

# Eliciting and Aggregating Forecasts When Information is Shared

by

Asa Benjamin Palley

Department of Business Administration  
Duke University

Date: \_\_\_\_\_

Approved:

---

Jack B. Soll, Co-Supervisor

---

Robert F. Nau, Co-Supervisor

---

Robert T. Clemen

---

Robert L. Winkler

Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in the Department of Business Administration  
in the Graduate School of Duke University  
2016

ABSTRACT

Eliciting and Aggregating Forecasts When Information is  
Shared

by

Asa Benjamin Palley

Department of Business Administration  
Duke University

Date: \_\_\_\_\_

Approved:

---

Jack B. Soll, Co-Supervisor

---

Robert F. Nau, Co-Supervisor

---

Robert T. Clemen

---

Robert L. Winkler

An abstract of a dissertation submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy in the Department of Business Administration  
in the Graduate School of Duke University

2016

Copyright © 2016 by Asa Benjamin Palley  
All rights reserved except the rights granted by the  
Creative Commons Attribution-Noncommercial Licence

# Abstract

Using the wisdom of crowds—combining many individual forecasts to obtain an aggregate estimate—can be an effective technique for improving forecast accuracy. When individual forecasts are drawn from independent and identical information sources, a simple average provides the optimal crowd forecast. However, correlated forecast errors greatly limit the ability of the wisdom of crowds to recover the truth. In practice, this dependence often emerges because information is shared: forecasters may to a large extent draw on the same data when formulating their responses.

To address this problem, I propose an elicitation procedure in which each respondent is asked to provide both their own best forecast and a guess of the average forecast that will be given by all other respondents. I study optimal responses in a stylized information setting and develop an aggregation method, called pivoting, which separates individual forecasts into shared and private information and then recombines these results in the optimal manner. I develop a tailored pivoting procedure for each of three information models, and introduce a simple and robust variant that outperforms the simple average across a variety of settings.

In three experiments, I investigate the method and the accuracy of the crowd forecasts. In the first study, I vary the shared and private information in a controlled environment, while the latter two studies examine forecasts in real-world contexts. Overall, the data suggest that a simple minimal pivoting procedure provides an effective aggregation technique that can significantly outperform the crowd average.

To my wife Charlotte, whose love and support made this possible, and to my family.

# Contents

Abstract	iv
List of Tables	ix
List of Figures	x
List of Abbreviations and Symbols	xii
Acknowledgements	xvi
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Review</b>	<b>5</b>
<b>3 Analysis of the Shared-Information Problem</b>	<b>10</b>
3.1 A Symmetric Model of Shared Information and Forecaster Behavior . . . . .	12
3.1.1 Eliciting Information from Forecasters . . . . .	14
3.1.2 Estimating $\theta$ When $w$ is Known . . . . .	15
3.1.3 A Behavioral Model of Responses . . . . .	17
3.2 What Happens Under Different Information Structures? . . . . .	18
3.2.1 Nested Information Model . . . . .	18
3.2.2 Nested-Symmetric Information Model . . . . .	19
3.2.3 General Estimation Procedure When $w$ and $p$ are Known . . . . .	20
3.2.4 Minimal Pivoting . . . . .	22
3.3 Practical Implementation of the Pivoting Procedures . . . . .	23

3.3.1	Estimating $\theta$ in the Symmetric Information Model When $w$ is Unknown . . . . .	23
3.3.2	Estimating $\theta$ in the Nested Information Model When $p$ is Unknown . . . . .	25
3.3.3	Estimating $\theta$ in the Nested-Symmetric Information Model When $w$ and $p$ are Unknown . . . . .	27
3.3.4	Minimal Pivoting . . . . .	29
3.3.5	Additional Information Models . . . . .	29
<b>4</b>	<b>Experimental Evidence</b>	<b>31</b>
4.1	Study 1: Predicting Coin Flips . . . . .	31
4.1.1	Methods . . . . .	33
4.1.2	Results . . . . .	36
4.2	Study 2: Estimating Grocery Prices . . . . .	44
4.3	Study 3: Forecasting NCAA Men’s Basketball Tournament Games . . . . .	45
<b>5</b>	<b>Discussion</b>	<b>54</b>
5.1	Pivoting as a Solution to Shared-Information Bias . . . . .	56
5.2	Limitations of the Method . . . . .	57
5.3	Additional Directions . . . . .	58
5.3.1	More General Information Models . . . . .	58
5.3.2	Estimating a Full Distribution for $X$ . . . . .	59
5.3.3	Further Empirical Evidence . . . . .	60
5.4	Conclusion . . . . .	61
<b>A</b>	<b>Proofs</b>	<b>62</b>
<b>B</b>	<b>Alternative Information Models</b>	<b>69</b>
B.1	A Beta-Binomial Model . . . . .	70
B.1.1	Symmetric Setting . . . . .	70

B.1.2	Nested Setting . . . . .	71
B.1.3	Nested-Symmetric Setting . . . . .	72
B.1.4	Minimal Pivoting . . . . .	73
B.2	A Gamma-Gamma Model . . . . .	74
B.2.1	Symmetric Setting . . . . .	76
B.2.2	Nested Setting . . . . .	77
B.2.3	Nested-Symmetric Setting . . . . .	77
B.2.4	Minimal Pivoting . . . . .	78
B.3	General Allocation of Information Across Forecasters . . . . .	80
B.3.1	Individual-Specific Information Weights $w_i$ . . . . .	80
	<b>Bibliography</b>	<b>83</b>
	<b>Biography</b>	<b>87</b>



# List of Tables

4.1	Experimental design and number of subjects in each condition. . . . .	33
4.2	Performance of different forecasting procedures in the different information settings using MAD and RMSE. . . . .	40
4.3	Brier score decomposition for the three forecasting methods using the discrete forecast categories 0%, 10%, . . . , 100%. . . . .	50

# List of Figures

3.1	Comparison of information structures across the models of shared information. . . . .	11
3.2	The signal generating process and belief updating for each forecaster $i$ in the Symmetric information model. . . . .	13
3.3	Estimating the private and shared signals in the Symmetric information model when $w$ is known. . . . .	16
3.4	Pivoting procedure for aggregating the crowd forecast in different information models for large $n$ when the parameters are known and $pw > 0$ . . . . .	22
3.5	Optimal hedge $h_{private}^*(\hat{w}, s_{\hat{w}}, n)$ as a function of $\hat{w}$ for different standard errors $s_{\hat{w}}$ when $n = 100$ . . . . .	26
4.1	Example of the experimental interface for a sophisticate in the Nested-Symmetric setting. . . . .	34
4.2	Winsorized implied weights placed on own private signal $t_i$ relative to the shared signal $s$ versus optimal weights across the 24 coins in the Symmetric information setting. . . . .	37
4.3	Performance of the respective pivoting method relative to the simple average forecast $\bar{f}$ in each of three information settings. . . . .	39
4.4	Reduction $ \bar{f} - \theta  -  \hat{\theta}_M - \theta $ in the absolute forecast error of the minimal pivoted forecast relative to the simple average forecast versus shared-information bias for all 72 coins across the three information settings. . . . .	42
4.5	Reduction $ \bar{f} - \theta  -  \hat{\theta}_M - \theta $ in the absolute forecast error of the minimal pivoted forecast relative to the simple average forecast versus bracketing rate for all 72 coins across the three information settings. . . . .	43
4.6	Example of the experimental interface for Study 3. . . . .	47

4.7 Calibration curves (hit rate versus forecast category) for the three forecasting methods. . . . .	51
---	----

# List of Abbreviations and Symbols

## Symbols

$X$	The random variable being forecasted.
$\theta$	The mean of $X$ .
$s$	The shared signal, observed by all forecasters.
$t_i$	A private signal observed only by forecaster $i$ .
$t$	An additional signal observed all sophisticated forecasters in the Nested information setting.
$\lambda$	The precision of $X$ in the normal-normal model.
$m$	The equivalent number of observations of $X$ contained in the shared signal $s$ .
$\ell$	The equivalent number of observations of $X$ contained in each of the additional signals $t$ and $t_i$ .
$w$	The relative weight on additional signals versus the shared signal, $w = \ell/(m + \ell)$ .
$n$	The number of forecasters.
$p$	The proportion of forecasters who are sophisticated.
$f_i$	The forecast provided by individual $i$ .
$g_i$	Individual $i$ 's estimate of the average forecast of the other $n - 1$ forecasters.
$\delta_i$	Error in the forecast provided by naïve forecaster $i$ .
$\varepsilon_i$	Error in the forecast provided by sophisticated forecaster $i$ .
$\gamma_i$	Error in sophisticate $i$ 's estimate of the average forecast of the other $n - 1$ forecasters.

$\bar{f}$	The average forecast, $\bar{f} = \sum_{i=1}^n \frac{f_i}{n}$ .
$\bar{g}$	The average estimate of others, $\bar{g} = \sum_{i=1}^n \frac{g_i}{n}$ .
$\bar{t}$	The average private signal, $\bar{t} = \sum_{i=1}^{pn} \frac{t_i}{n}$ , where $p = 1$ for the Symmetric setting.
$\bar{t}_{-i}$	The average private signal observed by all forecasters other than $i$ .
$\hat{t}$	The estimated average private signal.
$\hat{s}$	The estimated shared signal.
$\alpha$	The optimal Bayesian weight on the average private signal relative to the shared signal.
$\hat{\theta}_S$	The Symmetric pivoted estimate when $w$ is known.
$\tilde{f}_i$	The forecast provided by individual $i$ in the behavioral model.
$\tilde{g}_i$	Individual $i$ 's estimate of the average forecast of the other $n - 1$ forecasters in the behavioral model.
$\hat{\theta}_N$	The Nested pivoted estimate when $p$ is known.
$\hat{\theta}_{NS}$	The Nested-Symmetric pivoted estimate when $w$ and $p$ are known.
$\hat{\theta}_M$	The Minimal pivoted estimate, $\hat{\theta}_M = 2\bar{f} - \bar{g}$ .
$w^*$	The decision analyst's posterior expectation of $w$ .
$h_{private}^*$	The size of the pivot used by the decision analyst in estimating the average private signal when $w$ and $p$ are unknown.
$h_{shared}^*$	The size of the pivot used by the decision analyst in estimating the shared signal when $w$ and $p$ are unknown.
$\hat{w}$	The estimated slope coefficient from a simple least-squares regression of $g$ on $f$ .
$s_{\hat{w}}$	The standard error of the estimated slope coefficient from a simple least-squares regression of $g$ on $f$ .
$\hat{\theta}_S^*$	The Symmetric pivoted estimate when $w$ is unknown.

$d$	The number forecasters who provided a different guess about others than their own forecast, $g_i \neq f_i$ .
$\hat{p}$	The estimated proportion of forecasters who are sophisticated, $\hat{p} = d/n$ .
$\hat{\theta}_N^*$	The Nested pivoted estimate when $p$ is unknown.
$\hat{t}^*$	The estimated average private signal $w$ and $p$ are unknown.
$\hat{s}^*$	The estimated shared signal when $w$ and $p$ are unknown.
$\alpha_S^*$	The weight the decision analyst places on the estimated average private signal relative to the estimated shared signal in the Symmetric setting when $w$ and $p$ are unknown.
$q$	The product of $w$ and $p$ , $q = wp$ .
$\hat{q}$	The estimated slope coefficient from a simple least-squares regression of $g$ on $f$ within the subset of data $\{(f_i, g_i)   f_i \neq g_i\}$ .
$s_{\hat{q}}$	The standard error of the estimated slope coefficient from a simple least-squares regression of $g$ on $f$ within the subset of data $\{(f_i, g_i)   f_i \neq g_i\}$ .
$\alpha_{NS}^*$	The weight the decision analyst places on the estimated average private signal relative to the estimated shared signal in the Nested-Symmetric setting when $w$ and $p$ are unknown.
$q^*$	The decision analyst's posterior expectation of $q$ .
$\hat{\theta}_{NS}^*$	The Nested-Symmetric pivoted estimate when $w$ and $p$ are unknown.
$r_i$	A response from individual $i$ .
$y_r$	A crowd forecast of the probability that team 1 will win game $r$ .
$X_r$	An indicator variable for whether team 1 won game $r$ .
$o$	The decimal odds that a team will win a particular game.
$\kappa_r$	The implied market probability that team 1 will win game $r$ .
$\bar{X}$	The overall proportion of games won by team 1.
$p_c$	The probability forecast for category $c$ .
$n_c$	The number of crowd forecasts binned to category $c$ .

$\bar{X}_c$  The proportion of games where the crowd forecast was binned to category  $c$  that were won by team 1.

## Abbreviations

BTS	Bayesian truth serum.
LST	Least surprised by the truth.
S	Symmetric.
N	Nested.
NS	Nested-Symmetric.
M	Minimal.
MAD	Mean absolute deviation.
RMSE	Root mean squared error.
s.d.	Standard deviation.
n.s.	Not significant.

# Acknowledgements

I am deeply grateful to Jack Soll for advising me over the past four years and for introducing me to the forecasting problem studied in this dissertation. He has been incredibly generous and patient, and is always curious and eager to engage in an intellectual discussion. I am thankful to Bob Nau for his advice and support, and for joining me for many Duke basketball games at Cameron Indoor Stadium. I am also indebted to Bob Winkler and Bob Clemen for their guidance and for many helpful discussions and suggestions. Additionally, I would like to thank Professors Ralph Keeney, Casey Lichtendahl, Yael Grushka-Cockayne, Victor Jose, Saša Pekeč, Alex Belloni, Jim Smith, and Alessandro Arlotto for their encouragement and advice throughout my doctoral studies. Thank you also to Şafak Yücel, Noah Snyder-Mackler, Tres Lambert, Troy Campbell, Alex Wilson, Zack Abzug, Stefan Rampertshammer, Andrew Frazelle, Vinh Nguyen, Levi DeValve, Christina Rader, Dan Feiler, Jordan Tong, and Saed Alizamir for their help and friendship, and for helping make my time in Durham wonderful. I owe my gratitude to the Fuqua School of Business for providing me with the opportunity and financial support to pursue my doctoral degree and carry out my research. Finally, I would like to thank my family for their unwavering support and encouragement, especially my parents and grandparents for the many sacrifices they made to further my education.



# 1

## Introduction

Obtaining accurate forecasts of uncertain variables is an important problem across a broad variety of applications, ranging from managerial decision problems to macroeconomics to geopolitics. These forecasts often comprise a crucial input to decision making in practice, offering insight into questions such as “how many units of this new product will we sell next year?”, “how much oil can we expect to recover if we drill this well?”, “how much will GDP grow over the next three years if we implement this policy?”, or “what are the chances that Greece will exit the eurozone by end of 2016?”

Often, there are many pieces of useful information that can help generate a better forecast, but they may be scattered across different institutions and people. Combining many individual forecasts to obtain an aggregate estimate can therefore provide an effective technique for improving forecast accuracy. This idea is known as the wisdom of crowds (Surowiecki, 2005), and evidence for its effectiveness dates back more than a century (Galton, 1907). Since that time, many studies have shown that the simple average performs remarkably well in a wide variety of settings (Clemen & Winkler 1986; Clemen 1989; Makridakis & Winkler, 1983; Larrick, Mannes, & Soll,

2012). One reason for the success of averaging is that as long as forecasts bracket the truth (i.e., some are too high and others too low), averaging is guaranteed to be more accurate than the average individual (Larrick & Soll, 2006).

Theoretically, the simple average provides the best crowd forecast when individual forecast errors are independent and identically distributed. These conditions are rarely met in practice, and a variety of techniques have been proposed to improve accuracy. Some, such as trimming and Winsorizing (Armstrong, 2001; Jose & Winkler, 2008), have proven modestly beneficial because they are robust across different error distributions. Other, more formal approaches strive to exploit the covariance structure in past data to derive optimal weights (Winkler, 1981; Clemen & Winkler, 1986; Budescu & Chen, 2015). To be effective, these methods require a large amount of data and a sufficiently stable forecasting environment. In practice, perhaps because these conditions are usually not met, simple averaging often performs as well as and sometimes better than formal approaches (Clemen, 1989). A recently proposed technique that shows some promise is the “select crowd” method—forecasters are ranked according to a cue to expertise such as recent performance, and a simple average is taken of a subset, such as the top five (Mannes, Soll, & Larrick, 2014). This method outperforms the average of all judges when judges differ greatly in their ability, but still captures most of the benefits of aggregation when judges are more similar.

A shortcoming of the aforementioned methods is that they are severely limited by the lack of independence of information. Correlation in forecast errors, even to a small degree, greatly limits the ability of the wisdom of crowds to recover the truth. Clemen and Winkler (1985) examine the limits on the precision of an aggregated forecast when information is shared. By matching the precision of individual forecasters and the precision in the aggregate forecast, they introduce the notion of an equivalent number of independent experts, and demonstrate the severe limits that dependence

places on the value of aggregation. For example, if all forecast errors have a common variance and pairwise correlation of 0.25, the information that could be extracted from even a very large number of forecasts will be less than only four equivalent independent experts. In practice, this dependence often emerges because shared information leads to similar forecasts.

Sharing of information may occur naturally due to the way knowledge is distributed among forecasters, and may additionally arise as a result of social processes. For example, interaction between individuals can lead to informational or reputational cascades (Anderson & Holt, 1997; Hung & Plott, 2001; Wang & Wang, 2014) that may direct the crowd toward an inaccurate estimate. In addition, social influence can cause individual forecasts to converge, reducing the amount of private information contained in each of the responses (Lorenz et al., 2011). Likewise, forecasts that are made public may influence subsequent responses. Trueman (1994) finds that financial analysts exhibit herding behavior, tending to provide similar forecasts to those recently released.

As an illustrative example, consider a policy-maker who would like an accurate estimate of U.S. GDP growth in the coming quarter and the year ahead. She may ask for forecasts from a variety of individuals, including executives and macroeconomists at a variety of companies and organizations. These executives each possess useful private information from their firm, including internal projections, unreleased data on inventory levels and advance orders, and informal opinions from managers, salespeople, and suppliers. Likewise, macroeconomists may have information that is specific to their proprietary forecasting models. However, a large amount of information is also publicly available, such as projections by central banks, government organizations, and think tanks, recently released economic data, and compilations of other forecasts announced by economists. In these settings, even though the shared information may be highly informative, the benefits of averaging across many forecasters

are limited, because a significant proportion of the average forecast represents the same information being repeated over and over, a problem referred to as *shared-information bias*.

In this dissertation, I introduce a forecast aggregation method that can remove shared-information bias from the crowd forecast. To do this, I propose augmenting the standard elicitation procedure so that in addition to providing their own forecasts, individuals also guess how others will respond (Prelec, 2004). By asking this additional question, the decision analyst can estimate which part of each forecast is shared and which part of each forecast is private information. The aggregation method, which I call *pivoting*, then builds a crowd forecast by recombining each of these shared and private pieces of information in a Bayesian manner. This allows for a more precise estimate and reduces the aggregate forecast error to its minimum as the crowd size grows large. I introduce several variations on pivoting, each of which makes different assumptions about how information may be distributed across forecasters.

After developing the aggregation method, I present three studies that test the assumptions of this elicitation method and examine the accuracy of the aggregate forecasts. In the first experiment, shared and private information is varied in a controlled setting, which makes it possible to observe how individuals use their information to respond to the elicitation procedure, and also the ultimate effectiveness of the pivoting method in estimating the target variable. This is followed by two additional studies that examine responses and aggregate forecasts in settings with real-world outcomes.

## 2

### Literature Review

There have been several attempts to address shared-information bias in forecasts, which is a particularly challenging problem because it requires separating the private and shared information in some manner, either by the decision analyst or the forecasters themselves.

First, the incentives can be changed to encourage forecasters *themselves* to weight private information more heavily (Ottaviani & Sørensen, 2006; Marinovic, Ottaviani, & Sørensen, 2012). Lichtendahl, Grushka-Cockayne, and Pfeifer (2013) showed that this can be at least partially achieved by running a winner-take-all forecasting contest. The equilibrium strategy for each forecaster is to exaggerate the weight that they place on their own private information, because this allows them to sufficiently differentiate their forecast from others' in order to increase their chances of winning. As a result, shared information receives less weight in each response, and comprises a reduced proportion of the average forecast.

Second, assumptions can be made about the rate at which private information becomes public, which in turn can be capitalized upon in a multi-period setting where forecasters make predictions about the same quantity of interest over time.

Kim, Lim, and Shaw (2001) suggest using average past forecasts as a proxy for shared information, arguing that information that was formerly private may become commonly known over time. They propose a crowd forecast that reduces shared-information bias by overweighting the change in the mean forecast from the previous period.

A third type of approach is to elicit additional responses from each forecaster that can aid in distinguishing between shared and private information. Chen, Fine, and Huberman (2004) introduce a two-step aggregation procedure for forecasting discrete event probabilities. Their method first elicits individuals' probability estimates, and then identifies shared information through a coordination game. The method then combines individual forecasts by aggregating the probabilities corresponding to the residual private information after this shared information has been removed.

Prelec (2004) introduces an elicitation method that asks respondents to both select their own choice from among a discrete set of options and to provide an estimate of the frequencies of choices that others will make. Responses are scored according to a rule called the “Bayesian Truth Serum” (BTS), which incentivizes honesty for both types of responses. Building on this, Prelec, Seung, and McCoy (2013) propose a crowd estimation procedure based on a “least surprised by the truth” (LST) principle, which selects the forecast provided by those with the highest BTS scores. The LST choice can deviate from the majority choice when there is an asymmetry in predicting others. For example, suppose that the task is to decide whether the correct answer is A or B, and that the majority chooses B. Depending on the distribution of estimates, the LST may select A if A-responders predict that most people will pick B, and B-responders also think that others will pick B. At a high level, the A-responders are revealing that have additional information or meta-knowledge that leads them to knowingly select a relatively unpopular answer. In this dissertation, I propose a similar elicitation procedure, but use the guesses about others in a different way—to

identify shared information.

Although important theoretical advances have been achieved, past approaches to solving the shared-information problem have several limitations. The methods that have been proposed to change incentives are imperfect—at best they can eliminate a fraction of the shared-information bias, but not all of it. Similarly, Kim, Lim, and Shaw’s (2001) method promises to remove only a fraction of the bias. The LST method appears to work best in discrete choice scenarios in which some people know more than others (although it can potentially be extended to continuous settings by discretizing the responses). Other methods are also specialized to particular types of settings or forecasting tasks (Chen et al., 2004; Kim et al., 2001). Finally, although the aforementioned methods hold promise, more evidence is needed to demonstrate that they can reliably outperform the simple average in practice.

One might alternatively hope that forecasters can overcome shared-information bias through information exchange and discussion. However, many studies in judgment and decision making suggest that groups instead tend to “emphasize shared information at the expense of unshared information” (Sunstein, 2005). Precisely because it is held by more members, shared information is mentioned and repeated more frequently (Stasser & Titus, 1985), and bearers of shared information are perceived as more credible (Wittenbaum, Hubbell, & Zuckerman, 1999). In addition, Gigone & Hastie (1993) find that the influence of a piece of information on the collective judgment depends positively on the number of group members who held that information at the outset of discussion. As a result, just as in the algorithmic combination of forecasts, shared information in a group enters the aggregate multiple times and thus has an outsized impact on the group decision. Teams can benefit from employing a more structured deliberative group judgment process such as the Delphi method (Dalkey & Helmer, 1963; Hsu & Sandford, 2007), which takes care to independently and anonymously gather information, and provides individuals with the opportunity

to revise their opinions after observing the reasoning of others. Some recent work also provides a more optimistic view about teams, suggesting that both individual and aggregate forecasts are improved when teams of forecasters have the opportunity to exchange facts and opinions in an online chat platform (Chen & Budescu, 2015; Mellers et al., 2014).

Prediction markets provide another potential mechanism for harnessing knowledge from a collection of individuals in order to forecast uncertainties (Wolfers & Zitzewitz, 2004). Markets encourage the discovery and aggregation of information, and given sufficient participation, perform well in practice in a variety of settings (Plott, 2000; Sunstein, 2006). The price signal plays a key role in this process, directing forecasters to self-select the size of their trades based on the strength of their beliefs. This provides greater incentives for highly-informed individuals to participate, which in turn increases the weight assigned to their information. However, studies comparing prediction markets and mechanistic combination of individual forecasts have yielded mixed results (Gruca, Berg, & Cipriano, 2003; Van Bruggen et al., 2010; Atanasov et al., 2015), and it remains unclear whether markets can eliminate shared-information bias. In addition, prediction markets may still be vulnerable to herding behavior and market bubbles (Spyrou, 2013; Cipriani & Guarino, 2014). Further comparison of the effectiveness of markets versus mechanistic aggregation of individual judgments in combining information would be useful and offers a fruitful area for further research. In this dissertation, however, I will restrict my attention to combining individual forecasts through algorithmic mechanisms rather than through a market.

In the next section, I propose an aggregation method, called pivoting, that can be applied across a variety of settings and requires no prior data from individual forecasters. The elicitation procedure requires only two responses from each forecaster, and can be implemented directly without any interaction between individuals. In



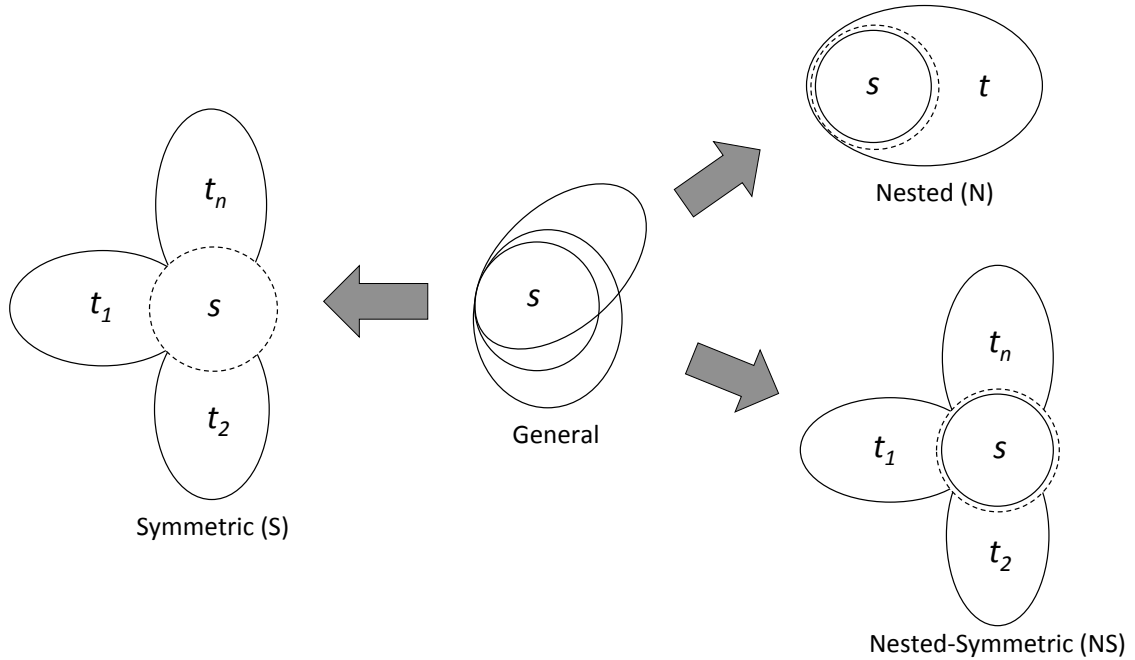
theory, pivoting has the potential to completely remove any shared-information bias from the average responses, providing the decision analyst with the most accurate crowd forecast possible.

## Analysis of the Shared-Information Problem

In the most general setting, information may be distributed haphazardly among forecasters as shown at the center of Figure 3.1. Some information set  $s$  is held by all forecasters, and additional information may be held uniquely or by various subsets of forecasters, represented by the overlapping ovals in the figure. Although averaging can be an effective way to combine forecasts, the shared signal  $s$  may bias the result in this setting—each forecaster will use  $s$  in his own forecast, meaning that  $s$  will comprise an outsized portion of the average forecast as well. The general model turns out to be extremely difficult to analyze (see Appendix B), because the possible sets of overlapping information are combinatorially complex. Therefore, I consider several idealized models that approximate a range of real-world situations. Whereas the models differ substantially in their assumptions about how information is distributed, they all imply a form of pivoting as the optimal aggregation policy.

I begin by considering a Symmetric (S) information model, versions of which have been studied by Kim, Lim, and Shaw (2001), Ottaviani and Sørensen (2006), and Lichtendahl, Grushka-Cockayne, and Pfeifer (2013), among others. This model assumes that all forecasters have access to the same shared information  $s$ , and ad-

FIGURE 3.1: Comparison of information structures across the models of shared information.



The solid line around  $s$  indicates that naïfs only observe the shared information, and the dashed line indicates that sophisticates observe both  $s$  and an additional signal as well.

ditionally each forecaster  $i$  receives an equal amount of private information that is uniquely held, represented by  $t_i$  in Figure 3.1. Thus, any given piece of information is held by either everyone or by only one person. The Symmetric model could be a plausible approximation to market situations in which forecasters supplement public information with their own research. Like the Symmetric model, the Nested (N) information model assumes that all forecasters, who are classified as either naïfs or sophisticates, have access to  $s$ .<sup>1</sup> The sophisticates also observe common information  $t$  in addition to  $s$ . The Nested model is meant to approximate situations in which

<sup>1</sup> Note that in the behavioral economics literature, the terms “naïf” and “sophisticate” are often used to specify differences in the rationality of an agent. In this dissertation I use the terms in a different way, to denote differences in the amount of information that forecasters possess. Naïfs observe only the shared signal  $s$  and are unaware of additional signals, while sophisticates observe  $s$  as well as an additional signal.

some people are very familiar with a forecasting context and others know less, such as predicting the outcome of a college basketball game. The decision analyst does not know which of the forecasters are sophisticates, yet must try to extract their extra knowledge based on everyone’s responses. A third model, the Nested-Symmetric (NS), combines elements of both structures. As in the Nested model, naïfs only observe  $s$ . Sophisticates each observe a unique  $t_i$  in addition to  $s$ , similar to the Symmetric model.

Although the assumptions of each model will rarely hold exactly in the real world, the range of models captures essential features of the shared information problem, while remaining simple enough to analyze and provide tractable solutions. Having several models also allows for examination of their robustness—how well do the aggregation procedures derived from each model perform when they deviate from the true information structure? To answer this question, the performance of each model is also compared to simple averaging and to a benchmark “minimal pivoting model” described below.

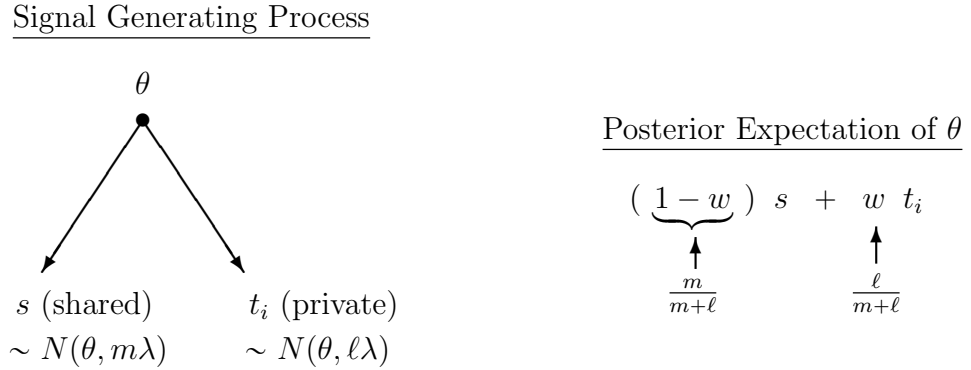
### 3.1 A Symmetric Model of Shared Information and Forecaster Behavior

I begin by presenting the Symmetric model of shared information. Let  $X$  denote the random variable being forecast, which is normally distributed with unknown mean  $\theta$  and known precision  $\lambda$ ,  $(X|\theta) \sim N(\theta, \lambda)$ . There are  $n$  forecasters, who share a diffuse prior belief about  $\theta$ . All forecasters observe the same shared signal  $s$ , where  $(s|\theta) \sim N(\theta, m\lambda)$ .<sup>2</sup> In addition, each forecaster  $i = 1, \dots, n$  receives his own private signal  $t_i$ , where  $(t_i|\theta) \sim N(\theta, \ell\lambda)$ . Conditional on  $\theta$ , private signals  $t_i$  are independent

---

<sup>2</sup> In this model, it would be equivalent for all forecasters to share a common non-diffuse prior, which would then be treated as a part of the shared signal. It is straightforward to show that all of the results that follow continue to hold with a common prior (see Lichtendahl, Grushka-Cockayne, & Pfeifer, 2013).

FIGURE 3.2: The signal generating process and belief updating for each forecaster  $i$  in the Symmetric information model.



The weight  $w = \ell/(m+\ell)$  placed on the forecaster's private information is determined by the relative precision of the shared and private signals.

across forecasters, and are independent of  $s$  as well.<sup>3</sup> One interpretation of the signals here would regard the shared signal  $s$  as equivalent to  $m$  independent observations of the random variable  $X$ , and each  $t_i$  as equivalent to  $\ell$  additional independent observations of  $X$ .

After receiving the pair of signals  $(s, t_i)$ , forecaster  $i$  updates his prior about  $\theta$  according to Bayes' rule. The forecaster combines information as shown in Figure 3.2, where the weight  $w = \ell/(m + \ell)$  represents the relative informativeness of private versus shared information. This in turn generates a posterior distribution for the random variable of interest  $X$ .<sup>4</sup> The forecaster's posterior distribution for  $\theta$  and resulting predictive distribution for  $X$  share the same mean of  $(1 - w)s + wt_i$ . This information structure and the parameters  $m$ ,  $\ell$ , and  $\lambda$  are commonly known among all forecasters.

---

<sup>3</sup> Assuming this independence of individual signals ensures that correlation between forecast errors is induced entirely by the structure of the information overlap across forecasters.

<sup>4</sup> Forecaster  $i$ 's posterior beliefs about  $\theta$  are given by  $(\theta|s, t_i) \sim N((1-w)s + wt_i, (m+\ell)\lambda)$  and his resulting posterior predictive distribution for  $X$  is  $(X|s, t_i) \sim N((1-w)s + wt_i, \frac{m+\ell}{m+\ell+1}\lambda)$  (Bernardo & Smith, 1994, p. 439).

### 3.1.1 Eliciting Information from Forecasters

The decision analyst would like to use responses from the forecasters to build a crowd forecast of  $X$ . In order to alleviate potential shared-information bias in the crowd forecast, I propose asking forecasters an additional question to help identify what part of their information is shared. Specifically, I propose asking, “What do you think will be the average forecast of the other  $n - 1$  forecasters?” By asking for both a forecast  $f_i$  of  $X$  and a forecast  $g_i$  of others’ forecasts, the decision analyst can distinguish between information that is common to all forecasts and information that is specific to forecaster  $i$ .

To incentivize forecasters to report their mean beliefs, the decision analyst may choose any two strictly proper scoring rules to elicit both  $f_i$  and  $g_i$ . Specifically, the forecaster can be rewarded with the sum of  $S_1(X, f_i)$  for his own forecast and  $S_2(\sum_{j \neq i} \frac{f_j}{n-1}, g_i)$  for his guess of others’ forecasts.<sup>5</sup> Even in the absence of a formal scoring mechanism during the elicitation process, this may be a reasonable approximation of a forecaster’s incentives when providing a response. He should respond to such a request by providing an honest assessment of both his own predictive mean for  $X$  and the mean forecast provided by others.

**Proposition 1.** *It is a Bayesian Nash equilibrium for each forecaster  $i$ , after observing the shared signal  $s$  and the private signal  $t_i$ , to provide responses of*

$$\begin{aligned} f_i &= (1 - w)s + wt_i, \\ g_i &= (1 - w^2)s + w^2t_i. \end{aligned} \tag{3.1}$$

Since the proposition holds for all forecasters, the shared signal will comprise the

---

<sup>5</sup> Each scoring rule should be strictly proper in the sense that reporting the mean of the forecaster’s subjective distribution for  $X$  provides him with the highest expected score among all possible reports. For example, the quadratic scoring rule (see e.g. Gneiting & Raftery 2007) offers a simple and frequently used proper scoring rule that can be rescaled in an appropriate positive affine manner to generate  $S_1$  and  $S_2$ .

fraction  $(1 - w)$  of a simple average of the individual forecasts  $f_i$ , as each individual is partially repeating  $s$  over and over again. The remaining fraction  $w$  reflects the forecasters' private information, which ideally would receive much more weight in the aggregate. As  $n$  grows large, the average of the private signals should converge to the mean  $\theta$  by the law of large numbers, while the specific realization  $s$  of the shared signal remains the same. As a result, any error resulting from  $s$  cannot be reduced by simply including more forecasters, leading to the problem known as shared-information bias.

The decision analyst can address the problem by considering the information in  $g_i$ . Although both optimal responses will lie between the shared and private signals,  $f_i$  lands closer to the private information while  $g_i$  lands closer to the shared information. This holds true on average as well, so that  $\bar{f}$  is shaded closer toward the average private signal while  $\bar{g}$  is shaded closer to  $s$ . Because  $s$  is overweighted in the simple average, I propose a method that combines information by starting with  $\bar{f}$  and pivoting away from  $\bar{g}$  in a way that removes the shared-information bias.

### 3.1.2 Estimating $\theta$ When $w$ is Known

If the relative information weight  $w$  is known, the decision analyst can estimate the shared and private signals by *pivoting* off of the average forecast  $\bar{f}$ , as shown in Figure 3. When estimating the forecasters' private signals, the decision analyst begins with the average forecast and pivots away from the average guess of others  $\bar{g}$  by a factor of  $1/w$ . Likewise, when estimating the shared signal, the decision analyst starts with the average forecast and pivots toward the guess of others by a factor of  $1/(1 - w)$ . After estimating all  $n + 1$  of these signals ( $n$  private signals and 1 shared signal), the decision analyst can combine them in the optimal Bayesian fashion. Details of these calculations are given in the following procedure:

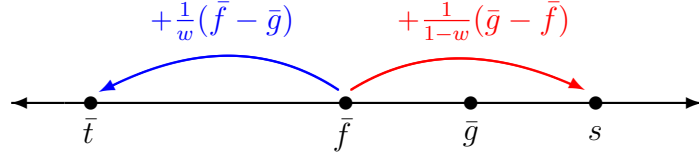


FIGURE 3.3: Estimating the private and shared signals in the Symmetric information model when  $w$  is known.

1. Estimate the average of the  $n$  private signals according to

$$\hat{t} = \begin{cases} \bar{f} + \frac{1}{w}(\bar{f} - \bar{g}) & \text{if } w > 0 \\ \bar{f} & \text{if } w = 0. \end{cases} \quad (3.2)$$

2. Estimate the shared signal according to

$$\hat{s} = \begin{cases} \bar{f} + \frac{1}{1-w}(\bar{g} - \bar{f}) & \text{if } w < 1 \\ \bar{f} & \text{if } w = 1. \end{cases} \quad (3.3)$$

3. Compute the aggregate forecast

$$\hat{\theta}_S = \alpha \hat{t} + (1 - \alpha) \hat{s}, \quad (3.4)$$

where

$$\alpha = \frac{nw}{nw + 1 - w}. \quad (3.5)$$

**Proposition 2.** *If forecasters report optimally, then the crowd forecast  $\hat{\theta}_S$  given in equation (3.4) is the minimum-variance linear weighting of the signals  $s$  and  $\{t_i\}_{i=1}^n$ .*

The weight  $\alpha$  in equation (3.5) captures the relative amount of information contained in the average private and shared signals, respectively. In two cases,  $\bar{f}$  and  $\bar{g}$  are equal, and the procedure simply returns the simple average  $\bar{f}$  as the crowd forecast. If  $w = 0$  then all information is shared and  $\bar{f} = \bar{g} = s$ . If  $w = 1$  then all information is privately held and  $\bar{f} = \bar{g} = \sum_{i=1}^n \frac{t_i}{n}$ . For all other weights  $w \in (0, 1)$ ,  $\alpha$  converges to 1 as the crowd size  $n$  grows large.



### 3.1.3 A Behavioral Model of Responses

In practice, responses may deviate from optimality for a variety of reasons. Cognitive limitations and mental costs often prevent people from fully processing all of the available information and working out the optimal responses, leading to the use of simplifying heuristics (Simon, 1955; Kahneman, 2011). In addition, forecasters may be inconsistent in attending to and weighting information, may incorporate heterogeneous prior beliefs into their responses, and may make other mistakes when thinking about the behavior of others. Relaxing the model of response behavior to account for these deviations from optimality is a necessary step to work with data from multiple forecasters, because not doing so would lead to an intractable, contradictory set of equations.

The behavioral model of responses assumes that equation (3.1) provides an accurate account of how individuals will behave *on average*. Each individual  $i$  provides a forecast

$$\tilde{f}_i = (1 - w)s + wt_i + \varepsilon_i, \quad (3.6)$$

where  $\varepsilon_i$  are independent mean-zero noise terms that are independent of  $s$  and  $t_j$  for all  $j = 1, \dots, n$ . Individual  $i$ 's guess of the average forecast of the other  $n - 1$  forecasters is given by

$$\tilde{g}_i = (1 - w)s + w\tilde{f}_i + \gamma_i, \quad (3.7)$$

$$= (1 - w^2)s + w^2t_i + w\varepsilon_i + \gamma_i, \quad (3.8)$$

where  $\gamma_i$  are independent mean-zero noise terms that are independent of  $s$ ,  $t_j$  and  $\varepsilon_j$  for all  $j = 1, \dots, n$ .

Applying the pivoting procedure detailed in equations (3.2) - (3.5) to these behavioral responses provides a forecast that outperforms the simple average for a sufficiently large crowd, yielding an unbiased forecast of  $X$  with minimum error.

**Proposition 3.** *If the behavioral model holds, both the average forecast  $\bar{f}$  and the crowd forecast  $\hat{\theta}_S$  in equation (3.4) provide unbiased estimates of the mean  $\theta$ . However, while the expected squared error  $\mathbb{E}[(\bar{f} - \theta)^2]$  is always greater than or equal to  $\frac{m}{(m+\ell)^2\lambda}$ , the expected squared error  $\mathbb{E}[(\hat{\theta}_S - \theta)^2]$  declines to 0 as the crowd size  $n$  grows large.*

Moreover, in practice it is unreasonable to expect that the relative information weight  $w$  will be known to the decision analyst. Using the behavioral model, I develop a procedure for the case of unknown  $w$  in Section 3.2.3.

## 3.2 What Happens Under Different Information Structures?

In general, the information structure could vary. In this section I describe the Nested and Nested-Symmetric information models displayed in Figure 3.1 and present a general pivoting procedure that is able to accommodate all three settings.

### 3.2.1 Nested Information Model

In many situations, forecasters may possess varying levels of expertise or access to information, with some knowing more than others. To study this setting, consider a simple model with two types of forecasters. A proportion  $1 - p$  of forecasters are *naïfs*, who only observe  $s$ , while the remaining proportion  $p$  of forecasters are *sophisticates*, who observe both  $s$  and an additional signal  $t$  that allows them to make better forecasts. As before,  $n$  forecasters share a diffuse prior belief about  $\theta$  and all observe the same shared signal  $s$ , where  $(s|\theta) \sim N(\theta, m\lambda)$ . Sophisticates receive an additional signal  $t$ , which is also normally distributed with mean  $\theta$  and precision  $\ell\lambda$ ,  $(t|\theta) \sim N(\theta, \ell\lambda)$ , and is independent of  $s$  conditional on  $\theta$ . Sophisticates are assumed to share a common expectation  $p$  for the proportion of other sophisticates in the

crowd.<sup>6</sup>

### *A Behavioral Model of Responses from Naïfs*

Naïve forecasters will on average forecast the shared signal  $s$  and, lacking additional information, will guess the same for the average response of others, so  $\tilde{f}_i = \tilde{g}_i = s + \delta_i$ , where  $\delta_i$  is a mean-zero error term that is independent of  $s$ ,  $t$ , and  $\delta_j$  for all  $j \neq i$ .

### *A Behavioral Model of Responses from Sophisticates*

Sophisticates will use both informative signals  $s$  and  $t$ , combining them according to the appropriate Bayesian weight  $w$ . In addition, they know that they possess more knowledge than the naïve types and expect that  $p$  of the other forecasters are informed and  $1 - p$  of the other forecasters are naïfs. As a result, sophisticates will provide a guess of others' forecasts of  $g_i = pf_i + (1 - p)s$ . A behavioral model for a sophisticate  $i$  can then be written as

$$\tilde{f}_i = (1 - w)s + wt + \varepsilon_i,$$

$$\tilde{g}_i = (1 - pw)s + pwt + p\varepsilon_i + \gamma_i,$$

where  $\varepsilon_i$  and  $\gamma_i$  are mean-zero error terms that are independent of  $s$ ,  $t$ , and each other.

### *3.2.2 Nested-Symmetric Information Model*

The Nested-Symmetric model combines elements of the previous two models. As in the Nested model, a proportion  $1 - p$  of the forecasters are naïfs, who observe only the shared signal  $s$ , and the remaining proportion  $p$  of forecasters are sophisticates,

---

<sup>6</sup> This premise simplifies the model to keep the informational and rationality assumptions about forecasters to a minimum. The parameter  $p$  serves as a reasonable expectation in a setting where the decision analyst samples randomly from a sufficiently large population of potential forecasters,  $p$  of whom are sophisticated. Otherwise, sophisticates would need to update their estimate of the proportion of other sophisticated forecasters in the crowd based on their own type, requiring that they possess additional knowledge about  $n$  and the identities of the other forecasters.

who observe an additional signal. However, this additional signal  $t_i$  is private, as in the Symmetric model.

#### *A Behavioral Model of Responses from Naïfs*

Following the same reasoning as in the Nested information setting, a response model for a naïve forecaster  $i$  can be written as  $\tilde{f}_i = \tilde{g}_i = s + \delta_i$ .

#### *A Behavioral Model of Responses from Sophisticates*

As in the Nested information setting, when guessing the average forecast given by others, sophisticates expect that  $1 - p$  of the other forecasters are naïfs and will provide an expected forecast of  $s$ . In addition, they expect that  $p$  of the other forecasters are sophisticates and will provide an average response of  $(1 - w)s + w\bar{t}_{-i}$ , where  $\bar{t}_{-i}$  is the average private signal observed by all other sophisticates. As in the Symmetric model, their best estimate of  $\bar{t}_{-i}$  is their own forecast  $f_i$ , which equals the posterior expectation of the mean  $\theta$  conditional on the observed signals  $s$  and  $t_i$ . After making the appropriate substitutions, a behavioral model of responses for a sophisticate  $i$  can be written as

$$\tilde{f}_i = (1 - w)s + wt_i + \varepsilon_i,$$

$$\tilde{g}_i = (1 - pw^2)s + pw^2t_i + p\varepsilon_i + \gamma_i.$$

#### *3.2.3 General Estimation Procedure When $w$ and $p$ are Known*

Below I provide a general procedure that accommodates all three of the information settings studied above. The Symmetric model was studied in the previous section and corresponds to the case where  $p = 1$ . The appropriate pivoting procedure and crowd forecast  $\hat{\theta}_N$  for the Nested model is obtained by setting  $w = 1$ .<sup>7</sup> Finally,

---

<sup>7</sup> In the Symmetric and Nested-Symmetric settings, the variables  $w$  and  $p$  in the estimation procedure correspond exactly to the parameters in the information model. However, in the Nested

the crowd forecast  $\hat{\theta}_{NS}$  for the Nested-Symmetric model can be estimated using the equation (3.11) with the appropriate values of  $w$  and  $p$  together.

1. Estimate the average of the  $pn$  private signals according to

$$\hat{t} = \begin{cases} \bar{f} + \frac{1}{pw}(\bar{f} - \bar{g}) & \text{if } pw > 0 \\ \bar{f} & \text{if } pw = 0. \end{cases} \quad (3.9)$$

2. Estimate the shared signal according to

$$\hat{s} = \begin{cases} \bar{f} + \frac{1}{1-pw}(\bar{g} - \bar{f}) & \text{if } pw < 1 \\ \bar{f} & \text{if } pw = 1. \end{cases} \quad (3.10)$$

3. Compute the aggregate forecast

$$\hat{\theta} = \alpha \hat{t} + (1 - \alpha) \hat{s}, \quad (3.11)$$

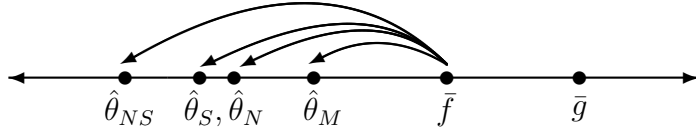
where

$$\alpha = \frac{npw}{npw + 1 - w}. \quad (3.12)$$

The optimal pivoting procedure for each of these information settings can be derived as in the Symmetric model; further details are provided in the proof of Proposition 4 in Appendix A. For each of the three respective information settings, if the behavioral model of reporting holds, then the expected squared error in the crowd forecast given by equations (3.9) - (3.12) outperforms the average forecast for sufficiently large  $n$ .

---

setting, the variable  $w$  in this general estimation procedure does not equal the parameter  $w$  in the information model. Nevertheless, it turns out that setting  $w = 1$  in this general procedure accommodates the optimal pivoting algorithm for the Nested setting when  $p$  is known.



If  $w, p$  known:

$$\hat{\theta}_M = \bar{f} + (\bar{f} - \bar{g})$$

$$\hat{\theta}_S \approx \bar{f} + \frac{1}{w}(\bar{f} - \bar{g})$$

$$\hat{\theta}_N \approx \bar{f} + \frac{1}{p}(\bar{f} - \bar{g})$$

$$\hat{\theta}_{NS} \approx \bar{f} + \frac{1}{pw}(\bar{f} - \bar{g})$$

FIGURE 3.4: Pivoting procedure for aggregating the crowd forecast in different information models for large  $n$  when the parameters are known and  $pw > 0$ .

### 3.2.4 Minimal Pivoting

Setting both parameters  $w = p = 1$  yields a Minimal (M) pivoting procedure. The minimal pivoting estimate, defined as  $\hat{\theta}_M \equiv 2\bar{f} - \bar{g}$ , provides a lower bound for the other three pivoting procedures. It pivots by a factor of 1, which is always smaller than the fraction for estimating the private information in the other pivoting procedures.

**Proposition 4.** *The crowd estimate  $\hat{\theta}_M$  obtained from the minimal pivoting procedure outperforms the simple average  $\bar{f}$  with a sufficiently large crowd in each of the Symmetric, Nested, and Nested-Symmetric information settings.*

The minimal estimate thus offers a robust and simple variant of the general pivoting procedure, which can be used in an off-the-shelf manner by a decision analyst regardless of the information setting. Minimal pivoting does not require parameter estimates for  $w$  or  $p$ , yet should still be effective in removing shared-information bias in a large crowd.

In settings with a large number of forecasters, if  $pw > 0$  then as  $n$  increases  $\alpha$  converges to 1 and the pivoting procedures are approximated by the expression for estimating the average private signal  $\hat{t}$ . Figure 3.4 depicts the pivoting process for each of the four procedures for a sufficiently large crowd. In the N and M

procedures, the equations given are exact for any  $n$ . In the S and NS procedures, the figure displays the expression given in equation (3.9), which approximates the exact formula for  $\hat{\theta}_S$  and  $\hat{\theta}_{NS}$  given by the full equation (3.11) when  $n$  is large.

### 3.3 Practical Implementation of the Pivoting Procedures

In practice, the decision analyst generally does not know the parameters of these information models, and must estimate them from response data. In the following section I present modified versions of the pivoting procedures presented in Section 3.2 that additionally account for uncertainty in these parameter values. It is also the case that the decision analyst generally will not know which model best corresponds to the forecasters' information setting. After developing a procedure for each of the S, N, NS, and M settings below, I apply the algorithms to empirical data in Chapter 4 and draw conclusions about which of the pivoting methods to use in practice.

#### 3.3.1 Estimating $\theta$ in the Symmetric Information Model When $w$ is Unknown

In the Symmetric information setting, the decision analyst can estimate a posterior distribution of  $w$  given the response data through a simple linear regression. The behavioral model assumes that  $g_i = (1 - w)s + wf_i + \gamma_i$ , where  $\gamma_i$  is a noise term with mean 0 and finite variance and is independent of  $s$ . If  $\gamma_i$  is assumed to be normal, then  $w$  can be estimated using the slope coefficient  $\hat{w}$  from a least-squares linear regression of  $g$  on  $f$ . Note, however, that if the number of forecasters  $n \leq 2$ , the decision analyst will not be able to estimate  $w$  and must revert to using the minimal pivoting procedure  $\hat{\theta}_M$ . If  $n > 2$ , this estimate  $\hat{w}$  follows a noncentral  $t$ -distribution with  $n - 2$  degrees of freedom, a mean of  $w$ , and standard error  $s_w$ . In other words, the likelihood function  $f(\hat{w}|w, s_w)$  is  $t_{n-2}^{w, s_w}(w)$ . Taking a uniform prior over the possible values of the weighting parameter  $w$  ( $\pi(w) = \mathbf{1}_{[0,1]}$ ) and using the standard error  $s_{\hat{w}}$  as an estimate of  $s_w$ , Bayes' theorem then allows the decision analyst to derive

a posterior distribution  $f(w|\hat{w}, s_{\hat{w}}) \propto t_{n-2}^{\hat{w}, s_{\hat{w}}}(w) \mathbf{1}_{[0,1]}$ . This means that the posterior estimate of the relative information weight is

$$w^* = \mathbb{E}[w|\{(f_i, g_i)\}_{i=1}^n] = \frac{\int_0^1 w t_{n-2}^{\hat{w}, s_{\hat{w}}}(w) dw}{\int_0^1 t_{n-2}^{\hat{w}, s_{\hat{w}}}(w) dw}. \quad (3.13)$$

However, the decision analyst should not use  $w^*$  directly in the pivoting procedure introduced in equation (3.4). In the general estimation procedure,  $w$  appears in the denominator when estimating the average private and shared signals in equations (3.2) and (3.3). This implies that any uncertainty in the estimate of  $w$  will cause “overpivoted” estimates of both  $\bar{t}$  and  $s$ , leading to a biased estimate of  $\hat{\theta}$  in equation (3.4). To solve this problem, the decision analyst should instead pivot according to a hedged weight  $h$  that minimizes the expected squared error in estimating these signals. The following algorithm provides an analogue of the optimal crowd forecast from Section 3.1.2 that accounts for this uncertainty in  $w$ :

1. If  $n \leq 2$ , set the aggregate forecast equal to  $\hat{\theta}_S^* = \bar{f}$ .
2. If  $n > 2$ , use the data  $\{(f_i, g_i)\}_{i=1}^n$  to estimate the slope coefficient  $\hat{w}$  and its standard error  $s_{\hat{w}}$  from a simple least-squares linear regression of  $g$  on  $f$ .
3. Estimate the average private signal according to

$$\hat{t}^* = \bar{f} + \frac{1}{h_{private}^*(\hat{w}, s_{\hat{w}}, n)} (\bar{f} - \bar{g}),$$

where

$$h_{private}^*(\hat{w}, s_{\hat{w}}, n) = \frac{\int_0^1 w^2 (1-w)^2 t_{n-2}^{\hat{w}, s_{\hat{w}}}(w) dw}{\int_0^1 w (1-w)^2 t_{n-2}^{\hat{w}, s_{\hat{w}}}(w) dw}.$$

4. Estimate the shared signal according to

$$\hat{s}^* = \bar{f} + \frac{1}{1 - h_{shared}^*(\hat{w}, s_{\hat{w}}, n)} (\bar{g} - \bar{f}),$$



where

$$h_{shared}^*(\hat{w}, s_{\hat{w}}, n) = \frac{\int_0^1 w^3(1-w)t_{n-2}^{\hat{w}, s_{\hat{w}}}(w)dw}{\int_0^1 w^2(1-w)t_{n-2}^{\hat{w}, s_{\hat{w}}}(w)dw}.$$

5. Compute the aggregate forecast

$$\hat{\theta}_S^* = \alpha_S^* \hat{t}^* + (1 - \alpha_S^*) \hat{s}^*, \quad (3.14)$$

where

$$\alpha_S^* = \frac{nw^*}{nw^* + 1 - w^*} \text{ and } w^* = \frac{\int_0^1 w t_{n-2}^{\hat{w}, s_{\hat{w}}}(w)dw}{\int_0^1 t_{n-2}^{\hat{w}, s_{\hat{w}}}(w)dw}.$$

**Proposition 5.** *Under the behavioral model of section 3.1.3, the hedged weights  $h_{private}^*(\hat{w}, s_{\hat{w}}, n)$  and  $h_{shared}^*(\hat{w}, s_{\hat{w}}, n)$  minimize the expected squared error in the estimated average private signal  $\hat{t}^*$  and the estimated shared signal  $\hat{s}^*$ , respectively.*

A graph of the hedging function for private signals for different levels of uncertainty is displayed in Figure 3.5. In step 5, the decision analyst simply replaces the information weight  $w$  in the expression for  $\alpha$  in equation (3.5) with its posterior estimate  $w^*$ . As  $n$  increases,  $\alpha^* \rightarrow 1$  and the crowd forecast converges to the estimated average private signal  $\hat{t}^*$ . While relating the crowd forecast in equation (3.14) to  $\bar{f}$  is much more difficult due to its complexity, the results proved in section 3.1.2 correspond closely to this hedged pivoting algorithm and strongly suggest that  $\hat{\theta}_S^*$  will outperform the simple average  $\bar{f}$  as the crowd grows large.

### 3.3.2 Estimating $\theta$ in the Nested Information Model When $p$ is Unknown

In the Nested model, the decision analyst must use the response data to estimate the proportion of sophisticates  $p$ . In this setting, the slope of a linear regression of  $g$  on  $f$  may not produce a unique estimate of  $p$ . Instead, I propose a much simpler approach, in which the decision analyst assumes that  $f_i = g_i$  if and only if forecaster  $i$

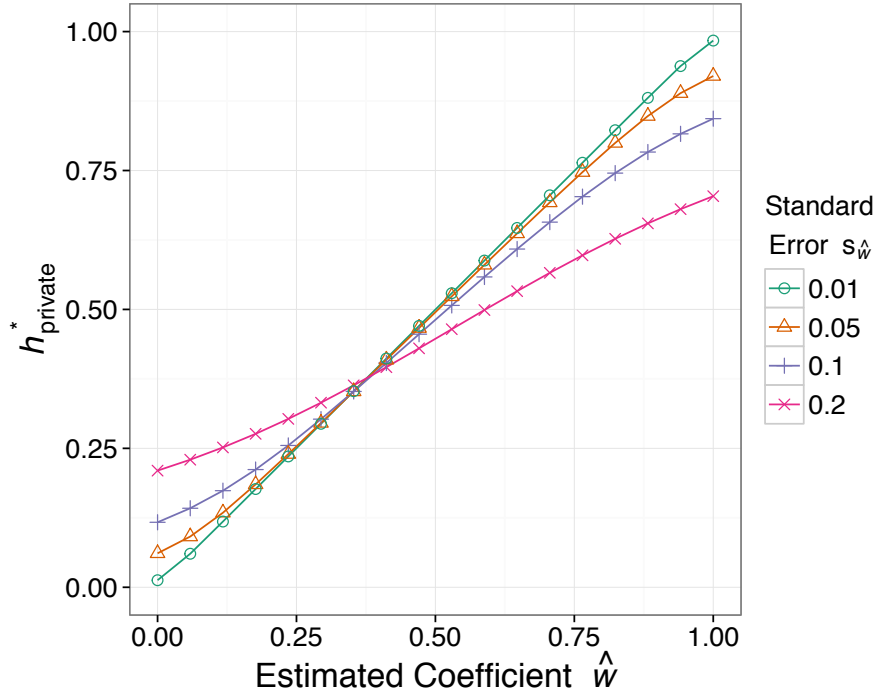


FIGURE 3.5: Optimal hedge  $h_{private}^*(\hat{w}, s_{\hat{w}}, n)$  as a function of  $\hat{w}$  for different standard errors  $s_{\hat{w}}$  when  $n = 100$ .

is a naïf. The decision analyst can then simply calculate the proportion of forecasters who provide a guess about others that differs from their own forecast and use this in place of  $p$  in equation (3.11).

There are two potential errors associated with this approach: First, a naïf may provides a guess  $g_i \neq f_i$ , which will cause the estimated proportion of sophisticates to be too high. Second, a sophisticate may provide a guess  $g_i = f_i$ , which will cause the estimated proportion of sophisticates to be too low. Overall, however, the assumption that  $f_i = g_i$  if and only if forecaster  $i$  is a naïf accords with the basic intuition of the Nested model: Naïfs generate their forecasts using limited information, which they believe all other forecasters observe as well, and cannot see how their forecast might differ from others'. Sophisticates, on the other hand, possess additional information that allows them to make a better forecast. The notion that this information is *extra*

is indicated by the forecasters discerning it as such; they recognize that it may be only available to some subset of forecasters, and that, as a result, the forecasts given by others will be different than their own. The crowd estimation procedure then proceeds as follows:

1. Count the number  $d$  of forecasters who provide a guess  $g_i \neq f_i$  and compute

$$\hat{p} = d/n. \quad (3.15)$$

2. Compute the crowd forecast according to<sup>8</sup>

$$\hat{\theta}_N^* = \begin{cases} \bar{f} + (\bar{f} - \bar{g})/\hat{p} & \text{if } \hat{p} > 0 \\ \bar{f} & \text{if } \hat{p} = 0. \end{cases} \quad (3.16)$$

### 3.3.3 Estimating $\theta$ in the Nested-Symmetric Information Model When $w$ and $p$ are Unknown

Here the pivoting procedure can be adapted by making use of two assumptions that are analogous to those discussed previously: First, forecaster  $i$  is assumed to be a naïf if and only if  $f_i = g_i$ , which allows the decision analyst to classify responses into either a *naïf* category or a *sophisticate* category. Next, the behavioral model for sophisticates given in Section 3.2.2 can be used to estimate  $pw$  by finding the slope of a linear regression of  $g$  on  $f$  among the *sophisticate* category of responses.

Specifically, defining  $q \equiv pw$ , the behavioral model for sophisticates can be written as  $g_i = (1 - q)s + qf_i + \gamma_i$ . As before, assuming that  $\gamma_i$  is normal, the estimate  $\hat{q}$  equals the slope coefficient from a least-squares linear regression of  $g$  on  $f$  among *sophisticates*. If  $d \leq 2$  of the forecasters are sophisticates, the decision analyst is forced to revert to the Nested procedure presented in equation (3.16). If  $d > 2$ , the sample weight  $\hat{q}$  follows a noncentral  $t$ -distribution with  $d - 2$  degrees of freedom,

<sup>8</sup> Step 2 (equation (3.10)) of the general pivoting procedure from Section 3.2 is unnecessary here because  $\alpha = 1$ . Note that  $\hat{p} = 0$  corresponds to the case where  $f_i = g_i$  for all  $i$ , so  $\bar{f} = \bar{g}$  and all forecasters are classified as naïfs.

a mean of  $q$ , and standard error  $s_q$ . Taking a uniform prior over the possible values of the weighting parameter  $q$  ( $\pi(q) = \mathbf{1}_{[0,1]}$ ) and using the standard error  $s_{\hat{q}}$  as an estimate of  $s_q$ , Bayes' theorem allows the decision analyst to derive a posterior distribution given the responses of  $f(q|\hat{q}, s_{\hat{q}}) \propto t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) \mathbf{1}_{[0,1]}$ . The following algorithm then provides a pivoting procedure that accounts for uncertainty in  $q$ :

1. Compute  $\bar{f}$  and  $\bar{g}$  using the full set of responses  $\{(f_i, g_i)\}_{i=1}^n$ .
2. Count the number of forecasters  $d$  who gave a different guess about others than their own forecast (count all  $i$  such that  $f_i \neq g_i$ ), and compute  $\hat{p} = d/n$ .
3. If  $d \leq 2$ , set the aggregate forecast equal to  $\hat{\theta}_{NS}^* = \begin{cases} \bar{f} + (\bar{f} - \bar{g})/\hat{p} & \text{if } \hat{p} > 0 \\ \bar{f} & \text{if } \hat{p} = 0. \end{cases}$
4. If  $d > 2$ , estimate the slope coefficient  $\hat{q}$  and its standard error  $s_{\hat{q}}$  from a least-squares linear regression of  $g$  on  $f$  within the subset of data  $\{(f_i, g_i) | f_i \neq g_i\}$ .
5. Estimate the average private signal according to

$$\hat{t}^* = \bar{f} + \frac{1}{h_{private}^*(\hat{q}, s_{\hat{q}}, d)} (\bar{f} - \bar{g}),$$

where

$$h_{private}^*(\hat{q}, s_{\hat{q}}, d) = \frac{\int_0^1 q^2 (1-q)^2 t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}{\int_0^1 q (1-q)^2 t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}.$$

6. Estimate the shared signal according to

$$\hat{s}^* = \bar{f} + \frac{1}{1 - h_{shared}^*(\hat{q}, s_{\hat{q}}, d)} (\bar{g} - \bar{f}),$$

where

$$h_{shared}^*(\hat{q}, s_{\hat{q}}, d) = \frac{\int_0^1 q^3 (1-q) t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}{\int_0^1 q^2 (1-q) t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}.$$

7. Compute the aggregate forecast

$$\hat{\theta}_{NS}^* = \alpha_{NS}^* \hat{t}^* + (1 - \alpha_{NS}^*) \hat{s}^*, \quad (3.17)$$

where<sup>9</sup>

$$\alpha_{NS}^* = \frac{nq^*}{nq^* + 1 - w^*}, \quad q^* = \frac{\int_0^1 q t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}{\int_0^1 t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}, \quad \text{and } w^* = \frac{\int_0^{\hat{p}} q t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}{\hat{p} \int_0^{\hat{p}} t_{d-2}^{\hat{q}, s_{\hat{q}}}(q) dq}.$$

### 3.3.4 Minimal Pivoting

The minimal pivoting procedure is simple and straightforward, since it does not require estimates of either  $w$  or  $p$ , which are both set equal to 1. It can be directly implemented by calculating

$$\hat{\theta}_M = 2\bar{f} - \bar{g}.$$

Because it does not use parameter estimates and can be applied in each of the S, N, and NS settings,  $\hat{\theta}_M$  is also robust to differences in the information environment.

### 3.3.5 Additional Information Models

So far I have only presented a careful analysis of the pivoting procedure in settings where signals and the random variable of interest are normally distributed. In Appendix B, I consider several alternative information models that study the optimal aggregation of forecasts with different statistical distributions and more general models of how information might be shared and distributed across forecasters. Broadly, the results lend additional support to pivoting as an aggregation procedure, and to minimal pivoting as a simple and robust variant for practical implementation.

Most importantly, the pivoting method can be extended from forecasting continuous random variables to a setting where the decision analyst seeks to estimate

---

<sup>9</sup> The estimation of  $w^*$  here is analogous to the estimation of  $q^*$  with one modification—by definition  $w$  must be in  $[0, 1]$ , and  $w = q/p$ , so  $q$  must lie in the interval  $[0, p]$ . Since  $\hat{p}$  is being used as an estimate of  $p$ , restricting the distribution of  $q$  to  $[0, \hat{p}]$  ensures that the estimate  $w^*$  will stay in  $[0, 1]$ .

the probability of a discrete event. Information here can be modeled according to a beta-binomial process, and the four pivoting methods work in an exactly analogous manner for aggregating probability forecasts in the S, N, and NS settings.<sup>10</sup> This extension greatly expands the applicability of the aggregation mechanism. For example, a pivoting procedure for estimating probabilities can be used to forecast distributions of continuous variables by assessing points on the cumulative distribution function. The empirical performance of the pivoting procedures in a probability forecasting context is explored in Study 3 of the next chapter.

---

<sup>10</sup> Note, however, that these crowd probability forecasts may require an additional correction (such as Winsorizing) to constrain the estimates to the interval  $[0, 1]$ .

## Experimental Evidence

In the next section, I present the results of three studies conducted to examine the theory presented previously. I study the accuracy of the individual and crowd forecasts in several settings, and demonstrate the benefits of the pivoting procedures in removing shared-information bias and improving the crowd forecast.

### 4.1 Study 1: Predicting Coin Flips

Study 1 is designed to test the main predictions of the preceding theory. The analysis of the results focuses on two critical questions—1) Does the behavioral model of forecasts provide a reasonable characterization of subjects' responses? and 2) Does the pivoting method offer an effective technique for recovering an accurate estimate of the true mean of the variable being forecast? In particular, does this aggregation method outperform the classic wisdom-of-crowds approach of simply averaging individual forecasts? To answer these questions, I developed three carefully controlled information settings that correspond closely to the theoretical framework in Chapter 3. The responses can then be used to study how participants respond to specific private and shared signals and how the pivoting method aggregates information em-

pirically.

The experiment asked participants to predict how many heads would appear in 100 flips of a *biased* two-sided coin. The bias (the probability of observing a head on each flip) of the coin was unknown to each subject, and they were told that it could be any number between 1% and 99%. In the notation of Chapter 3, the random variable  $X$  was the realized number of heads in the 100 flips of the coin, and the true mean  $\theta$  of  $X$  was the bias of the coin times 100. At the start of the experiment,  $\theta/100$  was drawn from a uniform distribution on  $[0.01, 0.99]$ , and this realized bias was then used to generate sample flips that made up the information that served as a basis for forecasting. Shared information was comprised of sample flips that everyone saw, and private information was comprised of sample flips seen by only one or some people. I examined nine information settings, each with a different group of participants, as shown in Table 4.1. Participants in each setting repeated the task for eight separate coins with different biases and sample flips. Altogether, this provided data for 72 different coins that could be used to test the pivoting method.

In the Symmetric (S) setting, the parameter  $w$  is varied through the number of shared and private flips. The weight of  $w = 0.25$  corresponds to a condition with 9 commonly observed flips and 3 private flips (each forecaster receives their own set of private flips). The weight of  $w = 0.5$  corresponds to a condition with 9 commonly observed flips and 9 private flips, and the weight of  $w = 0.75$  corresponds to a condition with 3 commonly observed flips and 9 private flips. Although the common flips dominate individual judgment in this case, overall the private flips greatly outnumber the common (shared) ones. Pivoting accounts for this by estimating the shared and private signals separately, and recombining them with an estimate of  $w$ . In the NS setting, I set the weight at  $w = 0.5$  and varied the proportion  $p$  of sophisticates who each received their own 9 private flips in addition to the 9 commonly observed flips. The N setting was similar, except that the 9 “private” flips were shared among



Table 4.1: Experimental design and number of subjects in each condition.

		Nested-Symmetric (NS)			Nested (N)		
S		$p = 0.25$	$p = 0.5$	$p = 0.75$	$p = 0.25$	$p = 0.5$	$p = 0.75$
$w = 0.25$	119						
$w = 0.5$	101	111	112	117	112	119	114
$w = 0.75$	125						

all the sophisticates. Therefore, the NS setting contains much more information than the N setting, because in the former setting there is a large quantity of private flips distributed across sophisticates. Because of this informational disparity, it is expected that pivoting will have the most benefit in the S and NS settings.

#### 4.1.1 Methods

##### *Participants*

Responses were gathered by running a number of online forecasting challenges on Amazon Mechanical Turk, each with a targeted size of 100 participants. Participants were recruited in three batches, one for each of the information settings. Within a batch, participants were either assigned to a low, medium, or high  $w$  setting (S), or to a low, medium, or high  $p$  setting (N and NS). In the latter two settings, participants were also randomly assigned to be *naïfs* or *sophisticates* in proportion to  $p$ . Several screens were used to make sure that participants understood the task and response scales—330 potential participants “screened out” and never performed the main task. An additional 222 participants started the study but did not finish it. This left 1,030 participants who provided complete data that were subjected to analysis (average age = 34, 60% male). Participants spent an average of 10 minutes on the task, and were compensated a flat fee of \$0.50 plus an average bonus of \$1.19.

### Coin 7 of 8

Here are the results for this coin.

- Everyone sees the commonly observed flips, but only you see your own private flips.
- 50% of the forecasters get their own 9 private flips. The rest only see the commonly observed flips.

**Commonly Observed Flips:** TTTHTTTT (1 heads out of 9 flips)

**Your Private Flips:** TTTHTHTT (3 heads out of 9 flips)

---

**All Flips That You Observed:** 4 heads out of 18 flips

Please use the first slider bar below to forecast the number of Heads ("H") that will be observed in **100 new flips** of the coin.

Then, please use the second slider bar to estimate the average forecast that other participants will give for this question.

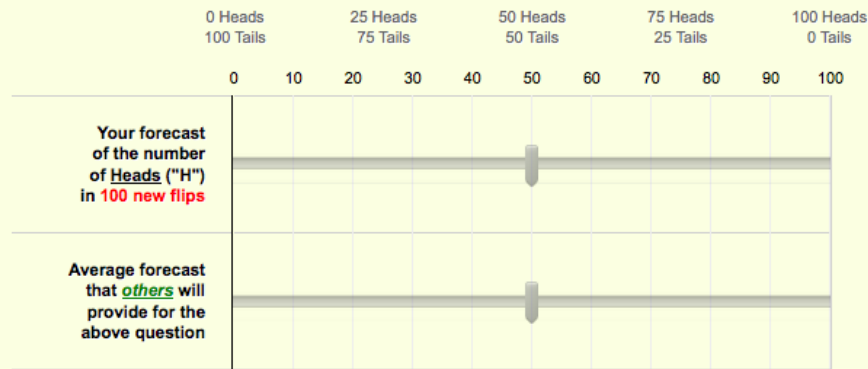


FIGURE 4.1: Example of the experimental interface for a sophisticate in the Nested-Symmetric setting.

### *Instructions and Procedures*

Participants were directed to an online survey interface created in Qualtrics, and read through instructions explaining the details of the forecasting challenge. They were told that they would be participating in eight separate coin prediction tasks, and that a new coin (with chances of heads anywhere between 1% and 99%) would be randomly selected for each task. To help develop an intuition for the task, they

were told the following:

*We will be showing you flips from biased coins. A biased coin is one that is tilted in favor of coming up either heads or tails. In case you're curious, these are virtual coins. The flipping is done by computer using a random number generator.*

They were also informed that after the experiment one of the coins would be randomly selected and flipped 100 times (virtually) to calculate their bonus payment. For their forecast of the number of heads, they started with \$1, and 5 cents were subtracted for every unit by which the forecast differed from the actual number of heads observed in 100 new flips of the coin. In addition, the same payment scheme was used to reward their accuracy in guessing the average prediction of others. Bonuses for each response were not allowed to fall below zero.

Participants provided answers for one practice coin to become familiar with the task, and eight additional coins, presented in a random order. For a given participant, the coins all shared the same information structure (e.g., Symmetric), as well as the same  $w$  and  $p$  values. The coins differed in their true biases  $\theta$  and the samples generated based on that  $\theta$ . In the N and NS settings, participants kept the same role as a sophisticate or a naïf for all coins (role labels were not communicated to participants). Participants in the naïf role were not aware that other participants would be getting additional flips, but participants in the sophisticate role were informed of the proportion of participants who were sophisticates. An example screen for a coin in the NS condition is shown in Figure 4.1. In the N condition the label for private flips was changed to “Additional Flips” to help clarify that these flips were shared by the fraction of participants who observed them.

Participants were asked to provide their own best forecast of the number of heads and a guess of the average forecast of others on sliding scales with an initial position of 50 (see Figure 4.1). After completing this task for all coins, participants answered several demographic questions, were thanked, and dismissed.

#### 4.1.2 Results

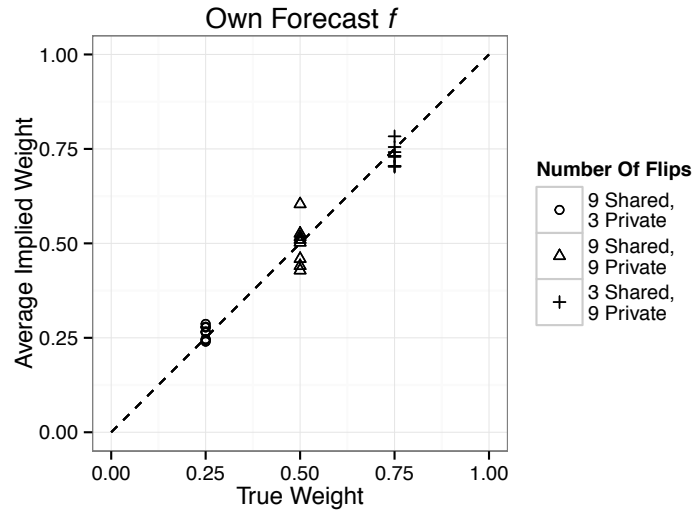
##### *Is There Evidence for the Behavioral Model?*

Although the forecasting setting in this experiment is somewhat stylized, the advantage of this setup is that it allows the signals participants receive to be controlled and observed exactly, and their forecasts can be compared to normative benchmarks. Given the information that each participant has seen, the relative weight that they are placing on private versus shared information can also be imputed, providing insight into whether the behavioral model reasonably characterizes their responses. For all sophisticates  $i$ , I estimated the implied weight on private information for each response  $r_i$  according to  $\frac{r_i - s}{t_i - s}$ , where  $t_i = t$  in the N setting. Implied weights were not calculated for naïfs, who receive only a shared signal  $s$ , or for cases in which the signals  $s$  and  $t_i$  equaled one another. Finally, weights outside the range  $[0, 1]$  were Winsorized (i.e., set equal to the nearer of 0 and 1), since responses outside the interval suggest that the two signals were not being combined.

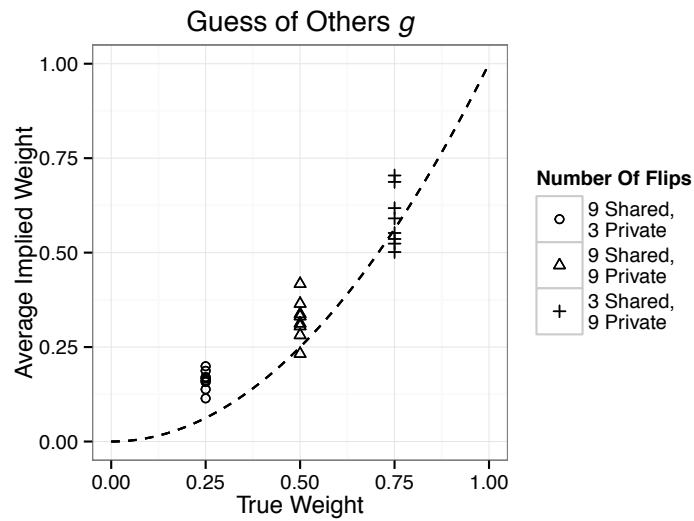
I first consider the average of these implied weights across all subjects for each coin in the S information setting, where the true relative information weight  $w$  was varied across conditions. The upper panel of Figure 4.2 shows the average weights implied by the individual forecasts  $f_i$  and the optimal weight  $w$ , represented by the dashed identity line. Not surprisingly, participants combined their private and shared information in a way that was close to optimal when providing their own forecast. The correlation between the average implied weight and the Bayesian optimal weight for forecasts across the 24 coins was 0.98.

However, to be effective, the pivoting method also requires forecasters to accurately guess others' forecasts, which depends crucially on their ability to *think vicariously*. As a result, a secondary but equally important question is whether participants were able to correctly use their *own* two pieces of information when guessing

FIGURE 4.2: Winsorized implied weights placed on own private signal  $t_i$  relative to the shared signal  $s$  versus optimal weights across the 24 coins in the Symmetric information setting.



(a) Average implied weight from own forecast.



(b) Average implied weight when guessing others.

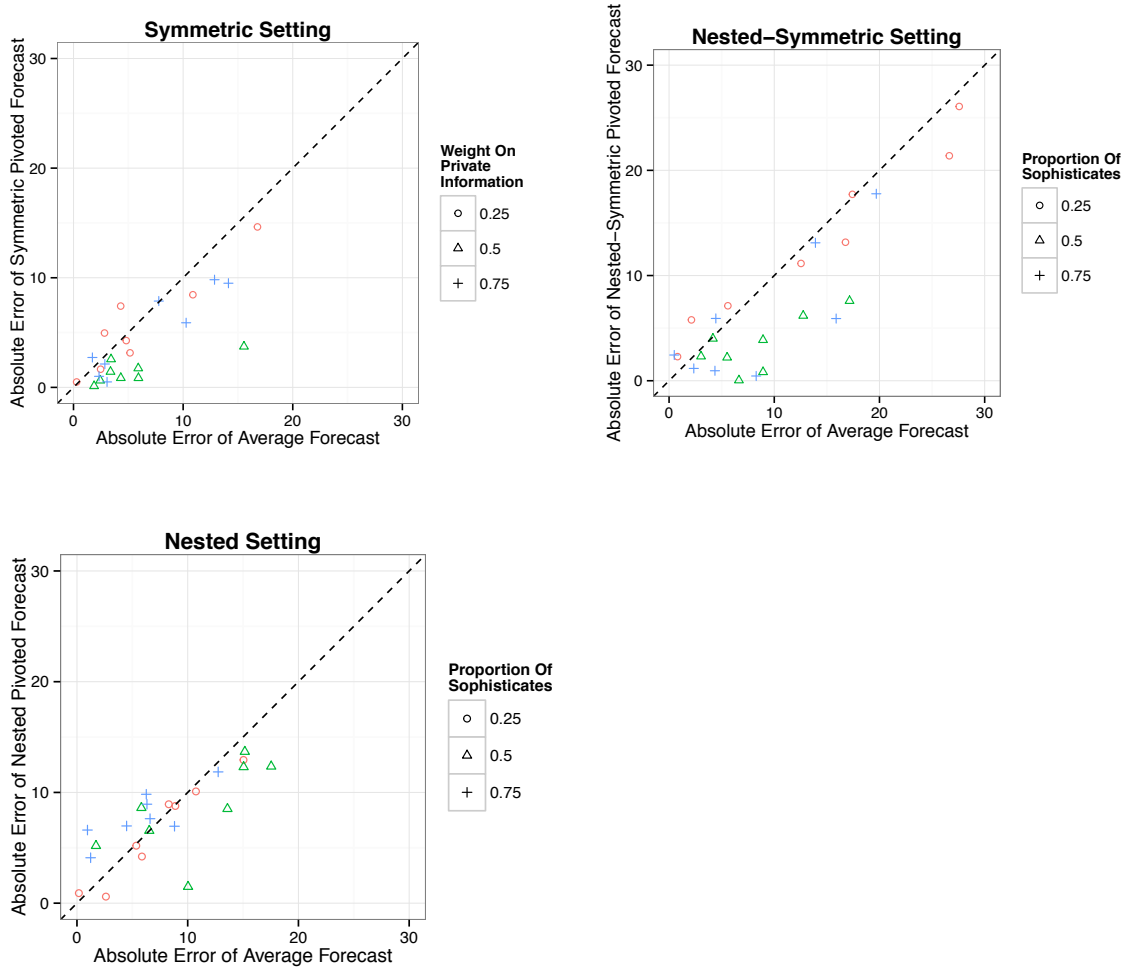
the average forecast of the other participants. In other words, did they on average place a weight of  $w^2$  on their own private signal when reporting  $g_i$ ? The lower panel of Figure 4.2 shows the average weights implied by the guesses about others  $g_i$  across each coin, and the dashed optimal weight  $w^2$  from Proposition 1. Although the individual weights varied substantially, participants on average used weights that were close to optimal here as well. The correlation between the average implied weight and the Bayesian optimal weight for guessing others across the 24 coins was 0.96.

Next, I considered the implied weights in the N and NS settings, where the relative information weight was fixed at  $w = 0.5$  and the proportion  $p$  of sophisticates was varied. On average, sophisticates put a weight of 0.29 on their private information in predicting others, compared to an average optimal weight of 0.19 across the coins in these settings. Contrary to the theory, participants were insensitive to  $p$  ( $r = -0.01, n.s.$ , across the two conditions). One potential reason for this is that a constrained optimal range (between 0.06 and 0.38) in these settings may have made it difficult to detect a relationship. Also, participants may have had difficulty coming up with a single estimate that aggregates over naïfs and sophisticates, leading to noisy estimates.

*Does the Pivoting Method Outperform Simple Averaging of Forecasts?*

Finally, and most importantly, I test how well the pivoting method performs at recovering the true mean  $\theta$  for each coin. I applied the four pivoting algorithms  $\hat{\theta}_S^*$ ,  $\hat{\theta}_N^*$ ,  $\hat{\theta}_{NS}^*$ , and  $\hat{\theta}_M^*$  to the response data, comparing the accuracy of the pivoted estimates against the accuracy of the simple average forecast  $\bar{f}$ . Figure 4.3 displays the absolute differences between both of these crowd forecasts and the true mean  $\theta$  for each coin, using the appropriate pivoting method in each information setting. Points below the dotted line indicate coins for which the absolute error of  $\bar{f}$  was higher than the absolute error of the pivoted forecast  $\hat{\theta}^*$  for that respective setting.

FIGURE 4.3: Performance of the respective pivoting method relative to the simple average forecast  $\bar{f}$  in each of three information settings.



In the S setting, the pivoted estimate  $\hat{\theta}_S^*$  had a significantly lower absolute error than  $\bar{f}$  in estimating  $\theta$  ( $|\bar{f} - \theta| = 6.06$  ( $s.d. = 4.76$ )  $>$   $|\hat{\theta}_S^* - \theta| = 4.02$  ( $s.d. = 3.81$ ), paired  $t(23) = 3.43$ ,  $p = 0.002$ ), and outperformed the average forecast for 19 of the 24 coins. In the NS setting, the pivoted estimate  $\hat{\theta}_{NS}^*$  also had a significantly lower absolute error than  $\bar{f}$  in estimating  $\theta$  ( $|\bar{f} - \theta| = 10.25$  ( $s.d. = 7.78$ )  $>$   $|\hat{\theta}_{NS}^* - \theta| = 7.48$  ( $s.d. = 7.22$ ), paired  $t(23) = 3.52$ ,  $p = 0.002$ ), outperforming the average forecast for 18 of the 24 coins.

However, the pivoted estimates provided only a small improvement in the N

Table 4.2: Performance of different forecasting procedures in the different information settings using MAD and RMSE.

Information Setting	Forecasting Procedure					
	$\hat{\theta}_S^*$	$\hat{\theta}_{NS}^*$	$\hat{\theta}_N^*$	$\hat{\theta}_M$	$\bar{f}$	$f_i$
Symmetric:	4.02	4.86	3.33	3.76	6.06	10.84
Nested-Symmetric:	7.65	7.48	8.05	8.34	10.25	13.83
Nested:	7.72	8.04	7.63	7.68	7.91	10.78
<b>Over All Settings:</b>	<b>6.46</b>	<b>6.79</b>	<b>6.34</b>	<b>6.59</b>	<b>8.07</b>	<b>11.82</b>

Each entry above provides the mean of the absolute deviation (MAD) between the estimate  $\hat{\theta}$  and the true  $\theta$  across the set of 24 coins in that setting. The final row provides the average over all 72 coins considered.

Information Setting	Forecasting Procedure					
	$\hat{\theta}_S^*$	$\hat{\theta}_{NS}^*$	$\hat{\theta}_N^*$	$\hat{\theta}_M$	$\bar{f}$	$f_i$
Symmetric:	5.48	6.54	4.99	5.28	7.75	15.03
Nested-Symmetric:	10.21	10.29	10.42	10.74	12.77	17.98
Nested:	8.77	9.21	8.45	8.53	9.29	15.66
<b>Over All Settings:</b>	<b>8.39</b>	<b>8.82</b>	<b>8.26</b>	<b>8.48</b>	<b>10.13</b>	<b>16.27</b>

Each entry above provides the square root of the mean squared error (RMSE) between the estimate  $\hat{\theta}$  and the true  $\theta$  across the set of 24 coins in that setting. The final row provides the average over all 72 coins considered.

setting, reducing the average absolute error from  $|\bar{f} - \theta| = 7.91$  (*s.d.* = 4.99) to  $|\hat{\theta}_N^* - \theta| = 7.63$  (*s.d.* = 3.71, paired  $t(23) = 0.42$ , *n.s.*) and outperforming the average forecast for only 13 of the 24 coins. This lower performance is partially expected due to the inherently restricted information in this setting—there are only two total signals available to the forecasters, so even perfectly removing shared-information bias offers limited benefit in forecast accuracy, particularly when the proportion of sophisticates is high.

In practice, the decision analyst may not know the appropriate information set-



ting, so to test the robustness of the pivoting method, I studied the performance of each estimation procedure across all three settings. In addition, I considered the average performance of the individual forecasts  $f_i$  for each coin. The average absolute errors for each forecasting procedure are displayed in Table 4.2. First, note the unsurprising result that the simple wisdom-of-crowds offers a sizable improvement in accuracy. Over all settings, the average forecast reduced error by 32% relative to the individual forecasts.

Moreover, each of the pivoting procedures offered a significant further reduction in error relative to this average forecast, ranging from an additional 16% reduction in MAD for  $\hat{\theta}_{NS}^*$  to an additional 21% reduction in MAD for  $\hat{\theta}_N^*$  across all 72 coins studied. As discussed earlier, using the minimal pivoting procedure  $\hat{\theta}_M$  is recommended due to its simplicity and robustness. These results suggest that the minimal procedure, which overall provides an additional 18% error reduction in MAD and an additional 16% error reduction in RMSE relative to the average forecast, captures most of the benefits of pivoting while performing well in all of the information settings. This additional error reduction was largest (38% in MAD, 32% in RMSE) in the S setting, in which participants received the highest number of total signals, and smallest (3% in MAD, 8% in RMSE) in the N setting, in which participants received the lowest number of total signals.

Finally, I studied the improvement of the minimal pivoting method as a function of the realized shared-information bias<sup>1</sup> that was present for each coin, displayed in Figure 4.4. Lower, and sometimes negative, improvements are observed from the minimal pivoting method when the shared information bias was small, and higher improvements are observed when the shared-information bias was high. This pattern of results is exactly what would be expected from a method that is designed

---

<sup>1</sup> Shared-information bias was calculated according to  $|\mathbb{E}[\bar{f}|s] - \theta| = (1 - pw)|s - \theta|$ , with  $p = 1$  for the Symmetric setting.

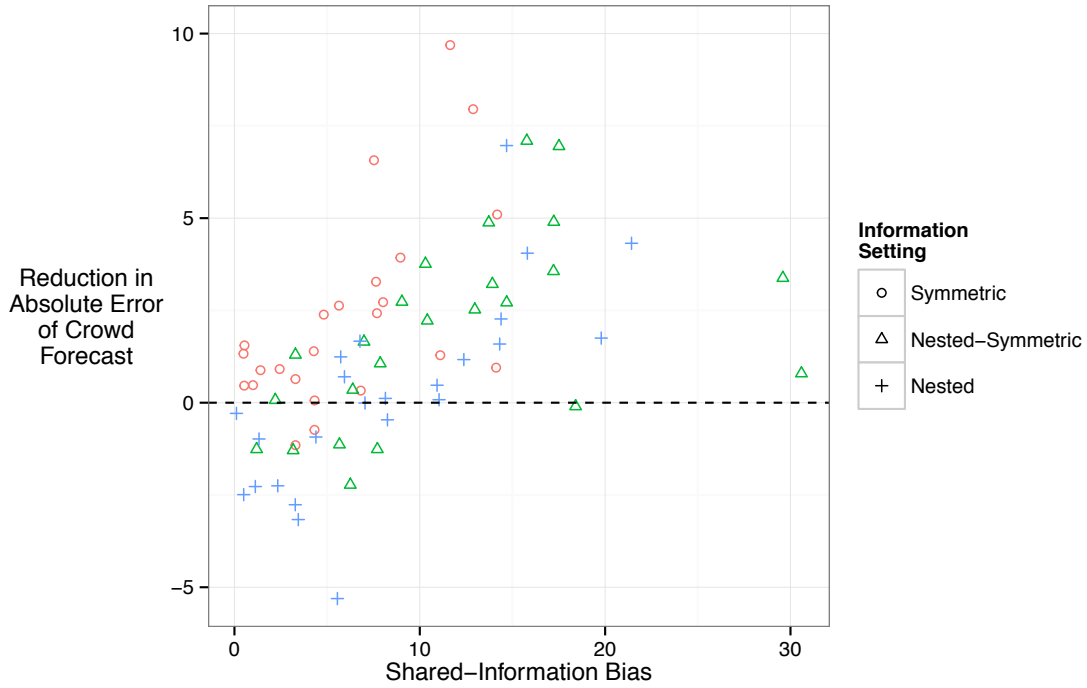


FIGURE 4.4: Reduction  $|\bar{f} - \theta| - |\hat{\theta}_M - \theta|$  in the absolute forecast error of the minimal pivoted forecast relative to the simple average forecast versus shared-information bias for all 72 coins across the three information settings.

to eliminate shared-information bias. When the shared signal is highly accurate with near-zero bias, pivoting cannot offer any improvement over averaging, and may even reduce accuracy in the crowd forecast due to noise introduced by its added complexity. However, as the shared-information bias grows larger, the pivoting procedure provides steadily increasing average benefits. The correlation between the improvement from minimal pivoting and the shared-information bias was 0.61 in the S setting, 0.50 in the NS setting, and 0.75 in the N setting. In practice, of course, the shared-information bias is not known to the decision analyst *a priori*. The data show that the pivoting method on average leads to improvement in all three information settings, with strongest performance when there is a high ratio of private to shared information. These results lend additional support to the robustness of the

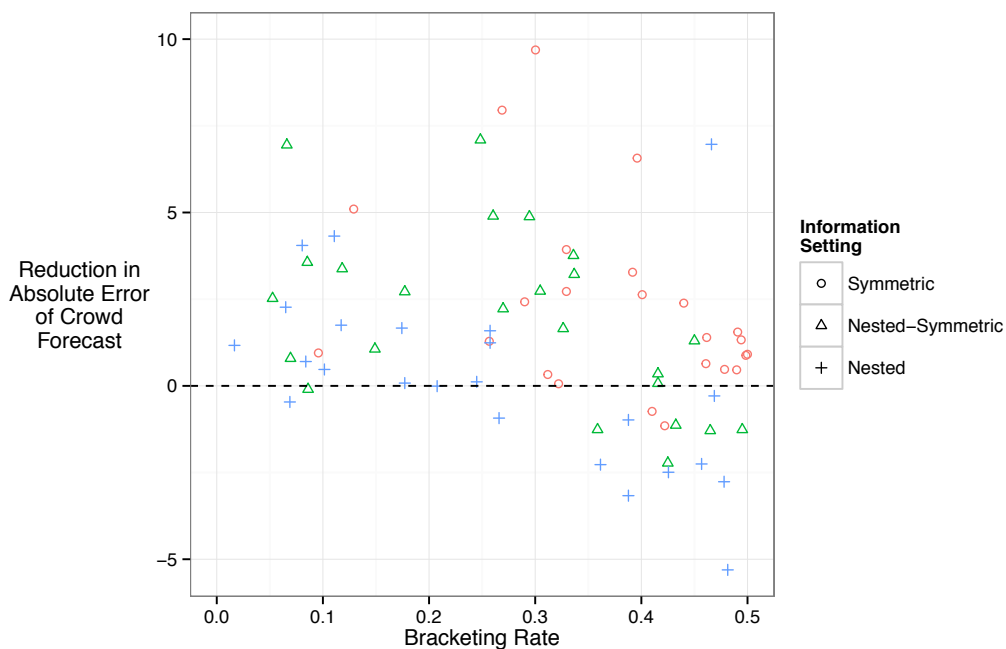


FIGURE 4.5: Reduction  $|\bar{f} - \theta| - |\hat{\theta}_M - \theta|$  in the absolute forecast error of the minimal pivoted forecast relative to the simple average forecast versus bracketing rate for all 72 coins across the three information settings.

simple minimal pivoting procedure  $\hat{\theta}_M$  in addressing shared-information bias across settings, and to the benefits of pivoting overall.

Bracketing (Larrick & Soll, 2006) can also provide some insight into the strong relationship between the effectiveness of pivoting and the size of the shared-information bias. Figure 4.5 displays the improvement from minimal pivoting versus the bracketing rate for each of the 72 coins. As expected, shared-information bias and bracketing were strongly negatively correlated ( $r = -0.74$ ). When the shared-information bias was low (less than 5), two randomly selected forecasters bracketed the truth 44% of the time (note that two independent judges can be expected to bracket 50% of the time). This implies that there were similar numbers of forecasts above and below the truth, and averaging eliminated a large proportion of the error. In

fact, simple averaging slightly outperformed minimal pivoting among these 24 coins ( $|\bar{f} - \theta| - |\hat{\theta}_M - \theta| = -0.34$ ). In contrast, when the shared-information bias was high (greater than 15), the bracketing rate dropped to 12%, leaving plenty of room for improving upon averaging, provided that pivoting adjusts the forecast in the correct direction. Minimal pivoting strongly outperformed the simple average among these 10 coins ( $|\bar{f} - \theta| - |\hat{\theta}_M - \theta| = 3.67$ ).

## 4.2 Study 2: Estimating Grocery Prices

The second study tested the pivoting method using real-world stimuli. I purchased 10 different bundles of nonperishable grocery items at a Target store near Duke University and recruited 49 volunteers passing through the student union to estimate the total price of the items in each bundle. Bundles were composed of three items, which were displayed on a table so that subjects had the opportunity to pick up and physically inspect the items while providing their answers. Participants were paid \$5 for answering the questions. This payment was not tied to their responses in any way, although they were asked to think carefully and provide their best responses. For each of the 10 bundles, participants were asked to first estimate the total cost to purchase the three items at regular price, and then to guess the average estimate that would be given by all of the other participants. All units were measured in U.S. dollars, and the actual prices of the bundles ranged from \$2.01 to \$33.21, with a mean price of \$15.00 and a median price of \$13.62.

For each of the 10 bundles, I compared individual forecasts  $f_i$ , the average forecast  $\bar{f}$ , and the estimates  $\hat{\theta}$  from the S, N, NS, and M pivoting procedures to the actual regular price  $\theta$  that was paid at the store. The mean absolute forecast error of a single individual ( $\sum_{i=1}^{49} |f_i - \theta|$ ) across the 10 questions was \$8.10.

The percentage improvement over the mean individual forecast error was 5.92%

for the average forecast  $\bar{f}$ . Each of the pivoting procedures yielded significantly better percentage improvements than the average forecast: 11.77% for the Minimal procedure  $\hat{\theta}_M$  (paired  $t(9) = 2.17, p = 0.058$ ), 13.34% for the S procedure  $\hat{\theta}_S^*$  (paired  $t(9) = 2.27, p = 0.049$ ), 14.25% for the N procedure  $\hat{\theta}_N^*$  (paired  $t(9) = 2.22, p = 0.054$ ), and 13.91% for the NS procedure  $\hat{\theta}_{NS}^*$  (paired  $t(9) = 2.19, p = 0.056$ ). Due to the small number of grocery bundles, I also ran Wilcoxon signed rank tests for each of the comparisons, and obtained similar results ( $p = 0.064$  for all comparisons). The pivoted estimates  $\hat{\theta}_M, \hat{\theta}_S^*, \hat{\theta}_N^*$ , and  $\hat{\theta}_{NS}^*$  each outperformed the average forecast  $\bar{f}$  on 7 out of the 10 questions.

When guessing what other forecasters would say, on average seventy percent ( $\hat{p} = 0.30$ ) of  $g_i$  were equal to the participant's own forecast  $f_i$ , and the average estimated weight on private information from the Symmetric information model was  $w^* = 0.77$ . In addition, the average forecast and average guess about others were generally not far apart ( $|\bar{f} - \bar{g}| = \$0.65, s.d. = \$0.54$ ). As a result, the pivoted estimates were fairly close to the simple average of participants' individual forecasts. Despite this, the small adjustments made to the average forecast still allowed the pivoting method to significantly outperform the average forecast, roughly doubling the improvement from simple averaging.

### 4.3 Study 3: Forecasting NCAA Men's Basketball Tournament Games

In the third study, I tested the pivoting procedures in a discrete-outcome setting by asking participants to forecast the probability that each team would win across 120 different games in the early rounds of the 2014, 2015, and 2016 NCAA Division I Men's Basketball Tournaments. Participants were recruited through Clearvoice in 2014 and through Amazon Mechanical Turk in 2015 and 2016 and were invited to participate in a web survey in exchange for a payment of \$0.50. To screen for

interest, subjects were initially asked about their interest in college basketball and informed that they would be predicting the outcomes of upcoming games in the NCAA tournament. Participants were presented with the two teams that would be playing in each game, and were instructed to pick a winner and then assign a probability on a scale from 50% to 100% that their chosen team would win. After providing this forecast, they were then asked to estimate the average probability of winning (on a scale from 0% to 100%) that other participants would assign to their chosen team. They were also informed that in addition to their payment for participation, they could earn a bonus of up to \$2 based on the accuracy of their forecasts. These payments were based on a rescaled quadratic scoring rule, which was displayed graphically above each game they were asked to predict. At the end of the survey, one game was random selected, and participants were based on the outcome of the game and their forecast for that game. Participants were not paid based on the accuracy of their guesses of other survey takers. An example screenshot from the study is displayed in Figure 4.6.

Altogether, a total of 712 individuals participated in the study. For those forecasting Round 2 of the NCAA tournament (the “Round of 64”), the four brackets of the tournament were divided among two sets of participants, such that each participant predicted 16 of the games. Participants forecasting Round 4 of the tournament (the “Round of 16”) each predicted all 8 games in that round. Because the data were collected over time, the number of participants varied for different sets of games:  $n = 73$  for games 1-16 (Round 2, 2014);  $n = 73$  for games 17-32 (Round 2, 2014);  $n = 164$  for games 33-40 (Round 4, 2014);  $n = 52$  for games 41-56 (Round 2, 2015);  $n = 48$  for games 57-72 (Round 2, 2015);  $n = 101$  for games 73-80 (Round 4, 2015);  $n = 65$  for games 81-96 (Round 2, 2016);  $n = 56$  for games 97-112 (Round 2, 2016);  $n = 80$  for games 113-120 (Round 4, 2016). Participants spent an average of 9.9 minutes completing the survey and received an average bonus of \$1.54.

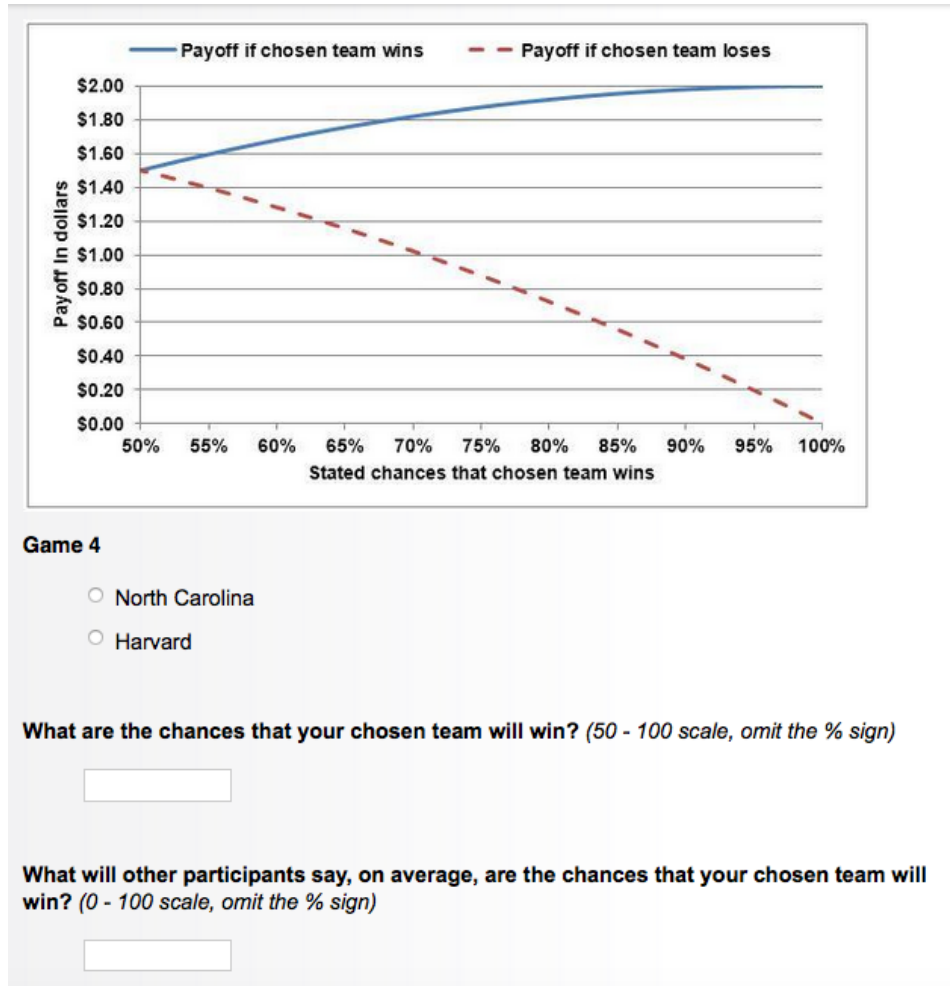


FIGURE 4.6: Example of the experimental interface for Study 3.

To measure the accuracy of the forecasts, I computed Brier scores (Brier, 1950) for different forecasting methods  $y$  (individual or crowd) across each of the 120 games.<sup>2</sup> The Brier score for each game  $r$  was calculated by comparing a particular forecast  $y_r$  of team 1 winning to the indicator variable  $X_r$  for whether team 1 won the game or not according to the formula  $(y_r - X_r)^2$ . Lower Brier scores indicate more accurate forecasts, with a score of 0.25 corresponding to a “know-nothing” forecaster who

<sup>2</sup> For simplicity, I avoided applying any corrections (such as Winsorizing) that would restrict the pivoted estimates to  $[0, 1]$ . Although this could be an important consideration in other data, this would only have affected one of the  $\hat{\theta}_{NS}^*$  forecasts, which came out to 1.047.

reports 50% for every game. As expected, individual forecasts  $f_i$  (mean Brier score = 0.232,  $s.d.$  = 0.121) were improved significantly by combining responses across individuals (mean Brier score for  $\bar{f}$  = 0.189,  $s.d.$  = 0.122), illustrating the benefits of using the wisdom of crowds (paired  $t(119) = 41.9, p < 0.001$ ). Next, I compared the performance of the pivoted crowd estimates to this simple average crowd forecast. Although each of the pivoted forecasts showed improvement over the simple average forecast (mean Brier score = 0.184 for  $\hat{\theta}_S^*$ , 0.185 for  $\hat{\theta}_N^*$ , 0.183 for  $\hat{\theta}_{NS}^*$ , 0.186 for  $\hat{\theta}_M$ ), none of these differences were statistically significant. The estimates  $\hat{\theta}_S^*$ ,  $\hat{\theta}_N^*$ ,  $\hat{\theta}_{NS}^*$ , and  $\hat{\theta}_M$  were each closer to the realized game outcome (represented by an indicator variable) than the average forecast  $\bar{f}$  in 78 out of the 120 games, which is significantly more than would be expected by chance (two-tailed binomial test  $p = 0.001$ ).

Since the probability of each team winning a particular game is unknown, I additionally examined the performance of the *market probabilities*  $\kappa_r$  of team 1 winning in game  $r$ , calculated according to the decimal betting market odds on the night after participants provided their responses.<sup>3</sup>

Specifically, I collected the decimal odds for bets on each team winning from the websites 5Dimes, Bovada, topbet, BetOnline, MyBookie.ag, BetDSI, BookMaker, GT Bets, SportBet, SportsBetting.com, and RealBet, when available. For each game, there are two decimal odds quoted—one for betting on the event that team 1 will win, and the other for betting on the event that team 2 will win. There is typically a small spread between the probabilities implied by each of these odds, which is part

---

<sup>3</sup> The payment rules under decimal odds differ slightly from the fractional odds that are familiar to many people. Under decimal odds, the bettor pays his stake up front, permanently parting with that money. If the event he bets on occurs, he then receives the decimal odds times the stake (decimal odds are always  $\geq 1$ ). For example, if he bets \$2 on an event with decimal odds of 4:1 ( $o = 4$ ), it means that if the event does not occur he will have lost \$2, and if the event does occur he will have lost \$2 but will win back \$8, giving him net winnings of \$6. Therefore if the bettor thinks that the probability of the event is  $\kappa$ , his expected return is  $-1 + o \cdot \kappa$ . The probability that makes a risk-neutral agent indifferent between taking the bet or not can be calculated by setting this expected return equal to 0, yielding an implied market probability of  $\kappa = 1/o$ .



of the bookmaker’s profit. Averaging these two decimals to get the probability of team 1 winning based on website  $k$ ’s odds accounts for the implied probabilities for the two teams summing to a number greater than 1. Letting  $o_{1rk}$  be the decimal odds on team 1 and  $o_{2rk}$  be the decimal odds on team 2, I computed the implied probability  $\kappa_{rk}$  for each website  $k$  by taking the average of  $1/o_{1rk}$  and  $1 - 1/o_{2rk}$ , and then computed the market probabilities  $\kappa_r$  of team 1 winning by averaging these implied probabilities  $\kappa_{rk}$  across all of the websites with available odds. As a result, one would expect these market probabilities to perform very well in predicting the game outcomes, and they do: the average Brier score obtained from using the market odds was 0.162 (*s.d.* = 0.190).

The performance of these forecasting methods can be further examined by “binning” them into discrete categories and decomposing the total Brier score into Reliability, Resolution, and Uncertainty (Murphy, 1973; Mason, 2004). To do this, forecasts must be sorted into categories that condense the full set of forecasts into a discrete set of probabilities that approximate the crowd’s estimate. I round each crowd forecast for each game to the nearest of 11 probabilities: 0%, 10%, . . . , 100%.<sup>4</sup> The total Brier score for each binned crowd forecasting method over the 120 games can then be calculated according to

$$BS = \frac{1}{120} \sum_{r=1}^{120} (y_r - X_r)^2 = \bar{X}(1 - \bar{X}) + \frac{1}{120} \sum_{c=1}^{11} n_c (p_c - \bar{X}_c)^2 - \frac{1}{120} \sum_{c=1}^{11} n_c (\bar{X}_c - \bar{X})^2,$$

where  $\bar{X} = \frac{1}{120} \sum_{r=1}^{120} X_r = 0.692$  equals the observed proportion of games won by

team 1,<sup>5</sup>  $p_c$  is the probability forecast for category  $c$ ,  $n_c$  is the number of crowd forecasts binned to  $p_c$ , and  $\bar{X}_c$  is the observed fraction of times team 1 won among all

<sup>4</sup> Specifically, all forecasts less than 5% are mapped to  $p_1 = 0\%$ , forecasts between 5% and 15% are mapped to  $p_2 = 10\%$ , . . . , and forecasts greater than 95% are mapped to  $p_{11} = 100\%$ .

<sup>5</sup> The fact that this winning proportion is significantly higher than 0.5 is not surprising, since team 1 was typically favored due to the seeding format of the tournament.

Table 4.3: Brier score decomposition for the three forecasting methods using the discrete forecast categories 0%, 10%, ..., 100%.

Information Setting	Forecasting Procedure		
	$\kappa$	$\hat{\theta}_M$	$\bar{f}$
Uncertainty:	0.213	0.213	0.213
Reliability:	0.005	0.013	0.012
Resolution:	0.060	0.046	0.035
<b>Total Brier Score:</b>	<b>0.159</b>	<b>0.181</b>	<b>0.190</b>
<b>Brier Skill Score:</b>	<b>0.256</b>	<b>0.151</b>	<b>0.108</b>

The total Brier score equals Uncertainty plus Reliability minus Resolution. The Brier Skill Score equals (Resolution - Reliability)/Uncertainty. Note that the scores may not add exactly due to rounding.

games where the crowd forecast was binned to  $p_c$ . These calculations were performed separately for each of three crowd forecasts—the market probabilities, the minimal pivoted estimate, and the simple average forecast—and the results are displayed in Table 4.3.

The first component, which is called Uncertainty and remains the same at 0.213 for each method, measures the performance that would be expected from using the empirical baseline probability of 0.692 as the forecast for all games. The second component, known as Reliability, measures the calibration of the probability forecasts relative to the empirical hit rate. The market probabilities display the highest reliability here (the score is closest to zero), and the average forecast is slightly more reliable than minimal pivoting. The negative of the third component is called Resolution, and measures how well the forecasts distinguish between events with different chances of occurring. As with the reliability component, the market probabilities offer the highest resolution, followed closely here by minimal pivoting. These com-

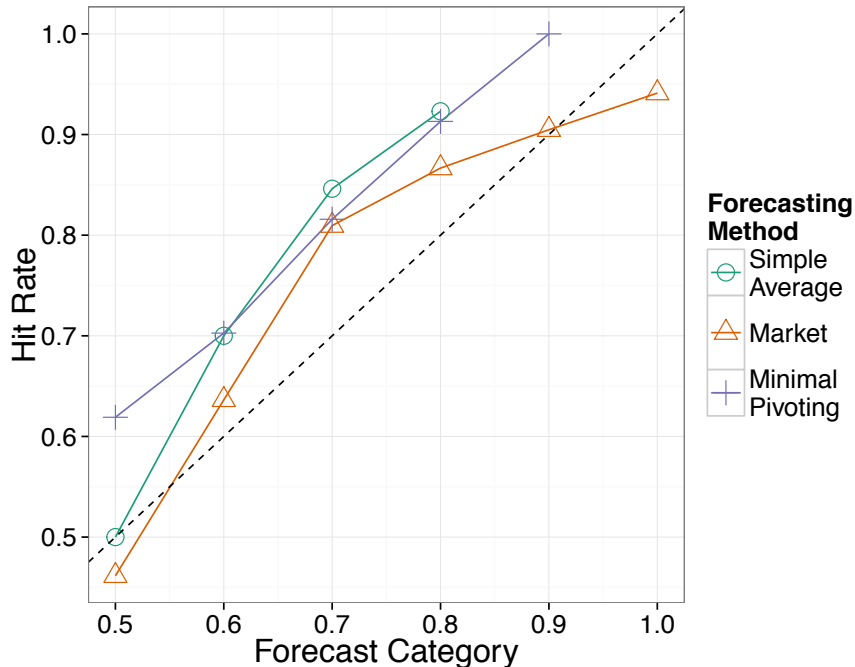


FIGURE 4.7: Calibration curves (hit rate versus forecast category) for the three forecasting methods.

ponents can then also be used to compute a skill score (Murphy & Winkler, 1992) for each crowd forecasting method, using the uncertainty component as a benchmark (Mason, 2004). Multiplying this skill score by 100 gives the percentage improvement in total Brier score relative to using the baseline probability for each game. Minimal pivoting offers an improvement of more than 15% here, higher than the almost 11% benefit from simple averaging, while the market probabilities provide the largest improvement of more than 25% over the baseline uncertainty.

Additionally, the binned crowd forecasts can be compared to the corresponding game outcomes to examine the calibration of the predictions. Specifically, I construct calibration curves that compare the probability forecast  $p_c$  for each category against the observed “hit rate”  $\bar{X}_c$  for that category—the fraction of times that team 1 won in games when the crowd forecasted a probability of  $p_c$  that team 1 would win the game. Note that the game outcomes are symmetric here, since participants

chose which team they favored to win and gave forecasts for that event, so crowd forecasts can be equivalently expressed in terms of the probability that the favored team will win, with outcomes renamed accordingly. Doing so allows the calibration curves to be condensed from the full set of 11 categories 0%, ..., 100% to the six categories 50%, ..., 100% and increases the sample size for estimating the hit rate in each category. Figure 4.7 displays these hit rates for each of three crowd forecasting methods. Both minimal pivoting and simple averaging tend to underestimate the probability that the favored team will win, which results in their higher (worse) reliability scores. The lower resolution of the simple average can also be seen in Figure 4.7, with no forecasts categories above 0.8, whereas minimal pivoting includes some forecasts of 0.9 and the market probabilities cover the full range of forecast categories.

It is clear from the Brier scores that the implied market probabilities provide the best forecasts for these games. This is not surprising, because they represent the aggregate views of thousands of highly motivated bettors who have placed significant sums of money on the line. There is good reason to believe that participants in these betting markets are more interested in the games and possess more information than participants in the study. As a result, the market probabilities provide a natural benchmark to measure the accuracy of other forecasts in the absence of objective probabilities for each game. Market probabilities may not be available in many other forecasting settings, and a decision analyst might need to choose between other forecasting procedures such as simple averaging and minimal pivoting. These alternatives can be compared based on how close they come to this benchmark  $\kappa$ . As before, individual forecasts are significantly improved by averaging ( $|f_i - \kappa| = 0.198 > |\bar{f} - \kappa| = 0.125$ , paired  $t(119) = 17.1, p < 0.001$ ). The pivoting procedures bring the crowd forecasts further toward the market probabilities than the simple average. The mean difference  $|\hat{\theta}_M - \kappa| = 0.116$  between the minimal estimates and

the market probabilities was significantly lower than the mean difference between the average forecast and the market probabilities (paired  $t(119) = 3.57, p < 0.001$ ). This difference was also significantly lower for the N estimates,  $|\hat{\theta}_N^* - \kappa| = 0.112$  (paired  $t(119) = 3.41, p < 0.001$ ), but was not significantly lower for the S estimates,  $|\hat{\theta}_S^* - \kappa| = 0.116$  (paired  $t(119) = 1.80, p = 0.075$ ), or the NS estimates,  $|\hat{\theta}_{NS}^* - \kappa| = 0.119$  (paired  $t(119) = 1.05, p = 0.298$ ). These results replicate the finding that minimal pivoting can improve upon the average forecast, and moreover can be extended to the aggregation of probabilities.

## Discussion

The fact that people rely on a combination of shared and private information poses a major challenge to the aggregation of forecasts. It is well established that one of the most effective ways to combine forecasts is to average responses. However, shared information induces correlation in forecast errors, which severely limits the ability to reduce error through averaging (Clemen & Winkler, 1985). For any given question, shared information will typically bias all forecasts in the same direction. This shared-information bias cannot be removed by averaging. I discussed this problem in the context of several information settings. Whereas past research has focused on a Symmetric setting in which each forecaster has both shared and private information, I have additionally studied two more settings—Nested and Nested-Symmetric. In both of these settings, some forecasters are *naïfs* who have access to only shared information, and others are *sophisticates* who have access to more information (which is private in the NS setting, and shared among sophisticates in the N setting).

For each setting, I developed a specialized pivoting procedure that under idealized conditions removes the entire shared-information bias. In the initial analytical exploration of the problem, I assumed that the decision analyst knows both the in-

formation setting and the weight  $w$  that forecasters put on private information (and the fraction of sophisticates  $p$  in the N and NS settings). I then generalized this to the case where these parameters are unknown and must be inferred from forecasters' judgments. The procedure involves asking individuals for both a forecast of the variable of interest and an estimate of what other forecasters will say on average. In principle, forecasters should weight shared information more heavily in predicting others than in predicting the variable itself. I showed that this difference can be exploited by starting at the mean forecast  $\bar{f}$  and pivoting away, by an appropriate amount, from the mean forecast of others  $\bar{g}$ . The logic behind pivoting is that the shared information contained in  $\bar{g}$  is overweighted in  $\bar{f}$  because it is part of everyone's forecast. Pivoting away from  $\bar{f}$  thus reduces the influence of the shared information to an appropriate level.

In practice, the implementation of pivoting is complicated by error-prone forecasters, and potentially imperfect diagnosis of the information setting. I addressed the first problem by proposing a behavioral model in each setting that accounts for judgmental error. By regressing the individual responses  $g_i$  on  $f_i$ , the decision analyst can recover the weight that forecasters put on private versus shared information, which is one of the inputs in determining how much to pivot. I address the problem of potentially mischaracterizing the information setting by offering a fourth procedure, minimal pivoting, that makes no assumptions about the parameter values or which of the three settings (S, N, or NS) best describes how information is distributed across forecasters. This procedure entails less correction than the others, and pivots by an amount equal to the distance between  $\bar{f}$  and  $\bar{g}$ , such that  $\hat{\theta}_M = 2\bar{f} - \bar{g}$ .

The three specialized pivoting procedures and minimal pivoting were tested in three studies. In Study 1 the settings were specifically constructed to match each of the stylized information models, so that in principle the corresponding specialized pivoting procedure should perform the best. Surprisingly, the most accurate proce-

cedure was not necessarily the one tailor-made for a given setting. This may be due to imprecision in estimating the parameters of the information setting and heterogeneity in how participants used the information that was available to them. Apparently, although pivoting can improve accuracy above and beyond simple averaging, it is less important to employ the theoretically optimal procedure. In fact, since some procedures pivot more than others, there is a significant risk of overcorrecting for shared information if the incorrect procedure is used. Minimal pivoting is recommended because it carries the least risk of overcorrecting, is simple to implement, and achieves similar performance relative to the other procedures. In each of the three studies, minimal pivoting outperformed the simple average.

## 5.1 Pivoting as a Solution to Shared-Information Bias

Pivoting may seem like an unlikely approach to aggregation, because it appears to contradict the common wisdom that an aggregate answer should remain inside the range of the individual answers. Ideally, forecasts will bracket the truth by falling on both sides of it, and thus error can be canceled out by averaging (Larrick & Soll, 2006). Bracketing is more likely to the extent that forecasters have different perspectives, rely on different information, and form judgments independently to avoid anchoring (Soll & Larrick, 2009). One approach to extracting more crowd wisdom, therefore, has been to focus on selection methods that recruit sufficient diversity of thought and group processes that facilitate independent judgment (Larrick, Mannes, & Soll, 2012). However, it may not always be possible to assemble a crowd with sufficient diversity and private information to negate the shared-information bias. In such low-bracketing situations averaging will make little headway in reducing error compared to the average individual. Pivoting can help here. Responses to the additional question about the opinions of others should point to the location of the shared information, which can be exploited by a pivoting algorithm.



## 5.2 Limitations of the Method

One drawback of the pivoting method is the added complexity of asking forecasters to guess what others will say. This question is not commonly asked, so answering may impose a significant cognitive burden on forecasters and slow the elicitation process. In addition, without these additional responses, a decision analyst cannot apply the pivoting method to data that has already been gathered. This means that the decision analyst must play an active role in the initial elicitation process to ensure that both responses are collected. Also, the pivoting method can only be applied to responses from human forecasters who can estimate how others will behave, and cannot be used on algorithmic responses that are generated automatically from a statistical model.

Pivoting should not be expected to yield improvement in every instance. It may be worse than simple averaging when the shared information happens by chance to be very accurate (this explains the left side of Figure 4.4). It also cannot be expected to outperform simple averaging when all information is either private or shared. However, in these cases pivoting would be expected to perform about as well as averaging.<sup>1</sup> When all information is shared, forecasters are likely to predict that others will give the same forecast as themselves. Even if they do deviate, it is unlikely to systematically be in a given direction. Thus  $\bar{f}$  and  $\bar{g}$  will be close together, and pivoting will produce a forecast very similar to  $\bar{f}$ . The same is likely to happen when all information is private. Thus, in situations where pivoting is unlikely to help, the method is unlikely to result in much pivoting anyway. As long as forecasters use their information to provide reasonably sensible estimates of both the variable of interest

<sup>1</sup> Another potential concern not considered in the formal models is the possibility of systematic bias in the forecasts. For example, a particular forecaster may be known to have a tendency to overestimate the variable of interest. If the decision analyst is aware of this, she may consider adjusting his responses accordingly before using them, so that his expected forecast error equals zero. However, in the absence of strong prior beliefs about a forecaster's errors, modeling information and individual forecasts as unbiased seems reasonable.

and what others will say, pivoting can correct for shared-information bias when it exists to improve aggregate forecasts.

Finally, it is possible that people may not be able to distinguish between private and shared information, or may only be able to make limited distinctions between what is shared and what is unique to themselves. To the extent that this is true, the average guess about others will be similar to the average forecast, and pivoting will give a similar result to simple averaging.

### 5.3 Additional Directions

Although the work presented so far offers important insights into addressing shared-information bias in a crowd forecast, a number of important questions remain open to future study.

#### *5.3.1 More General Information Models*

First, the models studied so far impose very specific assumptions on the statistical generating processes for signals and how this information may be distributed and shared across forecasters. These restrictions are useful for analytical tractability, but may be overly strong assumptions in practice. In Appendix B, I study two additional natural conjugate Bayesian models—a beta-binomial model and a gamma-gamma model—and show that the pivoting results continue to hold under these alternative distributions. The results appear to be driven by the linear structure of information, suggesting that pivoting will be effective in cases where posterior estimates are a convex combination of the signals observed. Although this is not true for all situations,<sup>2</sup> the settings where averaging forecasts is helpful are precisely

---

<sup>2</sup> For example, suppose the decision analyst wishes to estimate the upper bound of a uniformly distributed random variable, and forecasters observe random observations from the distribution. It would be reasonable to model inferences according to a Uniform-Pareto natural conjugate process, in which case the sufficient statistics and Bayesian updating are not linear.

the situations where pivoting should be effective as well.

In the latter part of Appendix B, I sketch how response behavior and the optimal aggregation procedure changes under a more general model of how information may be scattered across forecasters. As might be expected, calculating the optimal responses and aggregation procedure is significantly more complex, imposing an unrealistic burden on both the forecasters and the decision analyst. As a result, these more complex models do not appear to offer any additional practical insights for a decision analyst. The simplifying assumptions imposed in the S, N, and NS models presented previously strike a compromise between finely capturing all of the details of information overlap between forecasters and providing a parsimonious aggregation procedure that can be implemented without prior knowledge on the part of the decision analyst and allows for the possibility of noisy response behavior.

### *5.3.2 Estimating a Full Distribution for $X$*

Next, in many decision problems, the optimal choice depends not just on a single parameter  $\theta$ , but on the full distribution of the random variable  $X$ . In the main model of this dissertation, this issue is sidestepped by assuming that the decision analyst knows that  $X$  is normally distributed with known precision  $\lambda$ , and the only unknown is the mean  $\theta$ . In practice, the decision analyst may need to estimate the variance  $1/\lambda$  directly from the responses. For example, in the case of Sport Obermeyer, Ltd. (Hammond & Raman, 1994; Fisher & Raman, 1996), a set of  $n$  managers each provide a point forecast  $f_i$  of next year's demand for a new style of jacket. Demand is assumed to be normally distributed with unknown mean and standard deviation. The mean is estimated by taking the average of forecasts  $f_i$ , and the standard deviation is estimated by taking the standard deviation of the forecasts  $f_i$  and multiplying by two. This estimation procedure—using variance in individual forecasts as a cue to the variance of the underlying variable of interest—is consistent

with the Symmetric model presented in Chapter 3 with no error terms ( $\varepsilon_i = 0$ ). In this case, the variance of the individual forecasts can be expressed as a multiple of the variance of  $X$ . In a repeated forecasting context, the decision analyst could estimate this multiple using past responses and the error variance of the crowd forecast. In the case of Sport Obermeyer, for example, I would recommend estimating the mean using the minimal pivoting procedure  $\hat{\theta}_M$  and continuing to estimate the standard deviation by doubling the standard deviation of the forecasts.

However, this method is not feasible without sufficient past data, and the assumption of a specific distribution family may be restrictive. An alternative approach is to elicit forecasts  $f_{ik}$  of the cumulative probability  $p_k = \mathbb{P}(X \leq x_k)$  at a discrete set of outcome values  $x_k$ ,  $k = 1, \dots, K$ , and for the same questions also ask for a guess  $g_{ik}$  of the average response that will be given by others. The crowd estimate  $\hat{p}_k$  of the cumulative probabilities at each point  $x_k$  can then be calculated by using the minimal pivoting procedure, with a restriction to keep the estimates in  $[0, 1]$  as discussed in Appendix B. These probability estimates  $\hat{p}_k$  can then be interpolated to provide a piecewise linear approximation of the cumulative distribution function  $F(x)$ .<sup>3</sup>

### 5.3.3 Further Empirical Evidence

Finally, it would be useful to examine the performance of the minimal pivoting procedure in different contexts in order to more precisely understand when it offers benefits, and when and why it may not be effective. One question of particular interest is whether pivoting could be useful in conjunction with a market. For example, could pivoting be used to correct a market bubble by helping to distinguish

---

<sup>3</sup> An analogous approach would be to choose a discrete set of probabilities and ask forecasters to estimate the corresponding *quantiles* of the distribution as well as guess the average quantiles given by others. Although not supported directly by a formal information model, the underlying ideas about shared information suggest that applying the minimal pivoting procedure to these responses could also be reasonable.

private information from an overoptimistic public signal? Likewise, could pivoting be useful in counteracting information cascades or herding behavior in forecasting? Another question is how pivoting compares to other approaches such as winner-take-all contests (Lichtendahl, Grushka-Cockayne, & Pfeifer, 2013), which have appealing theoretical properties but have not been tested empirically. Additional data could help to answer these questions and provide more evidence on the effectiveness of the approach.

## 5.4 Conclusion

I have developed a novel method for dealing with shared-information bias in the aggregation of forecasts. By asking forecasters to also estimate the average forecast of others, it is possible to separate out shared and private signals, and then adjust for the fact that the simple average overweights shared information by pivoting off the mean forecast in a direction away from the shared signal. Although tailored formulations can be constructed for different information settings, I recommend a simple minimal pivoting procedure that achieves most of the benefits without a high risk of pivoting too far.

Overcoming shared-information bias is difficult, and the headway in addressing it to date has been primarily theoretical. The bias greatly limits the extent to which aggregating forecasts can reduce error, so the potential rewards from mitigating it are substantial. Future research should develop new approaches, refine existing ones, and seek to identify what is most effective. The results I have reported for pivoting suggest that it is a worthy contender.

# Appendix A

## Proofs

*Proof of Proposition 1:* Forecaster  $i$ 's optimization problem is

$$\max_{f,g} \mathbb{E}[S_1(X, f) + S_2(\sum_{j \neq i} \frac{f_j}{n-1}, g) \mid s, t_i].$$

Taking a first-order condition with respect to  $f$  yields  $\frac{\partial}{\partial f} \mathbb{E}[S_1(X, f) \mid s, t_i] = 0$ , which by properness of  $S_1$  yields a critical point of  $f^* = \mathbb{E}[X \mid s, t_i]$ . Likewise, the first-order condition with respect to  $g$  yields  $\frac{\partial}{\partial g} \mathbb{E}[S_2(\sum_{j \neq i} \frac{f_j}{n-1}, g) \mid s, t_i] = 0$ , which by properness of  $S_2$  yields a critical point of  $g^* = \mathbb{E}[\sum_{j \neq i} \frac{f_j}{n-1} \mid s, t_i] = \mathbb{E}[f_j \mid s, t_i]$ .

Taking the mean of the expert's posterior predictive distribution for  $X$  given his signals,  $f^* = (1-w)s + wt_i$ . If all other forecasters  $j$  follow this reporting strategy, then  $g^* = \mathbb{E}[f_j \mid s, t_i] = (1-w)s + w\mathbb{E}[t_j \mid s, t_i] = (1-w)s + w\mathbb{E}[\theta \mid s, t_i] = (1-w)s + w((1-w)s + wt_i) = (1-w)^2s + w^2t_i$ . Observe that  $\frac{\partial^2}{\partial f^2} \mathbb{E}[S_1(X, f)]|_{f=f^*} < 0$  and  $\frac{\partial^2}{\partial f^2} \mathbb{E}[\sum_{j \neq i} \frac{f_j}{n-1} \mid s, t_i]|_{g=g^*} < 0$  by strict properness of  $S_1$  and  $S_2$ . In addition,  $\frac{\partial^2}{\partial f \partial g} \mathbb{E}[S_1(X, f)] = 0$ , so the determinant of the Hessian is positive and  $(f^*, g^*)$  is

maximal.  $\square$

*Proof of Proposition 2:* Using equations (3.1),  $\bar{f} = (1 - w)s + w \sum_{i=1}^n \frac{t_i}{n}$  and  $\bar{g} = (1 - w^2)s + w^2 \sum_{i=1}^n \frac{t_i}{n}$ . Multiplying the first equation by  $w$  and subtracting, the shared signal is given by  $s = \frac{\bar{g} - w\bar{f}}{1 - w}$ . Substituting into the above expression for  $\bar{f}$  and simplifying yields equations (3.2) and (3.3). Now consider a general linear weighting of  $\hat{s}$  and  $\hat{t}$  given by  $\hat{\theta}_{y,z} = y\hat{t} + z\hat{s}$ . This can be written as  $\hat{\theta}_{y,z} = (\frac{1+w}{w}y - \frac{w}{1-w}z)\bar{f} + (\frac{1}{1-w}z - \frac{1}{w}y)\bar{g} = y \sum_{i=1}^n \frac{t_i}{n} + zs$ . Since both  $s$  and  $\sum_{i=1}^n \frac{t_i}{n}$  have mean  $\theta$ , we must have  $z = 1 - y$  in order for  $\hat{\theta}_{y,z}$  to be unbiased. Conditional on  $\theta$ ,  $s$  and  $\sum_{i=1}^n \frac{t_i}{n}$  are independent and normal with variance  $1/m\lambda$  and  $1/n\ell\lambda$ , respectively. The variance of  $\hat{\theta}_{y,1-y}$  is therefore  $\frac{y^2}{n\ell\lambda} + \frac{(1-y)^2}{m\lambda}$ , which is minimized by setting  $0 = \frac{2y}{n\ell\lambda} + \frac{-2+2y}{m\lambda}$ , or  $y^* = \frac{n\ell}{m+n\ell}$  (and  $z^* = 1 - y^* = \frac{m}{m+n\ell}$ ). Noting that  $w = \frac{\ell}{m+\ell}$ , we can restate the expression for the optimal linear weight  $y^*$  in terms of  $w$  as  $\alpha = y^* = \frac{nw}{nw+1-w}$ .  $\square$

*Proof of Proposition 3:* Conditional on  $\theta$ ,  $\bar{f} = (1 - w)s + w \sum_{i=1}^n \frac{t_i}{n} + \sum_{i=1}^n \frac{\varepsilon_i}{n}$  is the sum of independent variables. The relative information weight on each private signal relative to the shared signal is  $w = \frac{\ell}{m+\ell}$ ,  $s$  is normal with mean  $\theta$  and variance  $1/m\lambda$  by definition,  $\sum_{i=1}^n \frac{t_i}{n}$  is normal with mean  $\theta$  and variance  $1/n\ell\lambda$ , and  $\sum_{i=1}^n \frac{\varepsilon_i}{n}$  has mean 0 and variance  $\frac{Var(\varepsilon)}{n}$ . Therefore  $\bar{f}$  has a mean of  $(1 - w)\theta + w\theta + 0 = \theta$  and variance  $Var(\bar{f}) = (1 - w)^2 \frac{1}{m\lambda} + w^2 \frac{1}{n\ell\lambda} + \frac{Var(\varepsilon)}{n} = \frac{nm+\ell}{n(m+\ell)^2\lambda} + \frac{Var(\varepsilon)}{n}$ . As  $n \rightarrow \infty$ ,  $Var(\bar{f}) \rightarrow \frac{nm+\ell}{n(m+\ell)^2\lambda}$ , so the expected squared error  $\mathbb{E}[(\bar{f} - \theta)^2]$  from using the average forecast as the crowd grows large is always greater than or equal

to  $\lim_{n \rightarrow \infty} \text{Var}(\bar{f}) = \frac{m}{(m + \ell)^2 \lambda}$ . Now consider the pivoting method of Section 3.1.2:

The aggregate forecast can be written simply as  $\hat{\theta}_S = \frac{nw-w+n}{nw-w+1} \bar{f} + \frac{1-n}{nw-w+1} \bar{g} = \frac{1-w}{nw-w+1} s + \frac{nw}{nw-w+1} \sum_{i=1}^n \frac{t_i}{n} + \frac{n}{nw-w+1} \sum_{i=1}^n \frac{\varepsilon_i}{n} + \frac{1-n}{nw-w+1} \sum_{i=1}^n \frac{\gamma_i}{n}$ , so  $\hat{\theta}_S$  is normally distributed with mean  $\frac{nw-w+n}{nw-w+1} \theta + \frac{1-n}{nw-w+1} \theta = \theta$  and variance  $\text{Var}(\hat{\theta}_S) = \left(\frac{1-w}{nw-w+1}\right)^2 \frac{1}{m\lambda} + \left(\frac{nw}{nw-w+1}\right)^2 \frac{1}{n\ell\lambda} + \left(\frac{n}{nw-w+1}\right)^2 \frac{\text{Var}(\varepsilon)}{n} + \left(\frac{1-n}{nw-w+1}\right)^2 \frac{\text{Var}(\gamma)}{n}$ . As  $n \rightarrow \infty$ ,  $\text{Var}(\hat{\theta}_S) \rightarrow \frac{1}{n\ell\lambda} + \frac{1}{nw^2} (\text{Var}(\varepsilon) + \text{Var}(\gamma))$ , and the expected squared error  $\mathbb{E}[(\hat{\theta}_S - \theta)^2]$  as the crowd grows large is  $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_S) = 0$ .  $\square$

*Derivation of the General Estimation Procedure and Proof of Proposition 4:* Here I study the optimal aggregation procedure for the Nested and Nested-Symmetric settings, assuming as for the Symmetric setting in Proposition 2 that there is zero noise in responses. I then show for each setting that the Minimal pivoting procedure provides a lower variance than the average forecast in each of the three settings, even with noise in the responses.

In the Nested setting, optimal responses from the  $n - pn$  naïfs are given by  $f_i = s$  and  $g_i = s$ . Optimal responses from the  $pn$  sophisticates are given by  $f_i = (1 - w)s + wt$  and  $g_i = (1 - pw)s + pwt$ . Including noise in responses as specified in the behavioral

model, the average forecast is  $\bar{f} = (1 - pw)s + pwt + \sum_{i=1}^{n-pn} \frac{\delta_i}{n - pn} + \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn}$  and the

average guess of others is  $\bar{g} = (1 - p^2w)s + p^2wt + \sum_{i=1}^{n-pn} \frac{\delta_i}{n - pn} + p \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn} + p \sum_{i=1}^{pn} \frac{\gamma_i}{pn}$ .

The Nested crowd forecast is  $\hat{\theta}_N = \bar{f} + \frac{1}{p}(\bar{f} - \bar{g}) = (1 - w)s + wt + \sum_{i=1}^{n-pn} \frac{\delta_i}{n - pn} +$

$\sum_{i=1}^{pn} \frac{\varepsilon_i}{pn} - \sum_{i=1}^{pn} \frac{\gamma_i}{pn}$ . If the noise terms are zero, then  $\hat{\theta}_N$  provides the minimum-variance

weighting of the signals, with  $\text{Var}(\hat{\theta}_N) = \frac{1}{(m+\ell)\lambda}$ .



In the Nested-Symmetric setting, optimal responses from the  $n-pn$  naïfs are given by  $f_i = s$  and  $g_i = s$ . Optimal responses from the  $pn$  sophisticates are given by  $f_i = (1-w)s + wt_i$  and  $g_i = (1-pw^2)s + pw^2t_i$ . Including noise in responses as specified in the behavioral model, the average forecast is  $\bar{f} = (1-pw)s + pw \sum_{i=1}^{pn} \frac{t_i}{pn} + \sum_{i=1}^{n-pn} \frac{\delta_i}{n-pn} + \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn}$  and the average guess of others is  $\bar{g} = (1-p^2w^2)s + p^2w^2 \sum_{i=1}^{pn} \frac{t_i}{pn} + \sum_{i=1}^{n-pn} \frac{\delta_i}{n-pn} + p \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn} + p \sum_{i=1}^{pn} \frac{\gamma_i}{pn}$ . The Nested-Symmetric crowd forecast is  $\hat{\theta}_{NS} = \frac{npw}{npw+1-w}(\bar{f} + \frac{1}{pw}(\bar{f} - \bar{g})) + \frac{1-w}{npw+1-w}(\bar{f} + \frac{1}{1-pw}(\bar{g} - \bar{f})) = \frac{1-w}{npw+1-w}s + \frac{npw}{npw+1-w} \sum_{i=1}^{pn} \frac{t_i}{pn} + \sum_{i=1}^{n-pn} \frac{\delta_i}{n-pn} + \frac{n(1-p+pw)(1-pw) + p(1-w)^2}{(npw+1-w)(1-pw)} \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn} + \frac{(1-w)p - np(1-pw)}{(npw+1-w)(1-pw)} \sum_{i=1}^{pn} \frac{\gamma_i}{pn}$ . If the noise terms are zero, then  $\hat{\theta}_{NS}$  provides the minimum-variance weighting of the signals, with  $Var(\hat{\theta}_{NS}) = \frac{1}{(m+pn\ell)\lambda}$ .

Next, I use these results to prove Proposition 4: In the Symmetric setting,  $\hat{\theta}_M = 2\bar{f} - \bar{g} = (1-2w+w^2)s + (2w-w^2) \sum_{i=1}^n \frac{t_i}{n} + (2-w) \sum_{i=1}^n \frac{\varepsilon_i}{n} - \sum_{i=1}^n \frac{\gamma_i}{n}$ . Conditional on  $\theta$ , both  $\hat{\theta}_M$  and  $\bar{f}$  (as detailed in the proof of Proposition 3) can be expressed as a sum of independent variables with a mean of  $\theta$ . If  $w > 0$ , as  $n$  grows large the variance of all terms in these sums that do not include  $s$  decline to zero, and  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] = Var(\hat{\theta}_M) \rightarrow \frac{(1-w)^4}{m\lambda} < \mathbb{E}[(\bar{f} - \theta)^2] = Var(\bar{f}) \rightarrow \frac{(1-w)^2}{m\lambda}$ .

In the Nested setting,  $\hat{\theta}_M = 2\bar{f} - \bar{g} = (1-2pw+p^2w)s + (2pw-p^2w)t + \sum_{i=1}^{n-pn} \frac{\delta_i}{n-pn} + (2-p) \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn} - p \sum_{i=1}^{pn} \frac{\gamma_i}{pn}$ . Conditional on  $\theta$ , both  $\hat{\theta}_M$  and  $\bar{f}$  can be expressed as a sum of independent variables with a mean of  $\theta$ . Consider the function  $V(\alpha) = \alpha^2 Var(t) + (1-\alpha)^2 Var(s)$ , which provides the variance of the expression

$\alpha t + (1 - \alpha)s$ . Observe that  $V(\alpha)$  is a convex quadratic function of  $\alpha$  with a minimum at  $\alpha = w$ , as shown in the proof of Proposition 2. By showing that  $pw \leq pw(2 - p) \leq w$  it can therefore be established that  $V(pw) \geq V(2pw - p^2w) \geq V(w)$ . First,  $pw \leq pw(2 - p)$  since  $p \in [0, 1]$  and  $w \geq 0$ . Next, observe that  $p(2 - p)$  is a concave quadratic function of  $p$  that takes a maximum value of 1 at  $p = 1$ , so  $pw(2 - p) \leq w$  for  $p \in [0, 1]$  and  $w \geq 0$ . Then as  $n$  grows large,  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] = \text{Var}(\hat{\theta}_M) \rightarrow V(2pw - p^2w) \leq \mathbb{E}[(\bar{f} - \theta)^2] = \text{Var}(\bar{f}) \rightarrow V(pw)$ , with the inequality strict if  $p \in (0, 1)$  and  $w > 0$ .

In the Nested-Symmetric setting,  $\hat{\theta}_M = 2\bar{f} - \bar{g} = (1 - 2pw + p^2w^2)s + (2pw - p^2w^2) \sum_{i=1}^{pn} \frac{t_i}{pn} + \sum_{i=1}^{n-pn} \frac{\delta_i}{n - pn} + (2 - p) \sum_{i=1}^{pn} \frac{\varepsilon_i}{pn} - p \sum_{i=1}^{pn} \frac{\gamma_i}{pn}$ . Conditional on  $\theta$ , both  $\hat{\theta}_M$  and  $\bar{f}$  can be expressed as a sum of independent variables with a mean of  $\theta$ . If  $w, p > 0$ , as  $n$  grows large the variance of all terms in these sums that do not include  $s$  decline to zero, and  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] = \text{Var}(\hat{\theta}_M) \rightarrow \frac{(1-pw)^4}{m\lambda} < \mathbb{E}[(\bar{f} - \theta)^2] = \text{Var}(\bar{f}) \rightarrow \frac{(1-w)^2}{m\lambda}$ .  $\square$

*Proof of Proposition 5:* To derive the optimal hedged weight for estimating private signals, define the estimated average private signal when using a pivot of size  $h$  according to  $\bar{t}^h \equiv \bar{f} + \frac{1}{h}(\bar{f} - \bar{g})$ . The decision analyst wants to choose  $h$  to minimize the expected value of the squared error  $\mathcal{E}_{private}(h, w) = (\mathbb{E}[\bar{t}^h | \bar{t}, s, w] - \bar{t})^2$ , given her uncertainty about  $w$ . In other words, she wants to find the optimal hedged weight  $h_{private}^* = \arg \min_h \mathcal{D}_{private}(h)$ , where  $\mathcal{D}_{private}(h) = \mathbb{E}_w[\mathcal{E}_{private}(h, w) | \{(f_i, g_i)\}_{i=1}^n] = \int_0^1 \mathcal{E}_{private}(h, w) f(w | \{(f_i, g_i)\}_{i=1}^n) dw$ . Using the expressions for  $\bar{f}$  and  $\bar{g}$  from Proposition 2,  $\mathbb{E}[\bar{t}^h | \bar{t}, s, w] = \bar{f} + \frac{1}{h}(\bar{f} - \bar{g}) = (1 - w)s + w\bar{t} + \frac{w}{h}(1 - w)(\bar{t} - s)$ ,  $\mathcal{E}_{private}(h, w) = (\bar{t} - s)^2(1 - w)^2(1 - \frac{w}{h})^2$ , and the expression to be minimized as

$$\mathcal{D}_{private}(h) = (s - \bar{t})^2 \int_0^1 (1 - \frac{w}{h})^2 (1 - w)^2 f(w | \{(f_i, g_i)\}_{i=1}^n) dw.$$

The first derivative is  $\mathcal{D}'_{private}(h) = \frac{2}{h^2}(s - \bar{t})^2 \int_0^1 (1 - \frac{w}{h})w(1 - w)^2 f(w | \{(f_i, g_i)\}_{i=1}^n) dw$ .

The first-order condition then yields  $\int_0^1 \frac{w}{h} w(1-w)^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw = \int_0^1 w(1-w)^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw$ , or

$$h_{private}^* = \frac{\int_0^1 w^2(1-w)^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw}{\int_0^1 w(1-w)^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw}.$$

For  $h = 0$ ,  $\mathcal{D}_{private}(0)$  is undefined. For  $h \neq 0$ ,  $\mathcal{D}_{private}(h)$  is continuously differentiable with  $\lim_{\varepsilon \rightarrow 0} \mathcal{D}_{private}(\varepsilon) = \infty$  and  $\lim_{M \rightarrow \infty} \mathcal{D}_{private}(M) = \lim_{M \rightarrow \infty} \mathcal{D}_{private}(-M) = (s - \bar{t})^2 \int_0^1 (1-w)^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw$ . Observe that for  $h = 1$ ,  $(1 - \frac{w}{h})^2 = (1-w)^2 \leq 1$  for all  $w \in [0, 1]$ , so  $\mathcal{D}_{private}(1) \leq (s - \bar{t})^2 \int_0^1 (1-w)^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw$ . Since the only critical point is  $h_{private}^* > 0$ , it is straightforward to show that  $h_{private}^*$  is minimal.

To derive the optimal hedged weight for estimating shared signals, define the estimated shared signal when using a pivot of size  $h$  according to  $s^h \equiv \bar{f} + \frac{1}{1-h}(\bar{g} - \bar{f})$ . The decision analyst wants to choose  $h$  to minimize the expected value of the squared error  $\mathcal{E}_{shared}(h, w) = (\mathcal{E}[s^h | \bar{t}, s, w] - s)^2$ , given her uncertainty about  $w$ . In other words, she wants to find the optimal hedged weight  $h_{shared}^* = \arg \min_h \mathcal{D}_{shared}(h)$ , where  $\mathcal{D}_{shared}(h) = \mathbb{E}_w[\mathcal{E}_{shared}(h, w) | \{(f_i, g_i)\}_{i=1}^n] = \int_0^1 \mathcal{E}_{shared}(h, w) f(w|\{(f_i, g_i)\}_{i=1}^n) dw$ . Using the expressions for  $\bar{f}$  and  $\bar{g}$  from Proposition 2,  $\mathcal{E}[s^h | \bar{t}, s, w] = \frac{\bar{g} - h\bar{f}}{1-h} = s + w(\bar{t} - s) \frac{w-h}{1-h}$ ,  $\mathcal{E}_{shared}(h, w) = w^2(\bar{t} - s)^2 (\frac{w-h}{1-h})^2$ , and the expression to be minimized as

$$\mathcal{D}_{shared}(h) = (s - \bar{t})^2 \int_0^1 \left(\frac{w-h}{1-h}\right)^2 w^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw.$$

The first derivative is  $\mathcal{D}'_{shared}(h) = 2 \frac{(s-\bar{t})^2}{(1-h)^3} \int_0^1 (h-w)(1-w)w^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw$ .

The first-order condition then yields  $\int_0^1 h(1-w)w^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw = \int_0^1 (1-w)w^3 f(w|\{(f_i, g_i)\}_{i=1}^n) dw$ , or

$$h_{shared}^* = \frac{\int_0^1 (1-w)w^3 f(w|\{(f_i, g_i)\}_{i=1}^n) dw}{\int_0^1 (1-w)w^2 f(w|\{(f_i, g_i)\}_{i=1}^n) dw}.$$

For  $h = 1$ ,  $\mathcal{D}_{shared}(1)$  is undefined. For  $h \neq 1$ ,  $\mathcal{D}_{shared}(h)$  is continuously differentiable with  $\lim_{\varepsilon \rightarrow 0} \mathcal{D}_{shared}(1 + \varepsilon) = \infty$  and  $\lim_{M \rightarrow \infty} \mathcal{D}_{shared}(M) = \lim_{M \rightarrow \infty} \mathcal{D}_{shared}(-M) = (s - \bar{t})^2 \int_0^1 w^2 f(w | \{(f_i, g_i)\}_{i=1}^n) dw$ . Observe that for  $h = 0$ ,  $(\frac{w-h}{1-h})^2 = w^2 \leq 1$  for all  $w \in [0, 1]$ , so  $\mathcal{D}_{shared}(0) \leq (s - \bar{t})^2 \int_0^1 w^2 f(w | \{(f_i, g_i)\}_{i=1}^n) dw$ . Since the only critical point is  $h_{shared}^* < 1$ , it is straightforward to show that  $h_{shared}^*$  is minimal.  $\square$

# Appendix B

## Alternative Information Models

In this chapter, I consider individual behavior in two alternative natural conjugate statistical settings (Raiffa & Schlaifer, 1961) beyond the normal-normal information model studied in the main text. First, I consider a beta-binomial information setting, which is a natural model to study shared information when forecasters are predicting the probability of some event. Next, I use a gamma-gamma information model to extend the continuous forecasting task from a normal setting to consider potentially skewed distributions. For each statistical model, I present a Symmetric, Nested, and Nested-Symmetric version of how signals may be distributed across a crowd of forecasters, and show that the tailored pivoting procedures introduced previously allow the decision analyst to remove shared information bias in each case. In addition, I show that the minimal pivoting procedure maintains its status as a simple and robust approach under the alternative statistical models. Finally, I discuss more general models of how information might be distributed across forecasters.

## B.1 A Beta-Binomial Model

In this section, I study shared information in a setting where the decision analyst would like to estimate the probability of some event. Let  $X \in \{0, 1\}$  be the Bernoulli random variable indicating the event being forecast, which occurs randomly with probability  $\mathbb{P}(X = 1) = \theta$ . There are  $n$  forecasters, who share a diffuse prior belief about  $\theta$ , represented by an improper beta distribution  $\theta \sim \text{Beta}(0, 0)$ . All forecasters observe the same shared signal  $s$ , which is the average of  $m$  independent realizations  $X_k$  (the total number of 1's observed) from the Bernoulli process,  $s = \frac{1}{m} \sum_{k=1}^m X_k$ . In the Symmetric setting, each forecaster  $i = 1 \dots n$  also receives his own additional private signal  $t_i$ , which is the average of  $\ell$  independent realizations  $X_{ik}$  from the Bernoulli process,  $t_i = \frac{1}{\ell} \sum_{k=1}^{\ell} X_{ik}$ . In the Nested setting, a proportion  $1 - p$  of the forecasters are naïfs who only observe  $s$ , and the remaining proportion  $p$  of forecasters are sophisticates who observe both  $s$  and an additional signal  $t$ , which is the average of  $\ell$  independent realizations  $X_k$  from the Bernoulli process,  $t = \frac{1}{\ell} \sum_{k=1}^{\ell} X_k$ . In the Nested-Symmetric setting, the sophisticates observe both  $s$  and their own private signal  $t_i$ , which is the average of  $\ell$  independent realizations  $X_{ik}$  from the Bernoulli process,  $t_i = \frac{1}{\ell} \sum_{k=1}^{\ell} X_{ik}$ . All signals ( $s$ ,  $t_i$ , and  $t$ ) received by the forecasters are independent conditional on  $\theta$ . This information structure and the parameters  $m$  and  $\ell$  are commonly known among all forecasters, and the proportion  $p$  is known by sophisticates.

### B.1.1 Symmetric Setting

Forecaster  $i$ 's posterior beliefs about  $\theta$  are given by  $(\theta \mid s, t_i) \sim \text{Beta}(ms + \ell t_i, m(1 - s) + \ell(1 - t_i))$ , and his posterior predictive distribution for  $X$  is given by  $\mathbb{P}(X = 1 \mid s, t_i) = \mathbb{E}[X \mid s, t_i] = \frac{ms + \ell t_i}{m + \ell}$ . Conditional on his own information  $\{s, t_i\}$ , forecaster  $i$ 's expectation of the signals  $t_j$ ,  $j \neq i$ , observed by others can be calculated according

to  $\mathbb{E}[t_j | s, t_i] = \int_0^1 \mathbb{E}[t_j | \theta] f(\theta | s, t_i) d\theta = \int_0^1 \theta f(\theta | s, t_i) d\theta = \frac{ms + \ell t_i}{(m + \ell)}$ . Substituting this into the optimal forecast for forecaster  $j \neq i$ ,  $\mathbb{E}[f_j | s, t_i] = \mathbb{E}[(1 - w)s + wt_j | s, t_i] = (1 - w)s + w((1 - w)s + wt_i)$ , where  $w = \ell/(m + \ell)$ . It is therefore a Bayesian Nash equilibrium for each forecaster  $i$ , after observing the shared signal  $s$  and a private signal of  $t_i$ , to provide responses of

$$\begin{aligned} f_i &= \mathbb{E}[X | s, t_i] = (1 - w)s + wt_i, \\ g_i &= \mathbb{E}\left[\sum_{j \neq i} \frac{f_j}{n - 1} \mid s, t_i\right] = (1 - w^2)s + w^2 t_i. \end{aligned} \tag{B.1}$$

The posterior distribution for  $\theta$ , conditional on the signals observed by all  $n$  forecasters, is given by  $(\theta | s, \{t_i\}_{i=1}^n) \sim \text{Beta}(ms + \ell \sum_{i=1}^n t_i, m(1 - s) + \ell \sum_{i=1}^n (1 - t_i))$ , and the corresponding posterior predictive distribution for  $X$  is  $\mathbb{P}(X = 1 | s, \{t_i\}_{i=1}^n) = \frac{ms + \ell \sum_{i=1}^n t_i}{m + n\ell}$ . If forecasters respond optimally according to equation (B.1) above, the Symmetric pivoted estimate  $\hat{\theta}_S$  defined in equation (3.4) yields this optimal crowd forecast of  $\frac{1-w}{nw+1-w}s + \frac{w}{nw+1-w} \sum_{i=1}^n t_i = \frac{ms + \ell \sum_{i=1}^n t_i}{m + n\ell}$ .

### B.1.2 Nested Setting

A naïf's posterior beliefs about  $\theta$  are given by  $(\theta | s) \sim \text{Beta}(ms, m(1 - s))$ , with posterior predictive distribution  $\mathbb{P}(X = 1 | s) = s$ . Lacking additional information, naïfs should provide a forecast equal to the shared signal  $s$  and guess the same for the average response of others:  $f_i = g_i = s$ . A sophisticate's posterior beliefs about  $\theta$  are given by  $(\theta | s, t) \sim \text{Beta}(ms + \ell t, m(1 - s) + \ell(1 - t))$ , with posterior predictive distribution  $\mathbb{P}(X = 1 | s, t) = \frac{ms + \ell t}{m + \ell}$ . In addition, expecting that  $p$  of the other forecasters observe both  $s$  and  $t$  while  $1 - p$  of the other forecasters observe only  $s$ , he should provide responses of

$$\begin{aligned} f_i &= (1 - w)s + wt, \\ g_i &= (1 - pw)s + pwt. \end{aligned}$$

If  $p > 0$ , the posterior distribution for  $\theta$ , conditional on the signals observed by all forecasters, is given by  $(\theta \mid s, t) \sim \text{Beta}(ms + \ell t, m(1 - s) + \ell(1 - t))$ , and the corresponding posterior predictive distribution is  $\mathbb{P}(X = 1 \mid s, t) = \frac{ms + \ell t}{m + \ell}$ . If  $p = 0$ , no forecaster observes  $t$  and the posterior predictive distribution is simply  $\mathbb{P}(X = 1 \mid s) = s$ . In both cases, if forecasters respond as detailed above, substituting all of their forecasts and guesses about others into the Nested pivoted estimate  $\hat{\theta}_N$  defined in equation (3.11) and simplifying yields this optimal crowd forecast.<sup>1</sup>

### B.1.3 Nested-Symmetric Setting

As in the Nested setting, naïfs should provide a forecast equal to the shared signal  $s$  and guess the same for the average response of others:  $f_i = g_i = s$ . As in the Symmetric setting, sophisticate  $i$ 's posterior beliefs about  $\theta$  are given by  $(\theta \mid s, t_i) \sim \text{Beta}(ms + \ell t_i, m(1 - s) + \ell(1 - t_i))$ , his posterior predictive distribution is  $\mathbb{P}(X = 1 \mid s, t_i) = \frac{ms + \ell t_i}{m + \ell}$ , and his expectation of the signals  $t_j$ ,  $j \neq i$ , observed by other sophisticates is  $\mathbb{E}[t_j \mid s, t_i] = \frac{ms + \ell t_i}{(m + \ell)}$ . Expecting that  $p$  of the other forecasters observe both  $s$  and their own private signal  $t_j$ ,  $j \neq i$ , while  $1 - p$  of the other forecasters observe only  $s$ , he should provide responses of

$$\begin{aligned} f_i &= (1 - w)s + wt_i, \\ g_i &= (1 - pw^2)s + pw^2t_i. \end{aligned}$$

If  $p > 0$ , the posterior distribution for  $\theta$ , conditional on the signals observed by all forecasters, is given by  $(\theta \mid s, \{t_i\}_{i=1}^{pn}) \sim \text{Beta}(ms + \ell \sum_{i=1}^{pn} t_i, m(1 - s) + \ell \sum_{i=1}^{pn} (1 - t_i))$ , and the corresponding posterior predictive distribution is  $\mathbb{P}(X = 1 \mid s, \{t_i\}_{i=1}^{pn}) = \frac{ms + \ell \sum_{i=1}^{pn} t_i}{m + pn\ell}$ . If  $p = 0$ , all forecasters are naïve and the posterior predictive distribution is simply  $\mathbb{P}(X = 1 \mid s) = s$ . In both cases, if forecasters respond as detailed above,

---

<sup>1</sup> Note that to implement the algorithm for  $\hat{\theta}_N$ , the decision analyst sets  $w = 1$  and uses the appropriate value of  $p$  for the crowd.



substituting all of their forecasts and guesses about others into the Nested-Symmetric pivoted estimate  $\hat{\theta}_{NS}$  defined in equation (3.11) and simplifying yields this optimal crowd forecast.

#### B.1.4 Minimal Pivoting

In the analysis below, I show that the minimal pivoting procedure, given by  $\hat{\theta}_M = 2\bar{f} - \bar{g}$ , outperforms the simple average with a sufficiently large crowd in each of the S, N, and NS settings, analogous to Proposition 4. Note that conditional on  $\theta$ , the shared signal  $s$  has mean  $\theta$  and variance  $\frac{\theta(1-\theta)}{m}$  and private signals  $t_i$  have mean  $\theta$  and variance  $\frac{\theta(1-\theta)}{\ell}$ .

In the Symmetric setting, if forecasters respond optimally, the average forecast will be given by  $\bar{f} = (1-w)s + w \sum_{i=1}^n \frac{t_i}{n}$  and the minimal estimate will be given by  $\hat{\theta}_M = (1-2w+w^2)s + (2w-w^2) \sum_{i=1}^n \frac{t_i}{n}$ . Both  $\bar{f}$  and  $\hat{\theta}_M$  have mean  $\theta$ , but  $\bar{f}$  has variance  $(1-w)^2 \frac{\theta(1-\theta)}{m} + w^2 \frac{\theta(1-\theta)}{\ell n}$  while  $\hat{\theta}_M$  has variance  $(1-w)^4 \frac{\theta(1-\theta)}{m} + (2-w)^2 w^2 \frac{\theta(1-\theta)}{\ell n}$ . The second term of each variance expression declines to 0 as  $n \rightarrow \infty$ , so  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] \leq \mathbb{E}[(\bar{f} - \theta)^2]$  for sufficiently large  $n$ , with the inequality strict for  $w, \theta \in (0, 1)$ .

In the Nested setting, if forecasters respond optimally, the average forecast will be given by  $\bar{f} = (1-pw)s + pwt$  and the minimal estimate will be given by  $\hat{\theta}_M = (1+p^2w-2pw)s + (2pw-p^2w)t$ . Both  $\bar{f}$  and  $\hat{\theta}_M$  have mean  $\theta$ , but  $\bar{f}$  has variance  $(1-pw)^2 \frac{\theta(1-\theta)}{m} + p^2 w^2 \frac{\theta(1-\theta)}{\ell}$  while  $\hat{\theta}_M$  has variance  $(1-pw)^4 \frac{\theta(1-\theta)}{m} + (2-pw)^2 p^2 w^2 \frac{\theta(1-\theta)}{\ell}$ . Following the same argument found in the proof of Proposition 4,  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] \leq \mathbb{E}[(\bar{f} - \theta)^2]$  for all  $n$ , with the inequality strict if  $\theta \in (0, 1)$  and  $p > 0$ .

In the Nested-Symmetric setting, if forecasters respond optimally, the average forecast will be given by  $\bar{f} = (1-pw)s + pw \sum_{i=1}^{pn} \frac{t_i}{pn}$  and the minimal estimate will be given by  $\hat{\theta}_M = (1-2pw+p^2w^2)s + (2pw-p^2w^2) \sum_{i=1}^{pn} \frac{t_i}{pn}$ . Both  $\bar{f}$  and  $\hat{\theta}_M$

have mean  $\theta$ , but  $\bar{f}$  has variance  $(1 - pw)^2 \frac{\theta(1-\theta)}{m} + p^2 w^2 \frac{\theta(1-\theta)}{\ell p n}$  while  $\hat{\theta}_M$  has variance  $(1 - pw)^4 \frac{\theta(1-\theta)}{m} + (2 - pw)^2 p^2 w^2 \frac{\theta(1-\theta)}{\ell p n}$ . The second term of each variance expression declines to 0 as  $n \rightarrow \infty$ , so  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] \leq \mathbb{E}[(\bar{f} - \theta)^2]$  for sufficiently large  $n$ , with the inequality strict if  $pw, \theta \in (0, 1)$ .

Minimal pivoting therefore offers a robust alternative to the more complex tailored pivoting procedures, even when forecasting probabilities. One important additional consideration in this setting is the possibility that, even though the responses  $\bar{f}$  and  $\bar{g}$  must fall between 0 and 1, pivoting could push the crowd estimate outside its logical range. This means that the decision analyst may need to employ an additional correction to ensure that the estimated probability stays in the interval  $[0, 1]$ . Possible approaches include simply Winsorizing estimates to 0 and 1 or applying an  $S$ -shaped correction to the raw crowd estimates.

## B.2 A Gamma-Gamma Model

To study shared information in an analogous setting where the random variable of interest may have skewness, consider the following gamma-gamma information structure, adapted from Lichtendahl, Grushka-Cockayne, & Pfeifer (2014): Suppose we are forecasting a positive random variable  $X > 0$ , which follows a gamma distribution with known shape parameter  $\alpha$  and unknown rate parameter  $\theta$ ,  $X \sim Ga(\alpha, \theta)$ . A positive random variable  $Z$  follows a gamma distribution with shape  $\alpha > 0$  and rate  $\beta > 0$ ,  $Z \sim Ga(\alpha, \beta)$ , if it has the probability density function  $f(z; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{(\alpha-1)} \exp(-\beta y)$  over  $z > 0$ , with mean  $\frac{\alpha}{\beta}$  (Bernardo & Smith, 1994, p. 118). The gamma distribution is commonly used to model waiting times, and encompasses a large family of continuous random variables with skewness. For example, in the special case where  $\alpha = 1$ ,  $X$  is an exponential random variable with rate parameter  $\theta$ . If  $\alpha$  is an integer, the random variable  $X$  can be defined as the waiting time until

the  $\alpha$ th event in a homogeneous Poisson process with arrival rate  $\theta$ . Note that for ease of updating in the Bayesian information model, the mathematical characterization of the gamma distribution specified here uses a rate parameter rather than a scale parameter. It will be useful to define the compound gamma distribution which will describe the predictive distribution for  $X$ . A positive random variable  $Z$  follows a compound gamma distribution,  $Z \sim Ga(\alpha, \beta, \delta)$ , if it has the probability density function  $f(z; \alpha, \beta, \delta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\delta)}{\Gamma(\delta)} \frac{z^{\delta-1}}{(\beta+z)^{\alpha+\delta}}$  over  $z > 0$ , with mean  $\frac{\delta\beta}{\alpha-1}$ .

Assume that  $n$  forecasters share a common Gamma prior over  $\theta$  with shape 1 and rate 0,  $\theta \sim Ga(1, 0)$ . This specific prior is assumed for convenience so that, in conjunction with the other assumptions about the signal generating processes, the optimal forecast can be written as a linear combination of the two pieces of information,  $f_i = (1-w)s + wt_i$ , with the relative weight  $w = \ell/(m+\ell)$  defined as before. All forecasters observe the same shared signal  $s$ , which is distributed according to a Gamma distribution with shape  $m\alpha$  and rate  $m\theta$ ,  $(s | \theta) \sim Ga(m\alpha, m\theta)$ . In the Symmetric setting, each forecaster  $i = 1 \dots n$  also receives his own additional private signal  $t_i$ , which follows a Gamma distribution with shape  $\ell\alpha$  and rate  $\ell\theta$ ,  $(t_i | \theta) \sim Ga(\ell\alpha, \ell\theta)$ . In the Nested setting, a proportion  $1-p$  of the forecasters are naïfs who only observe  $s$ , and the remaining proportion  $p$  of forecasters are sophisticates who observe both  $s$  and an additional signal  $t$ , which follows a Gamma distribution with shape  $\ell\alpha$  and rate  $\ell\theta$ ,  $(t | \theta) \sim Ga(\ell\alpha, \ell\theta)$ . In the Nested-Symmetric setting, each sophisticate observes both  $s$  and his own private signal  $t_i$ ,  $(t_i | \theta) \sim Ga(\ell\alpha, \ell\theta)$ . All signals ( $s$ ,  $t_i$ , and  $t$ ) received by the forecasters are independent conditional on  $\theta$ . This information structure and the parameters  $\alpha$ ,  $m$ , and  $\ell$  are commonly known among all forecasters, and the proportion  $p$  is known by sophisticates.

### B.2.1 Symmetric Setting

In the Symmetric model, forecaster  $i$ 's posterior distribution for  $\theta$ , conditional on his signals, is given by  $(\theta \mid s, t_i) \sim Ga((m+\ell)\alpha+1, ms+lt_i)$ . His posterior predictive distribution for  $X$  is compound gamma,  $(X \mid s, t_i) \sim Gg((m+\ell)\alpha+1, ms+lt_i, \alpha)$ , with mean  $\frac{ms+lt_i}{m+\ell} = (1-w)s + wt_i$ . Conditional on his own information  $\{s, t_i\}$ , forecaster  $i$ 's expectation of the signals  $t_j$ ,  $j \neq i$ , observed by others can be calculated according to  $\mathbb{E}[t_j \mid s, t_i] = \int_0^\infty \mathbb{E}[t_j \mid \theta] f(\theta \mid s, t_i) d\theta = \int_0^\infty \frac{\alpha}{\theta} \frac{(ms+lt_i)^{(m+\ell)\alpha+1}}{\Gamma((m+\ell)\alpha+1)} \theta^{(m+\ell)\alpha} \exp(-(ms+lt_i)\theta) d\theta = \alpha \frac{ms+lt_i}{(m+\ell)\alpha} \int_0^\infty \frac{(ms+lt_i)^{(m+\ell)\alpha}}{\Gamma((m+\ell)\alpha)} \theta^{(m+\ell)\alpha-1} \exp(-(ms+lt_i)\theta) d\theta$ . The expression in the last integral is simply the density function for a  $Ga((m+\ell)\alpha, ms+lt_i)$  random variable, whose total probability integrates to 1 over its support  $(0, \infty)$ , so the conditional expectation simplifies to  $\mathbb{E}[t_j \mid s, t_i] = \frac{ms+lt_i}{(m+\ell)} = (1-w)s + wt_i$ . Substituting this into the optimal forecast for forecaster  $j \neq i$ ,  $\mathbb{E}[f_j \mid s, t_i] = \mathbb{E}[(1-w)s + wt_j \mid s, t_i] = (1-w)s + w((1-w)s + wt_i)$ .

Therefore it is a Bayesian Nash equilibrium for each forecaster  $i$ , after observing the shared signal  $s$  and the private signal  $t_i$ , to provide responses of

$$\begin{aligned} f_i &= \mathbb{E}[X \mid s, t_i] = (1-w)s + wt_i, \\ g_i &= \mathbb{E}\left[\sum_{j \neq i} \frac{f_j}{n-1} \mid s, t_i\right] = (1-w^2)s + w^2t_i. \end{aligned} \tag{B.2}$$

The posterior distribution for  $\theta$ , conditional on all of the signals observed by all  $n$  forecasters, is given by  $(\theta \mid s, \{t_i\}_{i=1}^n) \sim Ga((m+n\ell)\alpha+1, ms+\ell\sum_{i=1}^n t_i)$ , and the corresponding posterior predictive distribution for  $X$  is compound gamma,  $(X \mid s, \{t_i\}_{i=1}^n) \sim Gg((m+n\ell)\alpha+1, ms+\ell\sum_{i=1}^n t_i, \alpha)$ , with mean  $\frac{ms+\ell\sum_{i=1}^n t_i}{m+n\ell}$ . If forecasters respond optimally according to equation (B.2) above, the Symmetric pivoted estimate  $\hat{\theta}_S$  defined in equation (3.4) yields this optimal crowd forecast of

$$\frac{1-w}{nw+1-w}s + \frac{w}{nw+1-w} \sum_{i=1}^n t_i = \frac{ms+\ell\sum_{i=1}^n t_i}{m+n\ell}.$$

### B.2.2 Nested Setting

Naïfs' posterior beliefs about  $\theta$  are given by  $(\theta | s) \sim Ga(m\alpha + 1, ms)$ . His posterior predictive distribution for  $X$  is compound gamma,  $(X | s) \sim Gg(m\alpha + 1, ms, \alpha)$ , with mean  $s$ . Lacking additional information, naïfs should provide a forecast equal to the shared signal  $s$  and guess the same for the average response of others:  $f_i = g_i = s$ . Sophisticates' posterior beliefs about  $\theta$  are given by  $(\theta | s, t) \sim Ga((m + \ell)\alpha + 1, ms + \ell t)$ , with a compound gamma posterior predictive distribution for  $X$ ,  $(X | s, t) \sim Gg((m + \ell)\alpha + 1, ms + \ell t, \alpha)$ , with mean  $\frac{ms + \ell t}{m + \ell} = (1 - w)s + wt$ . Expecting that  $p$  of the other forecasters observe both  $s$  and  $t$  while  $1 - p$  of the other forecasters observe only  $s$ , sophisticates should provide responses of

$$\begin{aligned} f_i &= (1 - w)s + wt, \\ g_i &= (1 - pw)s + pwt. \end{aligned}$$

If  $p > 0$ , the posterior distribution for  $\theta$ , conditional on the signals observed by all forecasters, is given by  $(\theta | s, t) \sim Ga((m + \ell)\alpha + 1, ms + \ell t)$ , with a compound gamma posterior predictive distribution for  $X$ ,  $(X | s, t) \sim Gg((m + \ell)\alpha + 1, ms + \ell t, \alpha)$ , with mean  $\frac{ms + \ell t}{m + \ell} = (1 - w)s + wt$ . If  $p = 0$ , no forecaster observes  $t$  and the posterior predictive distribution is simply  $(\theta | s) \sim Ga(m\alpha + 1, ms)$ , with compound gamma posterior predictive distribution for  $X$ ,  $(X | s) \sim Gg(m\alpha + 1, ms, \alpha)$ , with mean  $s$ . In both cases, if forecasters respond as detailed above, substituting all of their forecasts and guesses about others into the Nested pivoted estimate  $\hat{\theta}_N$  defined in equation (3.11) and simplifying yields this optimal crowd forecast.<sup>2</sup>

### B.2.3 Nested-Symmetric Setting

As in the Nested Setting, naïfs should provide a forecast equal to the shared signal  $s$  and guess the same for the average response of others:  $f_i = g_i = s$ . As in the

<sup>2</sup> As before, to implement the algorithm for  $\hat{\theta}_N$ , the decision analyst sets  $w = 1$  and uses the appropriate value of  $p$  for the crowd.

Symmetric setting, sophisticate  $i$ 's posterior beliefs about  $\theta$  are given by  $(\theta \mid s, t_i) \sim Ga((m + \ell)\alpha + 1, ms + \ell t_i)$ , his posterior predictive distribution for  $X$  is compound gamma,  $(X \mid s, t_i) \sim Gg((m + \ell)\alpha + 1, ms + \ell t_i, \alpha)$ , with mean  $\frac{ms + \ell t_i}{m + \ell} = (1 - w)s + wt_i$ , and his expectation of the signals  $t_j$ ,  $j \neq i$ , observed by other sophisticates is  $\mathbb{E}[t_j \mid s, t_i] = \frac{ms + \ell t_i}{(m + \ell)}$ . Expecting that  $p$  of the other forecasters observe both  $s$  and their own private signal  $t_j$ ,  $j \neq i$ , while  $1 - p$  of the other forecasters observe only  $s$ , he should provide responses of

$$\begin{aligned} f_i &= (1 - w)s + wt_i, \\ g_i &= (1 - pw^2)s + pw^2t_i. \end{aligned}$$

If  $p > 0$ , the posterior distribution for  $\theta$ , conditional on the signals observed by all forecasters, is given by  $(\theta \mid s, \{t_i\}_{i=1}^{pn}) \sim Ga((m + pn\ell)\alpha + 1, ms + \ell \sum_{i=1}^p nt_i)$ , and the corresponding posterior predictive distribution for  $X$  is compound gamma,  $(X \mid s, \{t_i\}_{i=1}^{pn}) \sim Gg((m + pn\ell)\alpha + 1, ms + \ell \sum_{i=1}^{pn} t_i, \alpha)$ , with mean  $\frac{ms + \ell \sum_{i=1}^{pn} t_i}{m + n\ell}$ . If  $p = 0$ , all forecasters are naïve and the posterior predictive distribution is simply  $(X \mid s) \sim Gg(m\alpha + 1, ms, \alpha)$ , with mean  $s$ . In both cases, if forecasters respond as detailed above, substituting all of their forecasts and guesses about others into the Nested-Symmetric pivoted estimate  $\hat{\theta}_{NS}$  defined in equation (3.11) and simplifying yields this optimal crowd forecast.

#### B.2.4 Minimal Pivoting

As in Proposition 4, I show here that the minimal pivoting procedure  $\hat{\theta}_M$  outperforms the simple average with a sufficiently large crowd in each of the S, N, and NS settings. Observe that conditional on  $\theta$ , the shared signal  $s$  has mean  $\theta$  and variance  $\frac{\alpha}{m\theta^2}$  and private signals  $t_i$  have mean  $\theta$  and variance  $\frac{\alpha}{\ell\theta^2}$ .

In the Symmetric setting, if forecasters respond optimally, the average forecast will be given by  $\bar{f} = (1 - w)s + w \sum_{i=1}^n \frac{t_i}{n}$  and the minimal estimate will be given by

$\hat{\theta}_M = (1 - 2w + w^2)s + (2w - w^2) \sum_{i=1}^n \frac{t_i}{n}$ . Both  $\bar{f}$  and  $\hat{\theta}_M$  have mean  $\theta$ , but  $\bar{f}$  has variance  $(1-w)^2 \frac{\alpha}{m\theta^2} + w^2 \frac{\alpha}{\ell n \theta^2}$  while  $\hat{\theta}_M$  has variance  $(1-w)^4 \frac{\alpha}{m\theta^2} + (2-w)^2 w^2 \frac{\alpha}{\ell n \theta^2}$ . The second term of each variance expression declines to 0 as  $n \rightarrow \infty$ , so  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] \leq \mathbb{E}[(\bar{f} - \theta)^2]$  for sufficiently large  $n$ , with the inequality strict for  $w \in (0, 1)$ .

In the Nested setting, if forecasters respond optimally, the average forecast will be given by  $\bar{f} = (1 - pw)s + pwt$  and the minimal estimate will be given by  $\hat{\theta}_M = (1 + p^2w - 2pw)s + (2pw - p^2w)t$ . Both  $\bar{f}$  and  $\hat{\theta}_M$  have mean  $\theta$ , but  $\bar{f}$  has variance  $(1 - pw)^2 \frac{\alpha}{m\theta^2} + p^2w^2 \frac{\alpha}{\ell \theta^2}$  while  $\hat{\theta}_M$  has variance  $(1 - pw)^4 \frac{\alpha}{m\theta^2} + (2 - pw)^2 p^2w^2 \frac{\alpha}{\ell \theta^2}$ . Following the same argument found in the proof of Proposition 4,  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] \leq \mathbb{E}[(\bar{f} - \theta)^2]$  for all  $n$ , with the inequality strict if  $p > 0$ .

In the Nested-Symmetric setting, if forecasters respond optimally, the average forecast will be given by  $\bar{f} = (1 - pw)s + pw \sum_{i=1}^{pn} \frac{t_i}{pn}$  and the minimal estimate will be given by  $\hat{\theta}_M = (1 - 2pw + p^2w^2)s + (2pw - p^2w^2) \sum_{i=1}^{pn} \frac{t_i}{pn}$ . Both  $\bar{f}$  and  $\hat{\theta}_M$  have mean  $\theta$ , but  $\bar{f}$  has variance  $(1 - pw)^2 \frac{\alpha}{m\theta^2} + p^2w^2 \frac{\alpha}{\ell pn \theta^2}$  while  $\hat{\theta}_M$  has variance  $(1 - pw)^4 \frac{\alpha}{m\theta^2} + (2 - pw)^2 p^2w^2 \frac{\alpha}{\ell pn \theta^2}$ . The second term of each variance expression declines to 0 as  $n \rightarrow \infty$ , so  $\mathbb{E}[(\hat{\theta}_M - \theta)^2] \leq \mathbb{E}[(\bar{f} - \theta)^2]$  for sufficiently large  $n$ , with the inequality strict if  $pw \in (0, 1)$ .

As before, the minimal pivoting algorithm provides a robust alternative that reduces the error variance of the crowd forecast relative to the simple average in each of the three information settings. Overall, both the beta-binomial and gamma-gamma information models lend support to the idea that the pivoting procedures can be effective in removing shared-information bias and improving forecast accuracy. In the next section, I consider how varying assumptions about how information is distributed across forecasters may affect these results.

### B.3 General Allocation of Information Across Forecasters

The models studied so far make a number of restrictive assumptions about how information may be distributed and shared across forecasters. In the most general model, there might be  $k = 1, \dots, K$  different signals  $t_k$ , each distributed normally with mean  $\theta$  and precision  $\ell_k$ ,  $(t_k|\theta) \sim N(\theta, \ell_k\lambda)$ . Note that  $K$  and  $n$  need not be equal here, and in fact might be very different in magnitude. For each forecaster  $i$  and signal  $t_k$ ,  $i$  either observes  $t_k$  ( $e_{ik} = 1$ ) or does not observe  $t_k$  ( $e_{ik} = 0$ ). If these vectors of information access  $\mathbf{e}_i = \{e_{i1}, \dots, e_{iK}\}$  are common knowledge, each forecaster  $i$  can compute both an optimal forecast  $f_i$  and a guess about others  $g_i$  conditional on the signals they observe.

Although this setup could in theory encompass all possible combinations of information overlap and distribution across individuals, it imposes an extreme computational burden on the forecasters, who need to know all  $n \times K$  entries  $e_{ik}$  as well as the relative precisions  $\{\ell_k\}_{k=1}^K$  to estimate the average forecast that will be provided by others. The decision analyst must then also know the information access vectors  $\{\mathbf{e}_i\}_{i=1}^n$  and precisions  $\{\ell_k\}_{k=1}^K$  to optimally aggregate the responses. As a result, this general model proves to be analytically intractable and unrealistic without additional simplifying assumptions. Next I consider a more specific model of shared information that allows for heterogeneity in expertise across forecasters.

#### *B.3.1 Individual-Specific Information Weights $w_i$*

One restrictive assumption of the S, N, and NS models is that while the individual signals may vary, all sophisticated forecasters possess the same *amount* of information. The private signals observed by forecasters are equally informative, captured by a single relative information weight  $w$ . Below I consider a model where the amount of private information observed may vary from person to person. Specifically, re-



turning to the main normal-normal information structure studied previously, assume that  $(X|\theta) \sim N(\theta, \lambda)$ , and  $n$  forecasters, who share a diffuse prior belief about  $\theta$ , each observe a shared signal  $s$ , where  $(s|\theta) \sim N(\theta, m\lambda)$ . Each forecaster  $i = 1, \dots, n$  now additionally receives his own private signal  $t_i$ , where  $(t_i|\theta) \sim N(\theta, \ell_i\lambda)$ . As before, all signals are assumed to be independent conditional on  $\theta$  and the parameters  $\lambda, m, \ell_1, \dots, \ell_n$  are common knowledge. It will be useful to define the individual relative information weights  $w_i \equiv \ell_i/(m + \ell_i)$  here and  $\bar{w}_{-i} = \frac{1}{n-1} \sum_{j \neq i} w_j$ . The optimal forecast for individual  $i$  is given by

$$f_i = (1 - w_i)s + w_i t_i.$$

Conditional on his information  $(s, t_i)$ , his expectation of the forecast that will be provided by individual  $j$  is  $g_{ij} = (1 - w_j)s + w_j((1 - w_i)s + w_i t_i) = (1 - w_i w_j)s + w_i w_j t_i$ . His expectation of the average forecast of all other forecasters is then given by  $g_i = \frac{1}{n-1} \sum_{j \neq i} g_{ij}$ , which simplifies to

$$g_i = (1 - w_i \bar{w}_{-i})s + w_i \bar{w}_{-i} t_i.$$

If forecasters respond optimally and  $w_i > 0$  and  $\bar{w}_{-i} < 1$ , the signals can be recovered according to

$$t_i = f_i + \frac{1 - w_i}{w_i(1 - \bar{w}_{-i})}(f_i - g_i),$$

$$s_i = f_i + \frac{1}{1 - \bar{w}_{-i}}(g_i - f_i),$$

where  $s_i$  should equal  $s$  for all forecasters. If  $w_i = 0$ , the decision analyst should set  $t_i = \emptyset$  and if  $\bar{w}_{-i} = 1$  for any  $i$ , the decision analyst should set  $s = \emptyset$  and  $t_i = f_i$  for all  $i$ . The optimal crowd forecast can then be expressed as a form of pivoting by

re-aggregating these signals in the optimal Bayesian manner:<sup>3</sup>

$$\hat{\theta} = \begin{cases} \frac{1}{\sum_{i=1}^n \frac{1-w_i}{n} + \sum_{i=1}^n w_i} \sum_{i=1}^n w_i t_i + \frac{\sum_{i=1}^n \frac{1-w_i}{n}}{\sum_{i=1}^n \frac{1-w_i}{n} + \sum_{i=1}^n w_i} \sum_{i=1}^n \frac{s_i}{n} & \text{if } \bar{w}_{-i} < 1 \text{ for all } i \\ \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} t_i & \text{if } \bar{w}_{-i} = 1 \text{ for any } i. \end{cases}$$

Observe that this setup introduces a much higher level of complexity than the S, N, and NS settings studied previously because it requires all forecasters to possess detailed meta-knowledge about the weights  $w_j$  and relative informativeness of each other forecaster  $j$ 's private signals. In addition, it remains impractical since the decision analyst needs to know all of the weights  $w_i$  and has no way to estimate them from the response data. Given how sensitive both the optimal responses and the pivoting procedures are to these weights, the results seem unrealistic. The S and NS models studied previously parsimoniously embody this setting while using only a single parameter  $w$ , or a pair of parameters  $w$  and  $p$ . The complexity of the model can be greatly reduced by restricting  $\ell_i = 0$  for naïfs and  $\ell_i = \ell$  for sophisticates, letting additional heterogeneity be captured by the error terms in the behavioral model of responses.

---

<sup>3</sup> In the procedure below,  $s$  is calculated by averaging the each of the estimates  $s_i$  that would result from using only individual  $i$ 's responses  $(f_i, g_i)$ .

# Bibliography

- [1] ANDERSON, L. R., HOLT, C. A. (1997). Information Cascades in the Laboratory. *The American Economic Review* **87**(5): 847–862.
- [2] ARMSTRONG, S. J. (2001). Combining forecasts. In Scott J. Armstrong, (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, 417–439.
- [3] ATANASOV, P., RESCOBER, P., STONE, E., SWIFT, S. A., SERVANSCHREIBER, E., TETLOCK, P., UNGAR, L., MELLERS, B. (2015). Distilling the Wisdom of Crowds: Prediction Markets versus Prediction Polls. *Working Paper*.
- [4] BERNARDO, J. M. SMITH, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons, Ltd.
- [5] BRIER, G.W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**: 1–3.
- [6] BUDESCU, D. V., CHEN, E. (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science* **61**(2): 267–280.
- [7] CHEN, E., BUDESCU, D. V. (2015). The Contribution Weighted Model: Identification and Combination of Expertise. *Decision Analysis Today* **34**(1): 21–25.
- [8] CHEN, K., FINE, L. R., HUBERMAN, B. A. (2004). Eliminating Public Knowledge Biases in Information-Aggregation Mechanisms. *Management Science* **50**(7): 983–994.
- [9] CIPRIANI, M., GUARINO, A. (2014). Estimating a Structural Model of Herd Behavior in Financial Markets. *The American Economic Review* **104**(1): 224–251.
- [10] CLEMEN, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* **5**: 559–583.
- [11] CLEMEN, R. T., WINKLER, R. L. (1985). Limits for the Precision and Value of Information from Dependent Sources *Operations Research* **33**: 427–442.
- [12] CLEMEN, R. T., WINKLER, R. L. (1986). Combining economic forecasts. *Journal of Business & Economic Statistics* **4**(1): 39–46.

- [13] DALKEY, N., HELMER, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. *Management Science* **9**(3): 458–467.
- [14] FISHER, M., RAMAN, A. (1996). Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales. *Operations Research* **44**(1): 87–99.
- [15] GALTON, F. (1907). Vox Populi. *Nature* **1949**(75): 450–451.
- [16] GIGONE, D., HASTIE, R. (1993). The Common Knowledge Effect: Information Sharing and Group Judgment. *Journal of Personality and Social Psychology* **65**(5): 959–974.
- [17] GNEITING, T., RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**: 359–378.
- [18] GRUCA, T. S., BERG, J., CIPRIANO, M. (2003). The Effect of Electronic Markets on Forecasts of New Product Success. *Information Systems Frontiers* **5**(1): 95–105.
- [19] HAMMOND, J. H., RAMAN, A. (1994). *Sport Obermeyer, Ltd.* Harvard Business School.
- [20] HSU, C. C., SANDFORD, B. A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research & Evaluation* **12**(10): 1–8.
- [21] HUNG, A. A., PLOTT, C. R. (2001). Information Cascades: Replication and an Extension to Majority rule and Conformity-Rewarding Institutions. *The American Economic Review* **91**(5): 1508–1520.
- [22] JOSE, V. R. R., WINKLER, R. L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting* **24**: 163–169.
- [23] KAHNEMAN, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- [24] KIM, O., LIM, S. C., SHAW, K. W. (2001). The inefficiency of the mean analyst forecast as summary of forecast of earnings. *Journal of Accounting Research* **39**: 329–335.
- [25] LARRICK, R. P., SOLL, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science* **52**(1): 111–127.
- [26] LARRICK, R. P., MANNES, A. E., SOLL, J. B. (2012). Forecasters objectives and strategies. In *Krueger, J. I., ed. Frontiers of Social Psychology: Social Psychology and Decision Making*. New York: Psychology Press.

- [27] LICHTENDAHL, K. C., GRUSHKA-COCKAYNE, Y., PFEIFER, P. E. (2013). The Wisdom of Competitive Crowds. *Operations Research* **61**(6): 1383–1398.
- [28] LORENZ, J., RAUHUT, H., SCHWEITZER, F., HELBING, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences* **108**(22): 9020–9025.
- [29] MAKRIDAKIS, S., WINKLER, R. L. (1983). Averages of Forecasts: Some Empirical Results. *Management Science* **29**(9): 987–996.
- [30] MANNES, A. E., SOLL, J. B., LARRICK, R. P. (2014). The Wisdom of Select Crowds. *Journal of Personality and Social Psychology* **107**(2): 276–299.
- [31] MARINOVIC, I., OTTAVIANI, M., SØRENSEN, P. N. (2012). Forecasters objectives and strategies. In *Graham, E., Timmermann, A., eds. Handbook of Economic Forecasting, Vol. 2A*. Elsevier, Amsterdam.
- [32] MASON, S. J. (2004). On Using “Climatology” as a Reference Strategy in the Brier and Ranked Probability Skill Scores. *Monthly Weather Review* **132**: 1891–1895.
- [33] MELLERS, B. A., UNGAR, L., BARON, J., RAMOS, J., GURCAY, B., FINCHER, K., SCOTT, S. E., MOORE, D., ATANASOV, P., SWIFT, S. A., MURRAY, T., STONE, E., TETLOCK, P. E. (2014). Psychological strategies for winning geopolitical forecasting tournaments. *Psychological Science* **25**(5): 1106–1115.
- [34] MURPHY, A. H. (1973). A New Vector Partition of the Probability Score. *Journal of Applied Meteorology* **12**: 595–600.
- [35] MURPHY, A. H. WINKLER, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting* **7**: 435–455.
- [36] OTTAVIANI, M., SØRENSEN, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics* **81**: 441–466.
- [37] PLOTT, C. R. (2000). Markets as Information Gathering Tools. *Southern Economic Journal* **67**(1): 2–15.
- [38] PRELEC, D. (2004). A Bayesian Truth Serum for Subjective Data. *Science* **306**: 462–466.
- [39] PRELEC, D., SEUNG, H. S., MCCOY, J. (2013). Finding truth even if the crowd is wrong. *Working Paper*.
- [40] RAIFFA, H., AND SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Harvard University Press, Boston.

- [41] SIMON, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* **69**(1): 99–118.
- [42] SOLL, J. B., LARRICK, R. P. (2009). Strategies for Revising Judgment: How (and How Well) People Use Others’ Opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **35**(3): 780–805.
- [43] SPYROU, S. (2013). Herding in financial markets: a review of the literature. *Review of Behavioral Finance* **5**(2): 175–194.
- [44] STASSER, G., TITUS, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology* **48**(6): 1467–1478.
- [45] SUNSTEIN, C. R. (2005). Group Judgments: Statistical Means, Deliberation, and Information Markets. *NYU Law Review* **80**: 962–1049.
- [46] SUNSTEIN, C. R. (2006). Deliberating Groups versus Prediction Markets (or Hayek’s Challenge to Habermas). *Episteme* **3**(3): 192–213.
- [47] SUROWIECKI, J. (2005). *The Wisdom of Crowds*. Anchor Books, New York.
- [48] TRUEMAN, B. (1994). Analyst Forecasts and Herding Behavior. *The Review of Financial Studies* **7**(1): 97–124.
- [49] VAN BRUGGEN, G. H., SPANN, M., LILIEN, G. L., SKIERA, B. (2010). Prediction Markets as institutional forecasting support systems. *Decision Support Systems* **49**: 404–416.
- [50] WANG, T., WANG, D. (2014). Why Amazon’s Ratings Might Mislead You: The Story of Herding Effects. *Big Data* **2**(4): 196–204.
- [51] WINKLER, R. L. (1981). Combining Probability Distributions from Dependent Information Sources. *Management Science* **27**(4): 479–488.
- [52] WITTENBAUM, G. M., HUBBELL, A. P., ZUCKERMAN, C. (1999). Mutual Enhancement: Toward an Understanding of the Collective Preference for Shared Information. *Journal of Personality and Social Psychology* **77**(5): 967–978.
- [53] WOLFERS, J., ZITZEWITZ, E. (2004). Prediction Markets. *The Journal of Economic Perspectives* **18**(2): 107–126.

# Biography

Asa Benjamin Palley was born in Washington, DC in 1986. He graduated *magna cum laude* from Bowdoin College in 2007, received his M.S. in Applied Mathematics and Scientific Computation from the University of Maryland at College Park in 2009, and received his M.S. in Mathematics from Carnegie Mellon University in 2010. After completing a Ph.D. in Decision Sciences from The Fuqua School of Business at Duke University in 2016, he will join the Kelley School of Business at Indiana University as an Assistant Professor of Operations and Decision Technologies.

## Selected Honors and Awards:

Winner of the Decision Analysis Society Student Paper Award, 2015.

Doctoral Fellowship, The Fuqua School of Business, Duke University, 2010 – 2015.

Phi Beta Kappa, Bowdoin College, 2007.

Smyth Mathematical Prize, Bowdoin College, 2006.

## Publications:

OFFERMAN, T., PALLEY, A. B. (2016). Losses in Translation: An Off-the-Shelf Method to Recover Probabilistic Beliefs from Loss-Averse Agents. *Experimental Economics* **19**(1): 1–30.

PALLEY, A. B., KREMER, M. (2014). Sequential Search and Learning from Rank Feedback: Theory and Experimental Evidence. *Management Science* **60**(10): 2525–2542.

KEENEY, R. L., PALLEY, A. B. (2013). Decision Strategies to Reduce Teenage and Young Adult Deaths in the United States. *Risk Analysis* **33**(9): 1661–1676.