

No-Regret Forecasting with Egalitarian Committees

Jiun-Hua Su[†]

Institute of Economics
Academia Sinica

September 29, 2021

Abstract

The forecast combination puzzle is often found in literature: The equal-weight scheme tends to outperform sophisticated methods of combining individual forecasts. Exploiting this finding, we propose a hedge egalitarian committees algorithm (HECA), which can be implemented via mixed integer quadratic programming. Specifically, egalitarian committees are formed by the ridge regression with shrinkage toward equal weights; subsequently, the forecasts provided by these committees are averaged by the hedge algorithm. We establish the no-regret property of HECA. Using data collected from the ECB Survey of Professional Forecasters, we find the superiority of HECA relative to the equal-weight scheme during the COVID-19 recession.

Keywords: Hedge Egalitarian Committees Algorithm, No-Regret Forecasting, Mixed Integer Quadratic Programming, Forecast Combination Puzzle, ECB Survey of Professional Forecasters

JEL Classification: C22, C52, C53

[†]I thank Le-Yu Chen, Shiu-Sheng Chen, Yu-Chin Hsu, Chu-An Liu, and Shou-Yung Yin for helpful discussions. Address correspondence to Jiun-Hua Su, 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan; E-mail address: jhsu@econ.sinica.edu.tw.

1 Introduction

Big data has been a buzzword in social science in recent years, and its popularity is witnessed in surveys such as [Varian \(2014\)](#) and [Einav and Levin \(2014\)](#) in economics, [Lazer and Radford \(2017\)](#) in sociology, and [Brady \(2019\)](#) in political science. The importance of data in empirical studies is self-evident. However, as argued by [Lovell \(1983\)](#) almost forty years ago, “it is by no means obvious that reductions in the costs of data mining have been matched by a proportional increase in our knowledge of how the economy actually works.” The advance in theory is also important and indispensable for scientific progress. Data science — thus named perhaps because it emphasizes dialogues between theory and data — suggests an effective way to improve knowledge.

A dialogue between theory and data has been exemplified by the forecast combination puzzle in econometrics. This puzzle refers to the phenomenon that the equal-weight scheme, which is theoretically suboptimal in general, often outperforms the forecast combination with [Bates and Granger’s \(1969\)](#) optimal weights as well as other sophisticated combination methods in empirical studies. The early dialogue had sparked thought-provoking works on the combination of forecasts, as documented in [Clemen \(1989\)](#). More recent surveys on the forecast combinations are provided by [Timmermann \(2006\)](#) as well as [Elliott and Timmermann \(2016\)](#). Indeed, the dialogue on such a puzzle continues to this day. For example, acknowledging the equal-weight scheme as a high benchmark, [Diebold and Shin \(2019\)](#) propose the *egalitarian ridge regression*, which is the ridge regression with shrinkage toward the equal-weight scheme.¹ Their idea has heuristic appeal but leaves open the question of an appropriate computational method.

In this paper, we fortify the theoretical and computational foundations of [Diebold and Shin’s \(2019\)](#) shrinkage approach by developing a real-time forecasting algorithm under the decision-making framework given that the growing literature on forward-looking models in economics highlights the role of forecasting in decision making.² A decision maker first organizes *egalitarian committees*, that is, committees providing their forecasts, respectively, via a ridge regression with shrinkage toward the simple average of individual forecasts selected by mixed integer quadratic programming (MIQP). The application of MIQP and partition of a parameter space solve [Diebold and Shin’s](#)

¹Alternatively, [Diebold and Pauly \(1990\)](#) propose empirical Bayes forecasting procedures with shrinkage toward the equal-weight scheme. This Bayesian approach has been applied in, for example, [Stock and Watson \(2004\)](#) and [Aiolfi and Timmermann \(2006\)](#).

²As indicated in [Clarida et al. \(2000\)](#) and [Mavroidis \(2010\)](#), a forecast-based interest rate rule in response to future macroeconomic conditions can provide a guideline for a monetary policy maker. Forecasting matters in not only a public policy but a private agent’s decision as well. [Tanaka et al. \(2020\)](#) build a simple model concerning a firm’s decisions on inputs under uncertainty to rationalize the empirical evidence that its GDP forecast accuracy is a predictor of profitability and productivity.

computational difficulty in egalitarian ridge regression with simultaneous selection of individual forecasts. Next, the decision maker pools committee forecasts by applying a variant of [Freund and Schapire’s \(1997\)](#) hedge algorithm. This variant depends on an estimate of the maximal committee loss for the duration of its implementation. The decision maker’s two-stage implementation of real-time forecasting is referred to as *hedge egalitarian committees algorithm*, henceforth abbreviated to HECA.

We establish non-asymptotic upper bounds on the *average regret* attained by HECA, which is the decision maker’s average forecasting loss in excess of the smallest average forecasting loss accomplished by these egalitarian committees. First, these upper bounds indicate the decision maker’s own acumen of business cycles could pay off because given the committee forecasts, a more precise estimate of the maximal committee loss *ceteris paribus* yields a tighter upper bound on the average regret. Furthermore, these upper bounds show that HECA has *no-regret* property; that is, the decision maker’s long-run performance should be at least as good as the best long-run performance accomplished by the egalitarian committees. This result is in line with the findings in the online learning literature. An excellent overview of this literature is recently provided by [Cesa-Bianchi and Orabona \(2021\)](#). More importantly, this no-regret property implies the superiority of HECA relative to the equal-weight scheme in the long run. It is such ‘theoretical’ superiority that makes HECA eligible for the competition with the equal-weight scheme; however, its ‘empirical’ superiority remains to be examined.

To examine whether HECA outperforms the equal-weight scheme in an empirical study, we focus on the quarterly one-year-ahead forecasts of Euro-area real GDP growth in Survey of Professional Forecasters (SPF), which is conducted by the European Central Bank (ECB). The purpose of selecting this dataset is twofold. On the one hand, it generates the equal-weight scheme that has particularly hard-to-beat forecasting performance, as demonstrated in [Genre et al. \(2013\)](#), [Conflitti et al. \(2015\)](#), and [Diebold and Shin \(2019\)](#). On the other hand, it involves not-so-big data such that theoretical parts of data science (inclusive of domain knowledge, statistical methods, and computational techniques) are crucially important. Our empirical results show that HECA keeps pace with the equal-weight scheme before the outbreak of COVID-19 but wins the competition during the COVID-19 recession. Despite the superiority of HECA relative to the equal-weight scheme, HECA suffers from an upsurge in forecasting loss around the onset of COVID-19 pandemic. This pattern is consistent with previous research, as indicated in [Chauvet and Potter \(2013\)](#). We also find that the formation of egalitarian committees gives HECA an advantage over [Freund and Schapire’s](#) hedge algorithm during the COVID-19 recession.

In addition to the pursuit of forecasting performance, we are dedicated to credible forecasting in a spirit that only credible assumptions are maintained, as emphasized in Manski (2013). To achieve this goal, we treat the data generating process (DGP) of target variables and their individual forecasts as a *black box*; that is, no assumption on such DGP is imposed. Instead, the proposed HECA is a data-driven and adaptive approach: At each round, it outputs a combined forecast with more weights on forecasts provided by committees that have performed well in the past; furthermore, the built-in updating mechanism enables HECA to adapt to the environment in the presence of structural breaks that may make forecasters’ relative performance unstable over time. The unstable performance is called *model instability* in literature, and an excellent survey of this issue is provided by Rossi (2013). Additionally, the committee forecasts, as inputs of HECA, are obtained by the rolling egalitarian ridge regression scheme. The rolling scheme is used to guard against possible parameter drift, as indicated by West (2006), whereas the shrinkage toward the simple average of selected individual forecasts is supported by empirical evidence in literature.

HECA embodies interdisciplinary research, which is another marked characteristic of data science.³ Knowledge in econometric literature, statistical methods, numerical and computational techniques, and online learning modeling are woven into HECA for the decision maker’s real-time forecasting. The empirical findings in econometrics treat the equal-weight scheme as a high benchmark, to which the weights should shrink. In spite of alleviating numerical instability, the ridge regression in general fails to achieve the simultaneous selection of individual forecasts, which is implemented by mixed integer optimization. The hedge algorithm further allows for the adaptability to sequential data in real time. Equipped with these designs, HECA complements, but does not replace, existing forecasting methods. It is particularly useful in the situation where the decision maker has limited access to predictors. For surveys of data-rich methods, the reader is referred to Stock and Watson (2006) and Chauvet and Potter (2013).

Throughout this paper, we write $\|z\|_1$, $\|z\|_2$, and $\|z\|_\infty$ for the one-norm, two-norm, and infinity-norm of a generic column vector z in an Euclidean space, respectively. We denote the collection of positive integers and the collection of real numbers by \mathbb{N} and \mathbb{R} , respectively.

The structure of the remaining paper is as follows. Section 2 describes features of data collected from the ECB SPF and Eurostat. Section 3 presents the organization of egalitarian committees, the decision maker’s HECA, and the theoretical upper bounds

³This interdisciplinary characteristic is vividly illustrated in Drew Conway’s data science Venn diagram, which can be found at <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.

on the average regret of HECA. Section 4 discusses the empirical results of applying HECA to real-time forecasting of the year-on-year growth rate of euro area. Section 5 concludes. Technical proofs are deferred to the appendix.

2 Data from Eurostat and ECB SPF

The forecast target variables in this paper are the year-on-year Euro-area GDP growth estimates collected from Eurostat, the European Statistical Agency.⁴ Due to data revisions, several estimates for a given quarter are released by Eurostat. Following Genre et al. (2013), we focus on the $t + 45$ flash estimates, which are published about 45 days after the associated quarter, for our empirical study in Section 4. The evaluation sample runs from the first quarter of 2012 to the third quarter of 2020.

As in Genre et al. (2013), Conflitti et al. (2015), and Diebold and Shin (2019), we focus on the quarterly one-year-ahead forecasts of Euro-area real GDP growth in the ECB SPF. These one-year-ahead forecasts, however, are actually six to eight months ahead. For example, in the questionnaire for the third quarter of 2018, macroeconomic experts participating in the SPF are asked for the expected year-on-year real GDP growth for the first quarter of 2019 and provided with the GDP growth for the first quarter of 2018 as a reference.

One noticeable feature in the SPF is the frequent entry, exit, and reentry of experts so that an unbalanced panel arises. As pointed out in Genre et al. (2013), such an unbalanced panel may yield sampling distortions. To lessen the extent of undesirable distortions, we exclude experts who did not reply in two consecutive quarters during the evaluation period spanning from the first quarter of 2012 to the third quarter of 2020. After this removal, there remain 21 experts. Hereafter, we focus on the forecasts provided by this filtered panel of experts. In Figure 1, we mark a slot by the notation x if a forecast is provided by an associated expert for a specific quarter in the SPF; otherwise, we leave it blank. We further replace each missing value of a forecast with the simple average of the rest of the reported forecasts for the same quarter. For example, Expert 038 provides forecasts throughout the evaluation period except the one for the third quarter of 2015; this missing forecast is filled in with the simple average of forecasts provided by the other 19 experts, as the forecast associated with Expert 110 is also unreported.

Another well-known feature in the SPF is the forecast combination puzzle: It is hard for other sophisticated schemes to improve on the performance of equal-weight

⁴These estimates are available at <https://ec.europa.eu/eurostat/web/national-accounts/data/other>.

scheme, as indicated by [Genre et al. \(2013\)](#) and [Conflitti et al. \(2015\)](#). Some rationales behind this puzzle are proposed in literature. From a theoretical perspective, as shown in [Timmermann \(2006\)](#), the equal-weight scheme is optimal if individual forecast errors have the same variance and identical pairwise correlation. From a practical perspective, as noted in [Smith and Wallis \(2009\)](#) and [Conflitti et al. \(2015\)](#), the finite-sample error and numerical instability may make the estimated optimal weights inferior to the equal-weight scheme in terms of forecasting performance.

Both perspectives are crucial in an empirical study using the evaluation sample from the SPF. First, [Table 1](#) and [Figure 2](#) show the sample variances and pairwise correlation coefficients, respectively, of individual forecast errors for the 21 experts in the filtered panel. Since the sample variances are similar to each other whereas the sample correlation coefficients are centered around 0.995, the hard-to-beat performance of equal-weight scheme is unsurprising. In addition, the asymptotic approximation of estimated weights is arguably imprecise because the overall sample size — 35 quarters — is obviously small. More importantly, the numerical instability in the estimated optimal weights is severe; for example, the condition number associated with the ordinary least square regression using the evaluation sample is 31, 112.⁵ This finding yields a clue as to the application of ridge regression, a classical approach in literature to alleviating numerical instability, to the estimation of optimal weights. Recognizing the remarkable performance of equal-weight scheme, [Diebold and Shin \(2019\)](#) propose the egalitarian ridge regression, which is the ridge regression with shrinkage toward the equal-weight scheme. Following their approach, we further develop an algorithm in the next section that can select experts in each quarter and achieve some satisfactory objective in hindsight.

3 Forecasting with Egalitarian Committees

The fundamental importance of economic forecasting for forward-looking private agents and public policy makers motivates us to propose an algorithm that incorporates features of the SPF forecasts and outperforms the equal-weight scheme under the decision-making framework. Roughly speaking, we consider the situation where a single decision maker

⁵The (2-norm) condition number of a matrix A is defined as

$$\kappa_2(A) \equiv \frac{\max_{b: \|b\|_2=1} \|Ab\|_2}{\min_{b: \|b\|_2=1} \|Ab\|_2}$$

and often used to evaluate the stability of a linear system in numerical analysis. As argued in [Belsley et al. \(1980\)](#), “moderate to strong relations are associated with condition indexes of 30 to 100.”

is allowed access to forecasts provided by anonymous experts using either quantitative models or model-free judgments, and such forecasts generating processes are unknown to the decision maker.⁶ This decision maker is assumed to minimize the cumulative squared loss without discounting. The squared loss can be replaced with other loss functions, for example those documented in Section 2.2 of Elliott and Timmermann (2016), in the rest of this paper. We refrain from this replacement because the squared loss is used in common with the literature on forecast combination puzzle.

To elaborate on the proposed method, we now introduce notation. Suppose that there are M experts providing a forecast of the target variable y_t , respectively, before its realization. These individual forecasts are denoted by $f_t \equiv (f_{t,1}, \dots, f_{t,M})^\top$, where $f_{t,m}$ stands for the forecast of y_t provided by expert $m \in \{1, \dots, M\}$.⁷ Accessing the data encompassing current forecasts, realized target variables, and their corresponding forecasts, the decision maker announces his or her own forecast of y_t by a two-stage method: At the first stage, the decision maker imagines M committees $\{\mathcal{C}_c\}_{c=1}^M$, where committee \mathcal{C}_c consists of c members selected among M experts; subsequently, each committee provides a forecast $\hat{y}_{t,c}$, which is a combination of forecasts provided by its c members. At the second stage, this decision maker applies the hedge algorithm to $\{\hat{y}_{t,c}\}_{c=1}^M$ and then yields his or her own forecast of y_t . To complete the two-stage method, we explain how these committees $\{\mathcal{C}_c\}_{c=1}^M$ are formed at the first stage and how the hedge algorithm works at the second stage in the subsections below.

3.1 Egalitarian Committees

The imaginary committee \mathcal{C}_c is organized by solving the following optimization problem for a fixed rolling window $r \in \mathbb{N}$ and every λ in a pre-specified set Λ of grids:⁸

⁶These experts' potentially strategic behaviors are also ignored by the decision maker. For the strategic forecasting, we refer the reader to Marinovic et al. (2013) and references cited therein.

⁷In our empirical analysis in Section 4, the vector f_t of forecasts in the SPF are six to eight months prior to the release of the $t + 45$ flash estimate y_t by Eurostat.

⁸As indicated in Elliott and Timmermann (2016, p. 378), the length of estimation window can be selected by the cross-validation method, which is however rarely done.

(P1)

$$\begin{aligned} & \min_{b \in \mathbb{R}^M} \sum_{s=l}^{l+r-1} [y_{t-s} - f_{t-s}^\top b]^2 + \lambda \left\| b - \frac{1}{\|b\|_0} \mathbf{1} \right\|_2^2 \\ & \text{s.t. } 0 \leq b_j \leq 1, \text{ for } j = 1, \dots, M; \\ & \sum_{j=1}^M b_j = 1; \\ & \|b\|_0 = c \in \mathbb{N}, \end{aligned}$$

where $\|b\|_0$ is the number of nonzero elements in $b \equiv (b_1, \dots, b_M)^\top$ and $\mathbf{1}$ is the M dimensional column vector of ones. The rolling scheme is adopted to guard against possible parameter drift. Since all individual forecasts are measured on the same scale, they are not standardized in this ridge-type regression. In addition, the lag term can be set to be either $l = 1$ or $l = 2$ for real-time forecasting with the SPF forecasts. Let $\hat{\beta}_{t,c}(\lambda)$ be a solution to problem (P1) associated with λ , and $\iota_m \in \mathbb{R}^M$ be a unit vector with m -th element equal to one. The tuning parameter $\hat{\lambda}_{t,c}$ is selected by setting

$$\hat{\lambda}_{t,c} \equiv \arg \min_{\lambda \in \Lambda} \sum_{s=l}^{l+r_\lambda-1} [y_{t-s} - f_{t-s}^\top \hat{\beta}_{t-s,c}(\lambda)]^2,$$

where $r_\lambda \in \mathbb{N}$ denotes the number of periods for validation. The set

$$\mathcal{C}_c \equiv \left\{ m : \iota_m^\top \hat{\beta}_{t,c}(\hat{\lambda}_{t,c}) > 0 \right\}$$

is called the *egalitarian committee* with c members, for problem (P1) can be viewed as a subproblem of partial egalitarian ridge regression in [Diebold and Shin \(2019\)](#). To see this, let

$$\tilde{\beta}_t(\lambda) \equiv \arg \min \left\{ \sum_{s=l}^{l+r-1} [y_{t-s} - f_{t-s}^\top b]^2 + \lambda \left\| b - \frac{1}{\|b\|_0} \mathbf{1} \right\|_2^2 : b \in \left\{ \hat{\beta}_{t,c}(\lambda) \right\}_{c=1}^M \right\},$$

where $\hat{\beta}_{t,c}(\lambda)$ is a minimizer of (P1) for each $c \in \{1, \dots, M\}$ and a fixed λ . Although the objective function is discontinuous due to $\|b\|_0$, we have

$$\begin{aligned} \tilde{\beta}_t(\lambda) &= \arg \min_{b \in \mathbb{R}^M} \sum_{s=l}^{l+r-1} [y_{t-s} - f_{t-s}^\top b]^2 + \lambda \left\| b - \frac{1}{\|b\|_0} \mathbf{1} \right\|_2^2 \\ & \text{s.t. } 0 \leq b_j \leq 1, \text{ for } j = 1, \dots, M; \\ & \sum_{j=1}^M b_j = 1. \end{aligned}$$

Phrased differently, the partition of a parameter space according to the value of $\|b\|_0$ allows us to recover $\tilde{\beta}_t(\lambda)$. Therefore, Diebold and Shin’s (2019) ‘one-step’ partial egalitarian ridge regression can be equivalently implemented as long as problem (P1) is successfully solved for every c .⁹

To solve problem (P1), we recast it as the following MIQP:

(P2)

$$\begin{aligned} \min_{b \in \mathbb{R}^M, d \in \mathbb{R}^M} \quad & \sum_{s=l}^{l+r-1} [y_{t-s} - f_{t-s}^\top b]^2 + \lambda \left\| b - \frac{1}{c} \mathbf{1} \right\|_2^2 \\ \text{s.t.} \quad & d_j \epsilon \leq b_j \leq d_j, \quad \text{for } j = 1, \dots, M; \\ & d_j \in \{0, 1\}, \quad \text{for } j = 1, \dots, M; \\ & \sum_{j=1}^M b_j = 1; \quad \sum_{j=1}^M d_j = c \in \mathbb{N}. \end{aligned}$$

If ϵ is the smallest machine-representable positive real number,¹⁰ then problems (P1) and (P2) are computationally equivalent. The intuition is that under the constraints $d_j \epsilon \leq b_j \leq d_j$ and $d_j \in \{0, 1\}$, the dummy variable $d_j = \mathbb{1}_{[b_j > 0]}$ indicates whether b_j is positive; that is, expert j is selected in the committee \mathcal{C}_c with size $\|b\|_0$, which is equal to the sum of d_j ’s. We summarize the discussion in the following proposition.

Proposition 1. *Suppose that ϵ is the smallest machine-representable positive real number. The optimization problem (P1) and the MIQP (P2) are computationally equivalent in the following sense:*

- (i) *If the machine yields the minimizer b^* of (P1), then (b^*, d^*) is a minimizer of (P2), where $d^* \equiv (d_1^*, \dots, d_M^*)^\top$ and $d_j^* = \mathbb{1}_{[b_j^* > 0]}$ for each $j = 1, \dots, M$.*
- (ii) *If the machine yields the minimizer (b^*, d^*) of (P2), then b^* is a minimizer of (P1).*

A conceptually simple method of solving problem (P2) is exhaustive enumeration. To see this, note that there are $\binom{M}{c}$ feasible choices of $d \equiv (d_1, \dots, d_M)^\top$ in problem (P2). For any given feasible d , this optimization problem is essentially the constrained

⁹Although Diebold and Shin’s (2019) partial egalitarian ridge regression concerns the inclusion of Tibshirani’s (1996) one-norm regularization in the objective function rather than in the constraints, the idea of partitioning a parameter space still works *mutatis mutandis*.

¹⁰As defined in Judd (1998, p. 30), a machine zero is referred to as a quantity equivalent to zero on a machine. The positive real number ϵ is not a machine zero, but every positive real number less than ϵ is a machine zero.

ridge regression with the unknown parameter b . Implementing these $\binom{M}{c}$ ridge regressions thus suffices to solve problem (P2). We call this approach *complete subset ridge regressions* by analogy with complete subset regressions in Elliott et al. (2013). This exhaustive approach, however, may be computationally inefficient because every egalitarian committee is asked to provide a forecast $\hat{y}_{t,c} \equiv f_t^\top \hat{\beta}_{t,c}(\hat{\lambda}_{t,c})$; consequently, there are $\binom{M}{1} + \binom{M}{2} + \dots + \binom{M}{M} = 2^M - 1$ ridge regressions to be carried out for every $\lambda \in \Lambda$.

Instead of such an exhaustive search for $\{\hat{\beta}_{t,c}(\lambda)\}_{c=1}^M$ in the parameter space, the modern solver Gurobi can be used to implement the MIQP in problem (P2). The practical tractability of moderate-size MIQP, though NP-hard in nature, can be attributed to the rapid advances in computation power. According to Bertsimas and Dunn (2019), the overall speedup, inclusive of solvers and hardware, is approximately two trillion between 1991 and 2016. The advance of mixed integer optimization has sparked recent studies in statistics and econometrics, for example Bertsimas et al. (2016) and Chen and Lee (2018), among others. Details about algorithmic developments of mixed integer optimization can be found in Jünger et al. (2010), Conforti et al. (2014), and references cited therein.

Remark 1. *The constraint on total weight ($b^\top \mathbf{1} = 1$) and the range constraints ($0 \leq b_j \leq 1$ for each j) are included in problems (P1) and (P2) because they have the following nice properties. First, as pointed out in Granger and Ramanathan (1984), the combined forecast $\hat{y}_{t,c} \equiv f_t^\top \hat{\beta}_{t,c}(\hat{\lambda}_{t,c})$ provided by the egalitarian committee \mathcal{C}_c , under the constraint on total weight, remains unbiased if every member in this committee provides an unbiased forecast of y_t . As shown in Diebold (1988), the constraint on total weight also implies that the forecast errors $y_t - \hat{y}_{t,c}$ are serially uncorrelated if the individual forecast errors made by the members in this committee are serially uncorrelated. In addition, Jagannathan and Ma (2003) indicate the shrinkage effect of range constraints on reducing the estimation error of the covariances of experts' forecast errors. Finally, Conflitti et al. (2015) suggest that these constraints together can improve the numerical stability in computation.*

In econometrics and statistics, assumptions about the DGP of (y_t, f_t^\top) are usually imposed to establish theoretically nice properties of a forecasting method. Elliott and Timmermann (2016, p. 320), however, put it this way:

Interestingly, combination methods that attempt to explicitly model time variation in the combination weights often fail to perform well, suggesting

that regime switching or model “breakdown” can be difficult to predict or even track through time.

Additionally, the knowledge about how individual forecasts are generated by experts is unknown in principle to the decision maker. A practical example given in [Diebold \(2015\)](#) is that forecasts are purchased from a vendor using proprietary models, which are not revealed to the decision maker. Recognizing such limits to knowledge, the decision maker attempts to neither model nor assume the DGP of (y_t, f_t^\top) and is dedicated to credible forecasting.¹¹ This decision maker thus deals with the real-time forecasting problem by the hedge algorithm based on the past performance of the egalitarian committees, as described in the next subsection.

3.2 Hedge Egalitarian Committees Algorithm

After receiving forecasts $\hat{y}_t \equiv (\hat{y}_{t,1}, \dots, \hat{y}_{t,M})^\top$ made by all egalitarian committees, the decision maker announces his or her own forecast, which is a weighted average of $\{\hat{y}_{t,c}\}_{c=1}^M$. Subsequently, the nature announces the realization of the target variable y_t . The sequence of target variables can be generated as in statistical models. For example, it can represent business cycles undulating along a trend, either deterministic or stochastic; it can exhibit structural breaks with changing points, either known or unknown; and it can describe switching among different states, either observed or unobservable. Further examples about statistical modeling can be found in [Pesaran \(2015\)](#) and [Peña and Tsay \(2021\)](#). Alternatively, this sequence of target variables can be adversarially generated as in game-theoretic models, where the nature attempts to maximize the decision maker’s forecasting loss. Details about game-theoretic analysis can be found in [Cesa-Bianchi and Lugosi \(2006\)](#) and [Schapire and Freund \(2012\)](#). Briefly, the following happen in order for each round t :

1. Egalitarian committees announce their forecast combinations

$$\hat{y}_{t,c} \equiv f_t^\top \hat{\beta}_{t,c}(\hat{\lambda}_{t,c}), \text{ for } c = 1, \dots, M,$$

where $\hat{\beta}_{t,c}$ and $\hat{\lambda}_{t,c}$ are obtained by the method in the previous subsection;

2. Decision maker announces the forecast

$$\hat{y}_t \equiv \pi_t^\top \hat{y}_t$$

according to some distribution $\pi_t \in \Delta^M$;

¹¹As argued in [Manski \(2013\)](#), “[t]he fundamental difficulty of empirical research is to decide what assumptions to maintain.” He further argues that “[s]tronger assumptions yield conclusions that are more powerful but less credible.”

Algorithm 1 Hedge Egalitarian Committees Algorithm

Input:

$$\pi_1 = \pi_2 \equiv (1/M, \dots, 1/M)^\top \in \Delta^M;$$

$$\omega_{1,c} = \omega_{2,c} \equiv 1 \text{ for each } c = 1, \dots, M;$$

$$\eta_1 = \frac{2}{B_1} \sqrt{\log\{M\}}, \text{ where } B_1 \text{ is an assumed maximal committee loss;}$$

Output:

A sequence $\{\pi_t\}_{t=1}^T$ of distributions;

- 1: **for** each $t = 3, \dots, T$ **do**
 - 2: collect $\ell_{(t-2),c}$ for each $c = 1, \dots, M$;
 - 3: $\omega_{t,c} \leftarrow \omega_{(t-2),c} \exp\{-\eta_{(t-2)} \ell_{(t-2),c}\}$ for each $c = 1, \dots, M$;
 - 4: $\pi_{t,c} \leftarrow \frac{\omega_{t,c}}{\sum_{m=1}^M \omega_{t,m}}$ for each $c = 1, \dots, M$;
 - 5: $B_{(t-1)} \leftarrow \max\{B_{(t-2)}, \max_{c=1, \dots, M} \ell_{(t-2),c}\}$;
 - 6: $\eta_{(t-1)} \leftarrow \frac{2}{B_{(t-1)}} \sqrt{\frac{\log\{M\}}{t-1}}$;
 - 7: **end for**
-

3. Nature announces $y_t \in \mathbb{R}$;

4. Egalitarian committees incur loss

$$\ell_{t,c} \equiv (y_t - \hat{y}_{t,c})^2, \text{ for } c = 1, \dots, M.$$

Adopting the aforementioned strategy (i.e., announcement of \hat{y}_t in each round) the decision maker *ex ante* aims to obtain small average regret

$$R_T \equiv \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 - \min_{c \in \{1, \dots, M\}} \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_{t,c})^2,$$

by selecting the sequence $\{\pi_t\}_{t=1}^T$ of distributions. To achieve this goal, this decision maker selects $\{\pi_t\}_{t=1}^T$ by HECA, whose pseudocode is shown as Algorithm 1.

HECA, adapted from the hedge algorithm in Freund and Schapire (1997), incorporates features of the decision maker's real-time forecasting based on the SPF forecasts. Suppose that the decision maker announces his or her forecast \hat{y}_t immediately after receiving the

SPF forecasts. In this case, the decision maker does not know the realized forecasting loss $\{\ell_{t,c}\}_{c=1}^M$ until round $t + 2$. Hence, in the first two rounds, the decision maker has no information about any committee's performance and thus uses the uniform distributions π_1 and π_2 . In the third and ensuing rounds, the decision maker observes every committee's performance $\{\ell_{(t-2),c}\}_{c=1}^M$ of two rounds prior, thereby updating the distribution $\pi_t \equiv (\pi_{t,1}, \dots, \pi_{t,M})^\top$. Note that HECA requires an estimate of the maximal committee loss throughout the T rounds. The decision maker assumes B_1 to be this maximal loss, which might be a biased estimate, in the first two rounds, and updates it subsequently round by round.

HECA reflects the presumption embraced by the decision maker: A committee with relatively better performance (i.e., smaller $\ell_{(t-2)}$) would maintain the momentum to perform relatively well in the current round; therefore, its weight $\pi_{t,c}$ in the combined forecast \hat{y}_t should relatively increase. The idea of performance-based pooling of forecasts has been used in econometrics, for example forecasts weighted by inverse mean squared error in [Stock and Watson \(1999\)](#) and [Capistrán and Timmermann \(2009\)](#), and aggregated forecast through exponential reweighting in [Yang \(2004\)](#) and [Wei and Yang \(2012\)](#), among others. The built-in updating mechanism makes HECA adaptive to the environment. The adaptability of HECA is in sharp contrast to the constancy of equal-weight scheme even in the possibly ever-changing environment.

The following theorem gives upper bounds on the decision maker's average regret. These upper bounds are non-asymptotic; that is, they hold for every finite $T \in \mathbb{N}$.

Theorem 1. *Let $\bar{B}_T \equiv \max\{\ell_{t,c} : t = 1, \dots, T; c = 1, \dots, M\}$. HECA guarantees that for all $M, T \in \mathbb{N}$,*

$$R_T \leq \begin{cases} (1 + 2\gamma_u)\bar{B}_T\sqrt{\frac{\log\{M\}}{T}}, & \text{if } \bar{B}_T = \gamma_u B_1 \text{ for some real number } \gamma_u > 1; \\ 3\gamma_o\bar{B}_T\sqrt{\frac{\log\{M\}}{T}}, & \text{if } B_1 = \gamma_o\bar{B}_T \text{ for some real number } \gamma_o > 1; \\ 3\bar{B}_T\sqrt{\frac{\log\{M\}}{T}}, & \text{otherwise.} \end{cases}$$

The non-asymptotic upper bounds in [Theorem 1](#) hold *without* requirement of any assumption on the DGP. The decision maker may underestimate or overestimate the maximal loss \bar{B}_T . The biased estimation, however, is a contributing factor of an upper bound on the average regret. A sharper upper bound can be obtained if the magnitude of underestimation (overestimation), measured by γ_u (γ_o), is smaller. It turns out that given the same committee forecasts and duration of implementing HECA, decision makers having different evaluations of business cycle may achieve different forecasting

performances even if they use HECA. Calculation of these upper bounds is also easy *ex post*. In contrast, upper bounds in Yang (2004) and Wei and Yang (2012) involve nuisance parameters of the underlying DGP and need further evaluation.

Moreover, if the sequence $\{\bar{B}_T\}_{T=1}^\infty$ is bounded above,¹² then HECA exhibits no regret because these upper bounds on the average regret all shrink to zero as T tends to infinity. Phrased differently, the performance of HECA is at least close to that of the best egalitarian committee in the long run. The no-regret property is common in the online learning literature. We refer the reader to Cesa-Bianchi and Lugosi’s (2006) monograph for early findings and Cesa-Bianchi and Orabona’s (2021) survey for recent advances. From a pragmatic standpoint, the assumption about boundedness of $\{\bar{B}_T\}_{T=1}^\infty$ could be inconsequential. To see this, note that

$$\bar{B}_T \leq \max \{(y_t - f_{t,i})^2 : t = 1, \dots, T; i = 1, \dots, M\}$$

by Jensen’s inequality; additionally, as indicated by Elliott and Timmermann (2016, p. 17), “[i]n practice, forecasts are usually bounded and extremely large forecasts typically get trimmed as they are deemed implausible.”

More importantly, Theorem 1 establishes the intuition that the decision maker’s two-stage method should outperform the equal-weight scheme whenever T is so large that the upper bounds are small. Theorem 1 implies that in the long run, HECA should perform at least as well as the best egalitarian committee. In addition, the best egalitarian committee would dominate the egalitarian committee \mathcal{C}_M , which should in turn be weakly better than the equal-weight scheme. It follows from these arguments that HECA would outweigh the equal-weight scheme in terms of long-run forecasting performance.

If the decision maker regularly postpones announcing \hat{y}_t until the realization of $y_{(t-1)}$, then the information on $\{\ell_{(t-1),c}\}_{c=1}^M$ could be exploited. Because of the updated information, the decision maker’s forecasting performance is expected to improve. Indeed, HECA with delayed announcements (Algorithm 2) allows for tighter upper bounds on the average regret, as shown in the following theorem.

Theorem 2. *HECA with delayed announcements guarantees that for all $M, T \in \mathbb{N}$,*

$$R_T \leq \begin{cases} \frac{(1+2\gamma_u)}{\sqrt{2}} \bar{B}_T \sqrt{\frac{\log\{M\}}{T}}, & \text{if } \bar{B}_T = \gamma_u B_1 \text{ for some real number } \gamma_u > 1; \\ \frac{3\gamma_o}{\sqrt{2}} \bar{B}_T \sqrt{\frac{\log\{M\}}{T}}, & \text{if } B_1 = \gamma_o \bar{B}_T \text{ for some real number } \gamma_o > 1; \\ \frac{3}{\sqrt{2}} \bar{B}_T \sqrt{\frac{\log\{M\}}{T}}, & \text{otherwise,} \end{cases}$$

¹²The monotonicity of $\{\bar{B}_T\}_{T=1}^\infty$, together with its boundedness, implies that $\lim_{T \rightarrow \infty} \bar{B}_T < \infty$.

Algorithm 2 Hedge Egalitarian Committees Algorithm with Delayed Announcements

Input:

$$\pi_1 \equiv (1/M, \dots, 1/M)^\top \in \Delta^M;$$

$$\omega_{1,c} \equiv 1 \text{ for each } c = 1, \dots, M;$$

$$\eta_1 = \frac{1}{B_1} \sqrt{2 \log\{M\}}, \text{ where } B_1 \text{ is an assumed maximal committee loss;}$$

Output:

A sequence $\{\pi_t\}_{t=1}^T$ of distributions;

- 1: **for** each $t = 2, \dots, T$ **do**
 - 2: collect $\ell_{(t-1),c}$ for each $c = 1, \dots, M$;
 - 3: $\omega_{t,c} \leftarrow \omega_{(t-1),c} \exp\{-\eta_{(t-1)} \ell_{(t-1),c}\}$ for each $c = 1, \dots, M$;
 - 4: $\pi_{t,c} \leftarrow \frac{\omega_{t,c}}{\sum_{m=1}^M \omega_{t,m}}$ for each $c = 1, \dots, M$;
 - 5: $B_t \leftarrow \max\{B_{t-1}, \max_{c=1, \dots, M} \ell_{(t-1),c}\}$;
 - 6: $\eta_t \leftarrow \frac{1}{B_t} \sqrt{\frac{2 \log\{M\}}{t}}$;
 - 7: **end for**
-

where \bar{B}_T is defined in Theorem 1.

Remark 2. Unlike HECA, which updates $w_{t,c}$ according to the latest observed loss $\ell_{(t-2),c}$, an alternative updating mechanism makes smooth the adjustment to new weights by setting

$$\omega_{t,c} \leftarrow \omega_{(t-2),c} \exp\left\{-\frac{\eta_{(t-2)}}{t-2} \sum_{\tau=1}^{t-2} \ell_{\tau,c}\right\} \text{ for each } c = 1, \dots, M. \quad (1)$$

This updating mechanism relies on the latest observed empirical loss $(t-2)^{-1} \sum_{\tau=1}^{t-2} \ell_{\tau,c}$ and parallels [Fudenberg and Levine's \(1995\)](#) exponential fictitious play, which has the no-regret property, also known as Hannan-consistency in the literature on learning in games. For a book length treatment of this topic, see [Fudenberg and Levine \(1998\)](#) and [Cesa-Bianchi and Lugosi \(2006\)](#). Similarly, we can consider the exponential fictitious

play with delayed announcements by replacing (1) with the following updating mechanism

$$\omega_{t,c} \leftarrow \omega_{(t-1),c} \exp \left\{ -\frac{\eta(t-1)}{t-1} \sum_{\tau=1}^{t-1} \ell_{\tau,c} \right\} \text{ for each } c = 1, \dots, M. \quad (2)$$

4 An Empirical Study on the Growth Rate of Euro Area

Convinced of the theoretically asymptotic performance of HECA, we are now concerned with its forecasting performance for the evaluation sample mentioned in Section 2.

We first concentrate on the competition among the equal-weight scheme, HECA (Algorithm 1 with $l = 2$) and HECA with delayed announcements (Algorithm 2 with $l = 1$). The latter two algorithms involve the first-stage MIQP, which is implemented by the Gurobi Python interface, and the tuning parameters are set by $r = 16$, $r_\lambda = 1$, $\Lambda = \{0.01g\}_{g=1}^{200}$, and $\epsilon = 5 \times 10^{(-324)}$.¹³ We set B_1 to be the maximum of individual forecasting losses (observed by the decision maker) from the first quarter of 2012 to the quarter prior to $t = 1$. HECA, in comparison to its counterpart with delayed announcements, requires two extra rounds for ‘in-sample’ estimation of $\hat{\beta}$ ’s and validation of $\hat{\lambda}$ ’s. Thus, we consider $t = 1$ in HECA to be the fourth quarter of 2016 and $t = 1$ in HECA with delayed announcements to be the second quarter of 2016.

Table 2 reports their forecasting losses and associated differences per round for this competition. As can be seen, HECA and the equal-weight scheme are nearly neck and neck until the first quarter of 2020. Keeping abreast of the equal-weight scheme, which is a well-known high benchmark, HECA also performs well. HECA further outperforms the equal-weight scheme since the second quarter of 2021, in which “the fall in economic activity was unprecedented in depth, speed and scope”.¹⁴ HECA with delayed announcements even achieves better forecasting performance by exploiting updated information. Given the small-sample survey data from the SPF, we do not pursue statistical testing for the superiority of HECA because existing tests accounting for in-sample estimation error, for example the tests developed in Diebold and Mariano (1995) and Giacomini and White (2006), rely on out-of-sample asymptotic approximation to determine the critical value.

Moreover, Figure 3 implies that the number of experts in the committee performing best in a single round is not constant but time-varying. Despite this instability, the

¹³The number $5 \times 10^{(-324)}$ is equal to the product of `sys.float_info.min` and `sys.float_info.epsilon` in Python 3.7, and the mixed integer optimization are carried out by Gurobi 9.0.3, which is available at <https://www.gurobi.com/>.

¹⁴This description is given in the news published on 29th March 2021 by Euro Area Business Cycle Dating Committee. Further details are available at https://eabcn.org/sites/default/files/eabcdc_findings_29_march_2021.pdf.

theoretical results in Section 3 suggest that HECA could perform as well as the best committee over the entire evaluation period in hindsight. As can be seen from Table 3, the average regret is relatively small given the substantial impact of COVID-19 pandemic on the euro area economy. It is worth noting that the best egalitarian committee in the fourth quarter of 2019 and in the third quarter of 2020 are identical, and the average regret is less than 0.03 if HECA with/without delayed announcements terminates in the fourth quarter of 2019.

Finally, we turn the spotlight on the cousin and ancestor of HECA. As shown in Table 4, the exponential fictitious play with updating mechanism (1) performs almost the same as HECA; similarly, the exponential fictitious play with updating mechanism (2) very much resembles HECA with delayed announcements in terms of forecasting ability. The close resemblance gives a hint on the no-regret property of exponential fictitious play. Table 5 shows that although neither Freund and Schapire’s (1997) hedge algorithm nor HECA dominates each other before the fourth quarter of 2019, HECA wins the competition since the first quarter of 2020. Thus, combining the results in Tables 2 and 5, we find that during the COVID-19 recession, the formation of egalitarian committees gives HECA a competitive edge over the hedge algorithm, which beats the equal-weight scheme by adaptability.

5 Conclusion

The proposed HECA should be in the data scientist’s toolkit for three reasons as follows. From a theoretical perspective, HECA outputs credible forecasting because it relies on practically convincing assumptions and meanwhile achieves an asymptotically negligible upper bound on the average regret. From an empirical perspective, HECA outweighs the equal-weight scheme after the outbreak of COVID-19 in euro area, whereas the equal-weight scheme only outperforms HECA by a margin, if any, before such an outbreak. From a methodological perspective, HECA differs from other data-rich methods in that it is applicable in the context where no extra predictor of the target variables, except for the forecasts provided by the experts, is available for the decision maker.

We do not deal with the optimal timing of implementing HECA. Our empirical results seem to suggest that compared with the equal-weight scheme, HECA would be suitable for forecasting around business cycle turning points. It is also unclear whether the duration of implementing HECA should be determined at the very beginning. We delegate these fascinating issues for future work.

Appendix A Technical proofs

A.1 Proof of Proposition 1

For ease of notation, we define functions $Q_1 : \mathbb{R}^M \rightarrow \mathbb{R}$ and $Q_2 : \mathbb{R}^M \rightarrow \mathbb{R}$ to be

$$Q_1(b) \equiv \|Y - Fb\|_2^2 + \lambda \left\| b - \frac{1}{\|b\|_0} \mathbf{1} \right\|_2^2 \quad \text{and} \quad Q_2(b) \equiv \|Y - Fb\|_2^2 + \lambda \left\| b - \frac{1}{c} \mathbf{1} \right\|_2^2.$$

It is clear that $Q_1(b) = Q_2(b)$ whenever $\|b\|_0 = c$.

- (i) First, we show that (b^*, d^*) is feasible for problem (P2). Since b^* satisfies the constraints in (P1), $0 \leq b_j^* \leq 1$ for every j . By the construction of d_j^* ,

$$\sum_{j=1}^M d_j^* = \sum_{j=1}^M \mathbb{1}_{[b_j^* > 0]} = \|b^*\|_0 = c$$

and $d_j^* \in \{0, 1\}$ for each j . It remains to establish $d_j^* \epsilon \leq b_j^* \leq d_j^*$ for each j . Note that b_j^* is either zero or positive. If the machine yields $b_j^* = 0$, then $d_j^* = 0$ and

$$d_j^* \epsilon = 0 \leq b_j^* \leq 0 = d_j^*.$$

If the machine yields $b_j^* > 0$, then

$$d_j^* \epsilon = \epsilon \leq b_j^* \leq 1 = d_j^*$$

because ϵ is the smallest machine-representable positive real number.

Next, we prove that (b^*, d^*) is a minimizer of problem (P2) by contradiction. Suppose that (\tilde{b}, \tilde{d}) is feasible for problem (P2) and $Q_2(\tilde{b}) < Q_2(b^*)$. It follows from the constraints in (P2) that

$$\|\tilde{b}\|_0 = \sum_{j=1}^M \mathbb{1}_{[\tilde{b}_j > 0]} = \sum_{j=1}^M \tilde{d}_j = c.$$

Thus, \tilde{b} is feasible for problem (P1). It turns out that

$$Q_1(\tilde{b}) = Q_2(\tilde{b}) < Q_2(b^*) = Q_1(b^*),$$

contradicting the assumption that b^* is a minimizer of problem (P1).

- (ii) First, we show that b^* is feasible for problem (P1). Since (b^*, d^*) satisfies the constraints in (P2), we have

$$0 \leq d_j^* \epsilon \leq b_j^* \leq d_j^* \leq 1$$

for each j and

$$\|b^*\|_0 = \sum_{j=1}^M \mathbb{1}_{[b_j^* > 0]} = \sum_{j=1}^M d_j^* = c.$$

Thus, b^* satisfies the constraints in (P1).

Next, we prove that b^* is a minimizer of (P1) by contradiction. Suppose that \check{b} is computationally feasible for problem (P1) and $Q_1(\check{b}) < Q_1(b^*)$. Let $\check{d}_j \equiv \mathbb{1}_{[\check{b}_j > 0]}$ for each j . By construction,

$$\check{d}_j \in \{0, 1\} \quad \text{and} \quad \check{d}_j \epsilon \leq \check{b}_j \leq \check{d}_j$$

for each j . Additionally, we have

$$\sum_{j=1}^M \check{d}_j = \sum_{j=1}^M \mathbb{1}_{[\check{b}_j > 0]} = \|\check{b}\|_0 = c.$$

Hence, (\check{b}, \check{d}) is computationally feasible for problem (P2). It follows that

$$Q_2(\check{b}) = Q_1(\check{b}) < Q_1(b^*) = Q_2(b^*),$$

contradicting the assumption that (b^*, d^*) is a minimizer of (P2).

A.2 Proof of Theorem 1

Without loss of generality, let $\eta_t = \frac{2}{B_t} \sqrt{\frac{\log\{M\}}{t}}$ and $B_{t+1} = \max\{B_t, \max_{c=1, \dots, M} \ell_{t,c}\}$ for every $t \in \mathbb{N}$. We denote the Kullback-Leibler divergence of $x = (x_1, \dots, x_M)^\top$ and $z = (z_1, \dots, z_M)^\top$ by

$$D(x\|z) = \sum_{c=1}^M x_c \log \left\{ \frac{x_c}{z_c} \right\} \quad \text{for any } (x, z) \in \Delta^M \times \text{ri}(\Delta^M),$$

where $\text{ri}(\Delta^M)$ is the relative interior of probability M -simplex $\Delta^M \subseteq \mathbb{R}^M$. For each $t \in \mathbb{N}$, HECA outputs

$$\pi_{t+2} = \arg \min_{\pi \in \Delta^M} \left[\langle \ell_t, \pi - \pi_t \rangle + \frac{1}{\eta_t} D(\pi\|\pi_t) \right].$$

The Bregman proximal inequality given in Lemma 3.1 of [Teboulle \(2018\)](#) implies that for any $\pi \in \Delta^M$ and $t \in \mathbb{N}$,

$$\langle \ell_t, \pi_t - \pi \rangle \leq \frac{1}{\eta_t} [D(\pi\|\pi_t) - D(\pi\|\pi_{t+2})] + \langle \ell_t, \pi_t - \pi_{t+2} \rangle - \frac{1}{\eta_t} D(\pi_{t+2}\|\pi_t).$$

In addition, by Pinsker's inequality, we have

$$D(\pi \|\pi_t) \geq \frac{1}{2} \|\pi - \pi_t\|_1^2,$$

for all $\pi \in \Delta^M$. It follows from the two inequalities above that for any $\pi \in \Delta^M$ and $t \in \mathbb{N}$,

$$\begin{aligned} \langle \ell_t, \pi_t - \pi \rangle &\leq \frac{1}{\eta_t} [D(\pi \|\pi_t) - D(\pi \|\pi_{t+2})] + \|\ell_t\|_\infty \|\pi_{t+2} - \pi_t\|_1 - \frac{1}{2\eta_t} \|\pi_{t+2} - \pi_t\|_1^2 \\ &\leq \frac{1}{\eta_t} [D(\pi \|\pi_t) - D(\pi \|\pi_{t+2})] + \frac{\eta_t}{2} \|\ell_t\|_\infty^2. \end{aligned}$$

Since η_t is decreasing in t and $\|\ell_t\|_\infty \leq B_{t+1}$,

$$\begin{aligned} \sum_{t=1}^T \langle \ell_t, \pi_t \rangle - \sum_{t=1}^T \langle \ell_t, \pi \rangle &\leq \frac{1}{\eta_T} \sum_{t=1}^T [D(\pi \|\pi_t) - D(\pi \|\pi_{t+2})] + \frac{1}{2} \sum_{t=1}^T \eta_t \|\ell_t\|_\infty^2 \\ &\leq \frac{1}{\eta_T} [D(\pi \|\pi_1) + D(\pi \|\pi_2)] + \frac{1}{2} \sum_{t=1}^T \eta_t B_{t+1}^2 \\ &\leq \frac{2 \log\{M\}}{\eta_T} + \frac{1}{2} \sum_{t=1}^T \eta_t B_{t+1}^2, \end{aligned}$$

for any $T \in \mathbb{N}$. Substituting $\eta_t = \frac{2}{\bar{B}_t} \sqrt{\frac{\log\{M\}}{t}}$ into the last inequality yields

$$\sum_{t=1}^T \langle \ell_t, \pi_t \rangle - \sum_{t=1}^T \langle \ell_t, \pi \rangle \leq B_T \sqrt{T \log\{M\}} + \sqrt{\log\{M\}} \sum_{t=1}^T \frac{B_{t+1}^2}{B_t \sqrt{t}}. \quad (\text{A.1})$$

We first establish the upper bound on R_T for the case of $\bar{B}_T > B_1$. The right hand side of Inequality (A.1) is bounded above by

$$\bar{B}_T \sqrt{T \log\{M\}} + \gamma_u \bar{B}_T \sqrt{\log\{M\}} \int_0^T \frac{1}{\sqrt{s}} ds = (1 + 2\gamma_u) \bar{B}_T \sqrt{T \log\{M\}} \quad (\text{A.2})$$

because $B_{t+1}^2/B_t \leq \bar{B}_T^2/B_1 = \gamma_u \bar{B}_T$ for each $t = 1, \dots, T$. By Jensen's inequality, we have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 &= \frac{1}{T} \sum_{t=1}^T \left(\sum_{c=1}^M \pi_{t,c} (y_t - \hat{y}_{t,c}) \right)^2 \\ &\leq \frac{1}{T} \sum_{t=1}^T \sum_{c=1}^M \pi_{t,c} (y_t - \hat{y}_{t,c})^2 \\ &= \frac{1}{T} \sum_{t=1}^T \langle \ell_t, \pi_t \rangle. \end{aligned} \quad (\text{A.3})$$

Combining Inequalities (A.1)-(A.3), we obtain

$$R_T \leq \frac{1}{T} \sum_{t=1}^T \langle \ell_t, \pi_t \rangle - \min_{\pi \in \Delta^M} \frac{1}{T} \sum_{t=1}^T \langle \ell_t, \pi \rangle \leq (1 + 2\gamma_u) \bar{B}_T \sqrt{\frac{\log\{M\}}{T}}.$$

Let us now move on to the remaining two cases. In both cases, $B_t = B_1$ for all t . It follows that the right hand side of Inequality (A.1) is bounded above by

$$B_1 \sqrt{T \log\{M\}} + B_1 \sqrt{\log\{M\}} \int_0^T \frac{1}{\sqrt{s}} ds = 3B_1 \sqrt{T \log\{M\}}. \quad (\text{A.4})$$

Combining Inequalities (A.1), (A.3) and (A.4) yields

$$R_T \leq 3B_1 \sqrt{\frac{\log\{M\}}{T}}.$$

We complete the proof by noting that $B_1 = \gamma_o \bar{B}_T$ and $B_1 = \bar{B}_T$ correspond to the second and third case in the statement of this theorem, respectively.

A.3 Proof of Theorem 2

A simple modification of the proof of Theorem 1 yields the results.

References

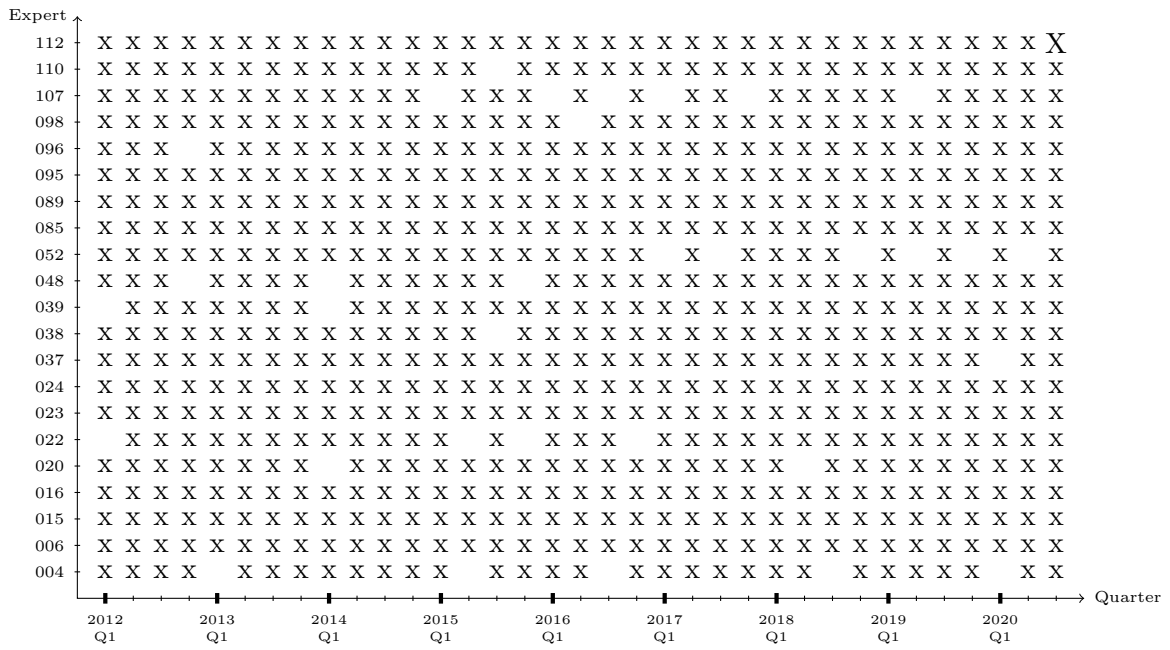
- AIOLFI, M. AND A. TIMMERMANN (2006): “Persistence in Forecasting Performance and Conditional Combination Strategies,” *Journal of Econometrics*, 135, 31–53.
- BATES, J. M. AND C. W. J. GRANGER (1969): “The Combination of Forecasts,” *Journal of the Operational Research Society*, 20, 451–468.
- BELSLEY, D. A., E. KUH, AND R. E. WELSCH (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley-Interscience.
- BERTSIMAS, D. AND J. DUNN (2019): *Machine Learning under a Modern Optimization Lens*, Dynamic Ideas LLC.
- BERTSIMAS, D., A. KING, AND R. MAZUMDER (2016): “Best Subset Selection via a Modern Optimization Lens,” *Annals of Statistics*, 44, 813–852.
- BRADY, H. E. (2019): “The Challenge of Big Data and Data Science,” *Annual Review of Political Science*, 22, 297–323.
- CAPISTRÁN, C. AND A. TIMMERMANN (2009): “Forecast Combination with Entry and Exit of Experts,” *Journal of Business & Economic Statistics*, 27, 428–440.
- CESA-BIANCHI, N. AND G. LUGOSI (2006): *Prediction, Learning, and Games*, Cambridge University Press.

- CESA-BIANCHI, N. AND F. ORABONA (2021): “Online Learning Algorithms,” *Annual Review of Statistics and Its Application*, 8, 165–190.
- CHAUVET, M. AND S. POTTER (2013): “Forecasting Output,” Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 3, 141–194.
- CHEN, L.-Y. AND S. LEE (2018): “Best Subset Binary Prediction,” *Journal of Econometrics*, 206, 39–56.
- CLARIDA, R., J. GALÍ, AND M. GERTLER (2000): “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,” *Quarterly Journal of Economics*, 115, 147–180.
- CLEMEN, R. T. (1989): “Combining Forecasts: A Review and Annotated Bibliography,” *International Journal of Forecasting*, 5, 559–583.
- CONFLITTI, C., C. DE MOL, AND D. GIANNONE (2015): “Optimal Combination of Survey Forecasts,” *International Journal of Forecasting*, 31, 1096–1103.
- CONFORTI, M., G. CORNUÉJOLS, AND G. ZAMBELLI (2014): *Integer Programming*, Springer, Cham.
- DIEBOLD, F. X. (1988): “Serial Correlation and the Combination of Forecasts,” *Journal of Business & Economic Statistics*, 6, 105–111.
- (2015): “Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests,” *Journal of Business & Economic Statistics*, 33, 1–9.
- DIEBOLD, F. X. AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 20, 134–144.
- DIEBOLD, F. X. AND P. PAULY (1990): “The Use of Prior Information in Forecast Combination,” *International Journal of Forecasting*, 6, 503–508.
- DIEBOLD, F. X. AND M. SHIN (2019): “Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian LASSO and its Derivatives,” *International Journal of Forecasting*, 35, 1679–1691.
- EINAV, L. AND J. LEVIN (2014): “Economics in the Age of Big Data,” *Science*, 346.
- ELLIOTT, G., A. GARGANO, AND A. TIMMERMANN (2013): “Complete Subset Regressions,” *Journal of Econometrics*, 177, 357–373.
- ELLIOTT, G. AND A. TIMMERMANN (2016): *Economic Forecasting*, Princeton University Press.
- FREUND, Y. AND R. E. SCHAPIRE (1997): “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” *Journal of Computer and System Sciences*, 55, 119–139.
- FUDENBERG, D. AND D. K. LEVINE (1995): “Consistency and Cautious Fictitious Play,” *Journal of Economic Dynamics and Control*, 19, 1065–1089.

- (1998): *The Theory of Learning in Games*, MIT Press.
- GENRE, V., G. KENNY, A. MEYLER, AND A. TIMMERMANN (2013): “Combining Expert Forecasts: Can Anything Beat the Simple Average?” *International Journal of Forecasting*, 29, 108–121.
- GIACOMINI, R. AND H. WHITE (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74, 1545–1578.
- GRANGER, C. W. J. AND R. RAMANATHAN (1984): “Improved Methods of Combining Forecasts,” *Journal of Forecasting*, 3, 197–204.
- JAGANNATHAN, R. AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *Journal of Finance*, 58, 1651–1683.
- JUDD, K. L. (1998): *Numerical Methods in Economics*, MIT Press.
- JÜNGER, M., T. M. LIEBLING, D. NADDEF, G. L. NEMHAUSER, W. R. PULLEYBLANK, G. REINELT, G. RINALDI, AND L. A. WOLSEY, eds. (2010): *50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art*, Springer, Berlin, Heidelberg.
- LAZER, D. AND J. RADFORD (2017): “Data ex Machina: Introduction to Big Data,” *Annual Review of Sociology*, 43, 19–39.
- LOVELL, M. C. (1983): “Data Mining,” *Review of Economics and Statistics*, 65, 1–12.
- MANSKI, C. F. (2013): *Public Policy in an Uncertain World*, Harvard University Press.
- MARINOVIC, I., M. OTTAVIANI, AND P. SORENSEN (2013): “Forecasters’ Objectives and Strategies,” Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 12, 690–720.
- MAVROEIDIS, S. (2010): “Monetary Policy Rules and Macroeconomic Stability: Some New Evidence,” *American Economic Review*, 100, 491–503.
- PESARAN, M. H. (2015): *Time Series and Panel Data Econometrics*, Oxford University Press.
- PEÑA, D. AND R. S. TSAY (2021): *Statistical Learning for Big Dependent Data*, Wiley.
- ROSSI, B. (2013): “Advances in Forecasting under Instability,” Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 21, 1203–1324.
- SCHAPIRE, R. E. AND Y. FREUND (2012): *Boosting: Foundations and Algorithms*, MIT Press.
- SMITH, J. AND K. F. WALLIS (2009): “A Simple Explanation of the Forecast Combination Puzzle,” *Oxford Bulletin of Economics and Statistics*, 71, 331–355.
- STOCK, J. AND M. WATSON (1999): “A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series,” in *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, ed. by R. Engle and H. White, Oxford University Press, chap. 1, 1–44.
- STOCK, J. H. AND M. W. WATSON (2004): “Combination Forecasts of Output Growth in a Seven-Country Data Set,” *Journal of Forecasting*, 23, 405–430.

- (2006): “Forecasting with Many Predictors,” Elsevier, vol. 1 of *Handbook of Economic Forecasting*, chap. 10, 515–554.
- TANAKA, M., N. BLOOM, J. M. DAVID, AND M. KOGA (2020): “Firm Performance and Macro Forecast Accuracy,” *Journal of Monetary Economics*, 114, 26–41.
- TEBOULLE, M. (2018): “A Simplified View of First Order Methods for Optimization,” *Mathematical Programming*, 1, 67–96.
- TIBSHIRANI, R. (1996): “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.
- TIMMERMANN, A. (2006): “Forecast Combinations,” Elsevier, vol. 1 of *Handbook of Economic Forecasting*, chap. 4, 135–196.
- VARIAN, H. R. (2014): “Big Data: New Tricks for Econometrics,” *Journal of Economic Perspectives*, 28, 3–28.
- WEI, X. AND Y. YANG (2012): “Robust Forecast Combinations,” *Journal of Econometrics*, 166, 224–236.
- WEST, K. D. (2006): “Forecast Evaluation,” Elsevier, vol. 1 of *Handbook of Economic Forecasting*, chap. 3, 99–134.
- YANG, Y. (2004): “Combining Forecasting Procedures: Some Theoretical Results,” *Econometric Theory*, 20, 176–222.

Figure 1: Entry and Exit of Experts



Notes: A slot is marked with the notation x if a forecast is provided but left blank otherwise.

Figure 2: Sample Correlation Coefficients of Forecast Errors

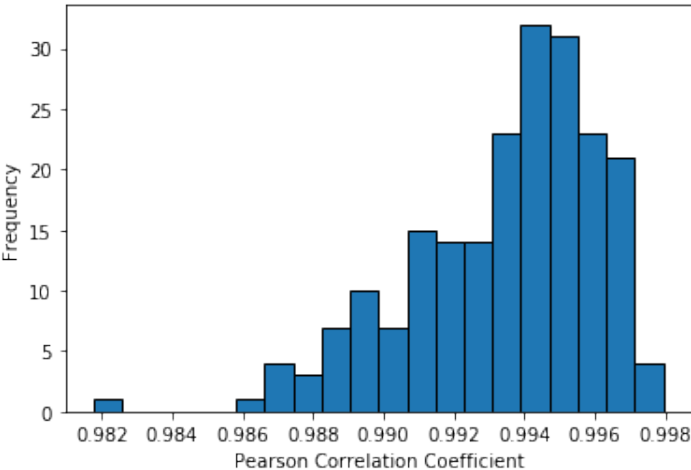
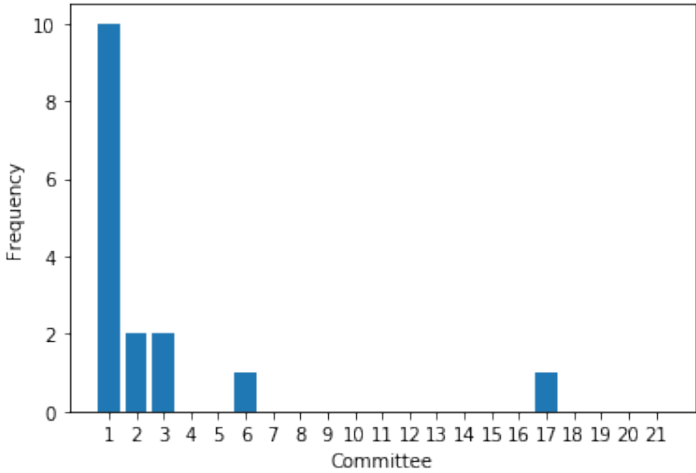


Figure 3: The Frequency of Number of Experts in the Committee Performing Best per Round

(a) HECA without delayed announcements (Algorithm 1 with $l = 2$)



(b) HECA with delayed announcements (Algorithm 2 with $l = 1$)

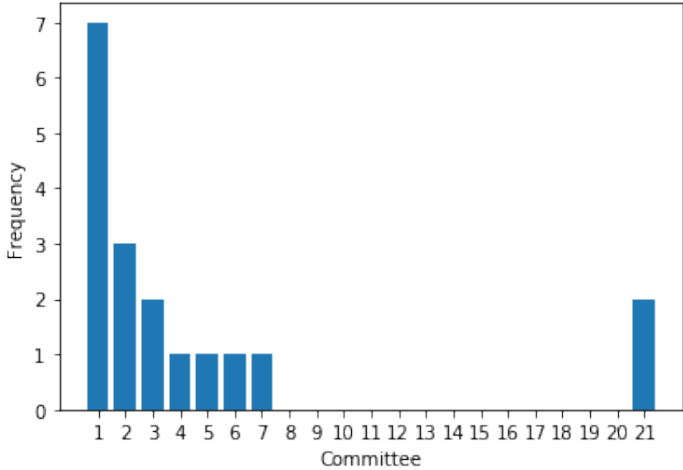


Table 1: Sample Variances of Forecast Errors

8.065	8.687	8.533	8.211	8.139	8.405	8.139
8.178	8.167	8.454	8.310	8.411	8.246	7.935
8.337	8.482	8.532	8.397	8.265	8.882	8.260

Table 2: Forecasting Loss of the Equal-Weight Scheme and HECA

	(1) Equal-Weight Scheme	(2) HECA Algorithm 1	(3) HECA Algorithm 2	Difference (1) - (2)	Difference (1) - (3)
2016Q2	0.0060	-	0.0003	-	0.0057
2016Q3	0.0060	-	0.0053	-	0.0007
2016Q4	0.0051	0.0003	0.0023	0.0048	0.0028
2017Q1	0.1630	0.1000	0.2629	0.0630	-0.0999
2017Q2	0.6170	0.7911	0.5335	-0.1741	0.0835
2017Q3	0.8522	0.9681	0.7981	-0.1160	0.0540
2017Q4	1.0559	1.0032	0.9484	0.0526	0.1074
2018Q1	0.4457	0.3646	0.2621	0.0812	0.1837
2018Q2	0.0393	0.0476	0.0926	-0.0083	-0.0532
2018Q3	0.2743	0.3592	0.4149	-0.0849	-0.1406
2018Q4	0.9174	0.9614	0.9729	-0.0441	-0.0556
2019Q1	0.5464	0.6747	0.6253	-0.1284	-0.0790
2019Q2	0.5781	0.4997	0.4996	0.0784	0.0785
2019Q3	0.1249	0.1667	0.1241	-0.0418	0.0008
2019Q4	0.1521	0.0827	0.0905	0.0694	0.0617
2020Q1	19.4562	19.3347	19.3059	0.1215	0.1502
2020Q2	250.3428	248.8660	244.8226	1.4769	5.5202
2020Q3	29.2636	28.7480	27.9170	0.5156	1.3465

Table 3: Cumulative Forecasting Loss and Average Regret of HECA

HECA			
	Algorithm 1	Best Committee	Difference
2016Q4	0.0003	0.0020	-0.0017
2017Q1	0.1003	0.0581	0.0422
2017Q2	0.8914	0.7823	0.1091
2017Q3	1.8595	1.7421	0.1174
2017Q4	2.8628	2.7036	0.1592
2018Q1	3.2274	3.2053	0.0220
2018Q2	3.2749	3.3391	-0.0642
2018Q3	3.6341	3.8457	-0.2116
2018Q4	4.5956	4.5637	0.0319
2019Q1	5.2703	5.2582	0.0121
2019Q2	5.7700	5.6740	0.0960
2019Q3	5.9367	5.6895	0.2472
2019Q4	6.0194	5.7567	0.2627
2020Q1	25.3541	25.2129	0.1412
2020Q2	274.2201	267.6883	6.5317
2020Q3	302.9680	295.6054	7.3627
Average Regret (2016Q4 ~ 2020Q3)			0.4602
HECA with Delayed Announcements			
	Algorithm 2	Best Committee	Difference
2016Q2	0.0003	0.0164	-0.0161
2016Q3	0.0056	0.0233	-0.0176
2016Q4	0.0079	0.0252	-0.0173
2017Q1	0.2708	0.0814	0.1894
2017Q2	0.8043	0.8056	-0.0013
2017Q3	1.6024	1.4135	0.1889
2017Q4	2.5509	2.3750	0.1795
2018Q1	2.8129	2.5417	0.2712
2018Q2	2.9055	2.6755	0.2300
2018Q3	3.3204	3.1821	0.1383
2018Q4	4.2933	4.3225	-0.0292
2019Q1	4.9186	4.7022	0.2165
2019Q2	5.4183	5.1179	0.3003
2019Q3	5.5424	5.1334	0.4090
2019Q4	5.6328	5.2006	0.4322
2020Q1	24.9388	24.6568	0.2820
2020Q2	269.7614	264.0279	5.7335
2020Q3	297.6784	291.9450	5.7335
Average Regret (2016Q2 ~ 2020Q3)			0.3185

Table 4: Forecasting Loss of the Equal-Weight Scheme, HECA, and Fictitious Play

	Equal-Weight Scheme	HECA Algorithm 1	Fictitious Play (1)	HECA Algorithm 2	Fictitious Play (2)
2016Q2	0.0060	-	-	0.0003	0.0003
2016Q3	0.0060	-	-	0.0053	0.0053
2016Q4	0.0051	0.0003	0.0003	0.0023	0.0023
2017Q1	0.1630	0.1000	0.1000	0.2629	0.2629
2017Q2	0.6170	0.7911	0.7911	0.5335	0.5331
2017Q3	0.8522	0.9681	0.9681	0.7981	0.7981
2017Q4	1.0559	1.0032	1.0033	0.9484	0.9484
2018Q1	0.4457	0.3646	0.3645	0.2621	0.2622
2018Q2	0.0393	0.0476	0.0476	0.0926	0.0925
2018Q3	0.2743	0.3592	0.3594	0.4149	0.4148
2018Q4	0.9174	0.9614	0.9618	0.9729	0.9731
2019Q1	0.5464	0.6747	0.6751	0.6253	0.6253
2019Q2	0.5781	0.4997	0.5000	0.4996	0.4996
2019Q3	0.1249	0.1667	0.1658	0.1241	0.1244
2019Q4	0.1521	0.0827	0.0828	0.0905	0.0905
2020Q1	19.4562	19.3347	19.3325	19.3059	19.3055
2020Q2	250.3428	248.8660	248.9044	244.8226	244.8175
2020Q3	29.2636	28.7480	28.7447	27.9170	27.9170

Table 5: Forecasting Loss of the Hedge Algorithm and HECA

	Freund and Schapire's Hedge Algorithm	HECA Algorithm 1	Difference
2016Q4	0.0051	0.0003	0.0048
2017Q1	0.1630	0.1000	0.0630
2017Q2	0.6184	0.7911	-0.1728
2017Q3	0.8484	0.9681	-0.1197
2017Q4	1.0461	1.0032	0.0428
2018Q1	0.4303	0.3646	0.0657
2018Q2	0.0372	0.0476	-0.0103
2018Q3	0.2904	0.3592	-0.0688
2018Q4	0.9266	0.9614	-0.0349
2019Q1	0.5574	0.6747	-0.1174
2019Q2	0.5640	0.4997	0.0643
2019Q3	0.1269	0.1667	-0.0398
2019Q4	0.1434	0.0827	0.0607
2020Q1	19.3987	19.3347	0.0640
2020Q2	250.1459	248.8660	1.2799
2020Q3	29.0959	28.7480	0.3479

	Freund and Schapire's Hedge Algorithm	HECA Algorithm 2	Difference
2016Q2	0.0060	0.0003	0.0057
2016Q3	0.0056	0.0053	0.0003
2016Q4	0.0053	0.0023	0.0030
2017Q1	0.1621	0.2629	-0.1008
2017Q2	0.6176	0.5335	0.0841
2017Q3	0.8499	0.7981	0.0518
2017Q4	1.0464	0.9484	0.0980
2018Q1	0.4281	0.2621	0.1661
2018Q2	0.0389	0.0926	-0.0537
2018Q3	0.2837	0.4149	-0.1312
2018Q4	0.9227	0.9729	-0.0502
2019Q1	0.5284	0.6253	-0.0970
2019Q2	0.5674	0.4996	0.0677
2019Q3	0.1248	0.1241	0.0007
2019Q4	0.1419	0.0905	0.0514
2020Q1	19.3722	19.3059	0.0663
2020Q2	249.5742	244.8226	4.7516
2020Q3	29.0456	27.9170	1.1286