL	0.986	0.979	0.974	0.965	0.948	0.970
CHECPICOR	0.962	0.951	0.933	0.935	0.926	0.941

Ratios below one indicate that the hedged random forest outperforms the standard random forest

C Comparing forecasting accuracy

To evaluate the statistical significance of the differences in forecasting accuracy between the (standard) random forest and the hedged random forest, we apply a modification of the Diebold–Mariano (DM) test proposed by Diebold and Mariano (1995). The original DM test as well as its (more conservative) HLN small-sample modification of Harvey et al. (1997) assume for the derivation of the standard errors used in the construction of the DM test statistic that forecast errors at horizon h are $(h-1)$-dependent; in particular, forecast errors at horizon $h = 1$ are independent. In our opinion, there is no convincing theoretical basis for such an assumption; moreover, we also find that it is violated in data on a regular basis. Therefore, we allow for arbitrary serial correlation in forecast errors, apart from a suitable weak-dependence condition, and use heteroskedasticity- and autocorrelation-consistent (HAC) standard errors. Specifically, our HAC standard errors employ the Quadratic Spectral (QS) kernel in conjunction with the automatic bandwidth selection of Andrews (1991).

The p-values presented in Fig. 4 and Table 17 are one-sided p-values for the null hypothesis that the hedged random forest does not outperform the (standard) random forest. In other words, the two underlying hypothesis-testing formulations are

$$H_0 : \mathbb{E}\big(\hat{e}^2_{t,\mathrm{RF}} - \hat{e}^2_{t,\mathrm{t,HRF}}\big) \leq 0 \quad \text{vs} \quad H_1 : \mathbb{E}\big(\hat{e}^2_{t,\mathrm{RF}} - \hat{e}^2_{t,\mathrm{t,HRF}}\big) > 0$$

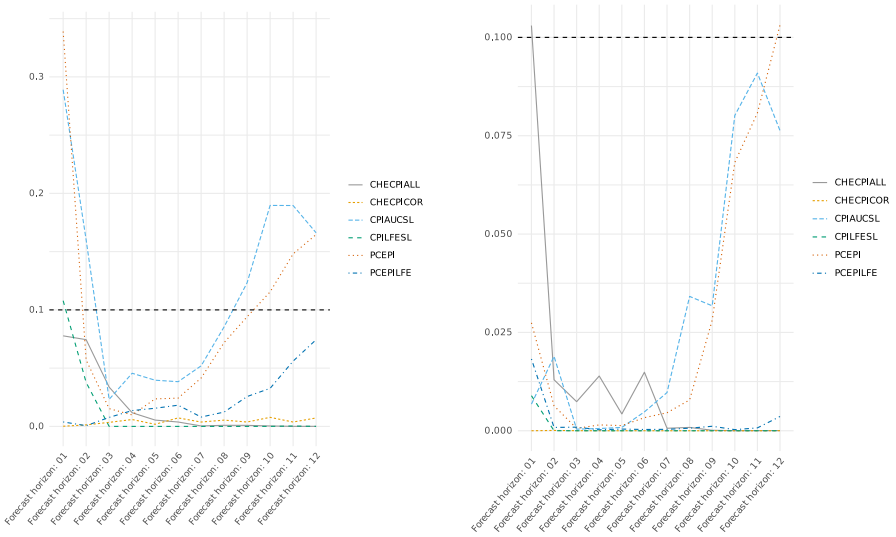

Fig. 4 One-sided p-values from the Diebold-Mariano test comparing the forecasting accuracy of the hedged random forest against the standard random forest across forecast horizons 1–12. Results are shown separately for squared-error loss (left) and absolute-error loss (right) loss functions. HAC standard errors are computed using the Quadratic Spectral kernel with automatic bandwidth selection following Andrews (1991). Small p-values constitute evidence of outperformance in favor of the hedged random forest

and

$$H_0 : \mathbb{E}\big(|\hat{e}_{t,\text{RF}}| - |\hat{e}_{t,\text{t,HRF}}|\big) \leq 0 \quad \text{vs} \quad H_1 : \mathbb{E}\big(|\hat{e}_{t,\text{RF}}| - |\hat{e}_{t,\text{t,HRF}}|\big) > 0 \,,$$

corresponding to squared-error loss and absolute-error loss, respectively. Readers who prefer two-sided p values can simply double our one-sided p values, since they are all below 0.5. As mentioned at the end of Sect. 5, switching to the original DM test or its small-sample HLN modification would yield even more significant results (in our favor).

D EWMA estimation & linear shrinkage

We provide a high-level description here that applies to a generic setting, not just the setting of estimating the mean and the covariance matrix of the vector of forecast errors corresponding to the individual trees of a random forest.

D.1 EWMA estimation

The data $\{x_t\}_{t=1}^T$ are assumed to be a (stretch of) a N-variate stationary time series with mean μ and covariance matrix Σ. It is further assumed that the autocovariance function of the process $\{x_t\}$ is absolutely summable, which implies that the process is ergodic.

Table 17 One-sided p values from the Diebold–Mariano test comparing the forecast accuracy of the hedged random forest against the default random forest across forecast horizons 1–12

Target	Forecast horizon 1	2	3	4	5	6	7	8	9	10	11	12
Panel (a): Squared-Error Loss												
CPIAUCSL	0.289	0.160	0.023	0.046	0.040	0.038	0.052	0.085	0.123	0.190	0.190	0.166
CPILFESL	0.108	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PCEPI	0.339	0.057	0.015	0.010	0.024	0.024	0.042	0.072	0.094	0.115	0.148	0.165
PCEPILFE	0.004	0.001	0.008	0.014	0.016	0.018	0.008	0.012	0.026	0.033	0.056	0.074
CHECPIALL	0.078	0.074	0.034	0.012	0.005	0.004	0.000	0.001	0.001	0.000	0.000	0.000
CHECPICOR	0.000	0.001	0.003	0.006	0.002	0.007	0.004	0.005	0.004	0.008	0.004	0.007
Panel (b): Absolute-Error Loss												
CPIAUCSL	0.007	0.019	0.000	0.001	0.001	0.005	0.010	0.034	0.032	0.080	0.091	0.076
CPILFESL	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PCEPI	0.027	0.006	0.001	0.001	0.001	0.003	0.005	0.008	0.028	0.068	0.081	0.103
PCEPILFE	0.018	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.004
CHECPIALL	0.103	0.013	0.007	0.014	0.004	0.015	0.001	0.001	0.000	0.000	0.000	0.000
CHECPICOR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Results are shown separately for squared-error (Panel (a)) and absolute-error (Panel (b)) loss functions. Standard errors are computed using the Quadratic Spectral kernel with automatic bandwidth selection following (Andrews 1991)

For a given $\lambda \in (0, 1)$ the exponentially-weighed moving average (EWMA) weights are defined as follows: $w_t := \lambda(1-\lambda)^{T-t}$ for $t = 1, \ldots, T$. The EWMA estimator of μ is then given by

$$\hat{\mu} := \sum_{t=1}^{T} w_t x_t \tag{D.1}$$

and the EWMA estimator of Σ is given by

$$\hat{\Sigma} := \sum_{t=1}^{T} w_t (x_t - \bar{x})(x_t - \bar{x})' \quad \text{with} \quad \bar{x} := \frac{1}{T} \sum_{t=1}^{T} x_t \, . \tag{D.2}$$

A more careful notation would use a subscript of λ for w_t, $\hat{\mu}$, and $\hat{\Sigma}$; however, in the interest of compactness of notation we abstain from doing so. Furthermore, denote the (i, j) element of $\hat{\Sigma}$ by $\hat{\sigma}_{ij}$, for $1 \leq i, j \leq N$.

It is worthwhile to point out that the EWMA scheme (with a fixed λ) does not lead to consistent estimators. To illustrate this fact, consider the case where $\{x_t\}_{t=1}^{T}$ is an independent and identically distributed (i.i.d.) univariate time series with mean μ and variance σ^2. It is well-known that the variance of the sample mean is equal to σ^2/T, which converges to zero as T tends to infinity. In contrast, the variance of the EWMA estimator of μ is given by

$$\begin{aligned} \mathrm{Var}(\hat{\mu}) &= \mathrm{Var}\Big(\sum_{t=1}^{T} w_t x_t\Big) = \sum_{t=1}^{T} \mathrm{Var}\big(w_t x_t\big) \\ &= \sum_{t=1}^{T} w_t^2 \sigma^2 = \sigma^2 \lambda^2 \sum_{t=1}^{T} (1-\lambda)^{2(T-t)} \overset{T\to\infty}{\longrightarrow} \sigma^2 \frac{\lambda^2}{1-(1-\lambda)^2} \, . \end{aligned} \tag{D.3}$$

Hence, the variance of the EWMA estimator does not vanish in the limit (for any fixed λ). The intuition here is that because of the exponentially fast decaying weights w_t the EWMA scheme, effectively, only uses a finite stretch of past data, where the length of the stretch is inversely proportional to the parameter λ: As λ approaches zero, EWMA gets closer and closer to equal weighting, and thus to using all the available data; as λ approaches one, EWMA gets closer and closer to only using x_T, that is, to using $w_T = 1$. Consequently, the factor on σ^2 in the final expression of (D.3) converges to zero as λ approaches zero whereas it converges to one as λ approaches one.

Some people might argue that an inconsistent estimator should never be used. However, the use of EWMA estimation is generally forecasting; for example, one uses $\hat{x}_{T+1} := \hat{\mu}$. To this end, arguably, it is not crucial for the estimator to be consistent; all one is after is a good forecast.

D.2 Shrinkage targets

We start with the estimation of the covariance matrix Σ. We will consider linear shrinkage of the EWMA estimator $\hat{\Sigma}$ to the two-parameter covariance matrix, also called constant-variance–covariance (CVC) matrix F. This matrix has constant diagonal (that is, variance) element f_1 and constant off-diagonal (that is, covariance) element f_2 given by

$$f_1 := \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{ii} \quad \text{and} \quad f_2 := \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\sigma}_{ij} \ ,$$

respectively. Therefore, with f_{ij} denoting the (i, j) element of F, it holds

$$f_{ij} = \begin{cases} f_1 & \text{if} \quad i = j \\ f_2 & \text{otherwise} \ . \end{cases}$$

A linear shrinkage estimator then is of the form

$$\Sigma^{\dagger} := \alpha F + (1-\alpha)\hat{\Sigma} \quad \text{for } \alpha \in [0, 1] \ . \tag{D.4}$$

Consequently, it constitutes a convex combination of the shrinkage target F and the EWMA estimator $\hat{\Sigma}$. The question that remains is how to construct a data-dependent shrinkage intensity $\alpha \in [0, 1]$.

For the estimation of the mean vector μ, we will consider linear shrinkage of the EWMA estimator $\hat{\mu}$ to the one-parameter constant-mean vector, which has constant element given by

$$\hat{\mu}_* := \frac{1}{p}\sum_{i=1}^{N}\hat{\mu}_i \ .$$

A linear shrinkage estimator then is of the form

$$\mu^{\dagger} := \alpha\hat{\mu}_* + (1-\alpha)\hat{\mu} \quad \text{for } \alpha \in [0, 1] \ , \tag{D.5}$$

where in some abuse of notation we let $\hat{\mu}_* := (\hat{\mu}_*, \ldots, \hat{\mu}_*)'$. Consequently, this estimator constitutes a convex combination of the shrinkage target $\hat{\mu}_*$ and the EWMA estimator $\hat{\mu}$. The question that remains is how to construct a data-dependent shrinkage intensity $\alpha \in [0, 1]$.

D.3 Shrinkage intensities

Again, we start with the estimation of the covariance matrix Σ. Linear shrinkage to the two-parameter shrinkage target has been considered before by Ledoit (1995, Appendix B.1), Schäfer and Strimmer (2005), and De Nard (2020). However, their starting point is always the sample covariance matrix instead of the EWMA estimator;

furthermore, they assume that the data $\{x_i\}$ are i.i.d. Therefore, their formulas for data-dependent shrinkage intensities do not apply in our setting.

Let

$$\phi_1 := \frac{1}{N}\sum_{i=1}^{N}\sigma_{ii} \quad \text{and} \quad \phi_2 := \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sigma_{ij} \, .$$

Furthermore, let Φ be an $N \times N$ matrix whose (i, j) element is denoted by ϕ_{ij} and defined as

$$\phi_{ij} = \begin{cases} \phi_1 & \text{if} \quad i = j \\ \phi_2 & \text{otherwise} \, . \end{cases}$$

In this way, Φ is the population counterpart to the shrinkage target F.

By Ledoit and Wolf (2003, Section 2.5) it follows that the optimal shrinkage intensity (under Frobenius-loss-based risk) satisfies

$$\alpha^* \approx \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\text{Var}(\hat{\sigma}_{ij}) - \text{Cov}(f_{ij}, \hat{\sigma}_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\text{Var}(f_{ij} - \hat{\sigma}_{ij}) + (\phi_{ij} - \sigma_{ij})^2} \, . \tag{D.6}$$

There only holds approximate equality in Eq. (D.6) because exact equality would require $\hat{\Sigma}$ to be an unbiased estimator of Σ, which is not the case for EWMA estimation. However, the bias of $\hat{\Sigma}$ is of order $1/T$, and therefore is negligible for large T. Furthermore, since Φ only has two parameters, the estimation uncertainty in F is also negligible compared to the estimation uncertainty in $\hat{\Sigma}$, which gives rise to the following second (round of) approximation:

$$\alpha^* \approx \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\text{Var}(\hat{\sigma}_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\text{Var}(\hat{\sigma}_{ij}) + (\phi_{ij} - \sigma_{ij})^2} =: \frac{\nu}{\nu + \gamma} \, . \tag{D.7}$$

The first goal is to find an estimator of $\nu_{ij} := \text{Var}(\hat{\sigma}_{ij})$, for $1 \leq i, j \leq N$. Letting $y_t := x_t - \bar{x}$, it holds $\hat{\sigma}_{ij} = \sum_{t=1}^{T} w_t y_{t,i} y_{t,j}$, which implies

$$\begin{aligned}
\text{Var}(\hat{\sigma}_{ij}) &= \sum_{t=1}^{T}\sum_{l=1}^{T}\text{Cov}(w_t y_{t,i} y_{t,j}, w_l y_{l,i} y_{l,j}) \\
&= \sum_{t=1}^{T} w_t^2 \text{Var}(y_{t,i} y_{t,j}) + 2\sum_{h=1}^{T-1}\sum_{t=h+1}^{T} w_{t-h} w_t \text{Cov}(y_{t-h,i} y_{t-h,j}, y_{t,i} y_{t,j}) \\
&= \sum_{t=1}^{T} w_t^2 \text{Var}(y_{t,i} y_{t,j}) + 2\sum_{h=1}^{T-1}\Big[(1-\lambda)^h \sum_{t=h+1}^{T} w_t^2 \text{Cov}(y_{t-h,i} y_{t-h,j}, y_{t,i} y_{t,j})\Big] \, ,
\end{aligned}$$

where the last equality follows from the definition of w_t, that is, from $w_t := \lambda(1-\lambda)^{T-t}$.

To arrive at an estimator of ν_{ij}, denote the sample autocovariance function of the process $\{y_{t,i} y_{t,j}\}_{t=1}^{T}$ by $\hat{\psi}_{ij}(\cdot)$ and introduce a bandwidth parameter $H \geq 0$. Noting

that $\sum_{t=1}^{\infty} w_t^2 = \frac{\lambda^2}{1-(1-\lambda)^2}$, the proposed estimator is then given by

$$\hat{v}_{ij} := \frac{\lambda^2}{1-(1-\lambda)^2}\left[\hat{\psi}_{ij}(0) + 2\sum_{h=1}^{H}(1-\lambda)^h \hat{\psi}_{ij}(h)\right], \tag{D.8}$$

giving rise to $\hat{v} := \sum_{i=1}^{N}\sum_{j=1}^{N} \hat{v}_{ij}$.

An estimator of $\gamma_{ij} := (\phi_{ij} - \sigma_{ij})^2$ is simply given by $\hat{\gamma}_{ij} := (f_{ij} - \hat{\sigma}_{ij})^2$, giving rise to $\hat{\gamma} := \sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\gamma}_{ij}$.

Finally, our data-dependent shrinkage intensity is defined as

$$\hat{\alpha} := \frac{\hat{v}}{\hat{v} + \hat{\gamma}} .$$

We now turn to the estimation of μ. The construction of a data-dependent shrinkage intensity is analogous to the estimation of Σ. Therefore, we only provide the major results and skip some intermediate steps.

The analog to approximation (D.7) now becomes

$$\alpha^* \approx \frac{\sum_{i=1}^{N} \mathrm{Var}(\hat{\mu}_i)}{\sum_{i=1}^{N} \mathrm{Var}(\hat{\mu}_i) + (\bar{\mu} - \mu_i)^2} =: \frac{v}{v+\gamma} \quad \text{with} \quad \bar{\mu} := \frac{1}{N}\sum_{i=1}^{N} \mu_i .$$

Denote the sample autocovariance function of the process $\{x_{t,i}\}_{t=1}^{T}$ by $\hat{\psi}_i(\cdot)$ and introduce a bandwidth parameter $H \geq 0$. Then an estimator of $v_i := \mathrm{Var}(\hat{\mu}_i)$ is given by

$$\hat{v}_i := \frac{\lambda^2}{1-(1-\lambda)^2}\left[\hat{\psi}_i(0) + 2\sum_{h=1}^{H}(1-\lambda)^h \hat{\psi}_i(h)\right], \tag{D.9}$$

giving rise to $\hat{v} := \sum_{i=1}^{N} \hat{v}_i$.

An estimator of $\gamma_i := (\bar{\mu} - \mu_i)^2$ is simply given by $\hat{\gamma}_i := (\hat{\mu}_* - \hat{\mu}_i)^2$, giving rise to $\hat{\gamma} := \sum_{i=1}^{N} \hat{\gamma}_i$.

Finally, our data-dependent shrinkage intensity is defined as

$$\hat{\alpha} := \frac{\hat{v}}{\hat{v} + \hat{\gamma}} .$$

References

Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59(3):817–858

Atkeson A, Ohanian LE (2001) Are Phillips curves useful for forecasting inflation? Q Rev 25(Win):2–11

Beck E, Kozbur D, Wolf M (2024) The hedged random forest. Working paper. Available at SSRN: https://ssrn.com/abstract=5032102
Borup D, Schütte ECM (2022) In search of a job: forecasting employment growth using Google trends. J Bus Econ Stat 40(1):186–200
Breiman L (2001) Random forests. Mach Learn 45:5–32
Chatfield C, Xing H (2019) The Analysis of Time Series: An Introduction With R, 7th edn. CRC Press, Boca Raton
Chen X, Yu D, Zhang X (2024) Optimal weighted random forests. J Mach Learn Res 25(320):1–81
De Nard G (2020) Oops! I shrunk the sample covariance matrix again: blockbuster meets shrinkage. J Financ Econom
Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–263
Faust J, Wright J (2013) Forecasting inflation. In: Handbook of Economic Forecasting, volume 2, chapter 1. Elsevier, pp 2–56
Garcia M, Medeiros M, Vasconcelos GF (2017) Real-time inflation forecasting with high-dimensional models: the case of Brazil. Int J Forecast 33(3):679–693
Goulet Coulombe P, Leroux M, Stevanovic D, Surprenant S (2021) Macroeconomic data transformations matter. Int J Forecast 37(4):1338–1354
Goulet Coulombe P, Goebel M, Klieber K (2024) Dual interpretation of machine learning forecasts. Working paper. Available at SSRN: https://ssrn.com/abstract=5029492
Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecast 13(2):281–291
Hastie TJ, Tibshirani R, Freedman JH (2017) The Elements of Statistical Learning, 2nd edn. Springer, New York
Ledoit O (1995) Essays on risk and return in the stock market. PhD thesis, Massachusetts Institute of Technology, Sloan School of Management. Available online at http://dspace.mit.edu/handle/1721.1/11875
Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J Empir Financ 10(5):603–621
Ledoit O, Wolf M (2022) Quadratic shrinkage for large covariance matrices. Bernoulli 28(3):1519–1547
Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2(3):18–22
Longerstaey J, Zangari P (1996) Risk metrics
McCracken MW, Ng S (2016) Fred-md: a monthly database for macroeconomic research. J Bus Econ Stat 34(4):574–589
Medeiros M, Mendes EF (2016) L1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. J Econom 191(1):255–271
Medeiros MC, Vasconcelos GF, Veiga Á, Zilberman E (2021) Forecasting inflation in a data-rich environment: The benefits of machine learning methods. J Bus Econ Stat 39(1):98–119
Pham H, Olafsson S (2020) On Césaro averages for weighted trees in the random forest. J Classif 37(1):223–236
Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol 4(1). Article 32
Stock JH, Watson MW (2007) Why has U.S. inflation become harder to forecast? J Money Credit Bank 39(s1):3–33
Winham SJ, Freimuth RR, Biernacka JM (2013) A weighted random forests approach to improve predictive performance. Stat Anal Data Min ASA Data Sci J 6(6):496–505
Wright MN, Ziegler A (2017) ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 77(1):1–17