

COMBINING COMBINED FORECASTS: A NETWORK APPROACH

MARCOS R. FERNANDES[†]

This version: June 2024.

ABSTRACT. This study investigates the practice of experts aggregating forecasts before informing a decision-maker. The significance of this subject extends to various contexts where experts inform their assessments to a decision-maker following discussions with peers. My findings show that, irrespective of the information structure, aggregation rules introduce no bias to decision-making in expected terms. Nevertheless, the concern revolves around variance. In situations where experts are equally precise, and pair-wise correlation of forecasts is the same across all pairs of experts, the network structure plays a pivotal role in decision-making variance. For classical structures, I show that star networks exhibit the highest variance, contrasting with d -regular networks that achieve zero variance, emphasizing their efficiency. Additionally, by employing the Poisson random graph model under the assumptions of a large network size and a small connection probability, the results indicate that both the expected Network Bias and its variance converge to zero as the network size becomes sufficiently large. These insights enhance the understanding of decision-making under different information, network structures and aggregation rules. They enrich the literature on combining forecasts by exploring the effects of prior network communication on decision-making.

JEL Classification: C44, C53, D81, D83, D85.

Keywords: Combining Forecasts, Pooling, Averaging, Social Networks, Social Learning, Bias, Bayesian Inference.

1. INTRODUCTION

The practice of combining forecasts is adopted to enhance forecast accuracy by aggregating information from diverse sources. Even simple combination methods have demonstrated efficacy comparable to more intricate approaches under specific circumstances (Wang, Hyndman, Li, and Kang, 2023; Timmermann, 2006). The foundational work of Bates and Granger (1969) played a pivotal role in popularizing the concept of forecast combinations, initiating a rich strand of research

[†] UNIVERSITY OF SÃO PAULO, DEPARTMENT OF ECONOMICS, E-MAIL: MARCOSROSS@USP.BR.

I thank Luis Alvarez and Matthew O. Jackson for very helpful conversations. I also thank the participants of the 23rd Annual SAET conference, and the 9th annual Network Science and Economics conference. This project has received funding from the University of São Paulo (USP) under the Incoming Assistant Professor Fellowship (Bolsa Novos Doutores).

that has spanned over half a century. Their seminal contribution firmly established the value of forecast combinations. However, the practice of combining forecasts presents several challenges, including situations characterized by consensus among forecasts, substantial divergence, forecast dependence, uncertainties regarding forecast characteristics, instability in the forecasting process, concerns about robustness, and group interaction.

Within the realm of *group interaction*, important drawbacks emerge, such as the susceptibility of the group to unwarranted influence from dominant agents. This exposure could undermine the collaborative benefits of combining forecasts. According to Winkler (1989), valuable insights can be gained by comparing the outcomes of group processes, mechanically combined forecasts, and sequential processes that integrate both mechanical combination and group interaction at different stages. Despite the extensive body of research, the current scholarship does not conclusively delineate the ultimate *consequences of forecast pooling practices under the possibility of prior experts interaction*. Thus, this study examines the effects on decision-making when experts engage in non-strategic pooling of forecasts before presenting them to the decision-maker. Particular emphasis is placed on scenarios where two classical pooling methods—*Linear and Bayesian*—are employed by experts and the decision-maker.

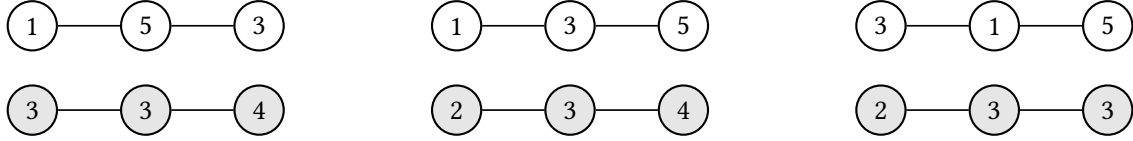
Consider the following *introductory example*, as depicted in Figure 1. Suppose we have three experts, each providing point forecasts to a decision-maker (*external* to and *unaware* of the network) in three different cases. The nodes in white represent the experts' forecasts without network communication, while the nodes in gray represent the experts' updated forecasts under network communication, where the updating process is a *simple average*.

In all cases, if the experts decide not to communicate with other experts, they will provide their original forecasts (i.e., numbers in the white nodes) to the decision-maker. Conversely, if they communicate and employ local simple averages of forecasts within their reach before communicating them to the decision-maker, they will first update their original forecasts (numbers in gray nodes), and then communicate these updated forecasts to the decision-maker.

For instance, consider *Case 1*, in which the experts whose forecasts are 1 and 3 (peripheral nodes) only have access to the expert whose forecast is 5, while the expert with the forecast of 5 (central node) has access to both other experts. If the experts employ simple averages to update their forecasts, the expert with the initial forecast of 1 will update its forecast to 3 $((1 + 5)/2)$, the expert with the initial forecast of 5 will update its forecast to 3 $((1 + 5 + 3)/3)$, and the expert with the initial forecast of 3 will update its forecast to 4 $((5 + 3)/2)$.

Without expert communication, the decision-maker would receive the numbers in the white nodes, whereas with expert communication, the decision-maker would receive the numbers in

the gray nodes. If the decision-maker also employs a simple average to aggregate the experts' forecasts, the final numbers would be different, with or without network communication, i.e., $(1 + 5 + 3)/3 = 3$ (white) and $(3 + 3 + 4)/3 = 10/3 \approx 3.33$ (gray). However, Cases 2 and 3, following the same protocol, illustrate that this difference, which I may refer to *Network Bias* later, may vary: 0 in Case 2 and $-1/3$ in Case 3.



(A) Case 1: diff.: $10/3 - 3 = +1/3$ (B) Case 2: diff.: $3 - 3 = 0$ (C) Case 3: diff.: $8/3 - 3 = -1/3$

FIGURE 1. Example of when all agents employ a simple average aggregation
 White nodes: no network communication (original forecasts).
 Gray nodes: with network communication (updated forecasts using simple average)

This specific introductory example illustrates that both network structure and the realization of forecasts may influence this difference. This paper aims to explain these dynamics in a more general setting where forecasts and network structure can be random.

The relevance of this topic extends to various contexts where experts provide their assessments to decision-makers after consulting with peers. For instance, in the field of health, experts play a crucial role in decisions related to the adoption of new technologies and determining staffing requirements for medical centers. In this context, experts may exchange information regarding their estimates before sharing them with medical centers. In law, expert knowledge serves as evidence in civil proceedings, and experts may establish communication with other experts before releasing their final statements to the judge. In economics, experts influence both key decisions of central banks through their insights into inflation expectations and budgetary decisions through their economic forecasts. In this context, professional forecasters may exchange their forecasts prior to informing central banks and governments. Finally, the relevance of this study is also evident in broader contexts, such as general reviews on platforms like Google and Yelp, where customers may exchange information with other customers before announcing their reviews on these platforms.

The results indicate that, in expected terms and under the same information structure assumption, prior network communication does not introduce bias into decision-making. However, variance remains a central concern. When experts are equally precise and the pair-wise correlation of forecasts is the same, any combination of aggregation rules simplifies to simple averages. In this case, the network structure becomes a crucial factor influencing decision-making variance.

Star networks exhibit the highest variance, which increases with the number of experts consulted, while d -regular networks achieve zero variance, making them particularly efficient structures. From a random graph perspective, sufficiently large networks mitigate both the bias and variance problems, regardless of the expected degree of experts, as long as experts' degrees become independent as the network size increases. These insights provide a deeper understanding of the dynamics associated with different information and network structures, as well as various aggregation rules.

2. LITERATURE REVIEW AND CONTRIBUTION

The method introduced by [Bates and Granger \(1969\)](#) aimed to determine optimal weights by minimizing the variance of the combined forecast error, focusing primarily on combinations of pairs of forecasts. Subsequently, [Newbold and Granger \(1974\)](#) extended this approach to encompass combinations of more than two forecasts. The optimal weight vector was defined as the one that minimizes the variance of the combined forecast error, with weights being indirectly proportional to these variances. In practical applications, the elements of the variance-covariance matrix are typically unknown, requiring estimation. In this context, the simple average of forecasts based on equal weights stands out as a popular and robust combination rule ([Makridakis, Andersen, Carbone, Fildes, Hibon, Lewandowski, Newton, Parzen, and Winkler, 1982](#); [Clemen, 1989](#); [Timmermann, 2006](#); [Hsiao and Wan, 2014](#); [Wang et al., 2023](#); [Makridakis, Spiliotis, and Assimakopoulos, 2020](#)). Attention has also been given to other pooling strategies, including the median, mode, and trimmed and winsorized means ([Genre, Kenny, Meyler, and Timmermann, 2013](#); [Jose, Grushka-Cockayne, and Lichtendahl Jr, 2014](#); [Grushka-Cockayne, Jose, and Lichtendahl Jr, 2017](#)).

The objective behind combining multiple individual forecasts is to enhance accuracy by aggregating information from different sources and to reduce data, parameter, and model uncertainties. A substantial body of literature has developed over the years concerning the combination of individual point forecasts. For a *comprehensive review* of this literature, refer to [Genest and Zidek \(1986\)](#); [Granger \(1989\)](#); [Clemen \(1989\)](#); [Jacobs \(1995\)](#); [Timmermann \(2006\)](#); [Mancuso and Werner \(2013\)](#); [Gneiting and Katzfuss \(2014\)](#); [Wallis \(2014\)](#); [Wang et al. \(2023\)](#). Forecast combinations have found *applications* in a wide range of fields, including energy ([Xie and Hong, 2016](#); [Nielsen, Nielsen, Madsen, Pindado, and Marti, 2007](#)), retail ([Ma and Fildes, 2021](#)), tourism ([Andrawis, Atiya, and El-Shishiny, 2011](#)), inflation forecasting ([Mitchell and Hall, 2005](#)), epidemiology, particularly in the context of COVID-19 ([Ray, Brooks, Bien, Biggerstaff, Bosse, Bracher, Cramer, Funk, Gerding, Johansson et al., 2023](#)), health ([Lipscomb, Parmigiani, and Hasselblad, 1998](#)), weather ([Gneiting](#)

and Raftery, 2005), environmental studies (Westerlund, Urbain, and Bonilla, 2014), and economics (Aastveit, Mitchell, Ravazzolo, and Van Dijk, 2018).

An important aspect of combining information from various sources is the potential presence of stochastic dependence among them. Initial efforts focused on Bayesian approaches for updating forecast combination weights in light of new information from various sources under conditions of statistical dependence. Assuming that the vector of forecast errors follows a normal distribution, Winkler (1981) and Clemen and Winkler (1986) developed a Bayesian approach using the conjugate prior for the variance-covariance matrix, represented by an inverted Wishart distribution. Subsequent work by Diebold and Pauly (1990) allowed for the incorporation of the standard normal-gamma conjugate prior, considering a normal regression-based combination method. More recently, efforts have been devoted to understanding situations where there is uncertainty about the correlation between sources and how aggregation procedures can be robust to this uncertainty (Arieli, Babichenko, and Smorodinsky (2018), Levy and Razin (2021), Levy and Razin (2022)).

While the forecast combination literature emphasizes optimal weight determination to minimize forecast errors and mitigate uncertainties, social learning studies focus on the mechanisms driving opinion (or forecast) formation within networks. Social networks act as conduits for information transmission, encompassing observations of other agents' decisions and engaging in conversations. Consequently, individuals assimilate information from their social environments, influencing their beliefs and, consequently, their forecasts. This field of research aims to explore questions related to the impact of initial opinions, the timing of agents' conformity, and the role of network structure in shaping final opinions.

An early branch of the social learning literature aimed at explaining herd behavior (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Smith and Sørensen, 2000). In these works, agents take single sequential actions after observing previous actions and signals. More recently, some works adapted the classical herding models to a network structure to allow partial observation of the history (Çelen and Kariv, 2004; Eyster and Rabin, 2010; Acemoglu, Dahleh, Lobel, and Ozdaglar, 2011; Lobel and Sadler, 2015). While sequential learning models provide valuable insights, they do not encompass the richness of network structures, where influence from peers flows in all directions.

In the context of repeated interactions in social learning, the focus is on understanding how network structure and initial opinions contribute to consensus, biases or misinformation, and disagreement or polarization. Early works in the repeated social learning literature aimed to explain consensus formation (DeGroot, 1974; Chatterjee and Seneta, 1977; Friedkin and Johnsen,

1999; DeMarzo, Vayanos, and Zwiebel, 2003; Golub and Jackson, 2010) or its absence (Hegselmann and Krause, 2002; Yildiz, Ozdaglar, Acemoglu, Saberi, and Scaglione, 2013; Acemoglu, Como, Fagnani, and Ozdaglar, 2013). Comprehensive assessments of this extensive literature are provided by Golub and Sadler (2017) and Grabisch and Rusinowska (2020). More recent research has adapted classical models of repeated interactions to incorporate Bayesian features, biases, conformism, and selective sharing in the opinion update process (Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi, 2012; Buechel, Hellmann, and Klößner, 2015; Li and Tan, 2020; Azzimonti and Fernandes, 2023; Fernandes, 2023; Buechel, Klößner, Meng, and Nassar, 2023; Bowen, Dmitriev, and Galperti, 2023).

This study links these two research strands by employing social learning within networks to examine the results of forecast pooling practices in situations where experts may interact before informing a decision-maker.

3. MODEL

3.1. Basic notation. In terms of notation, M_{i*} (or $[M]_{i*}$) represents the row i of any matrix M of size $n \times n$, whereas M_{*i} (or $[M]_{*i}$) represents the column i of the same matrix. Both the decision-maker and experts are denoted by indices i or j . In particular, $i = 0$ (or $j = 0$) represents the decision-maker, while experts are indexed by i (or j) $\in N = \{1, 2, \dots, n\}$.

3.2. Primitives.

Agents' prior assessments and point forecasts. Let θ be the random variable of interest, representing either a future observation or a parameter of a statistical model. Distributions for θ are evaluated by n experts, and for convenience, it is assumed that θ is real-valued and unbounded, and all n distribution functions are continuous. The prior density function for expert i is denoted by g_i , and \cdot . Thus, g_1, g_2, \dots, g_n are densities over \mathbb{R} representing the judgments of these n experts about θ . Consistent with the approaches in Winkler (1981), Clemen and Winkler (1985), and Clemen and Winkler (1986), I define the *mean* of expert i 's distribution as

$$x_i = \int_{-\infty}^{+\infty} \theta g_i(\theta) d\theta, \quad (1)$$

and interpret it as a *point forecast* for θ .¹ The forecast error for expert i is expressed as

$$\epsilon_i = x_i - \theta. \quad (2)$$

In this work, I assume that the forecast errors ϵ_i for each expert follow a normal distribution, specifically $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. Potential statistical dependence among these errors is represented

¹Here, I adopt a standard quadratic loss function. For further reference on forecast combinations under more general loss functions, refer to Elliott and Timmermann (2004), Timmermann (2006), Patton and Timmermann (2007).

by a variance-covariance matrix Σ . This matrix consists of variances σ_i^2 for all $i \in N$, as well as covariances $\rho_{ij}\sigma_i\sigma_j$ for all $i, j \in N$, where $i \neq j$. In this context, ρ_{ij} represents the correlation between any pair of forecast errors ϵ_i and ϵ_j . Finally, expert i 's precision, denoted as τ_i , is defined as the inverse of the variance, i.e., $\tau_i = 1/\sigma_i^2$.

Experts' network: a random graph approach. A group of experts $N = \{1, \dots, n\}$ is allowed to share forecasts before communicating it to the decision-maker. In order to generalize the information exchange process and results, I consider the classical *Poisson random graph model* in which each distinct pair of experts (nodes) meet with independent probability p . This Poisson random graph model on n experts with meeting probability p is denoted by $G(n, p)$. The probability of being connected to d nodes is

$$P(d) = \binom{n-1}{d} p^d (1-p)^{n-1-d}. \quad (3)$$

As the network structure is random, the expected degree of any node in $G(n, p)$ is $\langle d \rangle = \mathbb{E}[d] = \sum_d P(d)d = (n-1)p$ and, if we let the number of nodes grow large², the distribution of the degree of a neighboring node is

$$\tilde{P}(d) = \frac{P(d)d}{\langle d \rangle}. \quad (4)$$

If we fix the number of experts and some probability p of link formation, then every conceivable network has some positive probability of emerging. Any *realized* network, denoted by $G(N, A)$, consists of the group of experts, N , and a binary $n \times n$ adjacency matrix A , indicating whether agents i and j are participating in forecast exchange ($A_{ij} = 1$) or not ($A_{ij} = 0$). The network is considered undirected, ensuring that $A = A'$, i.e. all links are reciprocal. The neighborhood of expert i , defined as N_i , comprises the set of experts to which expert i is directly connected, i.e. $N_i = \{j \in N : A_{ij} = 1\}$. Furthermore, the degree of expert i , denoted as d_i , indicates the number of neighboring experts, i.e., $d_i = \sum_j A_{ij}$. Thus, the degree matrix, defined as $D = \text{diag}(d_1, d_2, \dots, d_n)$, is an $n \times n$ matrix with the degree of each expert in its main diagonal.

3.3. Classical Methods for Combining Forecasts. In the early development of mathematical aggregation of forecasts, probabilities, probability distributions, and opinions, the predominant focus was on formulating axiom-based aggregation rules. These studies employed a systematic approach, postulating essential properties for aggregation rules, which ultimately led to the derivation of functional forms for each method. In this work, the primary emphasis is not on examining these aggregation rules in detail. Instead, I take them as given and focus on two

²This is the case since the degrees of two neighbors are approximately independently distributed for large networks provided the largest nodes are not too large.

important cases: *Linear and Bayesian*, which are recognized in the academic literature as polar opposites.

3.3.1. *Linear Pooling*. For any agent i and any initial point forecast vector $\mathbf{x} = (x_1, \dots, x_n)$, the linear pool of point forecasts is represented by a linear pooling operator L_i defined as

$$L_i(\mathbf{x}) = \sum_{j=1}^n w_{ij} x_j, \quad (5)$$

where $w_{ij} \in \mathbb{R}$ represent the weight that agent i assigns to the forecast truthfully informed by agent j . The simple average pooling has been found to yield accurate forecasts (Clemen (1989), Hsiao and Wan (2014), Wang et al. (2023)) and it is widely assumed in the social learning literature (DeGroot (1974), Golub and Jackson (2010), Golub and Sadler (2017), Grabisch and Rusinowska (2020)). In this particular case of linear pooling, $w_{0j} = 1/n$ for every j for the decision-maker, and $w_{ij} = 1/d_i$ for any $i, j \in N$, and Equation (5) becomes

$$S_i(\mathbf{x}; A) = \begin{cases} \bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_j = (\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}'\mathbf{x} & , \text{if } i = 0 \text{ (DM)}, \\ \bar{x}_i = \frac{1}{d_i(A)} \sum_{j \in N_i(A)} x_j = [D^{-1}A]_{i*} \mathbf{x} & , \text{if } i \in N \text{ (Experts)}. \end{cases} \quad (6)$$

3.3.2. *Bayesian Pooling*. The information from experts can also be used in a Bayesian fashion to revise all agents' distributions for θ . If an arbitrary agent i consults an arbitrary number k of experts, then $f_i(\epsilon_1, \dots, \epsilon_k)$ can be taken as the likelihood function. In this case, the posterior distribution of θ for any agent i is

$$g_i(\theta | \mathbf{x}_i) \propto \begin{cases} f_i(\epsilon_i | \theta) & , \text{if } i = 0, \\ g_i(\theta) f_i(\epsilon_i | \theta) & , \text{if } i \in N, \end{cases} \quad (7)$$

where $\epsilon_0 = \epsilon = \{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ and $\mathbf{x}_0 = \mathbf{x}$, whereas $\epsilon_i = \{\epsilon_j | j \in N_i \setminus \{i\}\}$ and $\mathbf{x}_i = \{x_j | j \in N_i \setminus \{i\}\}$ for any $i \in N$. Thus, equation (7) implies that the decision-maker holds an improper diffuse prior density, while experts hold an initial prior density g_i , as previously described. Additionally, it specifies that experts will exclusively manage the forecasts shared by their direct neighbors.

The assumption of normality is often reasonable and highly tractable. Consequently, expert i 's distribution $g_i(\theta)$ is a normal distribution with a mean θ and a variance σ_i^2 . This implies that experts are calibrated (or unbiased) and differ solely in terms of sharpness (or precision). When $\sigma_i^2 < \sigma_j^2$, expert i is more precise than expert j .

Following equation (7), assuming normality, the posterior density of agent i for θ is

$$g_i(\theta | \mathbf{x}, A, \Sigma) \sim \mathcal{N}(B_i(\mathbf{x}; A, \Sigma), V_i(\Sigma, A)),$$

with

$$B_i(\mathbf{x}; A, \Sigma) = \begin{cases} \tilde{x}_i = \frac{\mathbf{1}'\Sigma^{-1}\mathbf{x}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} & , \text{ if } i = 0, \text{ (DM)} \\ \tilde{x}_i = \frac{A_{i*}\Sigma^{-1}\mathbf{x}(i)}{A_{i*}\Sigma^{-1}A_{*i}} & , \text{ if } i \in N, \text{ (Experts)} \end{cases} \quad (8)$$

where $\mathbf{x}(i) = A_{*i} \odot \mathbf{x}$ is a $n \times 1$ vector resulting of an entry-wise product of the $n \times 1$ vectors A_{*i} and \mathbf{x} ,³ and

$$V_i(\Sigma, A) = \begin{cases} \tilde{\sigma}_i^2 = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1} & , \text{ if } i = 0, \\ \tilde{\sigma}_i^2 = (A_{i*}\Sigma^{-1}A_{*i})^{-1} & , \text{ if } i \in N. \end{cases} \quad (9)$$

3.4. Combined Forecasts and Combining Combined Forecasts. In the study of forecast combinations, understanding the impact of network structures and expert interactions is instrumental to determine the consequences on decision-making. This subsection analyzes the intricate dynamics of combined forecasts and the subsequent process of combining these combined forecasts. Initially, I define combined forecasts within a networked context, where each expert selects a specific pooling rule. This approach enables an examination of how individual forecasts are aggregated within a network of experts.

Definition 1 (Combined Forecasts). *For any given network $G(N, A)$ and variance-covariance matrix Σ , let every expert $i \in N$ choose a pooling rule R_i , with the property that, for every vector \mathbf{x} of forecasts,*

$$R_i(\mathbf{x}) \in \{S_i(\mathbf{x}; A), B_i(\mathbf{x}; A, \Sigma)\}.$$

*The sequence $\{R_i\}_{i=1}^n$ is one particular combination of pooling rules. The set of all possible sequences of pooling rules to be employed by experts is denoted by \mathcal{R} , where each element $r \in \mathcal{R}$ corresponds to one possible combination of pooling rules. Thus, a vector of **combined forecasts**, denoted by \mathbf{x}^r , is a vector of updated forecasts following a given combination of pooling rules r .*

Building on this foundation, the analysis is extended to incorporate the decision-maker's role in combining the forecasts provided by the experts. This introduces the concept of combining combined forecasts, an essential definition for understanding forecast aggregation in networked environments.

³Thus, by construction, $x_j(i) = x_j$ if $j \in N_i$ and $x_j(i) = 0$ if $j \notin N_i$.

Definition 2 (Combining Combined Forecasts). *For any given network $G(N, A)$ and variance-covariance matrix Σ , let the decision-maker ($i = 0$) choose a pooling rule R_0 , with the property that, for any vector of forecasts \mathbf{x} informed by experts,*

$$R_0(\mathbf{x}) \in \{S_0(\mathbf{x}; A), B_0(\mathbf{x}; A, \Sigma)\}.$$

*We say the decision-maker is **combining combined forecasts** if the vector of forecasts informed by experts is the vector of combined forecasts \mathbf{x}^r , for any given sequence of pooling rules r , i.e., $R_0(\mathbf{x}^r)$.*

Finally, a key definition is the notion of *Network Bias*. This metric quantifies the impact of the prior network communication on decision-making relative to the baseline case of no experts' communication.

Definition 3 (Network Bias). *For any given network $G(N, A)$, variance covariance matrix Σ , and sequence of pooling rules r , we define the Network Bias in decision-making arising from experts communicating forecasts \mathbf{x} by*

$$\mathcal{B}(\mathbf{x}; A, \Sigma, r) = R_0(\mathbf{x}^r) - R_0(\mathbf{x}),$$

where $R_0(\cdot)$ represents the pooling rule adopted by the decision-maker, and \mathbf{x}^r is the vector of combined forecasts obtained by updating \mathbf{x} using the sequence of pooling rules r .

Thus, Network Bias is the difference between the decision-maker's ultimate pooled forecast under experts networked communication and the decision-maker's pooled forecast when such communication is absent. Four particular pooling rules combinations in \mathcal{R} and their Network Biases are

- (1) All agents use S : $\mathcal{B}(\mathbf{x}) = S_0(\bar{\mathbf{x}}) - S_0(\mathbf{x})$.
- (2) DM uses B and Experts use S : $\mathcal{B}(\mathbf{x}) = B_0(\bar{\mathbf{x}}) - B_0(\mathbf{x})$.
- (3) DM uses S and Experts use B : $\mathcal{B}(\mathbf{x}) = S_0(\tilde{\mathbf{x}}) - S_0(\mathbf{x})$.
- (4) When all agents use B : $\mathcal{B}(\mathbf{x}) = B_0(\tilde{\mathbf{x}}) - B_0(\mathbf{x})$.

4. MAIN RESULTS

4.1. Common positive correlation and common variance. Under common correlation ($\rho_{ij} = \rho$, for all $ij \in N$) and common variance ($\sigma_i^2 = \sigma^2$, for every $i \in N$), the precision matrix Σ^{-1} is

$$[\Sigma^{-1}]_{ij} = \begin{cases} \frac{1 + (n-2)\rho}{(1-\rho)[1+(n-1)\rho]\sigma^2} & , \text{ if } i = j, \\ \frac{-\rho}{(1-\rho)[1+(n-1)\rho]\sigma^2} & , \text{ if } i \neq j, \end{cases} \quad (10)$$

according to Equation (13) in Appendix A. In this setting, I show that regardless of the update rule, all sequences in \mathcal{R} reduce to simple average rules because the experts' precision is equal. This is established in the following lemma.

Lemma 1 (Equivalence of rules). *Under the assumptions of common positive correlation and common variance, the Bayesian pooling rule becomes a Simple Average pooling rule.*

Proof. By Lemma 2,

$$\mathbf{1}'\Sigma^{-1}\mathbf{x} = \left(\frac{1 + (n-2)\rho - (n-1)\rho}{(1-\rho)[1+(n-1)\rho]\sigma^2} \right) \mathbf{1}'\mathbf{x},$$

and,

$$\mathbf{1}'\Sigma^{-1}\mathbf{1} = \left(\frac{1 + (n-2)\rho - (n-1)\rho}{(1-\rho)[1+(n-1)\rho]\sigma^2} \right) \mathbf{1}'\mathbf{1}.$$

Similarly,

$$\begin{aligned} A_{i*}\Sigma^{-1}\mathbf{x}(i) &= \frac{(1 + (n-2)\rho - (d_i - 1)\rho) \sum_{j \in N_i} x_j(i) - d_i \rho \sum_{j \notin N_i} x_j(i)}{(1-\rho)[1+(n-1)\rho]\sigma^2} \\ &= \frac{(1 + (n-2)\rho - (d_i - 1)\rho) \sum_{j \in N_i} x_j(i)}{(1-\rho)[1+(n-1)\rho]\sigma^2} \\ &= \frac{(1 + (n-2)\rho - (d_i - 1)\rho) \sum_{j \in N_i} x_j}{(1-\rho)[1+(n-1)\rho]\sigma^2} \end{aligned}$$

and,

$$A_{i*}\Sigma^{-1}A_{*i} = \frac{d_i(1 + (n-2)\rho - (d_i - 1)\rho)}{(1-\rho)[1+(n-1)\rho]\sigma^2}.$$

Thus, Equation (8) can be written as

$$B_i^{\mathcal{N}}(\mathbf{x}) = \begin{cases} \tilde{x}_i = \frac{\mathbf{1}'\Sigma^{-1}\mathbf{x}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{x} = \bar{x}_i & , \text{ if } i = 0, \\ \tilde{x}_i = \frac{A_{i*}\Sigma^{-1}\mathbf{x}(i)}{A_{i*}\Sigma^{-1}A_{*i}} = \frac{1}{d_i} \sum_{j \in N_i} x_j = \bar{x}_i & , \text{ if } i \in N. \end{cases} \quad (11)$$

Therefore, $B_i^{\mathcal{N}}(\mathbf{x}) = L_i^S(\mathbf{x})$ for any i and \mathbf{x} . ■

Although the experts are symmetric in terms of forecast precision, meaning their forecast errors have the same variance and pair-wise correlation, they may differ in terms of their network location. This difference is captured by the *Attention Centrality* of each expert for any given network $G(N, A)$, defined below.⁴

Definition 4 (Attention Centrality). *Under the assumptions of common positive correlation and common variance, the **Attention Centrality** of an expert $i \in N$ within any network $G(N, A)$ is denoted by $\alpha_i(A)$ and defined as*

$$\alpha_i(A) = \sum_{j \in N_i(A)} \frac{1}{d_j} - 1. \quad (12)$$

In other words, the *Attention Centrality* of any agent i quantifies the net collective emphasis placed on its individual forecast x_i . With these definitions in hand, the following proposition demonstrates that, for any given network $G(N, A)$, the expected— in the sense that forecasts may be placed randomly over nodes— Network Bias is zero. However, the variance of the Network Bias is dependent on the network structure, and more specifically, it is a function of the Attention Centrality of the experts.

Example 1 (Attention Centrality). Consider the network explored in the introductory example, now represented in Figure 2. Suppose we have three experts, labeled A, B, and C, with degree centralities of 2, 3, and 2, respectively (self-loops omitted in the drawing). Experts A and C each retain 1/2 of their attention for themselves and devote the remaining 1/2 to expert B. In turn, expert B keeps 1/3 of the attention for herself and allocates 1/3 of her attention to each peripheral expert. Consequently, compared to the scenario in which they retain 100% of their attention for themselves (i.e., no network, represented by the -1 in the calculation), B experiences a net gain of 2/6 in the overall attention placed on her forecast, while A and C experience a net loss of attention of 1/6.

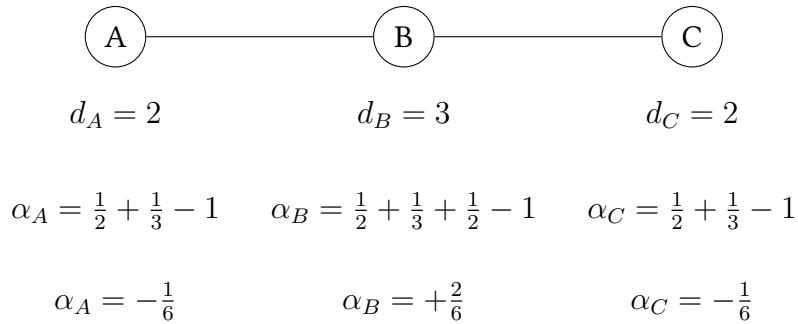


FIGURE 2. Attention Centrality in a star network of size 3 (self-loops omitted).

⁴Further properties of this new centrality measure can be found in Appendix B.

Having established the necessary definitions, I now show in the next proposition how the expectation and variance of the Network Bias are influenced by a given network structure $G(N, A)$.

Proposition 1. *Under the assumptions of common positive correlation and common variance,*

$$\mathbb{E}_\epsilon[\mathcal{B}(\mathbf{x})|A] = 0,$$

and

$$\text{Var}[\mathcal{B}(\mathbf{x})|A] = \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \alpha_i^2(A) + \frac{2\rho}{n} \sum_{i=1}^n \sum_{j:j>i} \alpha_i(A) \alpha_j(A) \right),$$

for any given network $G(N, A)$.

Proof. First,

$$\begin{aligned} \mathbb{E}_\epsilon[\mathcal{B}(\mathbf{x})] &= \mathbb{E}_\epsilon \left[\frac{1}{n} \sum_{i=1}^n \alpha_i(A) \epsilon_i \right] \\ &= \frac{1}{n} \sum_{i=1}^n \alpha_i(A) \mathbb{E}_\epsilon [\epsilon_i] \\ &= 0. \end{aligned}$$

Second,

$$\begin{aligned} \text{Var}[\mathcal{B}(\mathbf{x})] &= \text{Var} \left[\frac{1}{n} \sum_{i=1}^n \alpha_i(A) \epsilon_i \right] \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n \alpha_i^2(A) \text{Var} [\epsilon_i] + 2 \sum_{i=1}^n \sum_{j:j>i} \alpha_i(A) \alpha_j(A) \text{Cov} (\epsilon_i, \epsilon_j) \right) \\ &= \frac{1}{n^2} \left(\sigma^2 \sum_{i=1}^n \alpha_i^2(A) + 2\rho\sigma^2 \sum_{i=1}^n \sum_{j:j>i} \alpha_i(A) \alpha_j(A) \right) \\ &= \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \alpha_i^2(A) + \frac{2\rho}{n} \sum_{i=1}^n \sum_{j:j>i} \alpha_i(A) \alpha_j(A) \right) \end{aligned}$$

■

4.1.1. *Fixed Networks: Classical Network Structures.* To understand the effect of network structure on the variance of Network Bias, I will focus on classical network structures. By examining these classical networks, I aim to elucidate the fundamental impact of network structure on decision-making, which is captured by the Network Bias. This analysis sets the stage for a subsequent examination based on a random graph approach, which will generalize some of the findings.

Corollary 1 (from Proposition 1). *Under the assumptions of common positive correlation and common variance, if the network $G(N, A)$ is a star with size $n \geq 3$, then*

$$\text{Var}[\mathcal{B}(\mathbf{x})|A_{star}] = \frac{\sigma^2}{4n^3}(n-2)^2(n-1)(1-\rho).$$

Proof. For a star with $(n-1)$ peripheral nodes, the attentions scores are

$$\alpha_p = \left(\frac{1}{2} + \frac{1}{n}\right) - 1 = \frac{2-n}{2n}$$

for any peripheral node p , and

$$\alpha_c = \left(\frac{1}{n} + (n-1)\frac{1}{2}\right) - 1 = \frac{(n-2)(n-1)}{2n}$$

for the center node c . Thus,

$$\begin{aligned} \sum_{i=1}^n \sum_{j:j>i} \alpha_i \alpha_j &= (n-1) \frac{(n-2)(n-1)}{2n} \frac{(2-n)}{2n} + \frac{(n-1)(n-2)}{2} \left(\frac{2-n}{2n}\right)^2 \\ &= \frac{(n-1)(n-2)}{4n^2} \frac{(2-n)n}{2}, \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^n \alpha_i^2 &= \frac{(n-2)^2(n-1)^2}{4n^2} + (n-1) \left(\frac{2-n}{2n}\right)^2 \\ &= \frac{(n-2)^2(n-1)n}{4n^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}[\mathcal{B}(m(\mathbf{x}))] &= \frac{\sigma^2}{n} \frac{1}{n} \left(\frac{(n-2)^2(n-1)n}{4n^2} + 2\rho \frac{(n-1)(n-2)}{4n^2} \frac{(2-n)n}{2} \right) \\ &= \frac{\sigma^2}{4n^4} (n-2)(n-1)n \left((n-2) + 2\rho \frac{(2-n)}{2} \right) \\ &= \frac{\sigma^2}{4n^3} (n-2)^2(n-1)(1-\rho). \end{aligned}$$

■

Remark 1. *If network $G(N, A)$ is a star with size $n \geq 3$, then*

$$\frac{d}{dn} \text{Var}[\mathcal{B}(\mathbf{x})|A_{star}] > 0,$$

and

$$\lim_{n \rightarrow \infty} \text{Var}[\mathcal{B}(\mathbf{x})|A_{star}] = \frac{\sigma^2}{4}(1-\rho).$$

Corollary 2 (from Proposition 1). *Under the assumptions of common positive correlation and common variance, if network $G(N, A)$ is a line with size $n \geq 4$, then*

$$\text{Var}[\mathcal{B}(\mathbf{x})|A_{line}] = \frac{\sigma^2}{9n^2}(1 - \rho).$$

Proof. For any line of size $n \geq 4$, we have three "types" of nodes: 2 peripheral nodes (subscript p), 2 peripheral nodes' neighbors (subscript n), and $(n - 4)$ middle-nodes (subscript m).

Thus, the attention scores of each of these three types are

$$\alpha_p = \left(\frac{1}{2} + \frac{1}{3}\right) - 1 = -\frac{1}{6}$$

for any peripheral node,

$$\alpha_n = \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{3}\right) - 1 = \frac{1}{6}$$

for any peripheral node's neighbor, and

$$\alpha_m = \left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}\right) - 1 = 0$$

for any middle node.

Thus, $\alpha_p^2 = \alpha_n^2 = \frac{1}{6^2}$, $\alpha_m^2 = 0$, and $\sum_{i=1}^n \alpha_i^2 = 2\alpha_p^2 + 2\alpha_n^2 + (n - 4)\alpha_m^2 = \frac{4}{6^2}$, and

$$\begin{aligned} \sum_{i=1}^n \sum_{j:j>i} \alpha_i \alpha_j &= 4\alpha_p \alpha_n + 2(n - 4)\alpha_p \alpha_m + 2(n - 4)\alpha_n \alpha_m + \alpha_n^2 + \alpha_p^2 + \frac{(n - 4)(n - 5)}{2} \alpha_m^2 \\ &= -\frac{4}{6^2} - 2(n - 4)\frac{1}{6}0 + 2(n - 4)\frac{1}{6}0 + \frac{1}{6^2} + \left(-\frac{1}{6}\right)^2 + \frac{(n - 4)(n - 5)}{2}0 \\ &= -\frac{4}{6^2} + 2\frac{1}{6^2} \\ &= -\frac{2}{6^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}[\mathcal{B}(\mathbf{x})] &= \frac{\sigma^2}{n} \frac{1}{n} \left(\frac{4}{6^2} - 2\rho \frac{2}{6^2}\right) \\ &= \frac{\sigma^2}{9n^2}(1 - \rho). \end{aligned}$$

■

Remark 2. *If network $G(N, A)$ is a line with size $n \geq 4$, then*

$$\frac{d}{dn} \text{Var}[\mathcal{B}(\mathbf{x})|A_{line}] < 0,$$

and

$$\lim_{n \rightarrow \infty} \text{Var}[\mathcal{B}(\mathbf{x})|A_{line}] = 0.$$

Proposition 2. *Under the assumptions of common correlation and common variance, the Attention Centrality is zero for all experts $i \in N$ (i.e. $\alpha_i = 0$ for all i) if, and only if, $G(N, A)$ is a d -regular network of any size $n \geq 2$. Therefore, for any vector of forecasts \mathbf{x} , d -regular networks are the most efficient structure.*

Proof. If: If the network $G(N, A)$ is d -regular, then $d_i = |N_i(A)| = d$ for every node i . Thus, following Definition 4,

$$\begin{aligned}\alpha_i(A) &= \sum_{j \in N_i(A)} \frac{1}{d_j} - 1 \\ &= d \frac{1}{d} - 1 \\ &= 0\end{aligned}$$

for every i , and $\mathbb{E}[\mathcal{B}(\mathbf{x})] = \text{Var}[\mathcal{B}(\mathbf{x})] = 0$ for any vector of forecasts \mathbf{x} . *Only if:* proof under review. ■

4.1.2. *Random Networks: Poisson random graph model.* In this subsection, I extend the analysis to random networks using the Poisson random graph model, assuming a large network size n and a small connection probability p . This approach provides more general results by allowing both forecasts and network structure to be random. It illustrates that, under these assumptions, the Network Bias is zero and its variance converges to zero as n becomes sufficiently large. The intuition behind this result is that the network will either be too sparse (for a low expected degree), approximating the case of no network communication, or too dense (for a high expected degree), approximating the case of a complete network, which is a specific case of a d -regular network.

Proposition 3. *Under the **Poisson random graph** model $G(n, p(n))$, such that n is large and $p(n)$ - a decreasing function of n - is small, and the assumptions of common positive correlation and common variance,*

$$\mathbb{E}_d[\mathbb{E}_\epsilon[\mathcal{B}(\mathbf{x})|A]] = 0,$$

and

$$\mathbb{E}_d[\text{Var}[\mathcal{B}(\mathbf{x})|A]] \rightarrow 0,$$

as $n \rightarrow \infty$, both when $\langle d \rangle \rightarrow 1$ and $\langle d \rangle \rightarrow \infty$.

Proof.

$$\begin{aligned}\mathbb{E}_d[\mathbb{E}_\epsilon[\mathcal{B}(\mathbf{x})|A]] &= \mathbb{E}_d \left[\mathbb{E}_\epsilon \left[\frac{1}{n} \sum_{i=1}^n \alpha_i(A) \epsilon_i | A \right] \right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_d [\alpha_i(A)] \mathbb{E}_\epsilon [\epsilon_i] \\ &= 0\end{aligned}$$

For the variance term, we have

$$\begin{aligned}\mathbb{E} [\text{Var} [\mathcal{B}(\mathbf{x})|A]] &= \mathbb{E} \left[\frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \alpha_i^2(A) + \frac{2\rho}{n} \sum_{i=1}^n \sum_{j:j>i} \alpha_i(A) \alpha_j(A) \right) \right] \\ &= \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} [\alpha_i^2(A)] + \frac{2\rho}{n} \sum_{i=1}^n \sum_{j:j>i} \mathbb{E} [\alpha_i(A)] \mathbb{E} [\alpha_j(A)] \right) \quad (\text{by degree indep.}) \\ &= \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} [\alpha_i^2(A)] \right) \quad (\text{by Lemma 8}) \\ &= \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\sum_{j \in N_i} \frac{1}{d_j^2} + 2 \sum_{j \in N_i} \sum_{k \in N_i: k>j} \frac{1}{d_j} \frac{1}{d_k} - 2 \sum_{j \in N_i} \frac{1}{d_j} + 1 \right] \right) \\ &\approx \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^n \left(\langle d \rangle \frac{K(\langle d \rangle)}{\langle d \rangle} + 2 \frac{\langle d \rangle^2}{2} \frac{K(\langle d \rangle)}{\langle d \rangle} - 2 \langle d \rangle \frac{1}{\langle d \rangle} + 1 \right) \right) \quad (**) \\ &= \frac{\sigma^2}{n} (K(\langle d \rangle) [1 + \langle d \rangle] - 1),\end{aligned}$$

where

$$K(\langle d \rangle) \approx \frac{e^{-\langle d \rangle}}{1 - e^{-\langle d \rangle}} [\text{Ei}(\langle d \rangle) - \ln(\langle d \rangle) - \gamma],$$

according to Lemma 5, with $\gamma \approx 0.57721$ being the Euler–Mascheroni’s constant, and

$$\text{Ei}(\langle d \rangle) = \int_{-\infty}^{\langle d \rangle} \frac{e^t}{t} dt$$

is the Exponential Integral function. Therefore, according to Lemma 6

$$\lim_{\substack{n \rightarrow \infty \\ \langle d \rangle \rightarrow 1}} \mathbb{E} [\text{Var} [\mathcal{B}(\mathbf{x})|A]] = 0,$$

and

$$\lim_{\substack{n \rightarrow \infty \\ \langle d \rangle \rightarrow \infty}} \mathbb{E} [\text{Var} [\mathcal{B}(\mathbf{x})|A]] = 0.$$

(**) by Lemma 5, Wald’s Identity, and the assumption of n large, and small p . ■

Figure 3 shows a simulation of the Poisson random graph model for different network sizes with a fixed connection probability. The example demonstrates that, while the expectation remains zero, the variance of the Network Bias converges to zero as the network size increases.

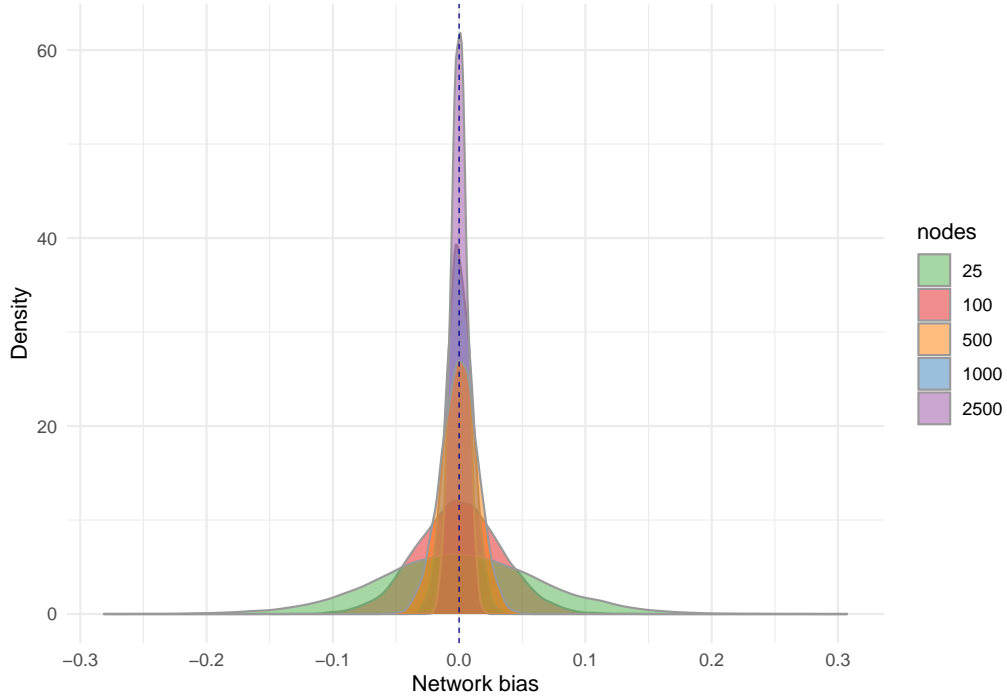


FIGURE 3. Poisson random graph: $\theta = 3, \sigma^2 = 1.2, \rho = 0, \langle d \rangle = 5, p = \frac{\langle d \rangle}{n - 1}$

5. CONCLUSION

In conclusion, this study contributes to the forecast combination literature by examining the implications of expert interaction within networks on decision-making processes. While the practice of combining forecasts has long been acknowledged for its potential to enhance accuracy, this work elucidates the interplay between network structure and pooling strategies, adding a nuanced perspective to this established field. In the context of non-strategic pooling of forecasts, this study highlights the potential impact on decision-making when employing two classical pooling methods: simple average pooling and Bayesian pooling.

Preliminary results reveal that, although the expected Network Bias remains unaffected by pooling rules and their mixtures, decision-making variance becomes a central concern. The structure of the network emerges as a crucial factor, with star networks exhibiting higher variance and d -regular networks achieving zero variance, demonstrating their efficiency. Furthermore, using the Poisson random graph model, it is shown that under the assumptions of a large network

size and a small connection probability, both the expected Network Bias and its variance converge to zero as the network size becomes sufficiently large. This occurs because the network becomes either too sparse, approximating no network communication, or too dense, approximating a complete network.

In the broader context of the forecast combination literature, this work also aligns with the social learning mechanisms within networks. By linking these two research strands, this study offers a novel perspective on decision-making processes influenced by both individual forecast precision and network structures.

As with any study, limitations exist, and future research opportunities abound. First, investigating the impact of experts' strategic interactions under different network structures and conditions can provide deeper insights into decision-making dynamics. Second, exploring the application of this work's findings in specific domains such as health, law, and economics can offer context-specific recommendations for practitioners. Finally, considering additional factors such as decision-maker preferences and the timing of expert interactions can further enhance applicability in real-world scenarios.

In summary, this study advances the understanding of forecast combination by incorporating social learning within networks prior to informing the decision-maker. These insights contribute to both academic scholarship and practical decision-making processes, paving the way for continued exploration and refinement in this evolving field.

REFERENCES

- AASTVEIT, K. A., J. MITCHELL, F. RAVAZZOLO, AND H. K. VAN DIJK (2018): “The evolution of forecast density combinations in economics,” Tech. rep., Tinbergen Institute Discussion Paper.
- ABRAMOWITZ, M. AND I. A. STEGUN (1968): *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55, US Government printing office.
- ACEMOGLU, D., G. COMO, F. FAGNANI, AND A. OZDAGLAR (2013): “Opinion fluctuations and disagreement in social networks,” *Mathematics of Operations Research*, 38, 1–27.
- ACEMOGLU, D., M. A. DAHLEH, I. LOBEL, AND A. OZDAGLAR (2011): “Bayesian learning in social networks,” *The Review of Economic Studies*, 78, 1201–1236.
- ANDRAWIS, R. R., A. F. ATIYA, AND H. EL-SHISHINY (2011): “Combination of long term and short term forecasts, with application to tourism demand forecasting,” *International Journal of Forecasting*, 27, 870–886.
- ARIELI, I., Y. BABICHENKO, AND R. SMORODINSKY (2018): “Robust forecast aggregation,” *Proceedings of the National Academy of Sciences*, 115, E12135–E12143.
- ARMSTRONG, J. S. (2001): *Combining forecasts*, Springer.
- AZZIMONTI, M. AND M. FERNANDES (2023): “Social media networks, fake news, and polarization,” *European journal of political economy*, 76, 102256.
- BANERJEE, A. V. (1992): “A simple model of herd behavior,” *The quarterly journal of economics*, 107, 797–817.
- BATES, J. M. AND C. W. GRANGER (1969): “The combination of forecasts,” *Journal of the operational research society*, 20, 451–468.
- BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): “A theory of fads, fashion, custom, and cultural change as informational cascades,” *Journal of political Economy*, 100, 992–1026.
- BOWEN, T. R., D. DMITRIEV, AND S. GALPERTI (2023): “Learning from shared news: When abundant information leads to belief polarization,” *The Quarterly Journal of Economics*, 138, 955–1000.
- BUECHEL, B., T. HELLMANN, AND S. KLÖSSNER (2015): “Opinion dynamics and wisdom under conformity,” *Journal of Economic Dynamics and Control*, 52, 240–257.
- BUECHEL, B., S. KLÖSSNER, F. MENG, AND A. NASSAR (2023): “Misinformation due to asymmetric information sharing,” *Journal of Economic Dynamics and Control*, 150, 104641.
- ÇELEN, B. AND S. KARIV (2004): “Observational learning under imperfect information,” *Games and Economic behavior*, 47, 72–86.
- CHAO, M. T. AND W. E. STRAWDERMAN (1972): “Negative Moments of Positive Random Variables,” *Journal of the American Statistical Association*, 67, 429–431.

- CHATTERJEE, S. AND S. CHATTERJEE (1987): "On combining expert opinions," *American Journal of mathematical and management sciences*, 7, 271–295.
- CHATTERJEE, S. AND E. SENETA (1977): "Towards consensus: Some convergence theorems on repeated averaging," *Journal of Applied Probability*, 14, 89–97.
- CLEMEN, R. T. (1989): "Combining forecasts: A review and annotated bibliography," *International journal of forecasting*, 5, 559–583.
- CLEMEN, R. T. AND R. L. WINKLER (1985): "Limits for the precision and value of information from dependent sources," *Operations Research*, 33, 427–442.
- (1986): "Combining economic forecasts," *Journal of Business & Economic Statistics*, 4, 39–46.
- DEGROOT, M. H. (1974): "Reaching a consensus," *Journal of the American Statistical association*, 69, 118–121.
- DEMARZO, P. M., D. VAYANOS, AND J. ZWIEBEL (2003): "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly journal of economics*, 118, 909–968.
- DIEBOLD, F. X. AND P. PAULY (1990): "The use of prior information in forecast combination," *International Journal of Forecasting*, 6, 503–508.
- ELLIOTT, G. AND A. TIMMERMANN (2004): "Optimal forecast combinations under general loss functions and forecast error distributions," *Journal of Econometrics*, 122, 47–79.
- EYSTER, E. AND M. RABIN (2010): "Naive herding in rich-information settings," *American economic journal: microeconomics*, 2, 221–243.
- FERNANDES, M. R. (2023): "Confirmation bias in social networks," *Mathematical Social Sciences*, 123, 59–76.
- FRIEDKIN, N. E. AND E. C. JOHNSEN (1999): "Influence networks and opinion change," *Advances in Group Processes*, 16, 1–29.
- GENEST, C. AND J. V. ZIDEK (1986): "Combining probability distributions: A critique and an annotated bibliography," *Statistical Science*, 114–135.
- 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.
- GNEITING, T. AND M. KATZFUSS (2014): "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, 1, 125–151.
- GNEITING, T. AND A. E. RAFTERY (2005): "Weather forecasting with ensemble methods," *Science*, 310, 248–249.
- GOLUB, B. AND M. O. JACKSON (2010): "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2, 112–149.
- GOLUB, B. AND E. SADLER (2017): "Learning in social networks," *Available at SSRN 2919146*.

- GRABISCH, M. AND A. RUSINOWSKA (2020): “A survey on nonstrategic models of opinion dynamics,” *Games*, 11, 65.
- GRANGER, C. W. (1989): “Invited review combining forecasts—twenty years later,” *Journal of forecasting*, 8, 167–173.
- GRUSHKA-COCKAYNE, Y., V. R. R. JOSE, AND K. C. LICHTENDAHL JR (2017): “Ensembles of overfit and overconfident forecasts,” *Management Science*, 63, 1110–1130.
- HEGSELMANN, R. AND U. KRAUSE (2002): “Opinion dynamics and bounded confidence: models, analysis and simulation,” *Journal of Artificial Societies and Social Simulation (JASSS)*, vol. 5.
- HSHIAO, C. AND S. K. WAN (2014): “Is there an optimal forecast combination?” *Journal of Econometrics*, 178, 294–309.
- JACOBS, R. A. (1995): “Methods for combining experts’ probability assessments,” *Neural computation*, 7, 867–888.
- JADBABAIE, A., P. MOLAVI, A. SANDRONI, AND A. TAHBAZ-SALEHI (2012): “Non-Bayesian social learning,” *Games and Economic Behavior*, 76, 210–225.
- JOSE, V. R. R., Y. GRUSHKA-COCKAYNE, AND K. C. LICHTENDAHL JR (2014): “Trimmed opinion pools and the crowd’s calibration problem,” *Management Science*, 60, 463–475.
- LEVY, G. AND R. RAZIN (2021): “A maximum likelihood approach to combining forecasts,” *Theoretical Economics*, 16, 49–71.
- (2022): “Combining forecasts in the presence of ambiguity over correlation structures,” *Journal of Economic Theory*, 199, 105075.
- LI, W. AND X. TAN (2020): “Locally Bayesian learning in networks,” *Theoretical Economics*, 15, 239–278.
- LIPSCOMB, J., G. PARMIGIANI, AND V. HASSELBLAD (1998): “Combining expert judgment by hierarchical modeling: an application to physician staffing,” *Management Science*, 44, 149–161.
- LOBEL, I. AND E. SADLER (2015): “Information diffusion in networks through social learning,” *Theoretical Economics*, 10, 807–851.
- MA, S. AND R. FILDES (2021): “Retail sales forecasting with meta-learning,” *European Journal of Operational Research*, 288, 111–128.
- MAKRIDAKIS, S., A. ANDERSEN, R. CARBONE, R. FILDES, M. HIBON, R. LEWANDOWSKI, J. NEWTON, E. PARZEN, AND R. WINKLER (1982): “The accuracy of extrapolation (time series) methods: Results of a forecasting competition,” *Journal of forecasting*, 1, 111–153.
- MAKRIDAKIS, S., E. SPILLOTIS, AND V. ASSIMAKOPOULOS (2020): “The M4 Competition: 100,000 time series and 61 forecasting methods,” *International Journal of Forecasting*, 36, 54–74.

- MANCUSO, A. C. B. AND L. WERNER (2013): "Review of combining forecasts approaches," *Independent journal of management & production*, 4, 248–277.
- MITCHELL, J. AND S. G. HALL (2005): "Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation," *Oxford bulletin of economics and statistics*, 67, 995–1033.
- NEWBOLD, P. AND C. W. GRANGER (1974): "Experience with forecasting univariate time series and the combination of forecasts," *Journal of the Royal Statistical Society: Series A (General)*, 137, 131–146.
- NIELSEN, H. A., T. S. NIELSEN, H. MADSEN, M. J. S. I. PINDADO, AND I. MARTI (2007): "Optimal combination of wind power forecasts," *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 10, 471–482.
- PATTON, A. J. AND A. TIMMERMANN (2007): "Properties of optimal forecasts under asymmetric loss and nonlinearity," *Journal of Econometrics*, 140, 884–918.
- RAY, E. L., L. C. BROOKS, J. BIEN, M. BIGGERSTAFF, N. I. BOSSE, J. BRACHER, E. Y. CRAMER, S. FUNK, A. GERDING, M. A. JOHANSSON, ET AL. (2023): "Comparing trained and untrained probabilistic ensemble forecasts of COVID-19 cases and deaths in the United States," *International journal of forecasting*, 39, 1366–1383.
- SMITH, L. AND P. SØRENSEN (2000): "Pathological outcomes of observational learning," *Econometrica*, 68, 371–398.
- SWAMEE, P. K. AND C. S. P. OJHA (1990): "Pump test analysis of confined aquifer," *Journal of Irrigation and Drainage Engineering*, 116, 99–106.
- TIMMERMANN, A. (2006): "Forecast combinations," *Handbook of economic forecasting*, 1, 135–196.
- WALLIS, K. F. (2014): "Combining forecasts—forty years later," in *Perspectives on Econometrics and Applied Economics*, Routledge, 33–42.
- WANG, X., R. J. HYNDMAN, F. LI, AND Y. KANG (2023): "Forecast combinations: an over 50-year review," *International Journal of Forecasting*.
- WESTERLUND, J., J.-P. URBAIN, AND J. BONILLA (2014): "Application of air quality combination forecasting to Bogota," *Atmospheric Environment*, 89, 22–28.
- WINKLER, R. L. (1981): "Combining probability distributions from dependent information sources," *Management Science*, 27, 479–488.
- (1989): "Combining forecasts: A philosophical basis and some current issues," *International Journal of Forecasting*, 5, 605–609.
- XIE, J. AND T. HONG (2016): "GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation," *International Journal of Forecasting*,

32, 1012–1016.

YILDIZ, E., A. OZDAGLAR, D. ACEMOGLU, A. SABERI, AND A. SCAGLIONE (2013): “Binary opinion dynamics with stubborn agents,” *ACM Transactions on Economics and Computation (TEAC)*, 1, 1–30.

APPENDIX A. AUXILIARY LEMMAS

Lemma 2. *Under common correlation and different variances, the inverse of the variance covariance matrix Σ is*

$$[\Sigma^{-1}]_{ij} = \begin{cases} \frac{1 + (n-2)\rho}{(1-\rho)[1+(n-1)\rho]\sigma_i^2} & , \text{ if } i = j, \\ \frac{-\rho}{(1-\rho)[1+(n-1)\rho]\sigma_i\sigma_j} & , \text{ if } i \neq j. \end{cases} \quad (13)$$

Proof. The proof is relatively straightforward. For further details, the reader may refer to Equations 5, 6 and 7 in [Clemen and Winkler \(1985\)](#). ■

Lemma 3. *Under the assumptions of common positive correlation and common variance, the precision of the decision-maker's posterior distribution for θ , when updated in a Bayesian fashion, is bounded above by $\frac{1}{\sigma^2\rho}$ as the number of consulted experts n grows large.*

Proof. Following equations (9) and (10), the variance of the decision-maker's posterior distribution for θ is

$$\tilde{\sigma}_0^2 = \frac{\sigma^2}{n} (1 + (n-1)\rho),$$

thus, $\lim_{n \rightarrow \infty} \tilde{\tau}_0 = \lim_{n \rightarrow \infty} \frac{1}{\tilde{\sigma}_0^2} = \frac{1}{\sigma^2\rho}$. ■

The intuition is that positive statistical dependence reduces the information content of a set or a subset of experts' estimate. Moreover, as there are no repeated interactions, the precision of the decision-maker's posterior distribution for θ is not dependent of the experts' network.

Lemma 4. *If the experts' network is formed according to a Poisson random graph Model, denoted as $G(n, p)$, then $\mathbb{E} \left[\binom{d}{2} \right] \approx \frac{\langle d \rangle^2}{2}$ as $n \rightarrow \infty$, since the degree of an expert, denoted by d , is a binomial random variable distributed according to $P(d)$ defined in Equation (3).*

Proof.

$$\begin{aligned}
\mathbb{E} \left[\binom{d}{2} \right] &= \mathbb{E} \left[\frac{d!}{2!(d-2)!} \right] \\
&= \mathbb{E} \left[\frac{d(d-1)}{2} \right] \\
&= \frac{1}{2} \sum_d P(d) (d^2 - d) \\
&= \frac{1}{2} (n(n-1)p^2 + np - (n-1)p) \\
&= \frac{1}{2} (np(n-1)p + np - (n-1)p) \\
&\approx \frac{1}{2} (\langle d \rangle^2 + \langle d \rangle - \langle d \rangle) \quad (\text{as } n \rightarrow \infty) \\
&= \frac{\langle d \rangle^2}{2}.
\end{aligned}$$

■

Lemma 5. *If the experts' network is formed according to a Poisson random graph Model, denoted as $G(n, p)$ with large n and small p , then for any expert $i \in N$, the expected value of the reciprocal of the degree of any neighboring expert $j \in N_i$ is*

$$\mathbb{E}_d \left[\frac{1}{d_j} \right] = \frac{1}{\langle d \rangle},$$

and the expected value of the reciprocal of the degree squared of any neighboring expert $j \in N_i$ is

$$\mathbb{E}_d \left[\frac{1}{d_j^2} \right] = \frac{K(\langle d \rangle)}{\langle d \rangle},$$

where

$$K(\langle d \rangle) \approx \frac{e^{-\langle d \rangle}}{1 - e^{-\langle d \rangle}} [Ei(\langle d \rangle) - \ln(\langle d \rangle) + \gamma],$$

$$Ei(\langle d \rangle) = \int_{-\infty}^{\langle d \rangle} \frac{e^t}{t} dt$$

is the Exponential Integral function, and $\gamma \approx 0.57721$ is the Euler-Mascheroni's constant.

Proof. The distribution of the degree of a neighboring node is $\tilde{P}(d_j)$, as defined in Equation (4), therefore

$$\begin{aligned}\mathbb{E}_d \left[\frac{1}{d_j} \right] &= \sum_{d_j} \tilde{P}(d_j) \frac{1}{d_j} \\ &= \sum_{d_j} \frac{P(d_j) d_j}{\langle d \rangle} \frac{1}{d_j} \\ &= \frac{1}{\langle d \rangle}.\end{aligned}$$

Second, for large n and small p , the degree of any expert, $d_i > 0$, is a random variable with Poisson distribution with parameter $\lambda = (n - 1)p = \langle d \rangle$. Thus, as per Equations (3.15) and (3.16) in [Chao and Strawderman \(1972\)](#), we have

$$\mathbb{E}_d \left[\frac{1}{d_i} \mid d_i > 0 \right] = \frac{e^{-\langle d \rangle}}{1 - e^{-\langle d \rangle}} [\text{Ei}(\langle d \rangle) - \ln(\langle d \rangle) - \gamma]$$

$$\begin{aligned}\mathbb{E}_d \left[\frac{1}{d_j^2} \right] &= \sum_{d_j} \tilde{P}(d_j) \frac{1}{d_j^2} \\ &= \frac{1}{\langle d \rangle} \sum_{d_j} P(d_j) \frac{1}{d_j} \\ &= \frac{K(\langle d \rangle)}{\langle d \rangle}\end{aligned}$$

■

Lemma 6. *The expression $K(\langle d \rangle) [1 + \langle d \rangle]$ tends to 0.7668 as $\langle d \rangle \rightarrow 1$, and to 1 as $\langle d \rangle \rightarrow \infty$.*

Proof. Following equation 5.1.53 in [Abramowitz and Stegun \(1968\)](#)⁵, we can approximate the Exponential Integral function for large $\langle d \rangle$ as

$$\text{Ei}(\langle d \rangle) \approx \frac{e^{\langle d \rangle}}{\langle d \rangle}.$$

In this case,

$$K(\langle d \rangle) [1 + \langle d \rangle] \approx e^{-\langle d \rangle} \left[\frac{e^{\langle d \rangle}}{\langle d \rangle} - \ln(\langle d \rangle) \right] [1 + \langle d \rangle].$$

This expression simplifies to

$$\frac{1 + \langle d \rangle}{\langle d \rangle} - e^{-\langle d \rangle} \ln(\langle d \rangle) [1 + \langle d \rangle].$$

⁵Equation 5.1.53 shows that $E_1(x) \approx \frac{e^{-x}}{x}$ for large positive x . Since $\text{Ei}(x) = -E_1(-x)$ for large negative x , we can use the asymptotic form of $E_1(-x)$ for large positive x . For large positive x , the asymptotic behavior of $\text{Ei}(x)$ is $\frac{e^x}{x}$.

The first term simplifies to $1 + \frac{1}{\langle d \rangle}$, and converges to 1 as $\langle d \rangle \rightarrow \infty$. The second term decays exponentially, making it negligible for large $\langle d \rangle$, thus

$$\lim_{\langle d \rangle \rightarrow \infty} e^{-\langle d \rangle} \ln(\langle d \rangle) [1 + \langle d \rangle] \rightarrow 0.$$

Therefore, $\lim_{\langle d \rangle \rightarrow \infty} K(\langle d \rangle) [1 + \langle d \rangle] = 1$.

Following equation 5.1.10 in [Abramowitz and Stegun \(1968\)](#), we can also approximate the Exponential Integral function for $\langle d \rangle > 0$ as

$$\text{Ei}(\langle d \rangle) \approx \gamma + \ln(\langle d \rangle) + \sum_{k=1}^{\infty} \frac{\langle d \rangle^k}{k k!}.$$

In this case,

$$K(\langle d \rangle) [1 + \langle d \rangle] \approx \frac{e^{-\langle d \rangle}}{1 - e^{-\langle d \rangle}} \left[\sum_{k=1}^{\infty} \frac{\langle d \rangle^k}{k k!} \right] [1 + \langle d \rangle],$$

and

$$\begin{aligned} \lim_{\langle d \rangle \rightarrow 1} K(\langle d \rangle) [1 + \langle d \rangle] &= \frac{e^{-1}}{1 - e^{-1}} \left[\sum_{k=1}^{\infty} \frac{1^k}{k k!} \right] [1 + 1] \\ &\approx 1.163954 \left[\sum_{k=1}^7 \frac{1}{k k!} + o(?) \right] \\ &\approx 0.766853. \end{aligned}$$

■

APPENDIX B. ATTENTION CENTRALITY: PROPERTIES

B.1. Same variance and covariance.

Lemma 7. For any given network $G(N, A)$, $\sum_{i=1}^n \alpha_i(A) = 0$.

Proof.

$$\begin{aligned} \sum_{i=1}^n \alpha_i(A) &= \sum_{i=1}^n \left(\sum_{j \in N_i(A)} \frac{1}{d_j} - 1 \right) \\ &= \sum_{i=1}^n \sum_{j \in N_i(A)} \frac{1}{d_j} - \sum_{i=1}^n 1 \\ &= \sum_{i=1}^n d_i \frac{1}{d_i} - n \\ &= n - n = 0. \end{aligned}$$

■

Lemma 8. *If the network is formed following a Poisson random graph Model $G(n, p)$, then $\mathbb{E}_d [\alpha_i(A)] = 0$ for any n and p .*

Proof.

$$\begin{aligned}
\mathbb{E}_d [\alpha_i(A)] &= \mathbb{E}_d \left[\sum_{j \in N_i(A)} \frac{1}{d_j} - 1 \right] \\
&= \mathbb{E}_d \left[\sum_{j \in N_i(A)} \frac{1}{d_j} \right] - 1 \\
&= \mathbb{E}_d [d_i] \mathbb{E}_d \left[\frac{1}{d_j} \right] - 1 \text{ (by Wald's Identity)} \\
&= \langle d \rangle \left(\sum_{d_j} \tilde{P}(d_j) \frac{1}{d_j} \right) - 1 \\
&= \langle d \rangle \left(\sum_{d_j} \frac{P(d_j) d_j}{\langle d \rangle} \frac{1}{d_j} \right) - 1 \\
&= \langle d \rangle \frac{1}{\langle d \rangle} \sum_{d_j} P(d_j) - 1 = 0
\end{aligned}$$

■