The University of Sydney Business School
The University of Sydney

BUSINESS ANALYTICS WORKING PAPER SERIES

Global combinations of expert forecasts

Yilin Qian, Ryan Thompson, Andrey L. Vasnev,_
Discipline of Business Analytics, The University of Sydney, NSW 2006 Australia
Department of Econometrics and Business Statistics, Monash University, VIC 3800 Australia

Abstract:
Expert forecast combination—the aggregation of individual forecasts from multiple subject-matter experts— is a proven approach to economic forecasting. To date, research in this area has exclusively concentrated on local combination methods, which handle separate but related forecasting tasks in isolation. Yet, it has been known for over two decades in the machine learning community that global methods, which exploit taskrelatedness, can improve on local methods that ignore it. Motivated by the possibility for improvement, this paper introduces a framework for globally combining expert forecasts. Through our framework, we develop global versions of several existing forecast combinations. To evaluate the efficacy of these new global forecast combinations, we conduct extensive comparisons using synthetic and real data. Our real data comparisons, which involve expert forecasts of core economic indicators in the Eurozone, are the first empirical evidence that the accuracy of global combinations of expert forecasts can surpass local combinations.

Keywords*:* Forecast combination, local forecasting, global forecasting, multi-task learning, European Central Bank, Survey of Professional Forecasters

July 2022

BA Working Paper No: BAWP-2022-02
http://sydney.edu.au/business/business_analytics/research/working_papers

Global combinations of expert forecasts

Yilin Qian[a], Ryan Thompson[b], Andrey L. Vasnev[a,*]

[a]*Discipline of Business Analytics, The University of Sydney, NSW 2006 Australia*
[b]*Department of Econometrics and Business Statistics, Monash University, VIC 3800 Australia*

Abstract

Expert forecast combination—the aggregation of individual forecasts from multiple subject-matter experts—is a proven approach to economic forecasting. To date, research in this area has exclusively concentrated on local combination methods, which handle separate but related forecasting tasks in isolation. Yet, it has been known for over two decades in the machine learning community that global methods, which exploit task-relatedness, can improve on local methods that ignore it. Motivated by the possibility for improvement, this paper introduces a framework for globally combining expert forecasts. Through our framework, we develop global versions of several existing forecast combinations. To evaluate the efficacy of these new global forecast combinations, we conduct extensive comparisons using synthetic and real data. Our real data comparisons, which involve expert forecasts of core economic indicators in the Eurozone, are the first empirical evidence that the accuracy of global combinations of expert forecasts can surpass local combinations.

Keywords: Forecast combination, local forecasting, global forecasting, multi-task learning, European Central Bank, Survey of Professional Forecasters

---

1. Introduction

Forecast combinations—aggregations of multiple individual forecasts—are one of the most persistently reported empirical successes in forecasting. As a key economic institution, the European Central Bank elicits economic forecasts every quarter for the Eurozone from more than one hundred forecasters, an exercise known as the Survey of Professional Forecasters (SPF). Each forecaster has unique expertise, and some possess private information, so combining is a means to a more accurate and robust projection of the economy than any one forecaster could alone produce. For this reason, the Federal Reserve Bank of Philadelphia runs a similar survey by the same name for the United States. Exactly how to combine forecasts from these surveys is a long-standing problem.

Bates and Granger (1969) and later Newbold and Granger (1974) and Granger and Ramanathan (1984) linearly combined forecasts using variance-minimising weights constrained to sum to one—so-called optimal

*Corresponding author.
Email address: andrey.vasnev@sydney.edu.au (Andrey L. Vasnev)

weights. When the forecasts are unbiased, these weights are optimal in a mean square error sense. In practice, however, they are often beaten by equal weights, a curious phenomenon Stock and Watson (2004) called the 'forecast combination puzzle'. This puzzle, explained theoretically by Claeskens et al. (2016), has spurred a formidable research effort to devise improved weighting schemes. Hansen (2008) studied weights that minimise Mallow's criterion, which adds a penalty for complexity. To guarantee a convex combination, Conflitti et al. (2015) added a restriction to prevent negative weights. Matsypura et al. (2018) performed a combinatorial search for the best subset of forecasts to equally weight, and a similar method was proposed in Diebold and Shin (2019) using an l_1-norm penalty. To handle highly correlated forecasts, Radchenko et al. (2021) allowed negative weights but subjected them to a trimming threshold. For other examples in this line of work, see Yang (2004), Aiolfi and Timmermann (2006), Capistrán and Timmermann (2009), Poncela et al. (2011), Genre et al. (2013), Bürgi and Sinclair (2017), Kourentzes et al. (2019), and Qian et al. (2022).

Common to all of the above papers is a focus on using local information to fit the weights, i.e., information that only concerns the forecast target. When just one variable needs forecasting, this approach is sensible. However, it is rare in economics to forecast only a single variable. Instead, forecasts of multiple variables are needed to paint a detailed picture of the economy, core examples being growth, inflation, and unemployment. In addition, policymakers often require forecasts of the economy at different time horizons to facilitate planning. The European Central Bank SPF indeed captures forecasts of multiple variables at multiple horizons, and each variable-horizon pair constitutes an individual forecasting task. Yet, these tasks are not independent; rather, they are highly related. For instance, Okun's law stipulates a strong negative correlation between growth and unemployment (Okun, 1962). The Phillips curve sets forth a similar relationship between unemployment and inflation (Phillips, 1958). It is not unrealistic to expect then that a forecaster's competence in predicting one variable might contain some signal about their competence in predicting another. This possibility motivates us to consider forecast combinations derived from global information shared across related tasks.

The idea of sharing information between prediction tasks emerged during the 1990s in the machine learning community, where it is known as multi-task learning (Caruana, 1997). A vast literature now exists on multi-task learning owing to its success; the interested reader is referred to Zhang and Yang (2021) for a comprehensive survey. Research in the forecasting community itself has lately trended towards multi-task learning (Laptev et al., 2017; Salinas et al., 2020; Godahewa et al., 2021; Montero-Manso and Hyndman, 2021). In the 2018 M4 competition, global methods that shared information across forecasting tasks took out the top-three places (Makridakis et al., 2020). Of these three, the second-place method by Montero-Manso et al. (2020) bears some relation to this work. Their method combined forecasts from a handful of classic time series models using weights from gradient boosted trees. The trees were grown on thousands of time series, enabling weights to be learned across tasks. Though similar, their problem is distinct from the expert forecast combination problem treated in this paper. Whereas Montero-Manso et al. (2020) combined a small number of forecasts for a large number of tasks, we combine a large number of forecasts for a small number of tasks. Elaborate models like boosted trees are not feasible in our setting.

In light of the preceding discussion, this paper proposes a new framework for globally combining expert forecasts. Our framework minimises a global loss function comprised of individual forecasting tasks. The framework is adaptable to the degree of relatedness among the different tasks. Specifically, using a task-coupling penalty, we interpolate between fully local combination, where all tasks are heterogeneous, and fully global combination, where all tasks are homogeneous. The best interpolation is determined in a data-driven fashion. Via this framework, we 'globalise' the weighting schemes of Bates and Granger (1969), Conflitti et al. (2015), and Matsypura et al. (2018). We then evaluate the new global combinations in both simulation and an application to the European Central Bank SPF. The results indicate neither fully local nor fully global combination uniformly performs best. Instead, combinations that lie somewhere between these extremes typically lead to the best out-of-sample performance.

The paper is organised into six sections. Section 2 introduces the proposed framework for globally combining expert forecasts. Section 3 addresses computation of the new combinations. Section 4 presents numerical experiments that gauge the benefits of globalisation. Section 5 describes empirical comparisons of the new methods in application. Section 6 closes the paper.

2. Global forecast combinations

2.1. Single-task forecast combination

To set the scene for our framework, we first describe the traditional single-task forecast combination problem. Let $y \in \mathbb{R}$ be the forecast target and $\boldsymbol{f} = (f_1, \dots, f_p)^\top \in \mathbb{R}^p$ be forecasts of y. Denote by $\boldsymbol{e} = y\mathbf{1} - \boldsymbol{f}$ the forecast errors. It is customary to assume the errors satisfy $\mathrm{E}(\boldsymbol{e}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{e}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a $p \times p$ positive-definite matrix. Consider the linear combination forecast $\tilde{f} = \boldsymbol{f}^\top \boldsymbol{w}$, where $\boldsymbol{w} = (w_1, \dots, w_p)^\top \in \mathbb{R}^p$ are unit sum weights controlling the contribution of individual forecasts to the combination forecast.

Since the forecasts are unbiased and the weights sum to one, the mean square error minimising forecast combination is that which minimises the combination forecast error variance $\mathrm{Var}(\boldsymbol{e}^\top \boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$. This minimisation is performed with respect to a constraint set $\mathcal{W}$:

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}.$$

The simplest configuration of the constraint set is $\mathcal{W}^{\mathrm{eql}} = \{\mathbf{1}/p\}$, yielding *equal weights*. Using $\mathcal{W}^{\mathrm{opt}} = \{\boldsymbol{w} \in \mathbb{R}^p : \mathbf{1}^\top \boldsymbol{w} = 1\}$ leads to *optimal weights* as proposed by Bates and Granger (1969). The constraint set $\mathcal{W}^{\mathrm{optcvx}} = \{\boldsymbol{w} \in \mathbb{R}^p : \mathbf{1}^\top \boldsymbol{w} = 1, \boldsymbol{w} \geq \mathbf{0}\}$, as studied by Conflitti et al. (2015), adds a nonnegativity condition to guarantee a convex combination. The resulting weights are referred to hereafter as *optimal* *convex weights*. A more elaborate configuration, $\mathcal{W}^{\text{opteql}} = \{\boldsymbol{w} \in \mathbb{R}^p : \mathbf{1}^\top \boldsymbol{w} = 1, \boldsymbol{w} = \boldsymbol{z}/(\mathbf{1}^\top \boldsymbol{z}), \boldsymbol{z} \in \{0,1\}^p\}$, produces equal weights restricted to an optimal subset of forecasts. These weights were investigated by Matsypura et al. (2018) and are referred to hereafter as *optimal equal weights*. Here, $\boldsymbol{z}$ is a vector of binary variables z_j ($j = 1, \ldots, p$) which assume the value one if a forecast is selected for inclusion in the combination and zero otherwise. The constraint $\boldsymbol{w} = \boldsymbol{z}/(\mathbf{1}^\top \boldsymbol{z})$ guarantees the selected forecasts are equally-weighted. Other weighting schemes can also be cast in this setup by appropriately choosing $\mathcal{W}$.

When the covariance matrix $\boldsymbol{\Sigma}$ is large-dimensional and estimated from data, it can be helpful to include a ridge penalty (Hoerl and Kennard, 1970) in the objective function (Roccazzella et al., 2022):

$$\min_{\boldsymbol{w} \in \mathcal{W}} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} + \lambda \|\boldsymbol{w}\|^2, \tag{1}$$

where $\lambda \geq 0$. The objective can be rearranged as $\boldsymbol{w}^\top(\boldsymbol{\Sigma} + \lambda \boldsymbol{I})\boldsymbol{w}$, so the ridge penalty has the effect of shrinking the covariance matrix towards the diagonal matrix $\boldsymbol{I}$, thereby stabilising the objective. Though there exist numerous covariance estimators that explicitly perform shrinkage (Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005; Touloumis, 2015), these do not accommodate missing data. Missing data is an important empirical consideration, discussed further in Section 5. On the other hand, it is straightforward to mimic the effect of shrinkage by plugging a standard missing-data covariance estimator into (1). Under all the aforementioned configurations of $\mathcal{W}$, the limiting shrinkage case ($\lambda \to \infty$) leads to equal weights as the optimal solution.

2.2. *Multi-task forecast combination*

The problem described above concerns one forecasting task y. Suppose now we have multiple tasks $\boldsymbol{y} = (y^{(1)}, \ldots, y^{(m)})^\top \in \mathbb{R}^m$. The m tasks may comprise, e.g., different variables or different forecast horizons. We index all quantities relating to the kth component by superscript (k). Hence, the combination forecast $\tilde{\boldsymbol{f}} = (\tilde{f}^{(1)}, \ldots, \tilde{f}^{(m)})^\top \in \mathbb{R}^m$ has elements $\tilde{f}^{(k)} = \boldsymbol{f}^{(k)\top} \boldsymbol{w}^{(k)}$, where $\boldsymbol{f}^{(k)} = (f_1^{(k)}, \ldots, f_p^{(k)})^\top$ and $\boldsymbol{w}^{(k)} = (w_1^{(k)}, \ldots, w_p^{(k)})^\top$. The errors are $\boldsymbol{e}^{(k)} = y^{(k)} \mathbf{1} - \boldsymbol{f}^{(k)}$ with $\operatorname{Var}(\boldsymbol{e}^{(k)}) = \boldsymbol{\Sigma}^{(k)}$.

Though the multi-task setup is typical of economics, research to date has treated the tasks in isolation, using weights fit on a per-task basis:

$$\min_{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)} \in \mathcal{W}} \sum_{k=1}^{m} \left(\boldsymbol{w}^{(k)\top} \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{(k)} + \lambda \|\boldsymbol{w}^{(k)}\|^2 \right). \tag{2}$$

This combination is local because the individual tasks are in no way linked, i.e., solving optimisation problem (1) for each task individually leads to the same weights as solving optimisation problem (2). Information from one task that might be relevant to other tasks is neglected. Instead, one can consider a single vector of weights that is a minimiser of the total loss across all tasks:

$$\min_{\boldsymbol{w} \in \mathcal{W}} \sum_{k=1}^{m} \left(\boldsymbol{w}^\top \boldsymbol{\Sigma}^{(k)} \boldsymbol{w} + \lambda \|\boldsymbol{w}\|^2 \right). \tag{3}$$

This combination is global insofar as the resulting weights take into account information contained in all tasks. Since the loss term in the objective can be expressed equivalently as $\boldsymbol{w}^\top(\sum_{k=1}^m \boldsymbol{\Sigma}^{(k)})\boldsymbol{w}$, this approach can be interpreted as averaging over the task-specific covariance matrices. When the covariance matrices are estimated by the sample covariance matrix, averaging is the same as estimating a single covariance matrix after aggregating data from different tasks. Unfortunately, an implicit assumption underlies this approach that the tasks are completely homogeneous. This assumption might be unreasonably strong in practice and could harm forecast performance.

Rather than committing to a fully local or fully global approach, one can consider bridging the two approaches using per-task weights that are globally regularised:

$$\min_{\substack{\boldsymbol{w}^{(1)},\ldots,\boldsymbol{w}^{(m)}\in\mathcal{W} \\ \bar{\boldsymbol{w}}\in\mathbb{R}^p}} \sum_{k=1}^{m} \left(\boldsymbol{w}^{(k)\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{w}^{(k)} + \lambda\|\boldsymbol{w}^{(k)}\|^2 + \gamma\|\bar{\boldsymbol{w}} - \boldsymbol{w}^{(k)}\|^2 \right). \tag{4}$$

Here, the regulariser $\gamma\sum_{k=1}^m \|\bar{\boldsymbol{w}} - \boldsymbol{w}^{(k)}\|^2$ with $\gamma \geq 0$ is a device to incorporate global information into the per-task weights. It achieves this goal by smoothly penalising departures from an auxiliary weight vector $\bar{\boldsymbol{w}}$ common to all tasks. Taking $\gamma \to \infty$ yields global combination (3), while taking $\gamma \to 0$ yields local combination (2). Hereafter, we refer to the limiting case $\gamma \to \infty$ as 'hard' global combination, and the case with finite non-zero values of γ as 'soft' global combination. These different cases are depicted in Figure 1. The value of γ should reflect the degree of relatedness among tasks—larger values encourage homogeneity,

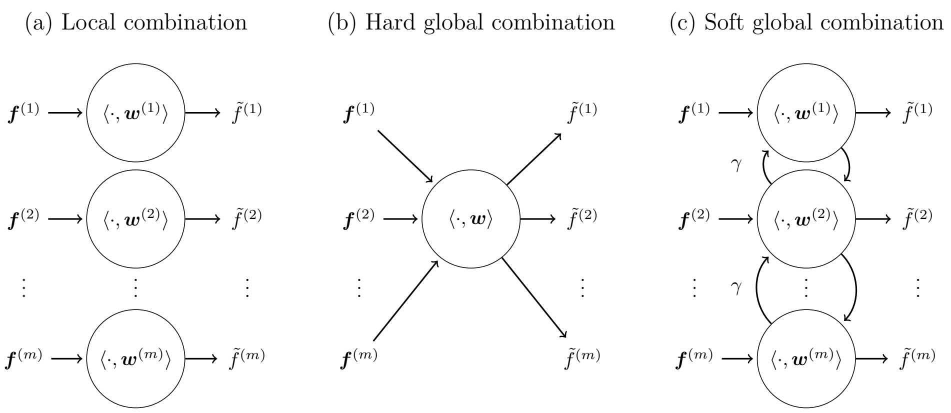

Figure 1: Global and local forecast combination frameworks. The notation $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^\top\boldsymbol{y}$ represents the dot product of two vectors $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{y} \in \mathbb{R}^p$. Local combination learns different weight vectors for each task independently of other tasks. Hard global combination learns one weight vector for all tasks. Soft global combination learns different weight vectors for each task while sharing information between tasks.

while smaller values promote heterogeneity. The best value in terms of out-of-sample forecast performance is usually unknown in application but is estimable from data.

2.3. Alternative formulations

The optimisation problem (4) can be cast solely in terms of the per-task weights $\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}$:

$$\min_{\boldsymbol{w}^{(1)},\ldots,\boldsymbol{w}^{(m)}\in\mathcal{W}} \sum_{k=1}^{m} \left(\boldsymbol{w}^{(k)\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{w}^{(k)} + \lambda\|\boldsymbol{w}^{(k)}\|^2\right) + \Omega_\gamma(\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}),$$

where

$$\Omega_\gamma(\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}) = \min_{\bar{\boldsymbol{w}}\in\mathbb{R}^p} \gamma \sum_{k=1}^{m} \|\bar{\boldsymbol{w}} - \boldsymbol{w}^{(k)}\|^2. \tag{5}$$

Regularisers like Ω_γ, which penalise departures from a common parameter vector, first appeared in the context of multi-task kernel learning (Evgeniou and Pontil, 2004; Evgeniou et al., 2005). When the departures are measured as sums of squared deviations, it is not difficult to obtain a closed-form solution:

$$\Omega_\gamma(\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}) = \gamma \sum_{k=1}^{m} \left\| \frac{1}{m} \sum_{l=1}^{m} \boldsymbol{w}^{(l)} - \boldsymbol{w}^{(k)} \right\|^2.$$

That is, the optimal value of the common parameter vector $\bar{\boldsymbol{w}}$ is the average of the individual parameter vectors $\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}$. One can thus interpret our approach as finding per-task weights within a certain distance of the average weight vector. Some additional algebra gives an alternative expression for Ω_γ:

$$\Omega_\gamma(\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}) = \frac{\gamma}{m} \sum_{k=1}^{m} \sum_{l=1}^{k} \|\boldsymbol{w}^{(l)} - \boldsymbol{w}^{(k)}\|^2.$$

This expression highlights that our approach explicitly penalises the distance between every weight vector. Our experience is that formulating soft global combination using either of the above analytical solutions yields computational performance similar to that of (4), provided the number of tasks m is not large. When m is large, these closed-form solutions lead to many more quadratic terms in the objective, which can impede computation.

2.4. Task grouping

Sometimes it can be useful to limit the flow of information between certain tasks, e.g., when one or more tasks are unrelated. For this purpose, denote by $\mathcal{G} := \{\mathcal{G}_1, \ldots, \mathcal{G}_g\}$ a collection of g groups of tasks, where $\mathcal{G}_l \subseteq \{1, \ldots, m\}$, $\mathcal{G}_1 \cup \cdots \cup \mathcal{G}_g = \{1, \ldots, m\}$, and $\mathcal{G}_l \cap \mathcal{G}_k = \emptyset$ for all $l \neq k$. Using this notation, one can modify Ω_γ to impose the restriction that only tasks within the same group share information:

$$\Omega_\gamma(\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)}) = \min_{\bar{\boldsymbol{w}}^{(1)},\ldots,\bar{\boldsymbol{w}}^{(g)}\in\mathbb{R}^p} \gamma \sum_{l=1}^{g} \sum_{k\in\mathcal{G}_l} \|\bar{\boldsymbol{w}}^{(l)} - \boldsymbol{w}^{(k)}\|^2,$$

where $\bar{\boldsymbol{w}}^{(l)}$ is an auxiliary weight vector for the lth group. When $\mathcal{G}$ consists of just one group, this grouped version of the regulariser reduces to (5). Conversely, when $\mathcal{G}$ consists of m groups, the grouped regulariser has no globalisation effect, i.e., it leads to local combination. The grouped version is helpful in our application to the SPF data in Section 5 where we study different groups of variables and forecast horizons.

2.5. Task scaling

If the tasks under consideration vary in difficulty, one or more tasks might dominate the objective function. To prevent this behaviour, we consider a scaled version of global combination:

$$\min_{\boldsymbol{w}^{(1)},\ldots,\boldsymbol{w}^{(m)}\in\mathcal{W}} \sum_{k=1}^{m} \frac{\boldsymbol{w}^{(k)\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{w}^{(k)} + \lambda\|\boldsymbol{w}^{(k)}\|^2}{\tau^{(k)}} + \Omega_\gamma(\boldsymbol{w}^{(1)},\ldots,\boldsymbol{w}^{(m)}),$$

where $\tau^{(1)},\ldots,\tau^{(m)} > 0$ are fixed scaling parameters. If the tasks are to be evenly balanced, a suitable value of $\tau^{(k)}$ is the optimal objective value from local combination:

$$\tau^{(k)} = \min_{\boldsymbol{w}\in\mathcal{W}} \boldsymbol{w}^\top\boldsymbol{\Sigma}^{(k)}\boldsymbol{w} + \lambda\|\boldsymbol{w}\|^2.$$

This configuration of $\tau^{(k)}$ places all tasks on equal footing.

3. Optimisation

Computation of forecast combinations in our framework varies in complexity according to the weighting scheme, i.e., the specific configuration of $\mathcal{W}$. We now describe methods for computation for some different weighting schemes: optimal weights of Bates and Granger (1969), optimal convex weights of Conflitti et al. (2015), and optimal equal weights of Matsypura et al. (2018).

3.1. Optimal (convex) weights

Optimal weights and optimal convex weights are natural candidates for our framework. The constraint sets $\mathcal{W}^{\text{opt}} = \{\boldsymbol{w}\in\mathbb{R}^p : \boldsymbol{1}^\top\boldsymbol{w} = 1\}$ and $\mathcal{W}^{\text{optcvx}} = \{\boldsymbol{w}\in\mathbb{R}^p : \boldsymbol{1}^\top\boldsymbol{w} = 1, \boldsymbol{w}\geq\boldsymbol{0}\}$ defining these combinations are convex. All the objective functions described in Section 2 are convex. The resulting convex optimisation problems are efficiently solvable using most mathematical programming solvers; we use `Gurobi`.

3.2. Optimal equal weights

The constraint set defining optimal equal weights is less tractable than that for optimal weights or optimal convex weights. Recall the set is defined by a mix of continuous and discrete variables:

$$\mathcal{W}^{\text{opteql}} = \{\boldsymbol{w}\in\mathbb{R}^p : \boldsymbol{1}^\top\boldsymbol{w} = 1, \boldsymbol{w} = \boldsymbol{z}/(\boldsymbol{1}^\top\boldsymbol{z}), \boldsymbol{z}\in\{0,1\}^p\}. \tag{6}$$

The integrality constraint $\boldsymbol{z}\in\{0,1\}^p$ is nonconvex but is amenable to a mixed-integer programming solver such as `Gurobi`. The constraint $\boldsymbol{w} = \boldsymbol{z}/(\boldsymbol{1}^\top\boldsymbol{z})$ is also nonconvex but cannot be handled directly by `Gurobi`. Matsypura et al. (2018) used the decomposition $\mathcal{W}^{\text{opteql}} = \cup_{s=1,\ldots,p}\mathcal{W}_s^{\text{opteql}}$, where $\mathcal{W}_s^{\text{opteql}} = \{\boldsymbol{w}\in\mathbb{R}^p : \boldsymbol{1}^\top\boldsymbol{w} = 1, \boldsymbol{w} = \boldsymbol{z}/s, \boldsymbol{z}\in\{0,1\}^p\}$ is the set of all vectors that equally weight s forecasts. Since s is fixed for $\mathcal{W}_s^{\text{opteql}}$, the constraint $\boldsymbol{w} = \boldsymbol{z}/s$ is linear. The authors sequentially optimise over $\mathcal{W}_1^{\text{opteql}},\ldots,\mathcal{W}_p^{\text{opteql}}$ and retain a solution with minimal objective value. This decomposition approach is, however, infeasible in our framework, because different tasks need not combine the same number of forecasts. To this end, we use a new one-step approach which directly optimises over $\mathcal{W}^{\text{opteql}}$. Though this new approach is proposed for the purpose of globally combining forecasts, it may be of independent interest for local forecast combination. We have found it to be to be uniformly faster than the approach in Matsypura et al. (2018) in the single-task setting, sometimes by an order of magnitude.

First, we rewrite the constraint $\boldsymbol{w} = \boldsymbol{z}/(\mathbf{1}^\top \boldsymbol{z})$ as the pair of constraints $\boldsymbol{w}s = \boldsymbol{z}$ and $s = \mathbf{1}^\top \boldsymbol{z}$, where $s \in \{1, \ldots, p\}$. The new constraint $\boldsymbol{w}s = \boldsymbol{z}$ is *bilinear* in $\boldsymbol{w}$ and s, meaning it is linear for fixed $\boldsymbol{w}$ or fixed s. Though this bilinear constraint remains nonconvex, it is amenable to spatial branch-and-bound techniques (Liberti, 2008) which are similar to classic branch-and-bound techniques used for handling integrality constraints. As of version 9, released in 2020, `Gurobi` can solve optimisation problems with bilinear constraints to global optimality. We now rewrite the constraint set (6) using the new bilinear constraint representation:

$$\mathcal{W}^{\text{opteql}} = \{\boldsymbol{w} \in \mathbb{R}^p : \mathbf{1}^\top \boldsymbol{w} = 1, \boldsymbol{w}s = \boldsymbol{z}, s = \mathbf{1}^\top \boldsymbol{z}, s \in \{1, \ldots, p\}, \boldsymbol{z} \in \{0, 1\}^p\}.$$

The constraint $s = \mathbf{1}^\top \boldsymbol{z}$ is, in fact, redundant in the above characterisation of $\mathcal{W}^{\text{opteql}}$ since it is implied by the remaining constraints. Our experience is that `Gurobi` benefits from excluding it.

4. Synthetic data experiments

4.1. Simulation design

We evaluate the possible gains from global forecast combination in simulation. We work directly with the forecast errors which are sampled from a p-dimensional Gaussian $\boldsymbol{e}_t^{(k)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{(k)})$ for $t = 1, \ldots, T$ and $k = 1, \ldots, m$. We fix $p = T = 50$, so the number of forecasters is of the same order as the number of samples. The number of tasks $m \in \{2, 5, 10\}$. The covariance matrices $\boldsymbol{\Sigma}^{(1)}, \ldots, \boldsymbol{\Sigma}^{(m)}$ are constructed element-wise as $\Sigma_{ij}^{(k)} = \sigma_i^{(k)} \sigma_j^{(k)} \rho^{|i-j|}$. The correlation parameter $\rho = 0.75$ to induce high correlations between forecasters, typical of forecaster surveys. For forecaster $j = 1, \ldots, p$, the standard deviations $\sigma_j^{(1)}, \ldots, \sigma_j^{(m)}$ are generated by drawing correlated random variates uniformly distributed on $[a, b]^m$ with correlation coefficient $\alpha \in \{0, 1/3, 2/3, 1\}$. The parameter α dictates the degree of task relatedness. As α approaches one, a forecaster's performance on one task is strongly indicative of their performance on other tasks. The converse is true as α approaches zero—a forecaster's performance on one task is weakly indicative of their performance on other tasks. The bounds $a = 1$ and $b = 3$ so the accuracy of the worst forecaster is up to three times poorer than that of the best forecaster.

As a measure of out-of-sample accuracy, we report the mean square forecast error on an infinitely large testing set relative to that from an oracle:

$$\text{MSFE relative to oracle} := \frac{(\hat{\boldsymbol{w}}^{(1)} - \boldsymbol{w}^{(1)})^\top \boldsymbol{\Sigma}^{(1)} (\hat{\boldsymbol{w}}^{(1)} - \boldsymbol{w}^{(1)})}{\boldsymbol{w}^{(1)\top} \boldsymbol{\Sigma}^{(1)} \boldsymbol{w}^{(1)}},$$

where $\hat{\boldsymbol{w}}^{(1)}$ denotes estimated weights for task one fit using an estimate $\hat{\boldsymbol{\Sigma}}^{(1)}$ of the true covariance matrix $\boldsymbol{\Sigma}^{(1)}$, and $\boldsymbol{w}^{(1)}$ denotes oracle weights fit using $\boldsymbol{\Sigma}^{(1)}$. We restrict our attention to the relative forecast error of the first task only to measure the marginal effect of adding additional tasks. The covariance matrices are estimated using the sample covariances $\hat{\Sigma}_{ij}^{(k)} = T^{-1}\sum_{t=1}^{T} e_{it}^{(k)} e_{jt}^{(k)}$ for all $(i,j) \in \{1,\ldots,p\}^2$.

The shrinkage parameter λ is swept over a grid of ten values evenly spaced on a logarithmic scale between 10^3 and 10^{-3}. For every value of λ, the globalisation parameter γ of soft global combination is swept over the same grid. The best values of λ and γ are chosen on a validation set constructed independently and identically to the training set, which we remark approximates the precision of leave-one-out cross-validation.

The simulations are run in parallel with `Gurobi` given a single core of an AMD Ryzen Threadripper 3970x and a 300 second time limit.

4.2. Forecast performance

Figure 2 reports the relative forecast errors from 30 simulations. The first row of plots is where the estimate $\hat{\boldsymbol{w}}$ and oracle $\boldsymbol{w}$ are fit under the sum to one constraint that defines optimal weights. The second and third rows correspond to the cases where $\hat{\boldsymbol{w}}$ and $\boldsymbol{w}$ are fit under the constraints that define optimal convex weights and optimal equal weights, respectively. The relative forecast error reported is not comparable across these three weighting schemes since the oracle is different in each case. Our goal is not to compare weighting schemes but rather to measure the benefits of globalisation.

Since local combination ignores information in additional tasks, its performance stays fixed as both the number of tasks and task relatedness increase. In contrast, the relative forecast error of hard global combination decreases roughly linearly with task relatedness, providing for substantial improvements when task relatedness is high. Yet, when task relatedness is low, hard global combination can underperform relative to local combination. This poor performance is made worse by adding additional tasks.

Soft global combination ameliorates the poor performance of hard global combination when the tasks are unrelated and nearly performs as well as hard global combination when the tasks are identical. There is, of course, a statistical cost to estimating the best level of globalisation. Between the extremes, soft global combination successfully adapts to the level of task relatedness to improve over both local and hard global combination. The greater the number of tasks, the greater the possibility for improvement.

Among the three weighting schemes, optimal weights benefit most from globalisation. The constraint set that defines optimal weights is unbounded, and thus its relative forecast error can be arbitrarily bad. Optimal convex weights and optimal equal weights are defined by bounded constraint sets, so there exist finite upper bounds on their relative forecast errors. Thus, the opportunity to improve these weights is somewhat less than for optimal weights, yet often still substantial.

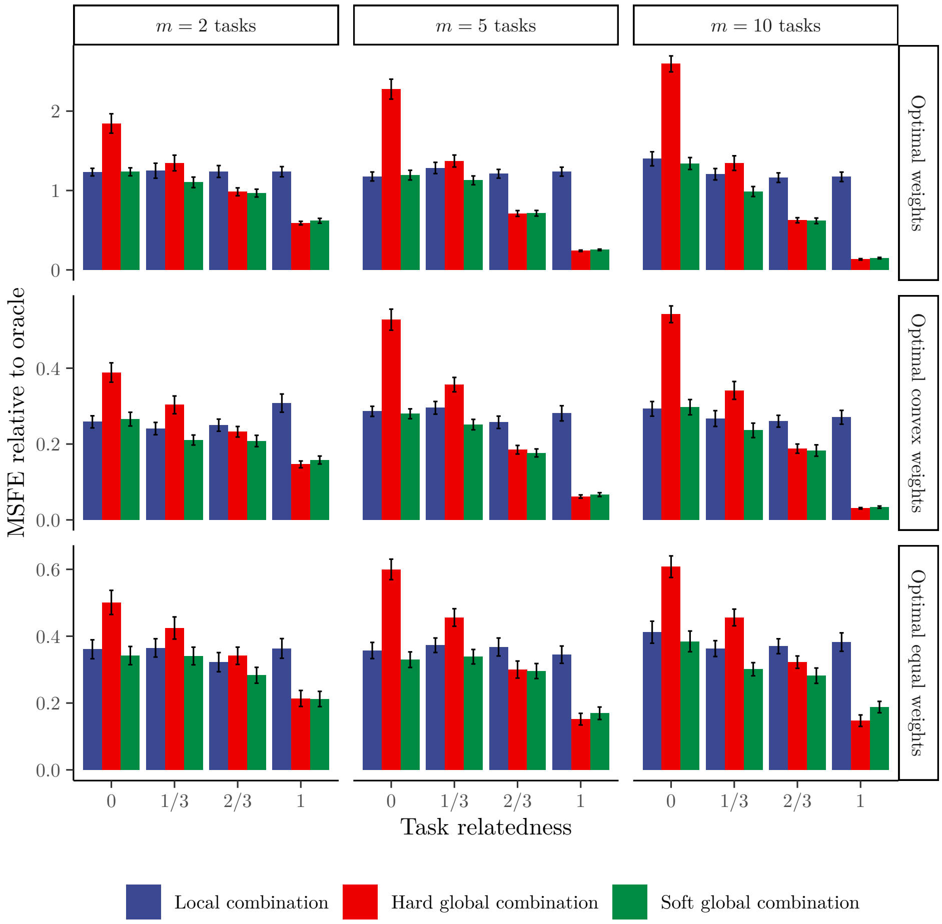

Figure 2: Mean square forecast error as a function of task relatedness parameter α for 30 synthetic datasets. Vertical bars represent averages and error bars denote one standard errors. All values are relative to oracle weights.

5. Survey of Professional Forecasters

5.1. Data and methodology

The European Central Bank SPF is an ongoing survey eliciting predictions for rates of growth, inflation, and unemployment from forecasters for the Eurozone. The survey has been conducted quarterly since 1999 Q1. In each round, the survey participants are asked to provide predictions of the three variables at several time horizons. We focus on the two rolling horizons in this paper, which are one and two years ahead of the latest available observation of the respective variable. For instance, in the 1999 Q1 survey, one-year forecasts corresponded to 1999 Q3 for growth, December 1999 for inflation, and November 1999 for unemployment.[1] The total number of forecasting tasks $m = 6$.

[1] To simplify exposition, forecasts of inflation and unemployment are referred to by the quarter they belong to, e.g., December 1999 inflation and November 1999 unemployment are called forecasts of 1999 Q4.

The SPF data is publicly available at the European Central Bank Statistical Data Warehouse (SDW). Actual values of inflation and unemployment are also available at the SDW. Actual values of growth are available from Eurostat. We access data at the SDW using the `R` package `ecb`, and data from Eurostat using the `R` package `eurostat`. The data used in this paper was retrieved on 17 April 2022. After merging the forecasts and actual values, between $T = 85$ and $T = 90$ observations are available. The first observations are 1999 Q3 (one-year growth), 1999 Q4 (one-year inflation and unemployment), 2000 Q3 (two-year growth), and 2000 Q4 (two-year inflation and unemployment). The last observation is 2021 Q4.

A notable feature of the SPF is that forecasters enter and exit the survey at different times. This aspect of the survey, coupled with periodic nonresponse, gives rise to a sizeable portion of missing data. To deal with this issue, we follow previous works (Matsypura et al., 2018; Radchenko et al., 2021) and filter the data to only include forecasters who respond for a reasonable number of periods. Specifically, the forecasters who provide a minimum of 40 forecasts (10 years) for every task over the full training set (1999 Q3 to 2019 Q4) are retained. This filtering criterion leads to a dataset comprising $p = 34$ forecasters. Figure 3 plots the filtered forecasts alongside actual values of the forecast targets.

To handle missing values that remain after filtering, the covariance matrices of forecast errors are estimated using all complete pairs of observations: $\hat{\Sigma}_{ij}^{(k)} = |\mathcal{T}_i^{(k)} \cap \mathcal{T}_j^{(k)}|^{-1} \sum_{t \in \mathcal{T}_i^{(k)} \cap \mathcal{T}_j^{(k)}} e_{it}^{(k)} e_{jt}^{(k)}$ for all $(i, j) \in \{1, \ldots, p\}^2$. Here, $\mathcal{T}_i^{(k)}$ denotes the periods in the training set where forecaster i provided a forecast for task k. Covariance matrices constructed in this manner are not guaranteed positive-definite. For this reason, we take the positive-definite matrix nearest to $\hat{\boldsymbol{\Sigma}}^{(k)}$ using `nearPD` from the `R` package `Matrix`. The forecast errors are standardised by the standard deviation of the forecast targets as estimated on the training set prior to estimating the covariance matrices.

5.2. Globalisation path

The first set of experiments study the evolution of out-of-sample forecast performance as the globalisation parameter γ is swept over its support (the 'globalisation path'). As a measure of out-of-sample accuracy, we report the mean square forecast error on a testing set relative to that from local combination:

$$\text{MSFE relative to local} := \frac{\sum_{t=\underline{T}-h}^{\bar{T}-h} (y_{t+h}^{(k)} - \tilde{f}_{t+h|t}^{(k)(\gamma)})^2}{\sum_{t=\underline{T}-h}^{\bar{T}-h} (y_{t+h}^{(k)} - \tilde{f}_{t+h|t}^{(k)(0)})^2},$$

where, for a given weighting scheme, $\tilde{f}_{t+h|t}^{(k)(\gamma)}$ is a global combination forecast of task k at time $t+h$ produced using a training set up to time t with $\gamma \in [0, \infty)$, and $\underline{T}$ and $\bar{T}$ are the first and last periods in the testing set. The denominator is the mean square forecast error from setting $\gamma = 0$, so this metric is the percentage improvement due to globalisation. We pick $\underline{T}$ and $\bar{T}$ so the testing set is the last five years to 2019 Q4. The period after 2019 Q4, covering the COVID-19 recession and 2021–2022 inflation surge, is considered in separate experiments in Section 5.3.

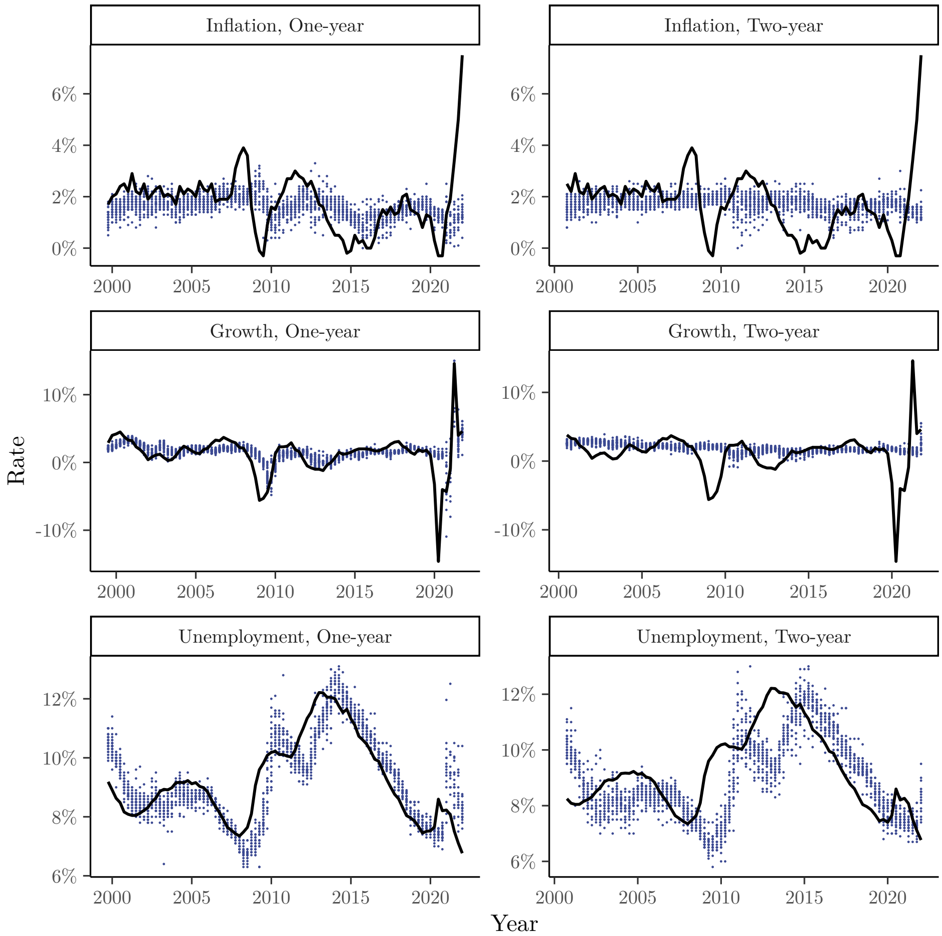

Figure 3: Data from the Survey of Professional Forecasters. Points represent forecasts and lines denote actual values of the forecast target. Forecasts from participants with low response rates are not included.

Figures 4, 5, and 6 report the globalisation paths of optimal weights, optimal convex weights, and optimal equal weights for fixed shrinkage parameter $\lambda = 10^{-1}$. The globalisation paths of optimal equal weights are step functions in γ due to the weights being discrete. Three ways of grouping the tasks are considered: grouping variable tasks (group 1: one-year growth, inflation, and unemployment; group 2: two-year growth, inflation, and unemployment); grouping forecast horizon tasks (group 1: one- and two-year growth; group 2: one- and two-year inflation; group 3: one- and two-year unemployment); and grouping all tasks (group 1: one- and two-year growth, inflation, and unemployment). The reader is reminded information flows only between tasks belonging to the same group.

Across all weighting schemes and tasks, there is always a globalisation path that attains its minimum at some positive amount of globalisation. The limiting case $\gamma \to \infty$, hard global combination, is sometimes helpful and sometimes harmful. For instance, growth and inflation realise roughly 15% improvement from

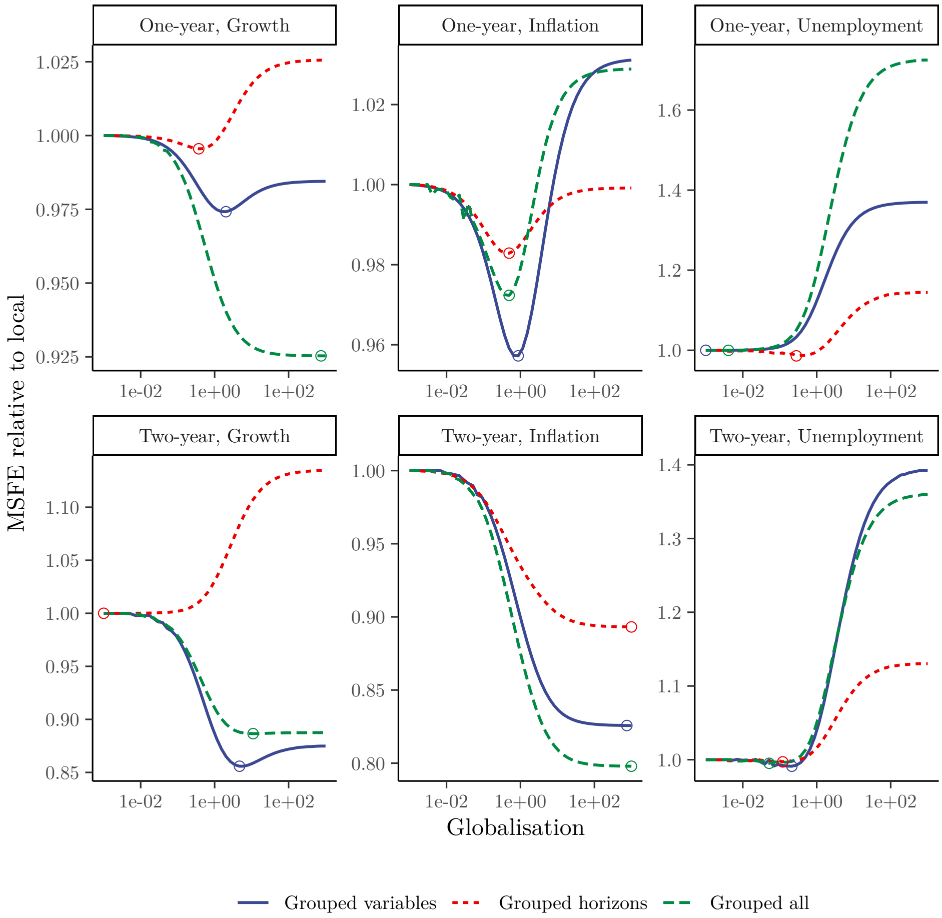

Figure 4: Mean square forecast error of *optimal weights* as a function of globalisation parameter γ for the Survey of Professional Forecasters. Testing period is 2015 Q1 to 2019 Q4. Minimum of each curve is marked by a circle. All values are relative to local combination ($\gamma \to 0$).

hard global combination (optimal weights, grouped variables) at the two-year horizon while unemployment deteriorates by about 40% at the same horizon. This behaviour might be attributable to growth and inflation being difficult tasks at the two-year horizon, thus providing a noisy signal to unemployment. However, for a suitable choice of γ, any negative effect on unemployment can be mitigated.

The degree to which different groupings produce good global combinations can vary by weighting scheme. For example, grouped horizons are the most useful grouping when forecasting growth under optimal convex weights and optimal equal weights. In contrast, grouped horizons are the least useful grouping under optimal weights. The additional structure imposed by the former weighting schemes (in particular, nonnegativity) appears critical to disentangling signal from noise here. Overall, no one grouping of the tasks dominates the other groupings. Though it does not uniformly perform best, grouping all tasks together seems a sensible

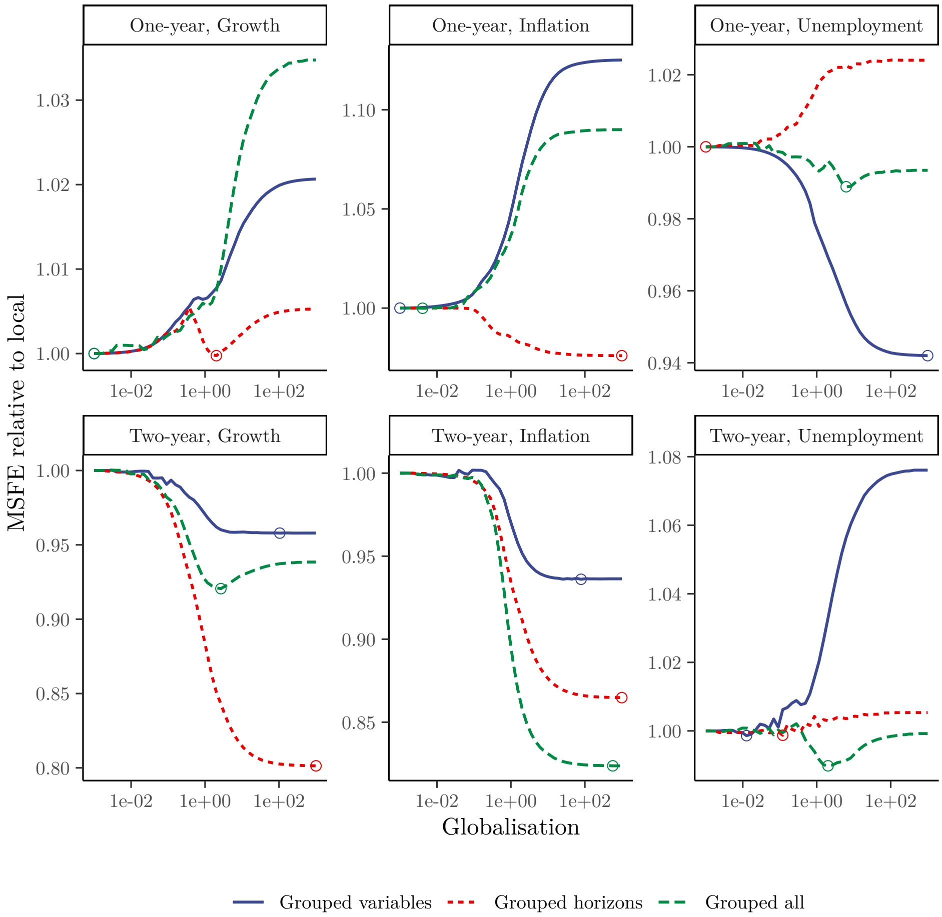

Figure 5: Mean square forecast error of *optimal convex weights* as a function of globalisation parameter γ for the Survey of Professional Forecasters. Testing period is 2015 Q1 to 2019 Q4. Minimum of each curve is marked by a circle. All values are relative to local combination ($\gamma \to 0$).

default provided γ is chosen judiciously on a task-by-task basis.

5.3. Tuned globalisation

The second set of experiments are broader comparisons that acknowledge the level of globalisation requires tuning in practice. For this purpose, we use leave-one-out cross-validation—a valid procedure provided the combination forecast errors are uncorrelated (Bergmeir et al., 2018). The value of γ is tuned per task, so different tasks need not use the same value. To allow for comparisons of forecast accuracy across weighting schemes, we report the mean square forecast error relative to that from equal weights, a common benchmark in practice:

$$\text{MSFE relative to equal} := \frac{\sum_{t=\underline{T}-h}^{\bar{T}-h}(y_{t+h}^{(k)} - \tilde{f}_{t+h|t}^{(k)})^2}{\sum_{t=\underline{T}-h}^{\bar{T}-h}(y_{t+h}^{(k)} - \bar{f}_{t+h|t}^{(k)})^2},$$

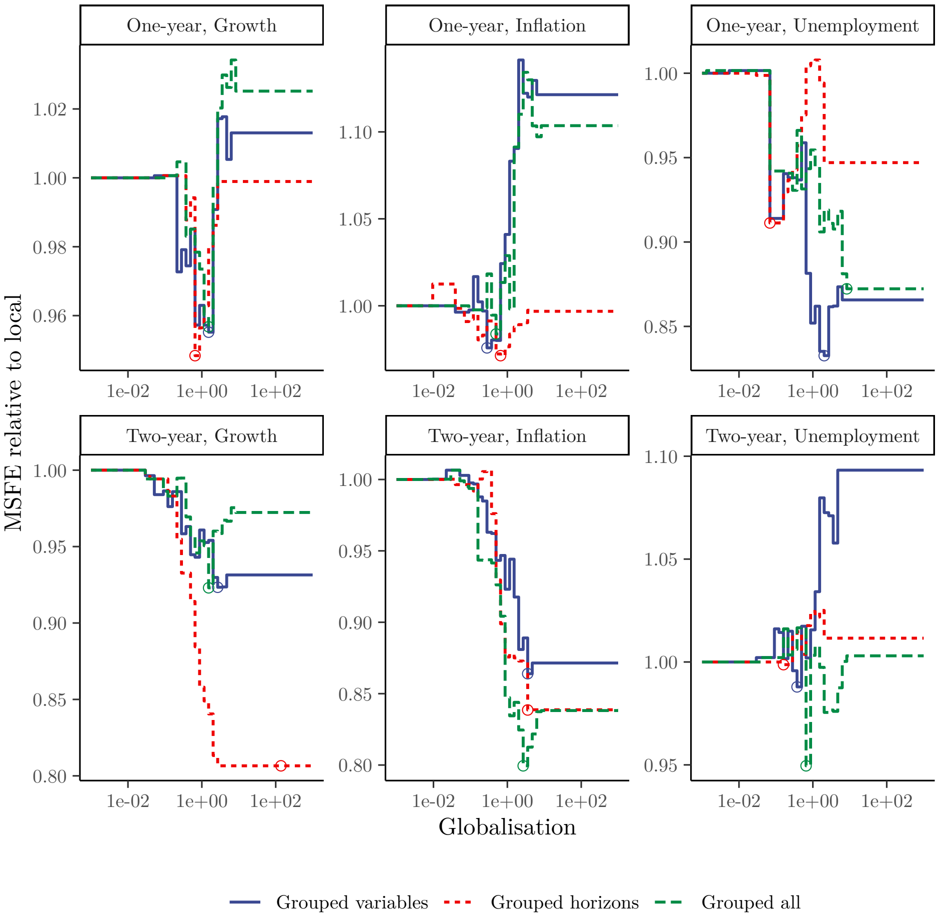

Figure 6: Mean square forecast error of *optimal equal weights* as a function of globalisation parameter γ for the Survey of Professional Forecasters. Testing period is 2015 Q1 to 2019 Q4. Minimum of each curve is marked by a circle. All values are relative to local combination ($\gamma \to 0$).

where $\tilde{f}^{(k)}_{t+h|t}$ is an arbitrary combination forecast and $\bar{f}^{(k)}_{t+h|t}$ is the equally-weighted combination forecast. Values of this metric less than one indicate superior performance to equal weights.

Table 1 reports the average value of the performance metric across the six tasks, with the minimal and maximal values among the tasks in brackets. The shrinkage parameter $\lambda = 10^{-1}$. The last five years of the data is again studied, but we now include the period 2020 Q1 to 2021 Q4 to evaluate recent performance during the COVID-19 recession and 2021–2022 inflation surge. Figure 3 highlights how the quarters on and after 2020 Q1 contain several outliers. To prevent these outliers dominating the performance metric, the testing set is split before and after 2020 Q1. Likewise, to avoid the outliers contaminating the estimated covariance matrices and thus the estimated weights, the training set is stopped at 2019 Q4.

With few exceptions, soft global combination improves on local combination. The improvements are

	Local combination	Hard global combination	Soft global combination
	2017 Q1 to 2019 Q4		
Optimal weights			
Grouped variables	1.059 [0.272, 2.354]	0.909 [0.512, 1.977]	0.907 [0.359, 2.207]
Grouped horizons	1.059 [0.272, 2.354]	0.956 [0.306, 1.830]	0.967 [0.299, 1.933]
Grouped all	1.059 [0.272, 2.354]	0.894 [0.592, 1.669]	0.856 [0.360, 1.720]
Optimal convex weights			
Grouped variables	0.991 [0.878, 1.159]	1.033 [0.878, 1.410]	0.968 [0.867, 1.111]
Grouped horizons	0.991 [0.878, 1.159]	0.957 [0.859, 1.178]	0.981 [0.879, 1.177]
Grouped all	0.991 [0.878, 1.159]	1.001 [0.893, 1.237]	0.980 [0.867, 1.183]
Optimal equal weights			
Grouped variables	1.005 [0.861, 1.223]	1.016 [0.861, 1.298]	0.959 [0.848, 1.097]
Grouped horizons	1.005 [0.861, 1.223]	0.962 [0.873, 1.188]	0.986 [0.866, 1.181]
Grouped all	1.005 [0.861, 1.223]	1.018 [0.880, 1.384]	0.972 [0.850, 1.146]
	2020 Q1 to 2021 Q4		
Optimal weights			
Grouped variables	1.046 [0.913, 1.269]	1.058 [0.959, 1.210]	1.023 [0.931, 1.189]
Grouped horizons	1.046 [0.913, 1.269]	1.002 [0.784, 1.171]	0.994 [0.718, 1.156]
Grouped all	1.046 [0.913, 1.269]	1.007 [0.959, 1.078]	0.995 [0.858, 1.089]
Optimal convex weights			
Grouped variables	1.006 [0.953, 1.072]	0.953 [0.815, 1.032]	0.992 [0.907, 1.044]
Grouped horizons	1.006 [0.953, 1.072]	1.009 [0.947, 1.048]	1.005 [0.957, 1.051]
Grouped all	1.006 [0.953, 1.072]	0.931 [0.764, 1.009]	0.999 [0.947, 1.042]
Optimal equal weights			
Grouped variables	1.012 [0.942, 1.112]	0.970 [0.772, 1.082]	1.017 [0.949, 1.112]
Grouped horizons	1.012 [0.942, 1.112]	1.029 [0.959, 1.140]	1.023 [0.954, 1.112]
Grouped all	1.012 [0.942, 1.112]	0.934 [0.681, 1.027]	0.997 [0.920, 1.112]

Table 1: Mean square forecast errors for the Survey of Professional Forecasters. Averages over all tasks are next to minimums and maximums over all tasks in brackets. All values are relative to equal weights.

generally greatest pre-2020. The more minor improvements post-2020 are possibly a consequence of the recent period of deteriorated economic conditions during which task relatedness could be less stable. In some instances, hard global combination outperforms both soft global combination and local combination. However, as in the previous section, it also sometimes underperforms. On the other hand, the data-driven determination of the globalisation level for soft global combination produces good combinations that consistently forecast well.

Optimal weights realise the most significant gains from globalisation among the three weighting schemes—soft global combination (grouped all) places first in terms of average performance across tasks (pre-2020) compared with local combination, which places last. Moreover, globalisation leads to smaller maximal loss for optimal weights. Though not always beating optimal weights according to average performance, optimal convex weights and optimal equal weights have more consistent performance across tasks, especially pre-2020. With a suitable amount of globalisation, each weighting scheme can beat the notoriously difficult benchmark of equal weights for one or more task groupings.

6. Concluding remarks

Expert forecasts are an indispensable source of forward-looking information on core economic variables. To date, the problem of combining these forecasts has been handled on a per-task basis, with the combination for each variable and forecast horizon learned independently of other variables and horizons. When the forecasting tasks are related, as economic theory and evidence suggest, this approach of learning the combinations using only local information is potentially suboptimal. This paper investigates the value of a global approach, where task-relatedness is directly exploited to improve the quality of combinations. At the heart of our approach is a principled framework that accounts for the level of homogeneity across tasks by adaptively interpolating between fully local and fully global combinations. In addition to unifying local and global approaches under one umbrella, the new framework accommodates many existing weighting schemes. Empirical evidence from the European Central Bank SPF suggests forecast combinations for rates of growth, inflation, and unemployment in the Eurozone benefit from some degree of globalisation, as do combinations of these same variables across one- and two-year horizons. A natural next step in this line of work is to investigate whether similar findings hold for the other economies. Another direction is to extend the ideas of this paper to interval and density forecasts, commonly used as measures of economic uncertainty.

An `R` implementation of the global forecast combinations in this paper is publicly available at

https://github.com/ryan-thompson/global-combinations.

Acknowledgements

Ryan Thompson acknowledges financial support by an Australian Government Research Training Program (RTP) Scholarship.

Conflict of interest

The authors declare no conflict of interest.

References

Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics 135, 31–53.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. OR 20, 451–468.

Bergmeir, C., Hyndman, R.J., Koo, B., 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics and Data Analysis 120, 70–83.

Bürgi, C., Sinclair, T.M., 2017. A nonparametric approach to identifying a subset of forecasters that outperforms the simple average. Empirical Economics 53, 101–115.

Capistrán, C., Timmermann, A., 2009. Forecast combination with entry and exit of experts. Journal of Business and Economic Statistics 27, 428–440.

Caruana, R., 1997. Multitask learning. Machine Learning 28, 41–75.

Claeskens, G., Magnus, J.R., Vasnev, A.L., Wang, W., 2016. The forecast combination puzzle: A simple theoretical explanation. International Journal of Forecasting 32, 754–762.

Conflitti, C., De Mol, C., Giannone, D., 2015. Optimal combination of survey forecasts. International Journal of Forecasting 31, 1096–1103.

Diebold, F.X., Shin, M., 2019. Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. International Journal of Forecasting 35, 1679–1691.

Evgeniou, T., Micchelli, C.A., Pontil, M., 2005. Learning multiple tasks with kernel methods. Journal of Machine Learning Research 6, 615–637.

Evgeniou, T., Pontil, M., 2004. Regularized multi–task learning, in: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 109–117.

Genre, V., Kenny, G., Meyler, A., Timmermann, A., 2013. Combining expert forecasts: Can anything beat the simple average? International Journal of Forecasting 29, 108–121.

Godahewa, R., Bandara, K., Webb, G.I., Smyl, S., Bergmeir, C., 2021. Ensembles of localised models for time series forecasting. Knowledge-Based Systems 233, 107518.

Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3, 197–204.

Hansen, B.E., 2008. Least-squares forecast averaging. Journal of Econometrics 146, 342–350.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Kourentzes, N., Barrow, D., Petropoulos, F., 2019. Another look at forecast selection and combination: Evidence from forecast pooling. International Journal of Production Economics 209, 226–235.

Laptev, N., Yosinski, J., Li, L.E., Smyl, S., 2017. Time-series extreme event forecasting with neural networks at Uber, in: ICML 2017 Time Series Workshop.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88, 365–411.

Liberti, L., 2008. Introduction to global optimization. Technical Report. URL: https://www.lix.polytechnique.fr/~liberti/teaching/globalopt-lima.pdf.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting 36, 54–74.

Matsypura, D., Thompson, R., Vasnev, A.L., 2018. Optimal selection of expert forecasts with integer programming. Omega 78, 165–175.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R.J., Talagala, T.S., 2020. FFORMA: Feature-based forecast model averaging. International Journal of Forecasting 36, 86–92.

Montero-Manso, P., Hyndman, R.J., 2021. Principles and algorithms for forecasting groups of time series: Locality and globality. International Journal of Forecasting 37, 1632–1653.

Newbold, P., Granger, C.W.J., 1974. Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society: Series A (General) 137, 131–165.

Okun, A.M., 1962. Potential GNP: Its measurement and significance, in: Proceedings of the Business and Economic Statistics

Section of the American Statistical Association.

Phillips, A.W., 1958. The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. Economica 25, 283–299.

Poncela, P., Rodríguez, J., Sánchez-Mangas, R., Senra, E., 2011. Forecast combination through dimension reduction techniques. International Journal of Forecasting 27, 224–237.

Qian, W., Rolling, C.A., Cheng, G., Yang, Y., 2022. Combining forecasts for universally optimal performance. International Journal of Forecasting 38, 193–208.

Radchenko, P., Vasnev, A.L., Wang, W., 2021. Too similar to combine? On negative weights in forecast combination. International Journal of Forecasting , In press.

Roccazzella, F., Gambetti, P., Vrins, F., 2022. Optimal and robust combination of forecasts via constrained optimization and shrinkage. International Journal of Forecasting 38, 97–116.

Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting 36, 1181–1191.

Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical Applications in Genetics and Molecular Biology 4.

Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Touloumis, A., 2015. Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings. Computational Statistics and Data Analysis 83, 251–261.

Yang, Y., 2004. Combining forecasting procedures: Some theoretical results. Econometric Theory 20, 176–222.

Zhang, Y., Yang, Q., 2021. A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering , In press.