A Machine Learning Approach to the Forecast Combination Puzzle

Antoine Mandel[a], Amir Sani[a,b,*]

[a] *Centre d'Économie de la Sorbonne*
Maison de sciences économiques
106-112 Boulevard de l'hôpital
75013 Paris (France)
[b] *CFM-Imperial Institute of Quantitative Finance*
Department of Mathematics
Imperial College London
SW7 2AZ

Abstract

Forecast combination algorithms provide a robust solution to noisy data and shifting process dynamics. However in practice, sophisticated combination methods often fail to consistently outperform the simple mean combination. This "forecast combination puzzle" limits the adoption of alternative combination approaches and forecasting algorithms by policy-makers. Through an adaptive machine learning algorithm designed for streaming data, this paper proposes a novel time-varying forecast combination approach that retains distribution-free guarantees in performance while automatically adapting combinations according to the performance of any selected combination approach or forecaster. In particular, the proposed algorithm offers policy-makers the ability to compute the worst-case loss with respect to the mean combination ex-ante, while also guaranteeing that the combination performance is never worse than this explicit guarantee. Theoretical bounds are reported with respect to the relative mean squared forecast error. Out-of-sample empirical performance is evaluated on the [56] seven-country dataset and the ECB Survey of Professional Forecasters.

Keywords: forecast combinations; machine learning; econometrics.

---

*Corresponding author
Email addresses: antoine.mandel@univ-pars1.fr (Antoine Mandel), reachme@amirsani.com (Amir Sani)

1. Introduction

Macroeconomic forecasts provide crucial inputs to decision-makers addressing monetary and fiscal policy issues. Forecast accuracy depends on a selected model's power to extract useful and meaningful information from available macroeconomic time series as each observation is received. The nature of macroeconomic time series limits forecasting models to a limited number of noisy aggregated samples across varying economic conditions within an unstable forecasting environment. As parameters are normally estimated over an interval of data, models generally suffer from misspecification, estimation errors and inconsistency. Even in the case where parameters are estimated in a "time-varying" manner, where parameters are estimated as new observations arrive, models generally make such strong assumptions on the process that their performance is inconsistent or unable to handle the real-time dynamics underlying the process [56, 57].

Forecast combination methods introduced by [4] offer a simple solution to these challenges. In particular, they often outperform forecasting approaches that estimate parameters on noisy data, structural breaks, inconsistent predictors and changing environmental dynamics [59]. Unfortunately forecast combination methods often fail to consistently outperform the mean combination over varying *pools* of forecasters and varying horizons. This paper offers the first automatic procedure to manage this so-called "forecast combination puzzle". Accordingly, a large body of research has focused on the theoretical and empirical development of complex forecast combination procedures that aim to fully exploit the information content within a pool of forecasters. However, empirical results in the literature demonstrate that existing forecast combination approaches fail to consistently outperform the mean [see e.g 57, for the case of output and inflation considered in this paper]. This negative result is often referred to as the mean forecast combination puzzle.

Building on recent advances in the machine learning literature [in particular 51], this paper introduces the only automatic procedure to manage this puzzle. First, we recast the forecast combination setting as a game of "prediction with expert advice" [8]. Next, we adapt the general structure of the $\mathcal{AB}$-Prod algorithm [see 51] to automatically hedge performance against the mean combination, inheriting its distribution-free performance guarantees [1]. Finally,

[1]Note that these distribution-free performance guarantees hold for any stochastic, non-stationary, noisy, shifting or real-time process, enabling decision-makers to deploy this automated procedure in real-time environments without concern for the forecast combination puzzle.

we refine the interpretation of the performance bound with explicit rates and the ability to compute ex-ante the risk of underperformance to the mean. Note that this is the only approach that guarantees the worst-case loss to the mean combination ex-ante. We illustrate this contribution by demonstrating a systematic real-time performance advantage against the mean in the Stock and Watson [57] seven-country dataset on output and inflation and the ECB survey of professional forecasters (SPF).

The paper proceeds as follows. In Section 2, we briefly review the relevant forecast combination and machine learning literature. In Section 3, we propose a theoretically guaranteed forecast combination approach that "hedges" performance against the mean with synthetic results. In Section 4, we illustrate the workings of these algorithms with synthetic data. In Section 5, we demonstrate the real-time performance of our approach for the forecast of output and inflation in the framework of Stock and Watson [57] and Euro-area SPF. Section 6 concludes.

2. Literature Review

Real macroeconomic data is observed at an aggregate level and often composed of a small sample of time series observations. Traditional least-squares forecasters often fail to forecast such series due to the limited number of samples, noise and model misspecification [59, 19, 22, 36]. Additionally, macroeconomic models often depend on the configuration of shocks hitting the economy, policy regimes and other institutional factors. This unstable real-world environment results in inconsistent forecasters. Forecast combination approaches offer a simple procedure for exploiting the information content of candidate forecasters, while ignoring the need for explicit model selection[2]. Generally, forecast combination approaches compute linear preference weights up until the most recent observation and fix the estimated parameters until some fixed re-estimation interval. "Time-varying" combination algorithms update parameter estimates as each observation arrives. The method presented in this paper subscribes to the later *time-varying* weights [see 59, for a recent survey].

Theoretical results from the literature demonstrate that gains achieved through combination weights are caused by forecast model instability, where increasing instability in individual forecasts results in a larger advantage (see e.g. [36, 21, 13, 14, 16, 47, 59, 1]). Combination weights provide a robust solution to small sample sizes, noise, regime shifts, model misspecification,

[2]This work deals with a finite pool of candidate forecasters. Other works address the case of a very large to infinite pools of forecasters. See [24, 25, 60]

diverse information sets, unstable forecasters and provide an efficient way to improve forecasting performance by diversifying over a pool of forecasts (for a survey, see [20, 43, 15, 12, 59, 39].). Practical successes in the forecast combination literature include output and inflation [56], interest rates [35], money supply [34], monetary policy [40], equity premiums [48], commodities [9] and realized volatility [45].

Unfortunately, many proposed combination approaches unrealistically assume a stationary, and often known, covariance structure [64, 49], relying on asymptotic guarantees that ignore the estimation error resulting from noisy aggregated data, small sample sizes and changing process dynamics [10]. Under real-world conditions, such as in macroeconomic series, estimation convergence guarantees from the theoretical literature fail and the divergence between empirical and actual weights can be substantial [61]. In fact, [23] found that optimizing combination weights based on error minimization resulted in a gain that is often overshadowed by estimation error. Accordingly, several recent works have stressed the need for combination approaches that can deal with real-world instabilities [see e.g 2, 26, 55, 46, 58, 3]. Finally, theoretical results prove that many of these methods fail to consistently outperform the simple average.

The inability to consistently outperform the mean is referred to as the "Forecast Combination Puzzle" and has been explained as the biased weighting of "optimal" weights due to the low predictive content of candidate forecasts [39]. This underperformance to the mean is further explained as the result of finite sample bias, model misspecification, unobserved variables, noise and changes in the underlying process [56, 57, 10, 54, 55, 11, 39]. Empirical and theoretical results demonstrate no consistent advantage in alternative means (geometric, trimmed, corrected) or the median over different horizons and endogenous variables [57]. [59, 38] illustrate the specific conditions where the relative gain from the true ex-post optimal weights over an unbiased mean combination are negligible. With regard to out-of-sample performance, [39] showed several simple cases where the mean combination approach even outperforms a linear model set to the data generating process.

Given the inconsistent temporal dynamics of macroeconomic data, the unbiased weights of the mean may not always provide the best performance. Time-varying combination weights have demonstrated great potential in outperforming the mean combination (See e.g. [44, 37, 41, 33, 64, 41, 53, 50, 49, 5, 59, 18]). Another problem cited in the literature is the restrictive nature of forecasts based solely on the mean, median and mode over the pool of forecasts [32]. One insight from [32, 28] is to forecast based on the distribution over the pool of forecasts. These works demonstrate the advantage of learning a distribution of weights according to the time-varying performance of the pool. Unfortunately, the statistical assumptions underlying many of these time-varying approaches are often too restrictive for real data. In particular, many of these approaches assume a model on the temporal dynamics, such as Markov switching [63], or a known or stationary covariance structure [64, 49, 33, 53, 41] or normality conditions on the residuals [see 10]. This results in inconsistent performance in real-time data environments. The machine learning literature on "prediction with expert advice" [see 8] exploits the density of forecaster performance, while also providing worst-case performance guarantees that make the least restrictive assumptions on the process generating the target time series. In particular, [51] provides distribution-free theoretical guarantees to both a benchmark algorithm $\mathcal{B}$ and alternate algorithm $\mathcal{A}$.

3. Theoretical Results

A large share of algorithms from the forecast combination literature focus on determining optimal weights from a fixed interval of observations. In particular, these approaches assume that the optimal weights can be estimated without any errors in estimation and that the in-sample fit is optimal for any out-of-sample forecasting exercise. In practice, these approaches make strong statistical assumptions to justify the in-sample estimation of linear combination weights $\mathbf{w}$. Unfortunately, these procedures fail to consistently outperform the unbiased mean combination, i.e. the forecast combination determined by fixed uniform weights $1/K$ over each of the K forecasters. A natural approach to manage this problem is to estimate time-varying combination weights $\mathbf{w}_t$ and exploit the distribution of performance observed over the K forecasters. This paper introduces a novel combination algorithm that provides an ex-ante characterization of risk, while automatically managing the forecast combination puzzle. In particular, once an acceptable level of worst-case performance to the mean combination is set, the learning algorithm automatically manages the relative mean-square forecast error of combination weights $\mathbf{w}_t$ according to the observed losses over the history of observations.

More formally, we consider a setting in which at a sequence of dates $t = 1 \cdots T$, a decision-maker has at his disposal K forecasts $(\hat{y}_{1,t+h}, \ldots, \hat{y}_{K,t+h}) \in \mathbb{R}^K$ of the value at horizon h of a variable of interest. As observation of the realized value y_t of the variable arrives, a time-varying forecast combination algorithm $\mathcal{G}$ learns convex decision weights $\mathbf{w}_t$ in the K-dimensional simplex $\mathcal{S} := \left\{\mathbf{w} \in \mathbb{R}_+^K : \sum_{i=1}^K w_i = 1\right\}$ as a function of the past performance of the forecaster [3]. The decision-maker then determines a point forecast of the variable of interest for the current period as follows.

1. Observe forecasts, $\hat{y}_{1,t+h}, \ldots, \hat{y}_{K,t+h}$ and realization y_t.
2. Update the vector of combination weights $\mathbf{w}_t$.
3. Compute a point forecast $\hat{y}_{t+h} = \sum_{i=1}^{K} w_{i,t}\hat{y}_{i,t+h}$.

In this paper, we focus on a class of algorithms inspired from the machine learning literature [see 8] where time-varying decision weights $\mathbf{w}_t$ are computed according to the following steps:

1. Compute the quadratic loss for each forecaster $i \in K$, $l_{i,t} = (y_t - \hat{y}_{i,t})^2$.
2. Compute a score $\lambda_{i,t}$ for each forecaster $i \in K$, as a function of the losses $l_{i,t}$.
3. Compute the forecasting weights $w_{i,t}$ by normalizing the scores according to:

$$w_{i,t} = \frac{\lambda_{i,t}}{\sum_{i=1}^{K} \lambda_{i,t}}, \forall i \in K.$$

In the forecasting literature, the performance of such an algorithm, which we denote by $\mathbf{w}$ is usually measured through the mean-square forecast error, defined as:

$$\mathbf{MSFE_w} = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{K} w_{i,t+h} l_{i,t+h}, \tag{1}$$

It is also standard to compare the performance to a benchmark, usually the simple AR model or a random walk. In line with our objective to address the mean forecast combination puzzle, we rather measure performance relative to the mean combination, which has proven to be an incredibly difficult comparator [see e.g. 59]. Hence we define the Relative Mean-Squared Forecast Error (RMSFE) of a combination algorithm $\mathbf{w}$ as:

$$\mathbf{RMSFE_w} = \frac{\mathbf{MSFE_w}}{\mathbf{MSFE}_\mu}.$$

In the machine learning literature, performance is rather measured through the notion of regret, which provides a cumulative measure of performance. Namely, given the cumulative performance of a combination algorithm $\mathbf{w}$

[3] This is in line with the bulk of the forecast combination literature [see 59] and the "Prediction with Expert Advice" framework [see 8]

$$L_{\mathbf{w},T} = \sum_{t=1}^{T}\sum_{i=1}^{K} w_{i,t+h} l_{i,t+h},$$

The regret with respect to a forecaster i is then defined as

$$\mathcal{R}_{\mathcal{G},t}(i) = L_{\mathcal{G},t} - L_{i,t},$$

where i stands for the algorithm that constantly puts full weight on the ith forecaster.

Regret is usually measured with respect to the ex-post optimal choice in hindsight, $i^* := \arg\min_{i\in K} L_{i,T}$ and an algorithm is said to be "learning" if its cumulative regret grows linearly with respect toi^* (note that the regret is negative if the algorithm outperforms i^*). In the following, we emphasize how one can build on existing regret bounds for machine-learning algorithms to construct a combination algorithm with guaranteed relative performance to the mean in terms of RMSE.

Let us first recall that the definition of the exponentially weighted average forecaster, or **Hedge** (see e.g. [29, 42, 62, 8]), in which scores are exponentially updates as:

$$\lambda_{i,t} = \lambda_{i,t-1} \exp(-\eta l_{i,t-1}), \forall i \in K, \tag{2}$$

Hedge achieves an "optimal" $\mathcal{O}(\sqrt{T \log K})$ regret to any forecaster i, including the ex-post optimal choice i^*, for all possible realization of the loss sequence (See Theorem 1). This bound can not be improved by any exponentially weighted forecaster [see 8, Theorem 2.2].

Theorem 1. *[Theorem 2.6 in 8] For any finite horizon T, forecasters K and optimized learning rate $\eta = \sqrt{\frac{8\log K}{T}}$, the regret upper bound for **Hedge** satisfies,*

$$\mathcal{R}_{\textbf{\textit{Hedge}},T}(i) \leq \sqrt{\frac{T}{2}\log K},$$

against any forecaster i.

Now, in the context of macroeconomic forecasting, the best forecaster in hindsight might not be an appropriate benchmark. Regime switches within macroeconomic time series offer a simple justification for this point. Hence, this paper aims to upper bound the worst-case regret with respect to the mean combination forecaster, while also maintaining an optimal $\mathcal{O}(\sqrt{T \log K})$ bound to a *pool* of alternative forecasters. A naive approach to achieve this objective is to run the **Hedge** algorithm over a pool of forecasters that includes the mean combination forecaster. Unfortunately, given the exponential updating mechanism within Hedge, the best we can hope to achieve is a *uniform* $\mathcal{O}(\sqrt{T \log K})$ upper bound to the pool [see 27]. **Hedge's** uniform bound fails to satisfy our objective.

In order to overcome this failure, one requires an analytic expression of the regret bound that is specific to a forecaster. Such an analytic expression can be obtained by replacing the exponential weight update of Hedge $\exp \mu x$ by its linear approximation, $1 + \mu x$. The resulting algorithm, usually referred to as Prod, is then defined by the following update rule

$$\lambda_{i,t+1} = \lambda_{i,t}(1 - \eta l_{i,t}), \forall i \in K. \tag{3}$$

It provides an explicit characterization of the regret with respect to a forecaster as a function of its losses. Namely, one has:

Theorem 2. *[Theorem 2.5 in 8] For any finite horizon T and learning rate $\eta \in (0, 1/2]$, **Prod** satisfies the following second-order regret bound,*

$$\mathcal{R}_{\textbf{\textit{Prod}},T}(i) \leq \eta \sum_{t=1}^{T} l_{i,t}^2 + \frac{\log K}{\eta},$$

for any forecaster i, and the following regret bound with optimized learning rate $\eta = \sqrt{\frac{\log K}{T}}$,

$$\mathcal{R}_{\textbf{\textit{Prod}},T}(i) \leq 2\sqrt{T \log K}. \tag{4}$$

Then, in order to obtain dual regret bounds, one with respect to a benchmark and the other with respect to the remaining forecasters, [27] introduce an alternative normalization rule in **Prod** for the scores of the different forecasts. The resulting algorithm **D-Prod** is defined over an extended set of $K + 1$ forecasters: the K original forecasts plus a fixed combination of these D. The score of the benchmark is kept fixed while the score of the other forecasters are updated as a function of their relative performance with respect to this of the benchmark, namely as:

$$\lambda_{i,t+1} = \lambda_{i,t}(1 - \eta(l_{i,t} - l_{D,t})), \text{for } i = 1, \ldots, K.$$

This modified update results in two bounds. A $\mathcal{O}\left(\sqrt{T \log K} + \sqrt{\frac{T}{\log K}} \log T\right)$ bound to the K experts and another constant regret bound to the fixed distribution D.

[51] extend this result by refining the algorithm to achieve optimal regret bounds for "easy" and "hard" sequences simultaneously. In particular, they modify the algorithm to bound performance in the "worst-case" through an algorithm $\mathcal{A}$ and the "easy" case with an algorithm $\mathcal{B}$, providing state-of-the-art regret performance in both easy and hard loss sequences. In particular, the score $\lambda_{\mathcal{B}}$ is fixed at initialization and the score for $\mathcal{A}$ is updated as,

$$\lambda_{\mathcal{A},t+1} = \lambda_{\mathcal{A},t}(1 + \eta(l_{\mathcal{B},t} - l_{\mathcal{A},t})), \tag{5}$$

where the deviation is no longer taken with respect to a fixed distribution D, but to $\mathcal{B}$, and both $\mathcal{A}$ and $\mathcal{B}$ are combination algorithms mixing over the same pool of forecasters. More precisely, $\mathcal{AB}$-Prod is computed as follows:

Input:

- Combination algorithms $\mathcal{A}$ and $\mathcal{B}$
- A history of observations for the target variable, in our case output or inflation, up to the current time t, of length T.
- Preference weight $\lambda_{\mathcal{B}} \in (0, 1)$ for the algorithm $\mathcal{B}$.

Initialization:

- $\lambda_{\mathcal{A},0} = 1 - \lambda_{\mathcal{B}}$
- Learning rate $\eta = \min\left(\sqrt{\frac{-\log(1-\lambda_{\mathcal{B}})}{T}}, \frac{1}{2}\right)$
- Set $s_0 = (TODOAMIRcomplete)$

Repeat the following for each observation from time $t = 0, \ldots, T$:

1. Compute combination weight $s_t = \dfrac{\lambda_{\mathcal{A},t}}{\lambda_{\mathcal{A},t} + \lambda_{\mathcal{B}}}$.
2. Observe the target variable y_t and compute the loss $l_{\mathcal{A},t}$ and $l_{\mathcal{B},t}$.
3. Compute the combination loss $l_{\mathcal{AB}-\text{Prod},t} = s_t l_{\mathcal{A},t} + (1 - s_t) l_{\mathcal{B},t}$.
4. Compute the deviation $\delta_t = l_{\mathcal{B},t} - l_{\mathcal{A},t}$
5. Update the Score $\lambda_{\mathcal{A},t+1} = \lambda_{\mathcal{A},t}(1 + \eta\delta_t)$

One the obtains the following performance guarantee:

Theorem 3 (cf. Theorem 1 in [51]). *Let $\mathcal{A}$ be any algorithm, $\mathcal{B}$ be any benchmark and D be an upper bound on the benchmark losses $L_{\mathcal{B},T}$. Then setting weight $\lambda_{\mathcal{B}} \in (0,1)$, $\lambda_{\mathcal{A}} = 1 - \lambda_{\mathcal{B}}$, Learning rate $\eta = C\sqrt{\frac{1}{T}} < \frac{1}{2}$, where $C = \sqrt{-\log(1-\lambda_{\mathcal{B}})}$ simultaneously guarantees,*

$$\mathcal{R}_{\mathcal{AB}\text{-}Prod,T}(i) \leq \mathcal{R}_{\mathcal{A},T}(i) + 2C\sqrt{D},$$

for any forecaster i and,

$$\mathcal{R}_{\mathcal{AB}\text{-}Prod,T}(i) \leq \mathcal{R}_{\mathcal{B},T}(i) + 2\log 2,$$

against any assignment of the loss sequence.

With the proper tools, we now focus on our stated objective to design an automatic approach to managing the forecast combination puzzle. First, we would like an approach that allows us to insure that any forecast combination approach we select is protected from underperformance to the mean combination μ. Second, the approach must be fully adaptive and automatic in managing its parameters after initialization and in real-time application. This later point is critical for practical consideration of the algorithm within a real-time policy-making environment. Third, the approach should stand up to worst-case and unexpected changes in the process, such as during the financial crisis. Finally, the approach must have a set of parameters that are easy to interpret and a clear measurable metric characterizing the level of protection it offers against the mean forecaster. Though the first three points can be claimed through an adaptation of the $\mathcal{AB}$-Prod algorithm, the later can not.

Here, we introduce the $\mathcal{A}\mu$-Prod algorithm and present results with regard to the RMSFE. More precisely, we set the mean combination forecaster μ as algorithm $\mathcal{B}$ to exploit the constant regret guarantee and select any alternative combination algorithm as $\mathcal{A}$, while running both μ and $\mathcal{A}$ on a fixed pool of forecasters. Given that the RMSFE is the standard performance metric in the forecast combination literature [see 59], we express the regret bounds in Theorem 3 in terms of the RMSFE. Further, we present the exact calculation of the constants in the bound to provide an ex-ante computable measure of this bound. Note that no other approach can provide such insight ex-ante. This paper provides the only automatic approach to managing the forecast combination puzzle with an explicit ex-ante calculation of the risk of any combination approach against the mean. It is also the only work that manages this risk without any assumption on the process, adaptively managing this risk over any possible realization of the loss sequence. We consider this to be a significant result.

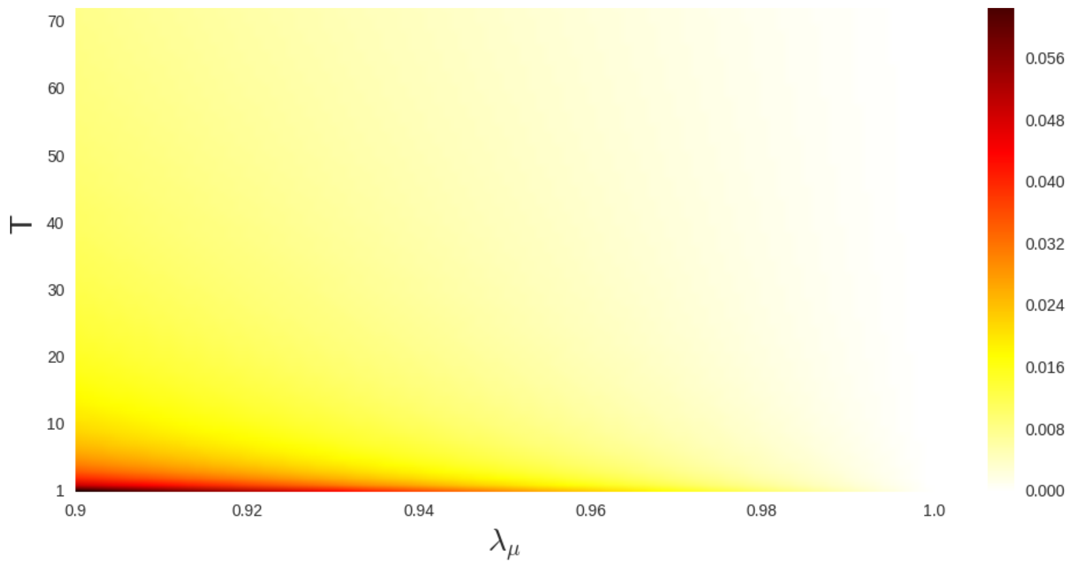

Figure 1: $\frac{\log(\lambda_\mu)}{T\eta}$ as a function of samples T and preference λ_μ.

For any assignment of the loss sequence, the total expected loss of $\mathcal{A}\mu$-Prod initialized with μ preference weight $\lambda_\mu \in (0,1)$, simultaneously satisfies,

Theorem 4. *Let $\mathcal{A}$ be any forecast combination algorithm and $\lambda_\mu \in (0,1)$ be the preference weight for the mean. Then setting $\eta = \sqrt{\frac{-\log(1-\lambda_\mu)}{T}} < \frac{1}{2}$ as the learning rate, simultaneously guarantees,*

$$\boldsymbol{RMSFE}_{\mathcal{A}\mu\text{-}Prod} \leq \boldsymbol{RMSFE}_{\mathcal{A}} + \sqrt{2}\eta,$$

and,

$$\boldsymbol{RMSFE}_{\mathcal{A}\mu\text{-}Prod} \leq 1 - \frac{\log(\lambda_\mu)}{T\eta},$$

against any assignment of the loss sequence.

The ability for $\mathcal{A}\mu$-Prod to manage the performance of any combination algorithm $\mathcal{A}$ depends on the setting for λ_μ. In particular, the second term of the second bound in Theorem 4, determines the hedging behaviour of $\mathcal{A}\mu$-Prod. It's clear from Figure 1 that setting λ_μ has less impact as additional samples are available. Unfortunately, in the case of macroeconomic data, it is often the case that the data is measured in Quarters or Months and this

T (Quarters)	$\mathcal{A}+\sqrt{2}\eta$	$1-\frac{\lambda_\mu}{T\eta}$
4 (1 Year)	$\mathcal{A}+2.62826$	1.00019
8 (2 Years)	$\mathcal{A}+1.85846$	1.000135
12 (3 Years)	$\mathcal{A}+1.51742$	1.00011
16 (4 Years)	$\mathcal{A}+1.31413$	1.000095
20 (5 Years)	$\mathcal{A}+1.17539$	1.000085
40 (10 Years)	$\mathcal{A}+0.83113$	1.000060
80 (20 Years)	$\mathcal{A}+0.587697$	1.000043

Table 1: A conservative $\lambda_\mu = 0.999$. Note that the performance of $\mathcal{A}\mu$-Prod is guaranteed to be the minimum with respect to each row.

limits the ability to set a loose λ_μ. At the level of 20 or more samples, we see that the performance is quite even. To better understand this behaviour, we construct a synthetic example with an increasing number of samples and explore the behaviour of this term further. In particular, Tables 1 and 2 present the impact of a conservative preference $\lambda_\mu = 0.999$ and more flexible preference of $\lambda_\mu = 0.90$ along an increasing number of samples. Note that $\mathcal{A}\mu$-Prod guarantees performance that is the minimum for each respective row. For example, Table 1 clearly illustrates that a more conservative preference requires many more samples (greater than 80 quarters or 20 years) or performance from the selected forecast combination algorithm $\mathcal{A}$ that is extraordinarily strong to be preferred. Conversely, we see in Table 2 that much less data is required for an algorithm $\mathcal{A}$ to have a chance of receiving preference. It follows that the performance requirements for $\mathcal{A}$ are also less stringent. In the next section, synthetic Monte-Carlo results will be presented to evaluate this performance further.

4. Synthetic Results

$\mathcal{A}\mu$-Prod offers an automatic approach to protecting a forecast combination algorithm $\mathcal{A}$ against underperformance to the mean. This section presents two scenarios that demonstrate this using synthetic loss sequences corresponding to situations where $\mathcal{A}$ and μ perform poorly. These scenarios are illustrated by setting $\mathcal{A}$ to **AdaHedge**, an adaptive variant of **Hedge** introduced in [see 17]. The two scenarios consist of 1000 losses observed from two forecasters. Losses are observed up until the current time step and have values in $\{0, 1\}$,, i.e. each expert can be right or wrong. Each forecast combination is produced

T (Quarters)	$\mathcal{A} + \sqrt{2}\eta$	$1 - \frac{\lambda_\mu}{T\eta}$
4 (1 Year)	$\mathcal{A} + 1.5174$	1.0347
8 (2 Years)	$\mathcal{A} + 1.0730$	1.0245
12 (3 Years)	$\mathcal{A} + 0.8761$	1.0200
16 (4 Years)	$\mathcal{A} + 0.7587$	1.0174
20 (5 Years)	$\mathcal{A} + 0.6786$	1.0155
40 (10 Years)	$\mathcal{A} + 0.4799$	1.0110
80 (20 Years)	$\mathcal{A} + 0.33931$	1.0078

Table 2: A balanced $\lambda_\mu = 0.90$. Note that the performance of $\mathcal{A}\mu$-Prod is guaranteed to be the minimum with respect to each row.

	Mean	AdaHedge	$\mathcal{A}\mu$-Prod(AdaHedge)
Scenario 0	0.5	13.187577	0.507855
Scenario 1	248.5	2.250753	89.459579

Table 3: Regret

for the time step $t + h$. The regret and RMSFE performance are presented for the mean forecast combination μ, **AdaHedge**, and $\mathcal{A}\mu$-Prod with $\mathcal{A}$ set to AdaHedge in Tables 3 and 4.

In Scenario 1, the mean combination clearly outperforms AdaHedge (see Tables I and II for the evolution in terms of regret and RMSFE). In fact, the performance is such that it is impossible to beat the mean combination approach, especially while changing weights at each round in time. Accordingly, **AdaHedge** is unable to exploit any additional information as new observations arrive, so it underperforms the mean. In contrast, $\mathcal{A}\mu$-Prod quickly recognizes the performance advantage of the mean combination approach as new observations arrive, so it pays a slight Regret in "learning" this advantage and shifts into the mean combination approach for safety.

In Scenario 2, the mean fails to demonstrate an explicit advantage and

	Mean	AdaHedge	$\mathcal{A}\mu$-Prod(AdaHedge)
Scenario 0	1.0	1.025401	1.000016
Scenario 1	1.0	0.507009	0.681601

Table 4: Relative MSE

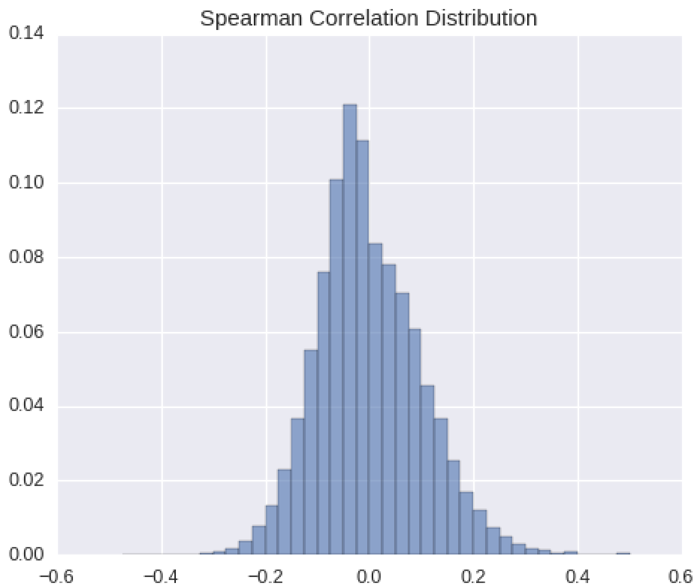

Figure 2: The Spearman correlation distribution corresponding to the loss tuples.

AdaHedge is clearly outperforming. This performance advantage demonstrates that there is clearly one forecaster loss sequence that dominates the other. $\mathcal{A}\mu$-Prod(AdaHedge) recognizes this advantage as observations arrive and shifts weights in each round to the alternative algorithm **AdaHedge** (see tables I and II).

We perform a second exercise to evaluate the hedging performance of $\mathcal{A}\mu$-Prod over a Monte-Carlo of synthetic losses. In particular, we generate $1,000,000$ loss sequences for two forecasters, each of length 20 (equivalent to 5 years), by drawing Random Bernoulli losses (ex. 00101000...) with a parameterization that varies with the Monte-Carlo simulation[4]. The aim of this setting is to investigate the performance of four forecast combination algorithms together with their $\mathcal{A}\mu$-Prod extensions in a conservative scenario, where $\lambda_\mu = 0.999$, and a more flexible scenario, where $\lambda_\mu = 0.90$.

The following baseline algorithms are considered: AdaHedge, the time-varying forecast combination approach of Bates-Granger (see Model 1 from [59]), the Recent Best forecaster (which chooses the forecaster that performed best last period) and the random forecaster (ie. choosing one of the forecasters uniformly at random). The corresponding extensions are denoted by $\mathcal{A}\mu$-

[4]The distribution of the Spearman correlations between the two forecasters over the Monte-Carlo simulations is reported in 2.

Prod(AdaHedge), $\mathcal{A}\mu$-Prod(Bates-Granger), $\mathcal{A}\mu$-Prod(Recent Best) and $\mathcal{A}\mu$-Prod(Random).

For all forecast combination algorithms considered, $\mathcal{A}\mu$-Prod demonstrates its ability to offer a real-time hedge to the mean. This advantage is demonstrated online and in multiple sequential out-of-sample evaluations. In particular, the maximum RMSFE for each baseline algorithm is notably greater than 1, while in the $\mathcal{A}\mu$-Prod instantiation, the maximum RSME is slightly greater than 1. In the latter case, the worst performance over all the algorithms is 1.0116. This corresponds to the ex-ante computable "cost" of using $\mathcal{A}\mu$-Prod in a real-time setting, which is bounded according to Section 3. Finally, note that a very strong preference was set for μ, resulting in a strong restriction in preferring $\mathcal{A}$.

5. Empirical Results

In order to illustrate the empirical value of our approach, we compare the performance of $\mathcal{A}\mu$-Prod to a set of standard forecast combination algorithms from the macro-economic and online learning literature in two independent experiments and over multiple online out-of-sample evaluations. Once again, note that the evaluations are performed online, so each forecast is made out-of-sample. The first forecasting exercise aims to forecast output and inflation using the seven-country dataset from [56]. The second forecasting exercise aims to forecast the Euro-area growth-rate using data from the survey of professional forecasters. The set of algorithms considered include:

- A set of basic combination methods: the mean forecaster (denoted by μ), trimmed mean forecasters with $\alpha = 0.05$ and $\alpha = 0.10$ and the median forecaster.

- Three benchmark time-varying combination methods: the AdaHedge algorithm, which is a version of **Hedge** with adaptive learning rate, the Bates Granger time-varying method 1 (BG) introduced in [59] and the Recent Best forecaster, which selects the forecaster with the lowest loss in the last round.

- The ex-post optimal forecaster, which can of course only be determined ex-post but provides a useful benchmark.

- The random forecaster, which selects a single forecaster at random at each round.

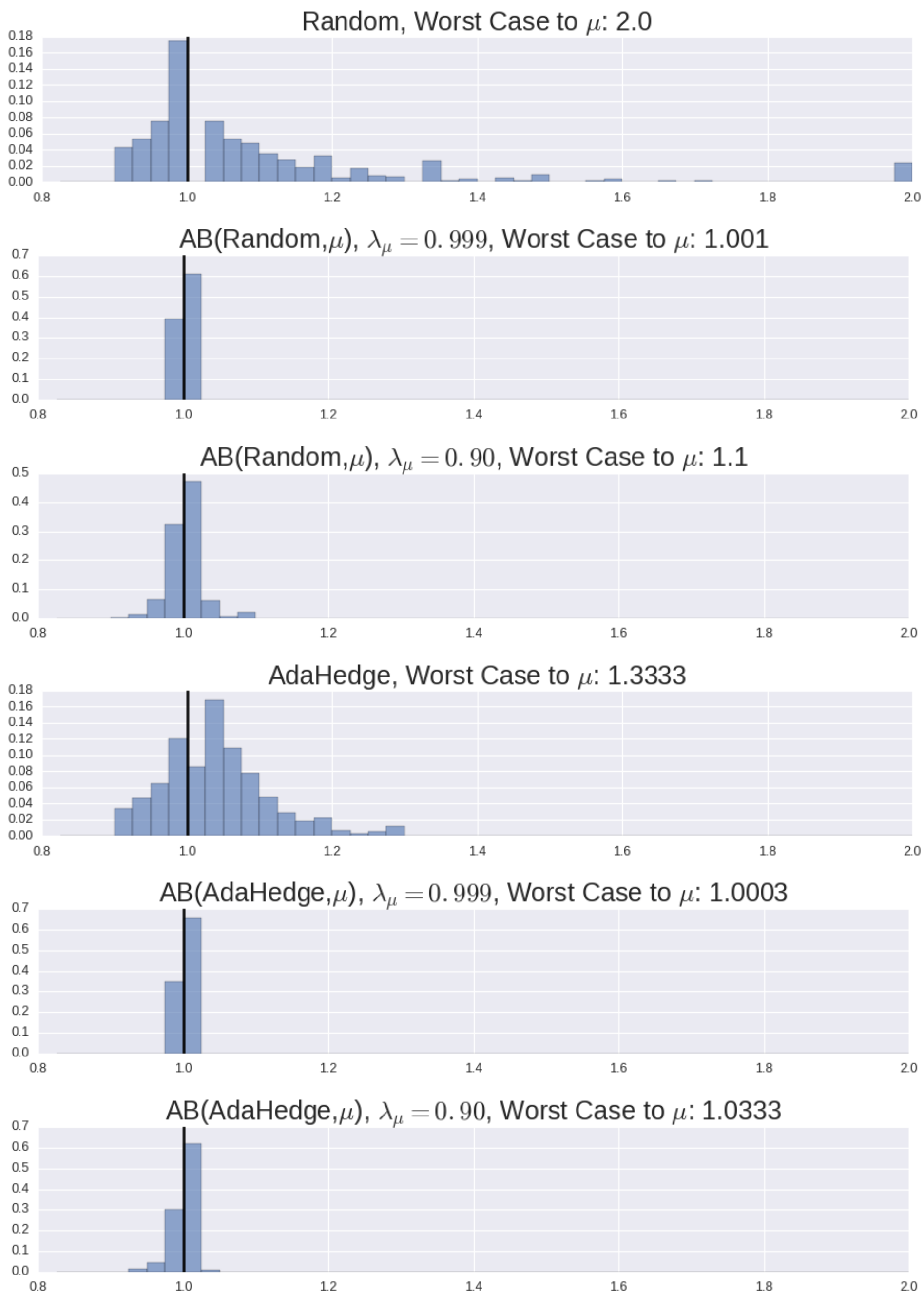

Figure 3: RMSFE over synthetic loss sequences.

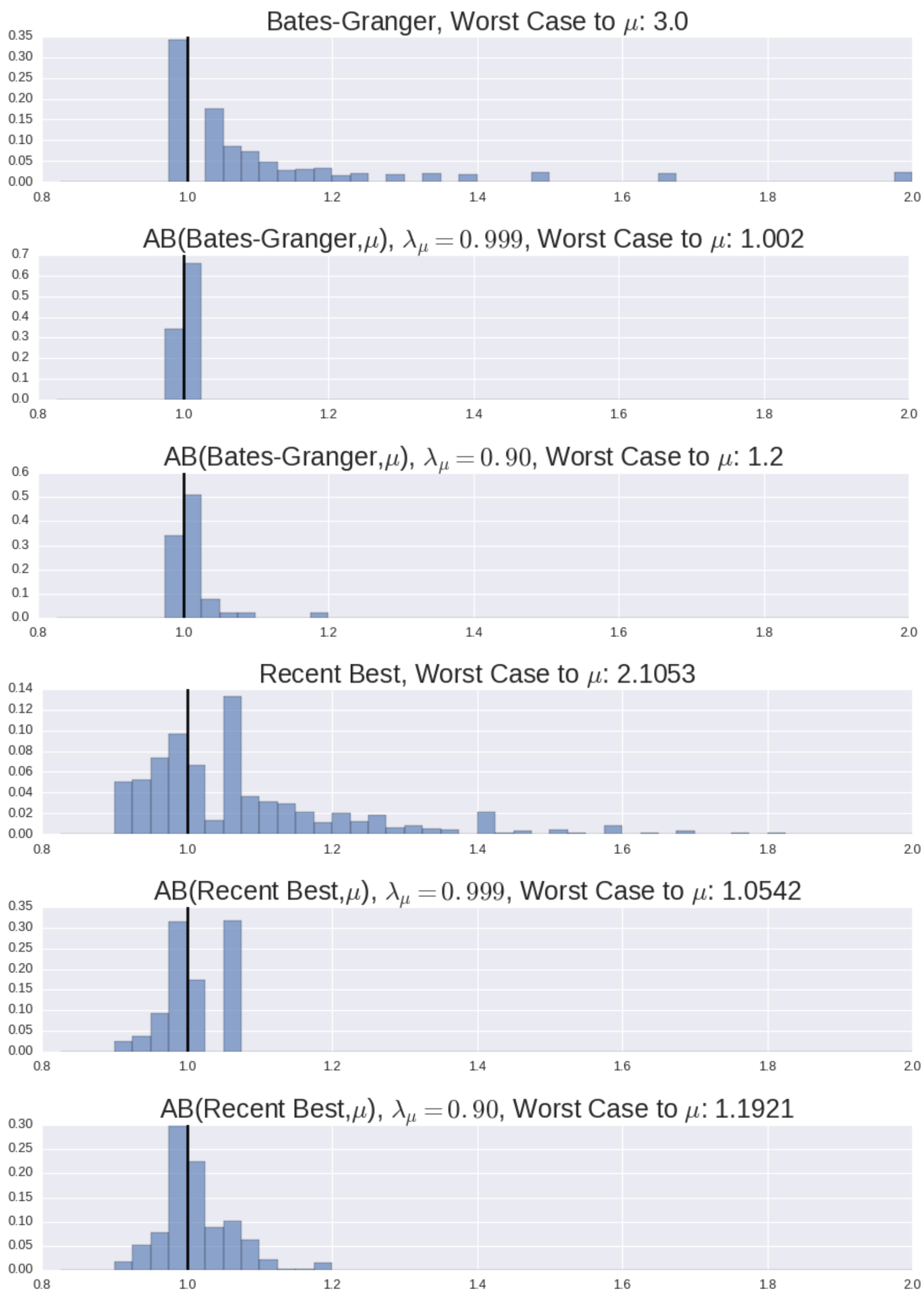

Figure 4: RMSFE over synthetic loss sequences.

- Three instantiations of $\mathcal{A}\mu$-Prod are presented with $\lambda_{\mathcal{B}} = 0.999$, heavily preferring μ.
  - $\mathcal{A}\mu$-Prod(AdaHedge): $\mathcal{A}$ =AdaHedge
  - $\mathcal{A}\mu$-Prod(Bates-Granger): $\mathcal{A}$ =the Bates-Granger method
  - $\mathcal{A}\mu$-Prod(Recent Best): $\mathcal{A}$ =the Recent Best forecaster over the previous round

Remark 1. *Macroeconomic data, and most particularly the SPF, often has missing values , resulting in missing losses that negatively bias otherwise well-performing forecasters. The solution we adopt in the following is to impute missing data with the mean, more precisely to lower-bound the performance of the missing forecaster performance by this of the mean.*

5.1. Seven country forecast combination dataset

The seven-country forecast combination dataset from [56], consists in 43 quarterly time series of macro-economic indicators available for seven different countries: Canada, France, Germany, Italy, Japan, the United Kingdom and United States. The time-series include asset prices, selected measures of real economic activity and money stock from 1959 to 1999. Each of these time-series is then used to produce independent forecasts of inflation and output by estimating an autoregressive model with one exogenous variable (ARX). These forecasts are then combined using our set of candidate algorithms with a burn-in period of 8 quarters. This experiment is then repeated independently for inflation and output for three different forecast horizons, $h = 2, 4$ and 8 quarters.

Remark 2. *The ARX forecasts are generated using observations up to time t for each exogenous variable with the Python Statsmodels library [52]. Coefficients are estimated according to the Akaike information criterion (AIC) on up to* 4 *lags, in accordance with the setting in [56], with ARX forecasts generated using a Broyden-Fletcher-Goldfarb-Shanno solver and maximum likelihood estimation. Failed forecasts due to failed maximum likelihood convergence are replaced with the preceding forecast. Note that this results in ARX forecasts up to each time step t and that forecasts of $t + h$, where h is the horizon, account for an out-of-sample forecast in this setting. This process is repeated until the end of the available data sequence.*

	Average RMSFE	Min RMSFE	Max RMSFE
AdaHedge	1.006743	0.727263	1.424006
Recent Best	1.288761	0.395634	18.447350
Bates-Granger	1.026792	0.726393	1.247406
Median	0.975889	0.723440	1.102208
Trimmed Mean(alpha=0.05)	0.957247	0.719920	1.024449
Trimmed Mean(alpha=0.10)	1.663990	0.659524	3.803069
$\mathcal{A}\mu$-Prod(AdaHedge)	0.952805	0.718049	0.999840
$\mathcal{A}\mu$-Prod(Bates-Granger)	0.952807	0.718049	0.999842
$\mathcal{A}\mu$-Prod(Recent Best)	0.952835	0.718046	0.999843
Random Forecaster	1.051770	0.735312	1.353242
Ex-Post Optimal	0.798261	0.557397	0.974834

Table 5: Average, Minimum and Maximum Ratio to the mean of the Mean Square Forecast Error over GDP, CPI, Horizons and Countries

Table 5 provides a summary of the main results (whose details are in the appendix). The experiment clearly illustrates the performance advantage and adaptive hedging capabilities of the $\mathcal{A}\mu$-Prod meta structure in a real-time forecast combination setting. The three $\mathcal{A}\mu$-Prod algorithms outperform the mean for every possible combination of indicator, country and horizon, i.e the maximal RMSFE ratio is less than 1. In terms of average performance, they outperform all but the ex-post optimal forecaster, which can only be determined ex-post. Average performance is also better than all other alternative time-varying combination algorithms (AdaHedge, Recent Best and Bates-Granger). Moreover, these traditional forecast combination algorithms do not systematically guarantee better performance than the mean. A detailed analysis of the results presented in the appendix shows that $\mathcal{A}\mu$-Prod outperforms AdaHedge, Recent Best and Bates-Granger almost systematically. This suggests that $\mathcal{A}\mu$-Prod manages, thanks to its meta-structure, to quickly recognize regime switches in the real-time data. It also demonstrates a preference for $\mathcal{A}$ when there is a comparative advantage to the mean. Further, this also demonstrates how well $\mathcal{A}\mu$-Prod hedges performance to the safety of the mean when $\mathcal{A}$ fails in a real-time setting. Note again that the preference weight λ_μ defines how aggressively $\mathcal{A}\mu$-Prod hedges the performance of $\mathcal{A}$. Here, we set this very aggressively to demonstrate a solution to the so-called forecast combination puzzle. In cases where safety is not critical, one might prefer to pay a larger cost to exploit any intermittent advantage provided by $\mathcal{A}$.

5.2. Survey of professional forecasters

The Euro-area Survey of Professional Forecasters has been conducted by the European Central Bank at a quarterly frequency since the inception of the European Monetary Union [see 6, 7, 30, for a detailed description]. There are around 75 survey participants, who are experts affiliated with financial and non-financial European institutions. The average number of respondents per survey is 59. Each participant[5] receives a survey of growth expectations for 1 and 2 year rolling horizons[6], with one week to reply. Survey results are published the following month. Further, the target forecast changes depending on the specific criterion by which the GDP is measured. These experts are asked to provide point forecasts for GDP and inflation at different horizons (we focus on the 1-year rolling GDP forecast horizon). Their answers provide a time-series of forecasts that are natural inputs for a forecast combination approach.

Remark 3. *The SPF suffers from a large number of missing values with less than 60 respondents in average. We have considered two approaches to overcome this issue: the reduction to a balanced panel, as is common in the SPF literature (see e.g. [26, 1, 31]), and imputation of the missing values through the mean of available forecasts at the specific time step. Both approaches give similar results. Due to these gaps and the frequency of mean imputations, forecaster performance is expected to be close to the mean of existing forecasts.* $\mathcal{A}\mu$ *instantiations are set to* $\lambda_\mu = 0.999$.

[5]Note that the specific expert at each institution is not necessarily the same in each survey and the data has many missing values.

[6]Note that the rolling horizons are set one and two years ahead of the latest period for which the variable in question is observed when the survey is conducted and not one or two years ahead of the survey date.

	Balanced	Imputed
AdaHedge	0.969944	0.969086
Recent Best	0.810353	0.808181
Bates-Granger	1.021676	1.010879
Median	0.995093	0.995307
Trimmed Mean(alpha=0.05)	0.997080	0.996030
Trimmed Mean(alpha=0.10)	0.985286	1.006134
$\mathcal{A}\mu$-Prod(AdaHedge)	0.995106	0.995319
$\mathcal{A}\mu$-Prod(Bates-Granger)	0.995108	0.995316
$\mathcal{A}\mu$-Prod(RB)	0.995831	0.996091
Random Forecaster	0.969570	0.963621
Ex-Post Optimal	0.862678	0.860968

Table 6: SPF Data: Imputed and Balanced

Table 6 reports the performance of the forecast combination algorithms in this setting. The three $\mathcal{A}\mu$-Prod algorithms are close to or outperform the mean combination forecast. The cost of protection is apparent in the gap between the $\mathcal{A}\mu$ instantiations and non-$\mathcal{A}\mu$ forms of the algorithms. This is most clear in the gap in performance between the $\mathcal{A}\mu$ and non-$\mathcal{A}\mu$ forms of the Recent Best algorithm. Given the lack of structure in the data resulting from gaps, changing forecasters and inconsistent forecast histories per forecaster, the problem of "learning" is expected. In each case, the $\mathcal{A}\mu$ instantiations learn to prefer the safety and consistency of the mean. If a larger margin of error was acceptable, one might consider reducing the weight λ_μ.

6. Discussion and Conclusions

This paper introduces the only algorithm that automatically manages the forecast combination puzzle. The proposed algorithm adapts the structure of the $\mathcal{AB}$-Prod algorithm introduced in [51] to solve the novel problem of automatically managing the forecast combination puzzle within macroeconomic forecasting. The result is the first distribution-free performance guarantees for both the mean combination and any alternative combination.

With the second bound in Theorem 4, the ex-ante risk of underperforming the mean and suffering from the "forecast combination puzzle", can now be computed in advance and set according to the user's preference. Not only does this explicit bound provide protection against underperformance, it also enables users to test new ideas and manage an acceptable risk ex-ante. The automatic algorithm proposed in this paper offers the first algorithmic mechanism for this level of ex-ante control in the forecast combination literature. Further, it does so without making strong statistical assumptions on the generating process.

Acknowledgments

We gratefully acknowledge the support of H2020-FET proactive project DOLFINS and hardware support from NVIDIA Corporation, AdapData SAS and the Grid5000 testbed that used for this research. We also thank Raffaella Giacomini, Marco Lippi and Gilles Stoltz for their helpful comments.

[1] Aiolfi, M., Capistrán, C., and Timmermann, A. G. (2010). Forecast combinations. *CREATES research paper*, (2010-21).

[2] Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1):31–53.

[3] Armstrong, J. S. (2001). Combining forecasts. In *Principles of forecasting*, pages 417–439. Springer.

[4] Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Or*, pages 451–468.

[5] Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, 177(2):213–232.

[6] Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., and Rautanen, T. (2007). The ecb survey of professional forecasters (spf)-a review after eight years' experience. *ECB Occasional Paper*, (59).

[7] Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., and Rautanen, T. (2010). An evaluation of the growth and unemployment forecasts in the ecb survey of professional forecasters. *OECD Journal. Journal of Business Cycle Measurement and Analysis*, 2010(2):63.

[8] Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games.* Cambridge university press.

[9] Chen, Y.-C., Rogoff, K., and Rossi, B. (2008). Can exchange rates forecast commodity prices? Technical report, National Bureau of Economic Research.

[10] Claeskens, G., Magnus, J. R., Vasnev, A. L., and Wang, W. (2014). The forecast combination puzzle: A simple theoretical explanation.

[11] Clark, T. E. and McCracken, M. W. (2009). Combining forecasts from nested models*. *Oxford Bulletin of Economics and Statistics*, 71(3):303–329.

[12] Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583.

[13] Clements, M. and Hendry, D. (1998). *Forecasting economic time series*. Cambridge University Press.

[14] Clements, M. and Hendry, D. (1999). Forecasting non-stationary economic time series: the zeuthen lectures on economic forecasting.

[15] Clements, M., Hendry, D. F., et al. (2002). *Pooling of forecasts*. Nuffield College.

[16] Clements, M. P. and Hendry, D. F. (2006). Forecasting with breaks. *Handbook of economic forecasting*, 1:605–657.

[17] de Rooij, S., van Erven, T., Grünwald, P. D., and Koolen, W. M. (2014). Follow the leader if you can, hedge if you must. *Accepted to the Journal of Machine Learning Research*.

[18] Del Negro, M., Hasegawa, R. B., and Schorfheide, F. (2016). Dynamic prediction pools: an investigation of financial frictions and forecasting performance. *Journal of Econometrics*, 192(2):391–405.

[19] Diebold, F. X. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*, 5(4):589–592.

[20] Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination.

[21] Diebold, F. X. and Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, 6(1):21–40.

[22] Diebold, F. X. and Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, 6(4):503–508.

[23] Elliott, G. (2011). Averaging and the optimal combination of forecasts. *University of California, San Diego.*

[24] Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2):357–373.

[25] Elliott, G., Gargano, A., and Timmermann, A. (2015). Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, 54:86–110.

[26] Elliott, G. and Timmermann, A. (2005). Optimal forecast combination under regime switching*. *International Economic Review*, 46(4):1081–1102.

[27] Even-Dar, E., Kearns, M., Mansour, Y., and Wortman, J. (2008). Regret to the best vs. regret to the average. *Machine Learning*, 72(1-2):21–37.

[28] Fawcett, N., Kapetanios, G., Mitchell, J., and Price, S. (2014). Generalised density forecast combinations.

[29] Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139.

[30] Garcia, J. A. (2003). An introduction to the ecb's survey of professional forecasters. *ECB Occasional Paper*, (8).

[31] Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, 29(1):108–121.

[32] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

[33] Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

[34] Granziera, E., Luu, C., St-Amant, P., et al. (2013). The accuracy of short-term forecast combinations. *Bank of Canada Review*, 2013(Summer):13–21.

[35] Guidolin, M. and Timmermann, A. (2009). Forecasts of us short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, 150(2):297–311.

[36] Hendry, D. F. and Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31.

[37] Hoogerheide, L., Kleijn, R., Ravazzolo, F., Van DijK, H. K., and Verbeek, M. (2010). Forecast accuracy and economic gains from Bayesian model averaging using time-varying weights. *Journal of Forecasting*, 29(1-2):251–269.

[38] Hsiao, C. and Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, 178:294–309.

[39] Huang, H. and Lee, T.-H. (2010). To combine forecasts or to combine information? *Econometric Reviews*, 29(5-6):534–570.

[40] Kapetanios, G., Labhard, V., and Price, S. (2008). Forecast combination and the bank of England's suite of statistical forecasting models. *Economic Modelling*, 25(4):772–792.

[41] LeSage, J. P. and Magura, M. (1992). A mixture-model approach to combining forecasts. *Journal of Business & Economic Statistics*, 10(4):445–452.

[42] Littlestone, N. and Warmuth, M. (1994). The weighted majority algorithm. *Information and Computation*, 108:212–261.

[43] Newbold, P. and Harvey, D. I. (2002). Forecast combination and encompassing. *A companion to economic forecasting*, pages 268–283.

[44] Novales, A. and de Fruto, R. F. (1997). Forecasting with periodic models a comparison with time invariant coefficient models. *International Journal of Forecasting*, 13(3):393–405.

[45] Patton, A. J. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.

[46] Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.

[47] Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129(1):183–217.

[48] Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2):821–862.

[49] Sancetta, A. (2010). Recursive forecast combination for dependent heterogeneous data. *Econometric Theory*, 26(02):598–631.

[50] Sánchez, I. (2008). Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, 24(4):679–693.

[51] Sani, A., Neu, G., and Lazaric, A. (2014). Exploiting easy data in online optimization. In *Advances in Neural Information Processing Systems*, pages 810–818.

[52] Seabold, J. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*.

[53] Sessions, D. N. and Chatterjee, S. (1989). The combining of forecasts using recursive techniques with non-stationary weights. *Journal of Forecasting*, 8(3):239–251.

[54] Smith, J. and Wallis, K. F. (2005). Combining point forecasts: The simple average rules, ok? *manuscript, University of Warwick*.

[55] Smith, J. and Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle*. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355.

[56] Stock, J. H. and Watson, M. W. (2001). Forecasting output and inflation: the role of asset prices. Technical report, National Bureau of Economic Research.

[57] Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

[58] Tian, J. and Anderson, H. M. (2014). Forecast combinations under structural break uncertainty. *International Journal of Forecasting*, 30(1):161–175.

[59] Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, 1:135–196.

[60] Uematsu, Y. and Tanaka, S. (2015). Macroeconomic forecasting and variable selection with a very large number of predictors: A penalized regression approach. *arXiv preprint arXiv:1508.04217*.

[61] Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25(3):655–686.

[62] Vovk, V. (1990). Aggregating strategies. In *Proceedings of the third annual workshop on Computational learning theory (COLT)*, pages 371–386.

[63] Waggoner, D. F. and Zha, T. (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics*, 171(2):167–184.

[64] Yang, Y. (2004). Combining forecasting procedures: some theoretical results. *Econometric Theory*, 20(01):176–222.

7. Appendix

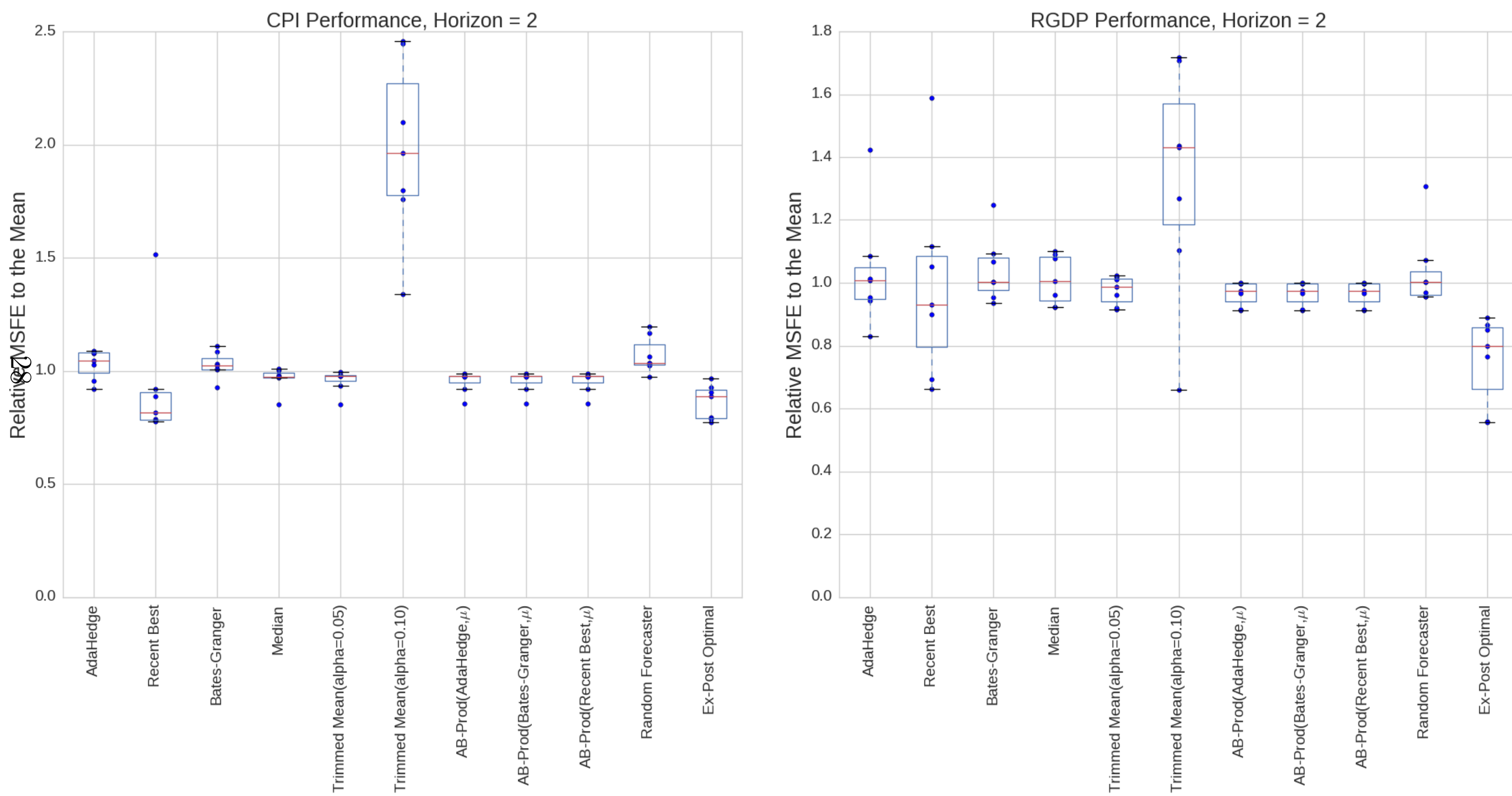

Figure 5: Relative MSFE, Horizon = 2

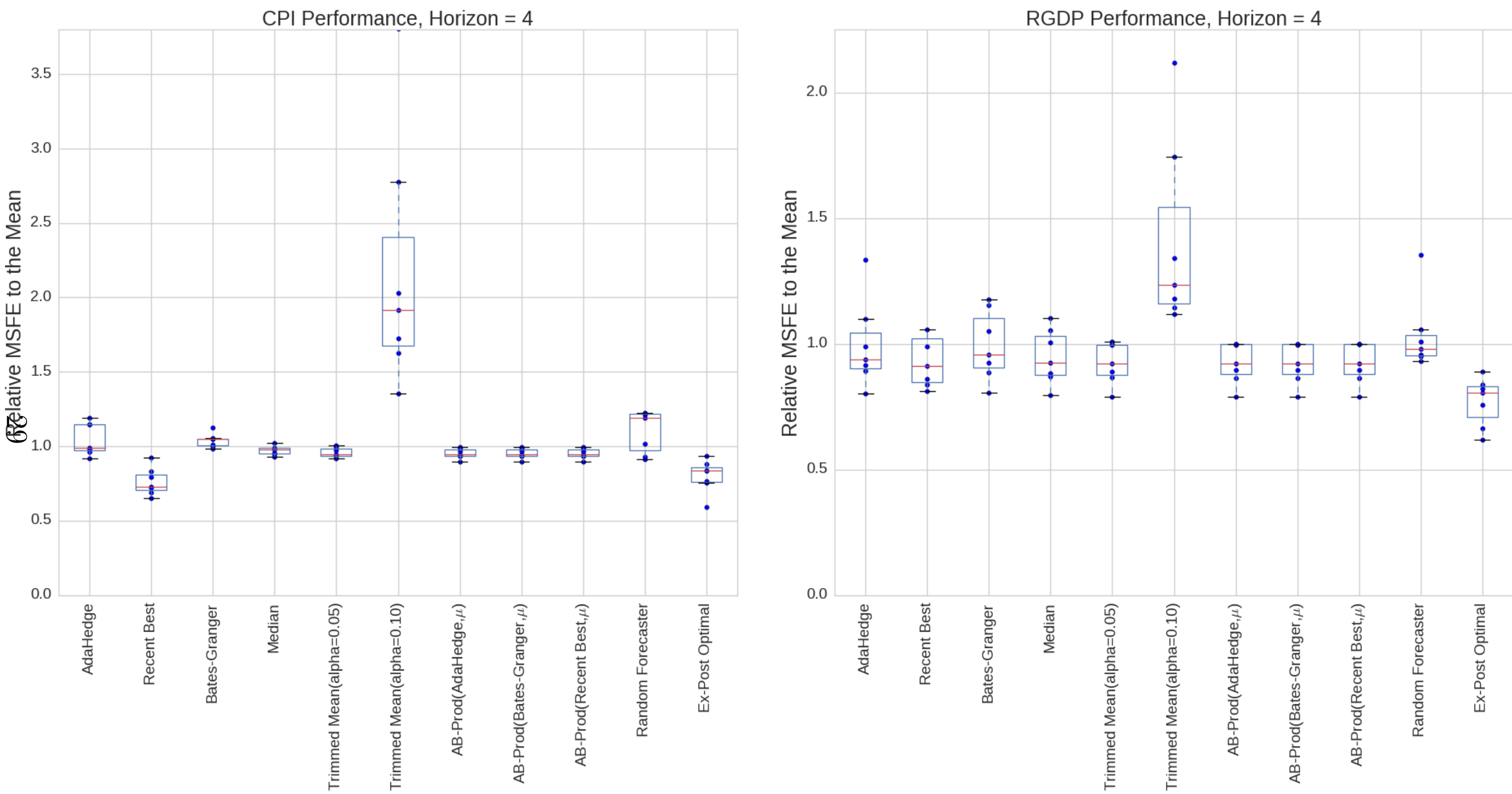

Figure 6: Relative MSFE, Horizon = 4

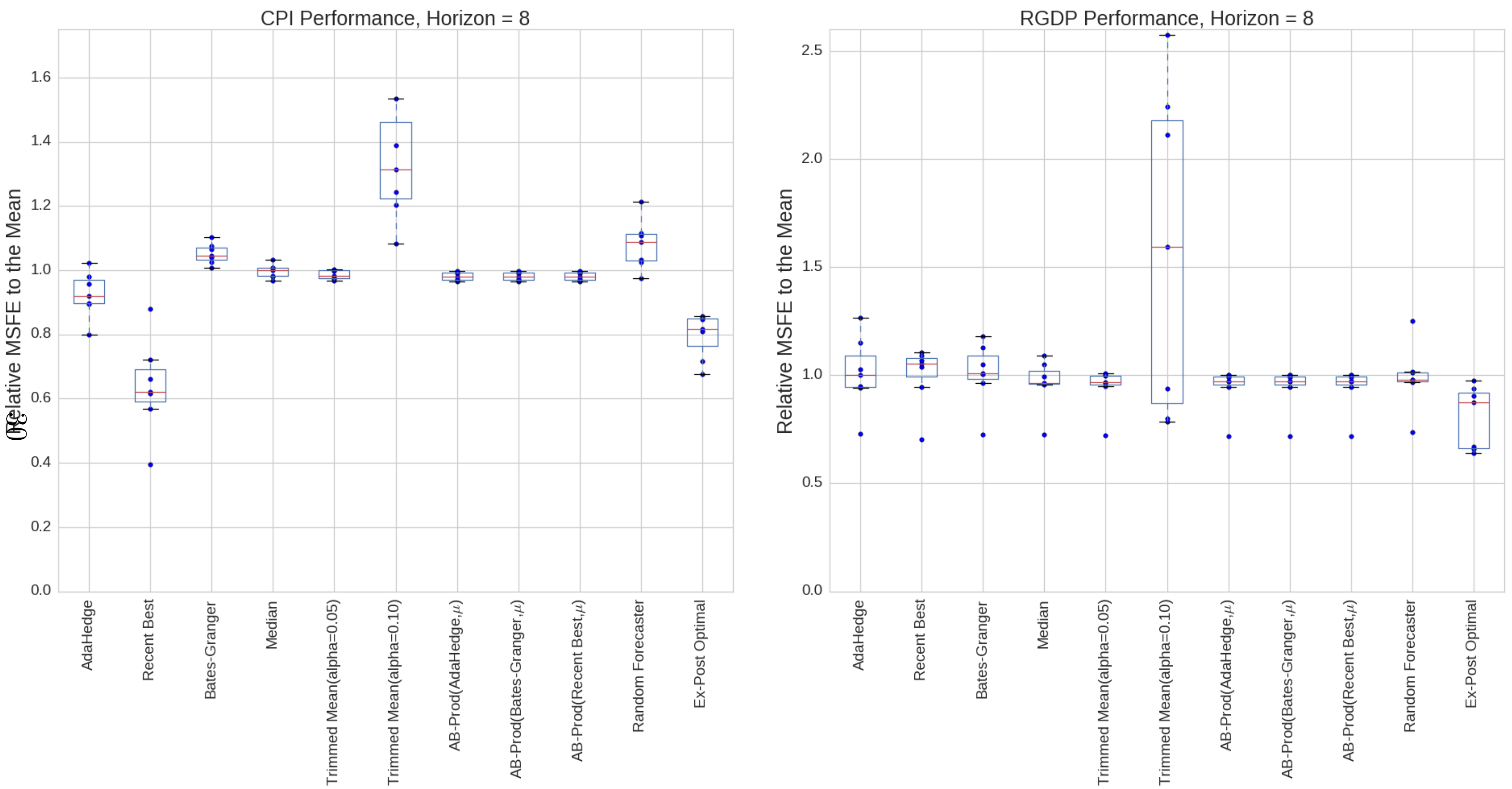

Figure 7: Relative MSFE, Horizon = 8

	China	France	Germany	Italy	Japan	UK	US
AdaHedge	1.027692	0.917716	0.955901	1.089603	1.078820	1.043915	1.079786
Recent Best	0.887602	0.785163	0.813654	0.778690	0.775533	0.919000	1.515963
Bates-Granger	1.028942	0.926949	1.006539	1.109468	1.007305	1.023833	1.084799
Median	0.967980	0.850313	0.979326	1.003207	0.971819	0.974722	1.009402
Trimmed Mean(alpha=0.05)	0.977863	0.850787	0.979695	0.979157	0.935158	0.975582	0.994010
Trimmed Mean(alpha=0.10)	1.759095	2.099013	1.337272	2.456080	1.962812	1.796089	2.445601
AB-Prod(AdaHedge,μ)	0.977713	0.854275	0.975541	0.975923	0.918655	0.974841	0.986960
AB-Prod(Bates-Granger,μ)	0.977713	0.854276	0.975546	0.975925	0.918648	0.974839	0.986961
AB-Prod(Recent Best,μ)	0.977699	0.854261	0.975526	0.975892	0.918625	0.974829	0.987004
Random Forecaster	1.023256	1.165608	1.032905	1.064479	1.193638	0.974331	1.034971
Ex-Post Optimal	0.903505	0.795443	0.788056	0.885872	0.774109	0.926796	0.965566

Table 7: CPI, Horizon = 2, Relative MSFE Results

	China	France	Germany	Italy	Japan	UK	US
AdaHedge	0.942731	1.008187	1.424006	1.084366	0.829337	1.013956	0.953742
Recent Best	0.898264	1.051995	1.587237	0.692809	0.663120	1.116999	0.931004
Bates-Granger	0.935480	1.002353	1.247406	1.092291	1.068171	1.002029	0.953713
Median	0.922007	1.005610	1.101678	1.089201	1.078190	0.962154	0.923201
Trimmed Mean(alpha=0.05)	0.915359	0.985940	1.018093	1.009618	1.024449	0.962236	0.919850
Trimmed Mean(alpha=0.10)	1.102148	1.435502	1.716145	1.705908	0.659524	1.430964	1.267201
AB-Prod(AdaHedge,μ)	0.912228	0.975225	0.997529	0.997485	0.999730	0.966550	0.916113
AB-Prod(Bates-Granger,μ)	0.912227	0.975224	0.997513	0.997486	0.999755	0.966549	0.916113
AB-Prod(Recent Best,μ)	0.912223	0.975229	0.997543	0.997445	0.999709	0.966560	0.916111
Random Forecaster	0.955954	1.001357	1.306822	1.072049	0.968677	1.001748	0.955778
Ex-Post Optimal	0.865022	0.850251	0.764778	0.559406	0.557397	0.887923	0.799327

Table 8: RGDP, Horizon = 2, Relative MSFE Results

	China	France	Germany	Italy	Japan	UK	US
AdaHedge	0.987378	1.148972	0.962878	1.188571	0.915973	0.981786	1.148653
Recent Best	0.690108	0.725732	0.831245	0.651087	0.923325	0.719829	0.791346
Bates-Granger	1.048280	0.997643	0.983153	1.047659	1.053514	1.008424	1.126311
Median	0.989902	0.944505	0.931294	0.956771	0.985518	0.975252	1.020092
Trimmed Mean(alpha=0.05)	0.995129	0.932900	0.946988	0.938937	0.915762	0.971114	1.005232
Trimmed Mean(alpha=0.10)	1.914956	1.725628	1.623579	2.029354	3.803069	1.351599	2.776898
AB-Prod(AdaHedge,μ)	0.990434	0.933375	0.946516	0.936971	0.894786	0.964997	0.996970
AB-Prod(Bates-Granger,μ)	0.990440	0.933360	0.946518	0.936957	0.894800	0.964999	0.996968
AB-Prod(Recent Best,μ)	0.990404	0.933333	0.946503	0.936917	0.894787	0.964970	0.996934
Random Forecaster	1.222143	0.910372	0.928015	1.211898	1.188929	1.015007	1.220682
Ex-Post Optimal	0.877543	0.835150	0.837161	0.753022	0.591105	0.762880	0.932993

Table 9: CPI, Horizon = 4, Relative MSFE Results

	China	France	Germany	Italy	Japan	UK	US
AdaHedge	0.892596	0.800332	1.099527	1.333724	0.987369	0.913697	0.938322
Recent Best	0.859455	0.809850	1.055864	18.447350	0.837273	0.987821	0.911919
Bates-Granger	0.885777	0.805800	1.174391	1.154563	1.048829	0.924879	0.955719
Median	0.868024	0.794325	1.004354	1.102208	1.053814	0.883106	0.923204
Trimmed Mean(alpha=0.05)	0.865283	0.790131	0.996797	0.995448	1.007904	0.888014	0.921476
Trimmed Mean(alpha=0.10)	1.143301	1.117660	1.744260	2.117454	1.233383	1.178624	1.339899
AB-Prod(AdaHedge,μ)	0.863956	0.789163	0.997695	0.995876	0.999673	0.894583	0.921555
AB-Prod(Bates-Granger,μ)	0.863955	0.789163	0.997703	0.995858	0.999679	0.894584	0.921557
AB-Prod(Recent Best,μ)	0.863953	0.789164	0.997691	0.997647	0.999658	0.894590	0.921553
Random Forecaster	0.955566	0.949688	1.007614	1.353242	1.057964	0.931288	0.979254
Ex-Post Optimal	0.821905	0.755434	0.887838	0.616751	0.663060	0.836903	0.806403

Table 10: RGDP, Horizon = 4, Relative MSFE Results

	China	France	Germany	Italy	Japan	UK	US
AdaHedge	0.799612	0.895583	0.919928	0.980815	0.956147	1.021483	0.896565
Recent Best	0.395634	0.568066	0.721221	0.621968	0.615896	0.879640	0.661627
Bates-Granger	1.044667	1.025635	1.007630	1.064509	1.102054	1.039898	1.076220
Median	1.006699	0.980125	0.967675	0.982552	1.008216	0.999470	1.033027
Trimmed Mean(alpha=0.05)	1.002034	0.968371	0.983280	0.973659	0.975715	0.997962	1.000458
Trimmed Mean(alpha=0.10)	1.244203	1.202384	1.312435	1.534758	2.803616	1.083087	1.388801
AB-Prod(AdaHedge,μ)	0.990934	0.964579	0.980968	0.971414	0.966193	0.994973	0.997491
AB-Prod(Bates-Granger,μ)	0.990959	0.964592	0.980977	0.971422	0.966207	0.994975	0.997509
AB-Prod(Recent Best,μ)	0.990894	0.964546	0.980948	0.971378	0.966159	0.994959	0.997467
Random Forecaster	0.975942	1.214248	1.025708	1.107808	1.087838	1.115666	1.032259
Ex-Post Optimal	0.716703	0.816942	0.852993	0.846011	0.675670	0.857574	0.809316

Table 11: CPI, Horizon = 8, Relative MSFE Results

	China	France	Germany	Italy	Japan	UK	US
AdaHedge	0.940745	0.727263	1.151296	1.264624	1.027481	1.001673	0.950453
Recent Best	0.946763	0.704056	1.037182	1.088804	1.053447	1.106216	1.068228
Bates-Granger	0.964636	0.726393	1.179088	1.128691	1.048486	1.009904	1.003223
Median	0.961519	0.723440	0.991709	1.090088	1.050152	0.956663	0.964825
Trimmed Mean(alpha=0.05)	0.949375	0.719920	0.995852	0.999659	1.008941	0.964136	0.966100
Trimmed Mean(alpha=0.10)	0.939316	0.800743	1.591848	2.573644	0.783951	2.112149	2.243635
AB-Prod(AdaHedge,μ)	0.945688	0.718049	0.999464	0.990341	0.999840	0.972429	0.970134
AB-Prod(Bates-Granger,μ)	0.945690	0.718049	0.999467	0.990327	0.999842	0.972430	0.970139
AB-Prod(Recent Best,μ)	0.945689	0.718046	0.999453	0.990323	0.999843	0.972440	0.970145
Random Forecaster	0.967656	0.735312	0.976700	1.249795	1.010907	1.014356	0.976907
Ex-Post Optimal	0.873424	0.668203	0.974834	0.638040	0.656253	0.936141	0.903260

Table 12: RGDP, Horizon = 8, Relative MSFE Results