Using Expert Assessments to Estimate Probability Distributions

Saurabh Bansal
The Pennsylvania State University, State College, 16803, sub32@psu.edu

Genaro J. Gutierrez
The University of Texas at Austin, Austin, 78712, genaro@austin.utexas.edu

John R. Keiser
Dow AgroSciences, Marshalltown, 50158, JRKeiser@dow.com

Previous decision analysis literature has established the presence of judgmental errors in the quantile judgments provided by an expert. However a systematic approach to using quantile judgments to deduce the mean and standard deviation of the underlying probability distribution while accounting for judgmental errors has not yet been established. This paper develops an optimization based approach to accomplish this task. The approach estimates both the mean and standard deviation as weighted linear combinations of quantile judgments, where the weights are explicit functions of the expert's judgmental errors. The approach is analytically tractable, and provides flexibility to elicit any set of quantiles from an expert. The structural properties of the solution explain a number of numerical observations made in the extant decision analysis literature. The approach also establishes that using an expert's quantile judgments to deduce the distribution parameters is equivalent to collecting data with a specific sample size. The approach allows combining the expert's judgments with data in a Bayesian fashion and with other experts' estimates in a manner that is consistent with the existing theory of combining point forecasts from noisy sources. The paper also contributes to practice by discussing a large-scale implementation of the theory at Dow AgroSciences to make an annual $800 million decision, and providing insights into the selection of quantiles to elicit from an expert. Results show that the judgments of the expert at Dow are equivalent to 5–6 years of data collection, with substantial monetary savings.

Acknowledgements: The authors thank the participants at the First Decision Analysis Conference held at Georgetown University in June 2014, and the Statistics Seminar Series at the Pennsylvania State University, and the University of Texas at Austin. Thanks to John Leichty, Wayne DeSarbo, Tom Sager, and Dennis Lin for their valuable comments during the revision. Thanks to Susan Gentry at Dow AgroSciences; Eric Bickel at the University of Texas at Austin; Jim Smith at Duke University; and Nagesh Gaverneni at Cornell University.

1. Introduction

1.1. Problem Context and Industry Motivation

A determination of the uncertainties in supply, demand, prices, or activity durations is critical for managing businesses. In stable business environments, the probability distributions for such uncertain quantities can be estimated using historical data. However historical data may not be available or may not be directly usable in dynamic business environments — for example, when there are frequent introductions of new products or when there are changes in competition. In such environments, one must rely on domain experts to estimate a quantity's probability distribution, typically through its mean and variance.

Unfortunately, experts do not perform well at directly providing estimates of the mean and variance of a distribution. Garthwaite et al. (2005) notes that experts have "serious misconceptions about variances" and acutely underestimate the variance (p. 686). This poor estimation of variance by experts is not surprising since variance is a second-order moment, it is difficult to compute mentally, and its physical meaning is difficult to grasp. Empirical evidence, such as that provided by Peterson and Miller (1964) and Zapata-Vázquez et al. (2014), shows that experts also provide poor estimates, to a lesser degree, of the mean — the first-order moment of a distribution.

In contrast, empirical evidence (see Abbas et al. 2008) suggests that experts provide good quantile judgments — subjective estimates for quantiles or fractiles[1] — of the distribution of an uncertain quantity. This is because quantiles have a more intuitive probabilistic interpretation, and experts find it simpler to relate quantiles to their contextual domain. For example, the median demand has the appealing interpretation of 50–50 odds of the demand being higher or lower than that value. The quantiles of a probability distribution are related to the distribution's mean and standard deviation. Therefore, quantile judgments provided by an expert contain information about the mean and standard deviation of the probability distribution.

However a systematic approach that uses quantile judgments to deduce the mean and standard

[1] The terms "fractile", "percentile", and "quantile" are used in the literature for the same quantity. We use "quantile" in the rest of the paper.

deviation of the probability distribution while accounting for judgmental errors has not yet been established. In this paper we develop an approach to accomplish this task. The approach is analytically tractable, it provides the flexibility of eliciting any set of quantiles from an expert, and it is amenable to combining an expert's quantile judgments with data in a Bayesian fashion, and with other experts' quantile judgments in a manner that is consistent with the extant theory on combining point forecasts from multiple experts. It also enables us to establish a new equivalence between the precision of an expert's judgments and the size of a sample of experimental data; this equivalence can be used to rank order experts. Finally, we discuss a large-scale implementation of the theory and insights for practice gleaned during the implementation.

The industry motivation and application for this research comes from a $800 million annual business decision at Dow AgroSciences. The firm produces and sells seed corn that farmers buy and plant to grow corn for human and animal consumption. The firm decides annually how many acres of land to plant with seed corn. The yield, or amount of seed corn obtained per acre of land, is random; contextual information suggests that the yield uncertainty is distributed Normally. Under this uncertainty, planting a very large area may result in a surplus with a large upfront production cost if the realized yield is high; planting a small area may result in costly shortages if the realized yield is low. Mathematical models that incorporate the yield distribution can determine the optimal area of land to plant, but the historical yield data are usually insufficient for obtaining a statistical distribution of the yield. This is due to the high turnover rate of seeds in this industry — a specific seed is grown fewer than five–six times before it is discontinued. Due to the lack of historical data for new hybrid seeds, the yield distributions must be estimated by a yield expert. Dow AgroSciences recognizes the need for a formal approach to estimating yield distributions that will calibrate the current and future experts' estimates, help make production decisions, and track improvement in judgments for yield distributions over time. Our analytical approach accomplishes these tasks.

1.2. Overview of Approach and Our Contributions

1.2.1. Overview We develop a two-step approach to deduce the mean and standard deviation from quantile judgments provided by an expert. In Step 1, the errors in his quantile judgments are quantified by comparing his judgments against the quantiles of multiple distributions obtained using historical data (e.g., observed yields for seeds grown repeatedly in the past). In Step 2, the error information is used as an input to a prescriptive model. The solution to this model assigns two sets of weights to the quantile judgments. The first set of weights is for estimating the mean as the weighted linear combination of quantile judgments. The second set of weights is used similarly to estimate the standard deviation. The weights are specific to the expert's judgmental errors. The estimates of mean and standard deviation obtained using these weights are unbiased and have the smallest possible variance.

Many behavioral reasons have been posited for the presence of judgmental errors in quantile judgments. These reasons include an expert's bounded understanding of the drivers of the uncertainty, use of approximate mental models to translate qualitative information into quantitative judgments, tiredness, and pure random variation (Blattberg and Hoch 1990, Keelin and Powley 2011). These errors can be decomposed into two components, bias and inconsistency. Bias represents an average overestimation or underestimation in the value of a quantile. Inconsistency represents the residual random variations in the estimation process. Our approach accounts for both components.

1.2.2. Contributions to Theory Our focus on obtaining estimates of the mean and standard deviation from quantile judgments addresses a gap between two streams of decision analysis literature: (i) judgmental error literature, and (ii) moment estimation literature. The existing judgmental error literature has focused on developing elicitation guidelines for reducing these errors through decomposition (e.g., Clemen and Ulu 2008), expert calibration (e.g., Budescu et al. 1997, Koehler et al. 2002), and combination of assessments from various experts (e.g., Clemen and Winkler 1999, Baker and Olaleye 2013). However, this literature has not addressed the problem of estimating moments from subjective quantile judgments.

In contrast, the moment estimation literature has explored the problem of deducing moments from quantiles *numerically* with a key assumption: *there are no judgmental errors.* Pearson and Tukey (1965), Keefer and Bodily (1983), Keefer and Verdini (1993), and Keefer (1994) follow this paradigm. They select specific test cases of means and standard deviations of a distribution and obtain the 5th, 50th, and 95th quantiles, or other specific symmetric quantiles, for these cases. Then they consider various sets of candidate weights. For each set of weights, they estimate the means of all test cases as weighted linear combinations of the quantile values. Finally, they identify the set of weights that results in the smallest deviations between the true and the estimated means over all cases. A similar process is followed to identify the set of weights to determine standard deviation or variance. Lau et al. (1996, 1998), Lau and Lau (1998), and Lau et al. (1999) have a similar focus and assume that experts can always provide error-free judgments for specific quantiles, typically the median and two/four/six specific symmetric quantiles.

This paper addresses this gap by developing a practical but analytically rigorous way to estimate mean and standard deviation from subjective quantile judgments, thereby extending the moment estimation literature that has ignored judgmental errors, and extending the judgmental error literature that acknowledges the errors but has not addressed the moment estimation-problem.

The approach developed in this paper also provides the following generalizations to the existing decision analysis literature.

- **Flexible Elicitation:** The approach provides weights for *any* set of quantile judgments that an expert can provide to estimate the mean and standard deviation, as opposed to the existing decision analysis literature that focuses on the elicitation of median and two/four/six specific symmetric quantiles. This is a useful generalization for decision analysis practice since an expert may not be willing to provide quantile judgments for specific symmetric quantiles. For example, the expert at Dow was habituated to seeing the 10th, 50th, and 75th quantiles for historical data on his software, and was willing to estimate only these quantiles.
- **Establishing Structural Summation Properties of Weights:** The prescriptive model shows that regardless of the magnitude of an expert's judgmental errors and the fractiles elicited, the weights for the estimation of the mean and standard deviation add up to 1 and 0, respectively. This structural property explains the numerical findings in Pearson and Tukey (1965), Lau et al. (1999, 1996) and other moment-estimation literature that has assumed judgmental errors not to exist. The assumption of absence of judgmental errors considered in these articles is a special case in our model when the judgmental errors are set equal to 0.

- **Novel Quantification of Expertise into Equivalent Sample Size:** Our approach establishes a new quantification of expertise: it specifies the size of a random sample that would provide estimates of mean and standard deviation with the same precision as that of the estimates obtained using expert's judgments for quantiles. This equivalence enables an objective comparison of experts.

- **Framework for Combining Estimates from Multiple Experts and Data:** In our approach, the optimal weights provide point estimates and the variability in the estimates of the moments. This information for variability enables us to combine quantile judgments from multiple experts in a manner that is consistent with the extant theory of combining point forecasts, and combine parameter estimates with data as they become available, in a Bayesian fashion.

1.2.3. Contribution to Practice: Section 6 focuses on practice. We provide a step-by-step approach to quantifying an expert's judgmental errors, and then discuss in detail some practical issues observed during this quantification at Dow. Specifically, we discuss a bootstrapping approach to separate judgmental errors from sampling errors during the error quantification process, and show that the information provided by the expert is equivalent to 5–6 years of data collection at the firm using our approach. We also discuss our experience with the potential drawbacks of eliciting extreme quantiles during the implementation. We found that experts sometimes are reluctant to provide judgments for extreme quantiles because of their (experts') inability to distinguish between random variations and systematic reasons as causes of extreme outcomes.

The rest of this paper is organized as follows: In Section 2, we specify a model of judgmental errors and use it to develop a prescriptive model for deducing mean and standard deviation from quantile judgments for distributions with a location-scale family. We derive the solution in Section 3 and show that it explains a number of numerical results in the extant literature. In Section 4, we establish an equivalence of expertise with randomly collected data, and then discuss combining judgments from multiple experts and combining judgments with data. Section 5 extends the results to Johnson distributions. Section 6 describes the error quantification process that feeds into the prescriptive model, its implementation at Dow AgroSciences, and analysis using Dow's data. Section 7 concludes with a summary.

2. Two-Step Approach

We consider a real valued continuous random variable X, whose distribution is to be estimated. The probability density function (pdf) of X is denoted as $\phi(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]^{\mathrm{T}}$ are the parameters of the pdf. Similar to Lindley (1987), O'Hagan (2006) and others, we assume that the distribution family is known from the application context, but the parameters are not known. The cumulative distribution function (cdf) of X is denoted as Φ. The expert provides her quantile judgments $\hat{x}_i$ corresponding to probability cdf values p_i for $i = 1, 2, ..., m; m > n$. In vector notation, we denote the quantile judgments as $\hat{\mathbf{x}} = [\hat{x}_1, \ldots, \hat{x}_m]^{\mathrm{T}}$ and probability values as $\mathbf{p} = [p_1, \ldots, p_m]^{\mathrm{T}}$. The quantile judgments $\hat{\mathbf{x}}$ are not always equal to the true values $\mathbf{x}$. Specifically, the expert's judgment $\hat{x}_i$ is composed of a true value x_i and a judgmental error e_i:

$$\hat{x}_i = x_i + e_i \tag{1}$$

In vector notation, the error model is $\hat{\mathbf{x}} = \mathbf{x} + \boldsymbol{e}$. This error model has been used by Wallsten and Budescu (1983), Ravinder et al. (1988), and others. The proposed approach to deduce the mean and standard deviation consists of two steps; in Step 1, we quantify the expert's judgmental errors, and in Step 2, we set up an optimization model that uses this error information to obtain the best set of weights for the quantile judgments. These two steps are discussed in Sections 2.1 and 2.2, respectively.

2.1. Step 1: Quantification of Judgmental Errors

Mathematically, the error e_i has two parts: bias δ_i and residual variation ϵ_i, such that $e_i = \delta_i + \epsilon_i$ and $E[\epsilon_i] = 0$. The bias δ_i captures the average deviation of the expert's judgments for quantile i from the true value. The residual variation ϵ_i captures the spread in the error due to random variations or noise in the expert's estimation process for quantile i. A comparison process using historical data can quantify the bias and residual variation. In this process, the expert provides his quantile judgments $\hat{x}_{il}$ for a number of distributions l=1,2,...,L of uncertainties. For expositional ease, we assume for now that the true values x_{il} of these quantiles are available (in Section 6.4 we consider the case when the available values x_{il} are subject to sampling variations). For each distribution l, we then compare x_{il} with the expert's judgments $\hat{x}_{il}$ and determine the errors as $\hat{e}_{il} = \hat{x}_{il} - x_{il}$, and determine the bias for the quantile i as $\hat{\delta}_i = \sum_{l=1}^{L} \hat{e}_{il}/L$. We subtract this bias from individual errors to obtain unbiased errors, $\hat{\epsilon}_{il} = \hat{e}_{il} - \hat{\delta}_i$. Then we determine the symmetric $m \times m$ variance-covariance matrix Ω of the unbiased errors $\hat{\epsilon}_{il}$. This matrix summarizes the residual variation in the expert's judgments for the m quantiles. The diagonal elements of this matrix $\omega_{ii} = Var(\epsilon_i)$ denote the variance in the unbiased error of quantile i. The off-diagonal elements are covariances of errors $\omega_{ij} = Cov(\epsilon_i, \epsilon_j)$. The unbiased quantile judgments are obtained as $\hat{\mathbf{q}} = \hat{\mathbf{x}} - \hat{\boldsymbol{\delta}}$ or using $\hat{\mathbf{x}} = \mathbf{x} + \hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\epsilon}}$ as $\hat{\mathbf{q}} = \mathbf{x} + \hat{\boldsymbol{\epsilon}}$. Then the error information Ω is input into the optimization model discussed next.

2.2. Step 2: Optimization Problem Given Judgmental Errors

We seek to derive weights $\mathbf{w}_\mu = [w_{\mu 1}, w_{\mu 2}, ..., w_{\mu m}]^{\mathrm{T}}$ and $\mathbf{w}_\sigma = [w_{\sigma 1}, w_{\sigma 2}, ..., w_{\sigma m}]^{\mathrm{T}}$ to obtain the estimates of the mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ as weighted linear functions $\hat{\mu} = {\mathbf{w}_\mu}^{\mathrm{T}}\hat{\mathbf{q}}$ and $\hat{\sigma} = {\mathbf{w}_\sigma}^{\mathrm{T}}\hat{\mathbf{q}}$. Since the unbiased judgments $\hat{\mathbf{q}}$ are subject to variations specified by Ω, the estimates $\hat{\mu}$ and $\hat{\sigma}$ have variances $Var[{\mathbf{w}_\mu}^{\mathrm{T}}\hat{\mathbf{q}}]$ and $Var[{\mathbf{w}_\sigma}^{\mathrm{T}}\hat{\mathbf{q}}]$, respectively. Smaller values of the variances of these estimates are desirable as it would imply that the estimates are more precise. Accordingly, we seek weights $\mathbf{w}_\mu$ and $\mathbf{w}_\sigma$ that minimize the variance in the estimates $\hat{\mu}$ and $\hat{\sigma}$, respectively. This can be accomplished by solving the following optimization problem twice, first with $a = \mu$, and then with $a = \sigma$.

$$\min_{\mathbf{w}_a} Var[{\mathbf{w}_a}^{\mathrm{T}}\hat{\mathbf{q}}] \tag{2}$$

$$s.t. \quad \mathrm{E}\left[{\mathbf{w}_a}^{\mathrm{T}}\hat{\mathbf{q}}\right] = a \tag{3}$$

The optimization constraint (3) imposes another property that is intuitively desirable: the estimates are unbiased — i.e. on average the estimated values $\hat{\mu}$ and $\hat{\sigma}$ are equal to the true value of the mean and standard deviation, respectively.

Although this problem formulation is intuitively appealing, it is not operational since the value of a is needed in (3) to obtain the weights $\mathbf{w}_a$, but the value of a is unknown and we seek weights $\mathbf{w}_a$ to estimate it. No analytical solution is known to exist for this problem in the decision analysis literature. Nevertheless, we establish in the next section that for random variables with distributions of a location–scale family, properties of these probability distributions can be used to restate the problem in a form that does not have this circularity and that is analytically solvable.

3. Weights for Quantile Judgments

In Section 3.1 we reformulate the problem (2)–(3) for distributions of a location-scale family which are used extensively to model symmetric and skewed uncertainties. The results for these distributions are also a basis for deducing the moments of the Johnson-Family of distributions that provide a greater flexibility in modeling uncertainties. The optimal weights for the reformulated problem and a discussion on the structural properties of the optimal weights is in Sections 3.2 and 3.3, respectively.

3.1. Reformulation for Distributions of a Location-Scale Family

We first transform the formulation (2)–(3) to an analytically tractable form using two properties of distributions of a location–scale family. The first property connects the quantiles of the distribution of a location–scale family to its parameters in a linear fashion: Let $\theta_1 \in \mathbb{R}$ and $\theta_2 \in \mathbb{R}_{++}$. Then if X is a location–scale random variable with pdf $\phi(\cdot;\boldsymbol{\theta})$ where θ_1 and θ_2 are location and scale parameters respectively, then a specific value x can be expressed as $x = \theta_1 + \theta_2 z$ where z denotes the value of *standardized* random variable Z with the *standardized* pdf $\phi(\cdot;\boldsymbol{\theta})$ at parameter values $\boldsymbol{\theta} = [0,1]^{\mathrm{T}}$(Casella and Berger 2002).

Using the property $x = \theta_1 + \theta_2 z$, we can write the relationship $\hat{q}_i = x_i + \epsilon_i$ between the unbiased quantile judgment and the true value as $\hat{q}_i = \theta_1 + \theta_2 z_i + \epsilon_i$, where z_i is the z value for probability p_i. In a more compact notation, we can write the relationship between unbiased judgments $\hat{\mathbf{q}}$ and the parameters $\boldsymbol{\theta}$ as

$$\hat{\mathbf{q}} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{4}$$

where $\mathbf{Z}$ is the $m \times 2$ matrix formed as $\mathbf{Z} = [\mathbf{1}, \mathbf{z}]$, $\mathbf{z}$ is the column vector of standardized quantiles corresponding to the probabilities $\mathbf{p}$, and $\mathbf{1}$ is a column vector of ones.

The second property below connects μ,σ to $\boldsymbol{\theta}$ linearly.

Lemma 1 *Let X be a random variable with finite j^{th} moments for $j=1,2,\ldots,k$, Then*

$$\begin{aligned}\mu &= \mu_1(\boldsymbol{\theta}) = [1,\kappa_1]\,\boldsymbol{\theta}, \quad \textit{and}\\ \sigma &= \sqrt{\mu_2(\boldsymbol{\theta})-\mu_1(\boldsymbol{\theta})^2} = \left[0,\sqrt{\kappa_2-\kappa_1^2}\right]\boldsymbol{\theta}\end{aligned} \tag{5}$$

if and only if X has a distribution with a location–scale family, where $\boldsymbol{\kappa}=[\kappa_0,\kappa_1,\kappa_2,\ldots]$ are the kth raw moments $\mu_k(0,1)$ of the corresponding standardized random variable.

The values of $\boldsymbol{\kappa}$ are documented in the literature, for example in Johnson et al. (1994). If X is a Normally distributed random variable, we have $(\kappa_0,\kappa_1,\kappa_2)=(1,0,1)$; other distributions of location–scale family include the Logistic distribution $(\kappa_0,\kappa_1,\kappa_2)=(1,0,\pi^2/3)$; the Laplace $(\kappa_0,\kappa_1,\kappa_2)=(1,0,2)$; and the Gumbel $(\kappa_0,\kappa_1,\kappa_2)=(1,-\gamma,\gamma^2+\frac{\pi^2}{6})$. We will use the generalized form $a=\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}; a\in\{\mu,\sigma\}$, where $\mathbf{a}^{\mathrm{T}}=[1,\kappa_1]$ for mean $a=\mu$, and $\mathbf{a}^{\mathrm{T}}=\left[0,\sqrt{\kappa_2-\kappa_1^2}\right]$ for standard deviation $a=\sigma$ in the subsequent discussion.

Using the two properties of distributions of a location-scale family, we now reformulate the problem (2)–(3). We start with constraint (3) and using (4) rewrite it as:

$$\mathrm{E}\big[\mathbf{w}_a^{\mathrm{T}}\hat{\mathbf{q}}\big]=\mathrm{E}\big[\mathbf{w}_a{}^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon})\big]=\mathbf{w}_a^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta}=\mathbf{a}^{\mathrm{T}}\boldsymbol{\theta};\forall\theta_1\in\mathbb{R},\theta_2\in\mathbb{R}_{++}\implies\mathbf{w}_a^{\mathrm{T}}\mathbf{Z}=\mathbf{a}^{\mathrm{T}} \tag{6}$$

The first simplification is obtained because the errors are unbiased, $E[\boldsymbol{\epsilon}]=0$, and the final simplification results because the third equality must be valid for all $\theta_1\in\mathbb{R},\theta_2\in\mathbb{R}_{++}$.

Similarly, substituting (4) in the objective function (2), we obtain:

$$\begin{aligned}\mathrm{Var}\big[\mathbf{w}_a^{\mathrm{T}}\hat{\mathbf{q}}\big]&=\mathrm{E}\Big[\big(\mathbf{w}_a^{\mathrm{T}}\hat{\mathbf{q}}-\mathrm{E}\big[\mathbf{w}_a^{\mathrm{T}}\hat{\mathbf{q}}\big]\big)^2\Big]=\mathrm{E}\Big[\big(\mathbf{w}_a^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon})-\mathbf{w}_a^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta}\big)^2\Big]\\&=\mathrm{E}\left[\mathbf{w}_a^{\mathrm{T}}\boldsymbol{\epsilon}\,\boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{w}_a\right]=\mathbf{w}_a^{\mathrm{T}}\Omega\,\mathbf{w}_a.\end{aligned} \tag{7}$$

With these two relationships, the optimization problem (2)–(3) of obtaining the weights leading to estimates of the mean and standard deviation is rewritten as follows:

$$\begin{aligned}\min_{\mathbf{w}_a}\quad&\mathbf{w}_a^{\mathrm{T}}\Omega\,\mathbf{w}_a\\ s.t.\quad&\mathbf{w}_a^{\mathrm{T}}\mathbf{Z}=\mathbf{a}^{\mathrm{T}}\end{aligned} \tag{8}$$

where the solution is $\hat{\mu}=\mathbf{w}_\mu^{*\,\mathrm{T}}\hat{\mathbf{q}}$ for $\mathbf{a}^{\mathrm{T}}=[1,\kappa_1]$ and $\hat{\sigma}=\mathbf{w}_\sigma^{*\,\mathrm{T}}\hat{\mathbf{q}}$ for $\mathbf{a}^{\mathrm{T}}=\left[0,\sqrt{\kappa_2-\kappa_1^2}\right]$. This problem is now tractable and we discuss its solution next.

3.2. Optimal Weights and Structural Properties

We first establish the uniqueness of the solution to problem (8). The matrix Ω is a covariance matrix, and therefore it is positive definite. It follows that the problem (8) is a quadratic–convex problem. The next result follows from convexity.

Lemma 2 *There is a unique set of optimizing weights* $\mathbf{w}_a^*$ *for problem (8).*

These weights are obtained by solving a Lagrange formulation of problem (8). The Lagrange formulation is necessary to incorporate the constraint that the estimates obtained for the mean and standard deviation be unbiased.

Theorem 1 *The weights that solve problem (8) are given by* $\mathbf{w}_a^{*\mathrm{T}} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}$.

The weights $\mathbf{w}_\mu^*$ are obtained by setting $\mathbf{a}^{\mathrm{T}} = [1, \kappa_1]$, and the weights $\mathbf{w}_\sigma^*$ are obtained by setting $\mathbf{a}^{\mathrm{T}} = \left[0, \sqrt{\kappa_2 - \kappa_1^2}\right]$ in Theorem 1. The conspicuous feature of the optimal weights $\mathbf{w}_\mu^{*\mathrm{T}}, \mathbf{w}_\sigma^{*\mathrm{T}}$ is that they are explicit functions of the expert's precision encoded in Ω. It follows that different experts may have different degrees of precisions in their judgmental estimates as captured in different matrices Ω. And these differences will lead to different sets of weights for the experts.

Example: At Dow AgroSciences, we elicited the 10th, 50th, and 75th quantiles of the Normally distributed yields. For these quantiles and the Normal distribution, $\mathbf{Z}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 \\ -1.285 & 0 & 0.674 \end{bmatrix}$. Suppose that the covariance matrix is $\Omega = \begin{bmatrix} 70 & 0 & 44 \\ 0 & 40 & 0 \\ 44 & 0 & 90 \end{bmatrix}$. Using Theorem 1, the weights for the three quantiles for obtaining the mean would be $\mathbf{w}_\mu^{*\mathrm{T}}$ =(0.13, 0.63, 0.24) and the weights for the computation of the standard deviation would be $\mathbf{w}_\sigma^{*\mathrm{T}}$=(-0.54,0.10, 0.44). If the covariance matrix of a different expert is $\Omega = \begin{bmatrix} 70 & 0 & 55 \\ 0 & 20 & 0 \\ 55 & 0 & 70 \end{bmatrix}$, the weights would change to $\mathbf{w}_\mu^{*\mathrm{T}}$ =(0.08, 0.76, 0.16) and $\mathbf{w}_\sigma^{*\mathrm{T}}$ =(-0.52,0.03,0.49).

Another desirable property of the weights $\mathbf{w}_\mu^*$ and $\mathbf{w}_\sigma^*$ obtained in Theorem 1 is that they are functions of the standardized quantiles $\mathbf{z}$ and hence of the elicited probabilities $\mathbf{p}$, but they do not depend on the actual values of μ, σ (as illustrated in the example above). Therefore, the variances in the estimates $\hat{\mu}, \hat{\sigma}$ do not depend on the true values of μ, σ. This feature means that the structural properties and the analysis of the weights discussed in the remainder of the paper are universally applicable for all true values of μ, σ.

3.3. Flexible Elicitation and Generalization of Results in Extant Literature

In this section, we discuss several flexibilities and generalizations provided by our approach over the existing decision analysis literature. The first generalization of our approach is that the expert can provide judgments for any set of quantiles that he is comfortable estimating, i.e. he is no longer restricted to providing his judgments for the median and 2/4/6 specific symmetric quantiles specified in extant literature such as Lau et al. (1996), Lau and Lau (1998), and Lau et al. (1999). This flexibility is also useful for a decision analyst who no longer needs to convince an expert to provide judgments for specific quantiles and instead can focus on understanding why the expert believes that he can provide better judgments for his chosen quantiles. We discuss one such example in Section 6.2. The second generalization of our approach is that it provides an analytical foundation to a numerical property observed consistently in the extant literature that the weights add up to a constant.

Proposition 1 *The weights for quantiles add up to constants. Specifically,* $\sum_{i=1}^{m} w^*_{\mu i} = 1$, *and* $\sum_{i=1}^{m} w^*_{\sigma i} = 0$.

This result is true regardless of the numerical values of Ω; therefore it holds true even when the judgmental errors are negligible or are ignored. This is the framework discussed in a number of articles in the decision analysis literature (see Pearson and Tukey 1965, Perry and Greig 1975, Johnson 1998, Keefer and Verdini 1993). This literature, after a numerical search, assigns specific sets of weights to the error-free estimates of symmetric quantiles (e.g., 5th, 50th, and the 95th quantiles), to estimate the mean and standard deviation. These weights add up to 1 and 0 for the mean and standard deviation, respectively. For example, Pearson and Tukey (1965) suggest weights of 2/3 for the median and weights of 1/6 for two symmetric quantiles (5th and 95th) to estimate the mean (equations 9 and 10 in their paper), and equal weights with opposite signs for two symmetric quantiles to estimate the standard deviation. Johnson (2002, 1998) also suggests weights for the median and four/six symmetric quantiles; their suggested weights for the mean add up to 1, and the weights for the standard deviation add up to 0. Proposition 1 establishes that the additivity properties observed numerically in these articles are structural properties of probability distributions and hold true even when quantile judgments are subject to unbiased judgmental errors.

Third, our approach validates observations made in the literature on the use of identical weights for the estimation of the first and second moment. Articles such as Keefer and Bodily (1983) use the same set of weights for the estimation of the mean and the variance. They report a good accuracy of these common weights to estimate the mean but not the standard deviation. In light of Proposition 1 and Theorem 1, this poor performance for the estimation of standard deviation is not surprising. Theorem 1 states that the optimal weights for estimating the mean and standard deviation are different; the weights for the standard deviation depend on κ_1, κ_2 but the weights for the mean depends only on κ_1. Their summation properties are also different as established in Proposition 1. Therefore, using the optimal weights for the estimation of the mean to estimate the standard deviation or vice-versa can result in a poor match between the assessed and true value.

A number of articles in the extant literature have relied on the assumption that judgmental errors are absent. The fourth generalization of our approach is that the weights for this assumption are obtained by substituting $\Omega = K\Omega_I$ where Ω_I is an identity matrix in Theorem 1, as a special case. We formalize this result in Proposition 5 in Section 4.4. Fifth, using Theorem 1, we can determine the optimal weights when Ω changes. For illustrative purposes, consider unbiased judgments for the median $\hat{q}_2$ and two other quantiles $\hat{q}_1, \hat{q}_3$, one in each tail, when the standardized value of the median $z_2 = 0$, as is the case with symmetric distributions. The weight on the median for the estimation of the mean is equal to $w_{\mu_2} = \frac{c_1 - \sqrt{Var(\hat{q}_2)}(z_1 - z_3)c_2}{c_3 Var(\hat{q}_2) - 2\sqrt{Var(\hat{q}_2)}(z_1 - z_3)c_2 + c_1}$, and for the estimation of the standard deviation is equal to $w_{\sigma 2} = \frac{c_4 - \sqrt{Var(\hat{q}_2)}(z_1 - z_3)c_2}{c_3 Var(\hat{q}_2) - 2\sqrt{Var(\hat{q}_2)}(z_1 - z_3)c_2 + c_1}$, where $c_1 = z_3^2 Var(\hat{q}_1) + z_1^2 Var(\hat{q}_3) - 2z_1 z_3 \rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)}$, $c_2 = z_1\rho_{23}\sqrt{Var(\hat{q}_3)} - z_3\rho_{12}\sqrt{Var(\hat{q}_1)}$, $c_3 = (z_1 - z_3)^2$, and $c_4 = z_3 Var(\hat{q}_1) + z_1 Var(\hat{q}_3) + (z_1 + z_3)\rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)}$ (details in the proof for Proposition 2).

Suppose that the expert's precision for the estimate of the median improves — i.e., $Var(\hat{q}_2)$ decreases. One would expect the weight $w^*_{\mu 2}$ to progressively converge to 1 and the remaining two weights $w^*_{\mu_1}, w^*_{\mu_3}$ for the 10th and the 75th quantiles respectively to converge individually to 0. Similarly, we would also expect the weight for the estimation of the standard deviation $w^*_{\sigma 2}$ to converge as $Var(\hat{q}_2)$ decreases. The following result validates this intuition.

Proposition 2 *(i) For the estimation of the mean,* $\lim_{Var(\hat{q}_2)\to 0} w^*_{\mu 2} = 1$. *Furthermore, this limit is reached from above if* $z_1\rho_{23}\sqrt{Var(\hat{q}_3)} \geq z_3\rho_{12}\sqrt{Var(\hat{q}_1)}$. *Otherwise it is reached from below.*

(ii) For the estimation of the standard deviation, $\lim_{Var(\hat{q}_2)\to 0} w^*_{\sigma 2} = c_4/c_1$*. Furthermore, this limit is reached from above if* $\rho_{23}\sqrt{Var(\hat{q}_3)} \geq \rho_{12}\sqrt{Var(\hat{q}_1)}$*. Otherwise it is reached from below.*

This process of reassigning weights to quantile judgments in response to an improved precision is consistent with the practice of assigning weights to multiple point assessments that are proportional to the precision of the assessments to obtain one point estimate, as discussed in Griffin and Tversky (2002). In addition, one can determine the impact of eliciting one specific quantile instead of another. For example, assuming that the expert can provide different quantiles with equal judgmental error, one could determine the slope $\frac{\partial([1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}}{\partial z_i}$ keeping Ω constant and determine the value of z_i at which the variance $[1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}$ is minimum.

4. Exploiting Quantification of Judgmental Errors

We now discuss four benefits of quantifying the expert's judgmental errors using our approach. In Section 4.1, we determine the size of a randomly drawn sample that is equivalent to the expert's expertise. In Sections 4.2 and 4.3, we discuss combining judgments of one expert with the judgments from other experts, and with data, respectively. In Section 4.4, we quantify the differences in precisions of the estimates obtained using our approach over approaches ignoring judgmental errors.

4.1. Equivalence Between Expertise and Size of a Random Sample

Expert input is sought for estimating probability distributions when collecting data is costly. The expert's quantile judgments, after using our approach, lead to point-estimates of μ, σ and the variances in these estimates just as if we had collected sample observations of the random variable X. Specifically, it is well known that a sample mean has a variance of σ^2/N_μ where N_μ is the sample size. In our quantile based approach, the variance in the estimate of the mean is equal to $Var[\hat{\mu}] = \mathbf{w}^{*\mathrm{T}}_\mu \Omega\, \mathbf{w}^*_\mu$, which can be simplified to $[1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}$ (details are in Section 6 of Appendix). By equating these two variances, we can determine the size of a randomly collected sample that would provide the same precision of the estimate of the mean as the expert does. We call this size an *equivalent sample size* for the mean and denote it by N_μ. We can do a similar analysis for standard deviation, and obtain the equivalent sample size for the standard deviation.

Proposition 3 *The precision of the estimates $\hat{\mu}, \hat{\sigma}$ obtained using an expert's quantile judgments with judgmental error matrix Ω is comparable to the precision of estimates obtained from an iid sample of size N_a, where $N_\mu = \frac{\sigma^2}{[1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}}$, and $N_\sigma \approx \frac{\sigma^2 \left(\frac{\sum_{j=0}^{4}(-\kappa_1)^j \kappa_{4-j}}{(\kappa_2-\kappa_1^2)^2} - \frac{\left(\sum_{j=0}^{2}(-\kappa_1)^j \kappa_{2-j}\right)^2}{(\kappa_2-\kappa_1^2)^2} \right)}{4[0,\sqrt{\kappa_2-\kappa_1^2}](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0,\sqrt{\kappa_2-\kappa_1^2}]^{\mathrm{T}}}$.*

This result has profound implications for quantifying expertise as various experts can be ranked objectively based on the magnitude of their equivalent sample size, which depends directly on the judgmental errors quantified in Ω. This ranking is obtained even though the true values μ, σ remain unknown. If a decision analyst has some idea on the range of the value of σ from contextual details then the upper bound and lower bound on equivalent sample sizes are obtained by substituting the upper and lower values of the range in the expressions for N_μ, N_σ. This quantification also provides a tangible metric to assess the value of increased precision in the expert's judgments, e.g. Akcay et al. (2011) consider the benefit of estimating distribution moments from a larger sample and show that an increase in sample size from 10 points to 20 points reduces the operating cost of a system by 10–20%. Our quantification of expertise into data points provides a natural connection to these results.

4.2. Combining Estimates from Multiple Experts

The technical development so far is readily extendible to multiple experts in a manner that is consistent with the extant theory of combining point forecasts from multiple sources. To use our approach for multiple $j = 1, 2, ..., n$ experts, we use the matrix $\Omega = \begin{bmatrix} \Omega_{11}\ \Omega_{12} ...\ \Omega_{1n} \\ \Omega_{12}\ \Omega_{22} ...\ \Omega_{2n} \\ \vdots \\ \Omega_{1n}\ \Omega_{2n} ...\ \Omega_{nn} \end{bmatrix}$ where Ω_{11} is the $m \times m$ matrix for residual errors of expert 1, the matrix Ω_{12} is the $m \times m$ covariance matrix for the errors of experts 1 and 2, and so on. The matrix Ω is used in Theorem 1 along with matrix $\mathbf{Z}^{\mathrm{T}}$ of size $2 \times mn$, $\mathbf{Z}^{\mathrm{T}} = [\mathbf{Z}_1{}^{\mathrm{T}}\ \mathbf{Z}_2{}^{\mathrm{T}}, ..., \mathbf{Z}_n{}^{\mathrm{T}}]$, to obtain the weights for each quantile provided by each expert.

The weights thus obtained can then be decomposed as a product of a marginal weight that is specific to expert j and conditional weights for the m quantiles for this expert. The values of the experts' marginal weights are consistent with the extant literature on combining point estimates obtained from noisy sources or experts. Bates and Granger (1969) and Dickinson (1975) combine point estimates by assigning them weights that minimize the variance of the weighted linear estimate, and show that the optimal weight assigned to a point estimate is inversely proportional to the variance of the noise in the estimate's source. Thus, for two sources where the first source has half the variance in noise than the second source and both noises are mutually independent, the optimal weights for the forecasts of the two sources are 2/3 and 1/3, respectively. In our context, even though we seek to aggregate multiple quantile judgments from multiple experts as opposed to point estimates from multiple experts, and to obtain an aggregated estimate of the mean and standard deviation as opposed to an aggregated point forecast, the expert-specific weights obtained after a decomposition are consistent with this point-estimation literature.

For illustration, we consider the special case where n experts *j=1, 2, ..., n* provide judgments for the same set of quantiles, i.e. $\mathbf{Z}^{\mathrm{T}} = [\mathbf{Z}_0{}^{\mathrm{T}}\, \mathbf{Z}_0{}^{\mathrm{T}}, ..., \mathbf{Z}_0{}^{\mathrm{T}}]$, and we assume that the covariance matrix of each expert j is given by $\Omega_{jj} = r_j \Omega_0 \ \forall j$ and further assume that the errors of any two experts are mutually independent, i.e. all elements of $\Omega_{ij}, i \neq j$ are equal to 0. The weights for the mn quantile judgments obtained using Theorem 1 are denoted as $\mathbf{w}_a^*$ with elements w_{ak}^* for $k = 1, 2, ..., mn$ and $a \in \{\mu, \sigma\}$. The first m weights are for expert 1, the next m weighs are for expert 2 and so on. We can write these weights as $\mathbf{w}_a^* = \left[\mathbf{w}_a^1, ..., \mathbf{w}_a^n\right]$ where $\mathbf{w}_a^j$ is the vector of weights of expert *j*. We can also decompose the weights $\mathbf{w}_a^*$ as the product of marginal weights α_j for expert j and a common vector of m conditional weights $\mathbf{w}_a^c$ that would be obtained if each expert was the only one available, i.e., $\mathbf{w}_a^* = \left[\alpha_1 \mathbf{w}_a^c, ..., \alpha_n \mathbf{w}_a^c\right]$. The values of α_j and the relationships between $\mathbf{w}_a^j$ and $\mathbf{w}_a^c$ are as follows:

Proposition 4 *Consider experts $j = 1, 2, ..., n$ whose covariance matrices are $r_j \Omega_0; \forall j$, and further assume that the judgmental errors across experts are mutually independent. Then,*

(i) If any expert j was the only expert available, his optimal weights would be $\mathbf{w}_a^{c\mathrm{T}} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}\mathbf{Z}_0)^{-1}\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}$ independent of the value of r_j.

(ii) When the quantile judgments of the n experts are considered simultaneously, the weights for each expert j are obtained as $\mathbf{w}_a^j = \alpha_j \mathbf{w}_a^c$, with marginal weights $\alpha_j = (1/r_j)/R$, where $R = \sum_{j=1}^{n}(1/r_j)$.

As an illustration, suppose that we have two experts with $r_1 = 1, r_2 = 2$. The constant $R = (1/1) + (1/2) = (3/2)$. It follows from part *(ii)* of the Proposition that the expert-specific marginal weights are $\alpha_1 = (1/r_1)/(3/2) = 2/3$ and $\alpha_2 = (1/r_2)/(3/2) = 1/3$ regardless of the quantiles estimated. These are the same weights that one would have used while combining point estimates from multiple sources as discussed in Bates and Granger (1969) and Dickinson (1975).

Further consider the case when $\Omega_0 = \begin{bmatrix} 80 & 30 & 35 \\ 30 & 22 & 30 \\ 35 & 30 & 68 \end{bmatrix}$ for the estimation of the 10th, 50th and the 75th quantiles. If the quantile judgments of only expert j are considered separately, the estimation weights would be obtained as $[\mathbf{w}_\mu^c, \mathbf{w}_\sigma^c]^{\mathrm{T}} = \begin{bmatrix} -0.167 & 1.484 & -0.317 \\ -0.576 & 0.190 & 0.386 \end{bmatrix}$ for either expert j by using $\mathbf{Z}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 \\ -1.285 & 0 & 0.674 \end{bmatrix}$ and Ω_0 in Theorem 1, as stated in part *(i)* of Proposition 4. When both experts are available, the optimal weights for their quantile judgments are obtained by multiplying the independent weights $\mathbf{w}_a^c$ with the expert-specific marginal weight α_j as $\mathbf{w}_a^j = \alpha_j \mathbf{w}_a^c$. For example, the weights for the mean are obtained as $\mathbf{w}_\mu^1 = (2/3) \times (-0.167, 1.484, -0.317) = (-0.111, 0.989, -0.211)$ and $\mathbf{w}_\mu^2 = (1/3) \times (-0.167, 1.484, -0.317) = (-0.055, 0.495, -0.105)$ for the first and second expert, respectively. The same weights are also obtained directly by first constructing the combined matrix $\Omega = \begin{bmatrix} r_1\Omega_0 & \mathbf{0} \\ \mathbf{0} & r_2\Omega_0 \end{bmatrix}$ and using it in Theorem 1 with matrix $\mathbf{Z}^{\mathrm{T}} = [\mathbf{Z}_0{}^{\mathrm{T}}, \mathbf{Z}_0{}^{\mathrm{T}}] = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1.285 & 0 & 0.674 & -1.285 & 0 & 0.674 \end{bmatrix}$, which gives $\mathbf{w}_\mu^* = [-0.111, 0.989, -0.211, -0.055, 0.495, -0.105]$.

Finally, we note that the equivalent sample sizes N_μ, N_σ for a group of experts can be determined using Proposition 3. This size can be compared with equivalent sample sizes of individual experts to quantify the additional value provided by every expert.

4.3. Copula based Approach for Bayesian Updating Using Data

We now explain how to combine the quantile judgments with data, using results for Bayesian updating discussed in Gelman et al. (2003). The parameter estimates $\hat{\theta}_1, \hat{\theta}_2$ for the distribution are obtained from $\hat{\mu}, \hat{\sigma}$ using the linearity property discussed in Lemma 1. The Lemma states that $\mu = \theta_1 + \kappa_1\theta_2$ and $\sigma = \sqrt{\kappa_2 - \kappa_1^2}\theta_2$. Solving this linear system of equations provides $\theta_1 = \mu - \frac{\kappa_1}{\sqrt{\kappa_2 - \kappa_1^2}}\sigma$ and $\theta_2 = \frac{1}{\sqrt{\kappa_2 - \kappa_1^2}}\sigma$. Substituting the values of the estimates $\hat{\mu}$ and $\hat{\sigma}$ from Theorem 1, we obtain

$$\hat{\theta}_1 = \hat{\mu} - \frac{\kappa_1}{\sqrt{\kappa_2 - \kappa_1^2}}\hat{\sigma} = ([1, \kappa_1] - [0, \kappa_1])\,(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}} = [1, 0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}}$$

$$\hat{\theta}_2 = \hat{\sigma}\frac{1}{\sqrt{\kappa_2 - \kappa_1^2}} = [0, 1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}}$$

That is, $\hat{\theta}_i = \mathbf{a}_{\theta_i}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}} = {\mathbf{w}_{\theta_i}^{*}}^{\mathrm{T}}\hat{\mathbf{q}}; i = 1,2$ with $\mathbf{a}_{\theta_1}^{\mathrm{T}} = [1,0]$ and $\mathbf{a}_{\theta_2}^{\mathrm{T}} = [0,1]$. The expected values of the estimates are $E[\hat{\theta}_i] = \mathbf{a}_{\theta_i}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}}$, the variances are $\mathrm{Var}\left[\hat{\theta}_i\right] = \mathbf{a}_{\theta_i}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a}_{\hat{\theta}_i}$, and the covariance is $Cov[\hat{\theta}_1, \hat{\theta}_2] = \mathbf{a}_{\theta_1}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a}_{\theta_2}$ (see the details in Section 6 of Appendix). The correlation ρ of the two parameter estimates is $\rho = \frac{Cov[\hat{\theta}_1,\hat{\theta}_2]}{\sqrt{\mathrm{Var}\left[\hat{\theta}_1\right]\mathrm{Var}\left[\hat{\theta}_2\right]}}$. Therefore, we have available the moments of the distributions of the estimates of $\boldsymbol{\theta}$. These distributions are the prior distributions for $\boldsymbol{\theta}$ before any data collection. New data, when available, can be combined with these priors to obtain posterior distributions for $\boldsymbol{\theta}$.

To do this we first further characterize the prior distributions. We know the means, variances, and correlation of the prior distributions for θ_1, θ_2 as noted above. The parametric forms of the prior distributions are obtained during the error quantification process described in Section 2.1. For each distribution used in the error quantification process *l=1,2,...,L* with true values θ_{1l}, θ_{2l} we determine the unbiased errors in the estimation of θ_1, θ_2 as $e_{1l} = \theta_{1l} - {\mathbf{w}_{\theta_1}^{*}}^{\mathrm{T}}\hat{\mathbf{q}}_l$ and $e_{2l} = \theta_{2l} - {\mathbf{w}_{\theta_2}^{*}}^{\mathrm{T}}\hat{\mathbf{q}}_l$, respectively, where $\hat{\mathbf{q}}_l$ are unbiased quantile judgments for distribution l. Then we assume parametric families to accurately represent these two error streams. Let the distributions' cdfs be specified as F_1 and F_2, respectively. The parametric families of the two distributions may be different. The Normal Copula approach explained next is a general approach to estimate the joint prior mass on the domain of θ_1, θ_2 for such situations (for details, see Wang and Dyer 2012).

We first create an equi-spaced grid with points $\theta_1 = \theta_{1i}, \theta_2 = \theta_{2j}$; *i=1,2,...,I; j=1,2,...,J*, over the domain of $\theta_1 \in [\underline{\theta_1}, \bar{\theta}_1], \theta_2 \in [\underline{\theta_2}, \bar{\theta}_2]$. The domain can be guessed at first and then refined iteratively by removing those end regions that do not have any mass in the posterior. To determine the probability mass for the prior distribution on each point in grid, we use a Normal Copula based simulation (for details see Wang and Dyer 2012). We draw sets *r=1,2...,R* of two uncorrelated standard Normal values v_1^r, v_2^r and then translate them into draws θ_1^r, θ_2^r from the correlated prior distributions as

$$\theta_1^r = F_1^{-1}\left(\Phi_N\left(v_1^r\sqrt{(1+\rho)/2} + v_2^r\sqrt{(1-\rho)/2}\right); E[\hat{\theta}_1], Var(\hat{\theta}_1)\right) \tag{9}$$

$$\theta_2^r = F_2^{-1}\left(\Phi_N\left(v_1^r\sqrt{(1+\rho)/2} - v_2^r\sqrt{(1-\rho)/2}\right); E[\hat{\theta}_2], Var(\hat{\theta}_2)\right) \tag{10}$$

where Φ_N denotes the standard Normal cdf. Then we determine the grid-point where this draw belongs. After R iterations, we count the number of draws in each point in the grid to obtain the

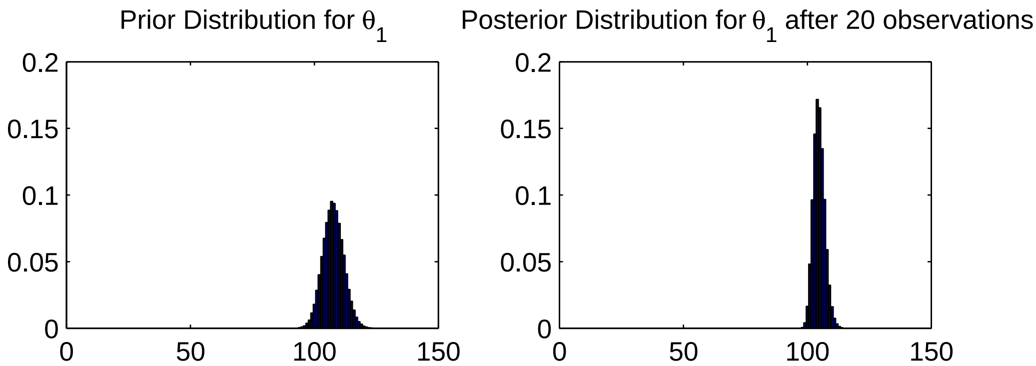


Figure 1 Bayesian updating of θ_1 for $\Omega = \begin{bmatrix} 78.99 & 29.38 & 33.33 \\ 29.38 & 21.92 & 30.30 \\ 33.33 & 30.30 & 68.17 \end{bmatrix}$ obtained at Dow with Gamma prior distributions. The expert was assumed to provide values of 74,105 and 115 for the 10th, 50th, and 75th quantiles. The updating was done with 20 points. See Section 9 of Appendix for details.

frequency of each point in the grid. Dividing this frequency by R, we obtain the prior probability mass $Pr(\theta_1 = \theta_{1i}, \theta_2 = \theta_{2j})$. The Bayesian updating process follows p. 284 of Gelman et al. (2003). When a sample of k=*1,2.,...,*K data points $\boldsymbol{y} = [y_1, y_2, ..., y_K]$ is available, the joint posterior mass function is obtained by visiting each point in the grid, multiplying its prior mass $Pr(\theta_1 = \theta_{1i}, \theta_2 = \theta_{2j})$ with the likelihood $\Pi_{k=1}^{K} \phi(y_k|\theta_{1i}, \theta_{2j})$ and then scaling by the total mass as,

$$Pr\left(\boldsymbol{\theta} = (\theta_{1i}, \theta_{2j})|\boldsymbol{y}\right) = \frac{\Pi_{k=1}^{K} \phi(y_k|\theta_{1i}, \theta_{2j}) Pr(\theta_1 = \theta_{1i}, \theta_2 = \theta_{2j})}{\sum_{i=1}^{I} \sum_{j=1}^{J} \Pi_{k=1}^{K} \phi(y_k|\theta_{1i}, \theta_{2j}) Pr(\theta_1 = \theta_{1i}, \theta_2 = \theta_{2j})} \tag{11}$$

The marginal posterior mass functions for θ_1, θ_2 are obtained from the joint mass function. Figure 1 provides an illustration of the Bayesian updating of θ_1 using 20 data points for our application context at Dow. A complete set of details along with the results for θ_2 is in Section 9 of Appendix.

4.4. Benefit of Incorporating Judgmental Errors

We now quantify the benefit of using the information on judgmental errors for estimating $\hat{\mu}, \hat{\sigma}$ over two benchmarks that ignore these errors.

Benchmark I: Ignoring Judgmental Errors and Eliciting m>*2* Quantiles: In the first benchmark, we ignore judgmental errors but elicit m quantile judgments (m>*2*) and find the weights that minimize the sum of squared errors: $\left\{ \sum_{i=1}^{m} \left(\mu + z_i \sigma - \hat{q}_i\right)^2 \right\}$. It can be shown that this problem provides estimates $\hat{a}^I = \mathbf{w}_a^{I*^{\mathrm{T}}} \hat{\mathbf{q}}$ for $a \in \{\mu, \sigma\}$ with the weights $w_{\mu i}^{I*} = \frac{(S_2 - S_1 z_i)}{m S_2 - S_1^2}$ and $w_{\sigma i}^{I*} = \frac{S_1 (S_2 - S_1 z_i)}{S_2 (m S_2 - S_1^2)} -$

$\frac{z_i}{S_2}$ where $S_1 = \sum z_i$ and $S_2 = \sum z_i^2$ (superscript I denotes Benchmark I), and that the problem is a special case of our optimization problem in (8) when the expert is equally proficient at all judgments:

Proposition 5 *The problem $Min_{\mu,\sigma}\left\{\sum_{i=1}^{m}\left(\mu + z_i\sigma - \hat{q}_i\right)^2\right\}$ is equivalent to problem* (8) *and its solution is obtained by assuming a non-informative covariance matrix $\Omega = K\Omega'$ where $K > 0$ is a scalar, the diagonals elements of Ω' are equal to 1, and the off-diagonal elements are equal to the correlation value ρ.*

Therefore the weights $\mathbf{w}_\mu^{I*}$ and $\mathbf{w}_\sigma^{I*}$ have all the structural properties discussed earlier in Section 3.3, including summation to 1 and 0 for the mean and standard deviation, respectively.

A pertinent issue here is when do the weights for this benchmark $\mathbf{w}_a^{I*}$ differ significantly from the optimal weights $\mathbf{w}_a^*$ discussed in Theorem 1? And what are the implications of using $\mathbf{w}_a^{I*}$ instead of $\mathbf{w}_a^*$? We answer the first question here and the second question in Proposition 6 where we establish that the estimates $\hat{a}^I = \mathbf{w}_a^{I*}\hat{\mathbf{q}}$ are less precise as compared to estimates $\hat{a} = \mathbf{w}_a^*\hat{\mathbf{q}}$. Proposition 5 implies that the weights $\mathbf{w}_a^{I*}$ coincide with the optimal weights $\mathbf{w}_a^*$ only when all variances of judgmental errors in Ω and all covariances of judgmental errors in Ω are equal — i.e. when the expert is equally proficient at estimating all quantiles. Therefore the optimal weights $\mathbf{w}_a^*$ capture <u>both</u> the quantile-specific information (through $\mathbf{Z}$) and the expert's variability information through Ω. When Ω is non-informative as is the case when the expert is equally proficient at estimating all quantiles, the weights $\mathbf{w}_a^{I*}$ capture <u>only</u> the quantile-specific information in $\mathbf{Z}$. It follows that the more information present in Ω (or equivalently the larger the differences in the expert's ability to estimate various quantiles) the larger the gap between the optimal weights $\mathbf{w}_a^*$ and the error ignoring weights $\mathbf{w}_a^{I*}$.

Figure 2 provides a numerical illustration for the covariance matrix $\Omega = \begin{bmatrix} 80 & 0 & 0 \\ 0 & 216 & 0 \\ 0 & 0 & 80 \end{bmatrix}$ when the 10th, 50th, and 75th quantiles are elicited. The figure shows the variations in weights $\mathbf{w}_\mu^*$ and $\mathbf{w}_\mu^{I*}$ for the estimates of the 10th and 50th quantiles as the variance in judgmental error for the median $Var(\hat{q}_2)$ decreases from 216. The values of $\mathbf{w}_\mu^*$ and $\mathbf{w}_\mu^{I*}$ coincide when the variance in judgmental error for the median is equal to 80. At this point, the expert is equally good at estimating all quantiles.

Benchmark II: Eliciting *m = 2* Quantiles: This benchmark builds on the previous one and was suggested by an anonymous reviewer: if judgmental errors are not to be considered, then elicit two

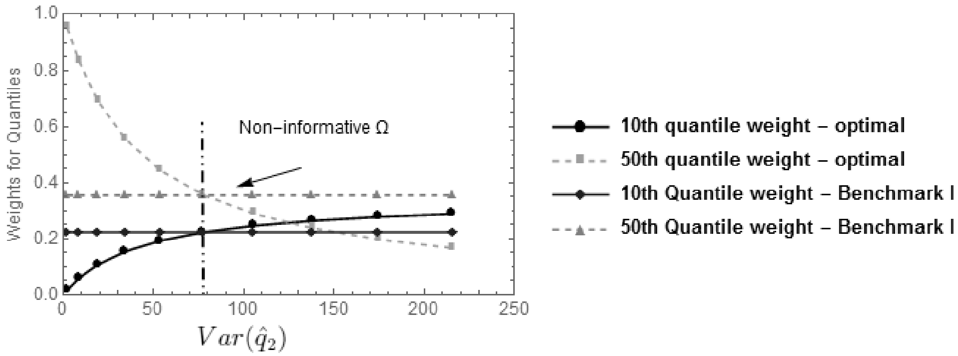


Figure 2 Change in the weights for the 10th, and 50th quantiles for the estimation of the mean as the precision of the expert's judgment of the 50th quantile changes.

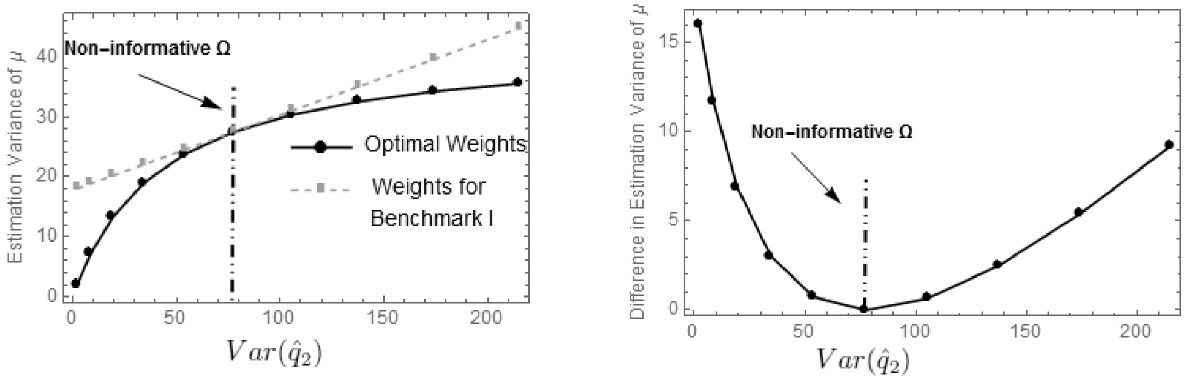
quantiles and deduce the two parameters from these quantiles by solving two simultaneous equations. For example, for quantile judgments $\hat{q}_i; i = 1, 2$ of a Normal distribution, we can write $\hat{q}_1 = \hat{\mu}^{II} + z_1\hat{\sigma}^{II}$ and $\hat{q}_2 = \hat{\mu}^{II} + z_2\hat{\sigma}^{II}$ (superscript II denotes Benchmark II). From these two equations, we obtain the estimates $\hat{\mu}^{II} = \frac{\hat{q}_1 z_2 - \hat{q}_2 z_1}{z_2 - z_1}$ and $\hat{\sigma}^{II} = \frac{\hat{q}_2 - \hat{q}_1}{z_2 - z_1}$. These estimates imply the weights $\mathbf{w}_\mu^{II*} = [\frac{z_2}{z_2 - z_1}, -\frac{z_1}{z_2 - z_1}]^{\mathrm{T}}$ for $\hat{q}_1$ and $\hat{q}_2$, respectively, to estimate the mean, and the weights $\mathbf{w}_\sigma^{II*} = [\frac{1}{z_2 - z_1}, -\frac{1}{z_2 - z_1}]^{\mathrm{T}}$ for $\hat{q}_1$ and $\hat{q}_2$, respectively, to estimate the standard deviation. These weights add up to 1 for the mean and to 0 for the standard deviation, consistent with the output of our model.

We now quantify the effect of using weights $\mathbf{w}_a^{I*}$ and $\mathbf{w}_a^{II*}$ from Benchmarks I and II, respectively, instead of the optimal weights $\mathbf{w}_a^*$. The variances of the estimates $\hat{a}, \hat{a}^I, \hat{a}^{II}$ for $a \in \{\mu, \sigma\}$ are equal to $\mathbf{w}_a^{*\mathrm{T}}\Omega\mathbf{w}_a^*, \mathbf{w}_a^{I*\mathrm{T}}\Omega\mathbf{w}_a^{I*}$ and $\mathbf{w}_a^{II*\mathrm{T}}\Omega\mathbf{w}_a^{II*}$ respectively. The following result establishes that the estimate $\hat{a}$ has the smallest variance. The equivalent sample size for the expert obtained using our approach N_a is also larger than the sizes N_a^I, N_a^{II} for the benchmarks.

Proposition 6 *(i)* $Var(\hat{a}^I) \geq Var(\hat{a})$ *and* $Var(\hat{a}^{II}) \geq Var(\hat{a})$ *for* $a \in \{\mu, \sigma\}$.

(ii) $N_a^I \leq N_a$ *and* $N_a^{II} \leq N_a$ *for* $a \in \{\mu, \sigma\}$.

The differences in the variances of the estimates and in the equivalent sample sizes from our approach and the two benchmarks can be significant. For brevity, we focus on the estimation of the mean. Figure 3(a) shows the variances of the estimates of mean obtained from our error-including approach and Benchmark I for the variance-covariance matrix $\Omega = \begin{bmatrix} 80 & 0 & 0 \\ 0 & 216 & 0 \\ 0 & 0 & 80 \end{bmatrix}$ when the variance in



(a) Variances in $\hat{\mu}$ and $\hat{\mu}^I$

(b) Difference in the variances of $\hat{\mu}$ and $\hat{\mu}^I$

Figure 3 Changes in the variance of the estimated mean as the expert's residual errors in the judgment for the 50th quantile change.

the judgmental errors for the median $Var(\hat{q}_2) = 216$ decreases (on the x-axis). As evident in the figure, the difference in the estimation variances varies significantly and is lowest when $Var(\hat{q}_2) = 80$. At this point, the matrix $\Omega = \begin{bmatrix} 80 & 0 & 0 \\ 0 & 80 & 0 \\ 0 & 0 & 80 \end{bmatrix}$ is non-informative, and the weights in our approach and Benchmark I coincide as shown in Figure 2. Figure 3(b) shows the numerical value of the difference in the variances. The effect is directionally even stronger for Benchmark II.

It follows from these figures that at $Var(\hat{q}_2) = 80$, $N_a^I = N_a$, but otherwise $N_a^I < N_a$. In Section 6.5, we show for Dow's application that $Var(\hat{a}) \approx Var(\hat{a}^I)/2 \approx Var(\hat{a}^{II})/4$ and $N_\mu \approx 2N_\mu^I \approx 4N_\mu^{II}$ suggesting that the benefits of using our approach over the two benchmarks can be significant.

5. Extensions to Johnson Distributions

The technical development discussed so far for distributions with a location–scale family also enables us to estimate the parameters of Johnson distributions used to model probability distributions with a greater degree of flexibility. The key connection between a distribution with a location–scale family and Johnson distributions is that a random variable X with Johnson distribution with parameters $\boldsymbol{\theta}$ results in a Normal variable Y with parameters $\boldsymbol{\theta}$ after a non-linear monotonic transformation, g – i.e., $g(X) = Y$ (Johnson 1949). For example, if X is a Lognormal random variable (type 2 Johnson variable), then $Y = ln(X)$ is a Normal random variable. In the absence of judgmental errors, the estimation of the parameters of X is straightforward. For any $x \in \mathbb{R}$ we have $\Pr\{X \leq x\} = \Pr\{g(X) \leq g(x)\}$ where $g(x)$ is non-decreasing in x; hence the p_i-quantiles of X and $g(X)$, denoted respectively as $x(p_i)$ and $x^g(p_i)$, satisfy the relationship $g(x(p_i)) = x^g(p_i)$. Therefore, one could simply take the inverse g^{-1} of the unbiased quantile judgments $\hat{q}_i$ to transform them on the underlying Normal distribution, and then estimate the parameters of this underlying Normal distribution using the results developed earlier. However, this process does not carry over to when the expert's judgments have errors. As an example consider the Lognormal distribution, any elicited quantile satisfies $ln(\hat{q}_i) = ln(x_i + \epsilon_i)$, but clearly $ln(\hat{q}_i) \neq ln(x_i) + ln(\epsilon_i)$. However, we can approximate $g(X)$ using the second–order Taylor series expansion of $g(x_i)$ about $\hat{q}_i = x_i + \epsilon_i$ as:

$$g(x_i) \approx g(\hat{q}_i) + g'(\hat{q}_i)(x_i - \hat{q}_i) + \frac{g''(\hat{q}_i)}{2}(x_i - \hat{q}_i)^2. \tag{12}$$

We first write this equation as $g(\hat{q}_i) = g(x_i) + e_i^g$ where $e_i^g = g'(\hat{q}_i)\epsilon_i - \frac{g''(\hat{q}_i)}{2}\epsilon_i^2$. Hence $\mathrm{E}\left[\, e_i^g \,\right] = -\frac{1}{2}g''(\hat{q}_i)Var(\hat{q}_i)$. We correct this bias by adding $\frac{1}{2}g''(\hat{q}_i)Var(\hat{q}_i)$ on the L.H.S. to obtain $g(\hat{q}_i) + \frac{g''(\hat{q}_i)}{2}Var(\hat{q}_i) = g(x_i) + e_i^g$, and then substitute $g(x_i) = \theta_1 + \theta_2 z_i$ to obtain

$$g(\hat{q}_i) + \frac{g''(\hat{q}_i)}{2}Var(q_i) = \theta_1 + \theta_2 z_i + e_i^g. \tag{13}$$

We can estimate the parameters θ_1 and θ_2 of the Normal pdf of $g(X)$ from (13). The variance-covariance matrix Ω' for the errors in (13) is approximated as

$$\begin{aligned}
\omega'_{ii} &= \mathrm{E}\left[\,\left(e_i^g - \mathrm{E}\left[\,e_i^g\,\right]\right)^2\,\right] = \mathrm{E}\left[\,\left(g'(\hat{q}_i)\epsilon_i - \frac{1}{2}g''(\hat{q}_i)\epsilon_i^2 + \frac{1}{2}g''(\hat{q}_i)Var(\hat{q}_i)\right)^2\,\right] \\
\omega'_{ij} &= \mathrm{E}\left[\,\left(e_i^g - \mathrm{E}\left[\,e_i^g\,\right]\right)\left(e_j^g - \mathrm{E}\left[\,e_j^g\,\right]\right)\,\right] \\
&= \mathrm{E}\left[\,\left(g'(\hat{q}_i)\epsilon_i - \frac{1}{2}g''(\hat{q}_i)\epsilon_i^2 + \frac{1}{2}g''(\hat{q}_i)Var(\hat{q}_i)\right)\left(g'(\hat{q}_j)\epsilon_j - \frac{1}{2}g''(\hat{q}_j)\epsilon_j^2 + \frac{1}{2}g''(\hat{q}_j)Var(\epsilon_j)\right)\,\right]
\end{aligned}$$

Since $g(x)$ transformations for Johnson distributions are tractable, each term in the expressions for ω'_{ii} and ω'_{ij} admits algebraic simplification. Once the matrix Ω' is determined, the estimates $\hat{\mu}, \hat{\sigma}$ of the underlying Normal distribution are obtained by using Theorem 1 on $g(\hat{q}_i)$, replacing Ω with Ω'.

6. Implementation at Dow AgroSciences: Protocol and Bootstrapping for Quantification of Expert's quantile judgments

In this section, we discuss a step-by-step approach used at Dow AgroSciences for quantifying an expert's judgmental errors. In Section 6.1, we provide the context for expert elicitation. In Sections

6.2 – 6.4 we discuss the four-step process followed. Section 6.5 analyzes the data and quantifies the value of using our approach in estimating yield distributions from quantile judgments.

Dow produces and sells several hundred varieties of seed corn, which generates an annual revenue of $800 million. The seed corn yield obtained during the production of the seed corn is measured as number of bags (each with 80,000 seeds) per acre of land and it is random. Given this random seed production yield, Dow faces the classical newsvendor's tradeoff in determining the optimal area of land on which to produce the seed corn. Producing the seed corn on a very large area leads to a high cost; using a small area is risky since the production might be insufficient to meet demand. Every year, Dow determines the optimal acreage for this trade-off for each variety of seed corn. The specification of the yield distribution is a critical input for making this decision.

6.1. Need for Quantile Judgments for Yield Distribution Estimation and Expert's Mental Model

At Dow, the yield distributions are estimated using expert judgment due to the following biological reason. Dow has a parent pool of approximately 125 types of corn with specific genes. Corn plants have male as well as female parts. Therefore, to obtain the parent plants with a specific gene, the seeds with this gene are planted in a field. Self-pollination on these plants provides seeds with the same gene. This inbreeding is carried out regularly to replenish the stock of parent seeds. Statistical distributions for yields obtained from this inbreeding process are available from historical data.

However almost 100% of the seed corn sold by Dow is hybrid seed that is obtained by cross mating. This cross mating occurs when two different types (or parents) of seed are planted in the field. Plants of one type, say X (not to be confused with random variable X), are treated chemically and physically to make them act as female, and the plants of the other type, say Y, are made to act as male. The cross pollination between these parents provides the hybrid seeds. Among the several hundred varieties of hybrid seed corn sold every year, only a few have been sold continuously in the last decade. The average life of hybrid varieties in market is less than two years. Most hybrid seeds are produced only three or four times. Therefore, sufficient historical yield data necessary to obtain statistical distribution are not available for most hybrid seeds.

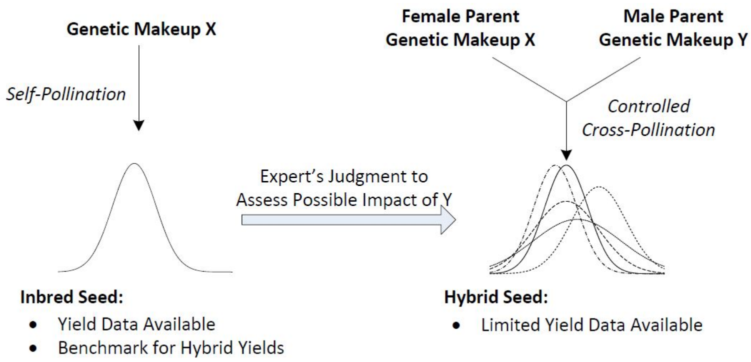


Figure 4 During cross-pollination the male Y changes the inbred yield distribution of X shown on the left. The expert's mental model involves judgments about changes in the location and/or spread of the distribution due to Y. Possible distributions after cross breeding are shown in dotted lines on the right.

In the absence of these data, the firm relies on a yield expert to estimate the distributions. Figure 4 summarizes the expert's mental model for estimating the yield distribution of a hybrid seed. Female plants provide the body on which the hybrid seed grows; the male plants provide the pollen to fertilize the female plant. Since the female plant nurtures the seed, the available statistical distribution for the inbreeding for type X provides a statistical benchmark (on the left) for the hybrid seed. The male affects this distribution during cross-pollination, leading to various likely distributions as shown in dotted lines (on the right). The expert's contextual knowledge provides him with insights into how the distribution will change during cross pollination. In the past, the yield expert has adjusted the median of the inbreeding distribution higher or lower to provide an estimate of the median yield for the hybrid seed. Thus, the estimate of the median yield for the distribution of the hybrid seed production yield is grounded in statistics but is judgmental in nature. There is still a need to estimate the spread of the yield distribution for the hybrid seed to determine the number of acres on which the hybrid seed should be grown.

We used the theory developed earlier for the determination of the yield distributions. For determining the matrix Ω of judgmental errors, we used the following four-step approach. In Steps 1 and 2, we let the expert select the quantiles to estimate and obtained historical data for seeds that were grown repeatedly. In Steps 3 and 4, we elicited quantile judgments from the expert and quantified his judgmental errors into Ω by comparing his judgments with the historical data. We then used Ω to obtain the optimal weights to estimate the mean and standard deviation for a future use.

6.2. Steps 1 and 2: Selection of Quantiles for Elicitation and Data Collection

In Step 1, we asked Dow to identify a set of hybrid seeds that has been produced repetitively in the last few years. Overall, Dow found $L = 22$ such hybrid seeds indexed by l=*1,2,...,22*, and provided us with the historical yield data for these seeds. The expert did not see or analyze these data. Analysis of these data and prior experience at Dow suggest that the yields are distributed Normally. Theoretical justification also supports this conclusion: when the total area of land is divided into smaller pieces of land (as is the case at Dow), the total output obtained can be considered as an aggregation of smaller random outputs, and the Central Limit Theorem is applicable.

In Step 2, for each of the hybrid seeds l=*1,2,...,L*, we asked the expert to select three quantiles to estimate. The selection of three quantiles (rather than more than three) was motivated by existing literature that suggests that three quantiles perform almost as well as five quantiles (Wallsten et al. 2013), as well as the time constraints faced by the expert. The expert is a scientist, he is well-trained in statistics, and has worked extensively with yield data. His quantitative background and experience were helpful as he clearly understood the probabilistic meaning and implications of quantiles. The first quantile he selected was the 50th quantile, since he has estimated this quantile regularly in the last few years (Table 1 shows that the expert's assessment of this quantile was very precise indeed). The extant literature also has established that estimating this quantile has the intuitive 50–50 high–low interpretation that managers understand well (O'Hagan 2006).

We asked the expert to provide us with his quantile judgments for two other quantiles, one in each tail of the yield distribution, that he was comfortable estimating. The yield expert chose to provide his judgments for the 10th and the 75th quantiles, for several reasons. First, he has developed familiarity with these quantiles in the last few years: his statistical software typically provides a limited number of quantile values including these two quantiles during data analysis, and he is accustomed to thinking about them. Second, the expert suggested the use of these asymmetric quantiles because if asked for symmetric quantiles, he would intuitively use his statistical knowledge and "estimate one tail quantile and calculate the other symmetric quantile using the properties of the Normal distribution." This would lead to spurious errors in the deduced quantile judgment.

Finally, the expert was not comfortable in providing judgments for quantiles that were further out in the tails, such as the 1st and the 95th quantiles. This reluctance was interesting and highlighted some subtle disconnects between theory and practice. Some articles (e.g., Lau et al. 1998, Lau and Lau 1998) have suggested weights for extreme quantiles such as 1 percentile assuming no judgmental errors. However, the expert found it difficult to estimate extreme quantiles. Specifically, he was concerned that he might not be able to differentiate between random variations (that we seek to capture) and acts of nature such as floods (that we seek to exclude) that lead to extreme outcomes.

6.3. Step 3: Elicitation Sequence and Consistency Check

In Step 3, for each distribution l, we obtained the three quantile judgments $\hat{x}_{il}(p_i); i = 1,2,3; l = 1,2,...,22; p_i = 0.1, 0.5, 0.75$ from the expert. We obtained these judgments in two rounds. In Round 1, for each hybrid l, the expert followed his usual procedure for studying the yield distribution for the female, looking at the properties of the male, and providing his judgment for the median. We then asked the expert to provide his judgment for the 10th and 75th quantiles, in that order, in an Excel file where the expert could enter only the three inputs for the quantiles. This customized sequence is consistent with the extant literature that suggests first obtaining an assessment for 50–50 odds (Garthwaite and Dickey 1985), and then focusing further on quantiles in the tails. In Round 2 of estimation, to encourage a careful reconfirmation of the judgments provided in Round 1, we used a feedback mechanism. We used the information from two quantile judgments to make simple deductions and then asked the expert to validate these deductions. If the expert did not concur with the deductions, we encouraged him to fine tune the quantile judgments.

As an illustrative example, suppose the expert provided values of 15, 70, and 100 for the 10th, 50th, and the 75th quantiles respectively. The stated values of the 10th and 50th quantiles imply a mean yield of 70 and standard deviation of 42.92. These two values imply that there is a 50 percent chance that the yield will be between 41 and 99 (the implied 25th and the 75th quantile). We asked the yield expert the following question: "Your estimate of the 10th quantile implies that there is a 50 percent chance that the yield will be between 41 and 99. If you think that this range should be narrower, please consider increasing the estimate of the 10th quantile. If you think the range should be wider, please consider decreasing the estimate of the 10th quantile." We implemented this feedback in an automated fashion so that the values in the feedback question were generated automatically using his quantile estimates. The expert could revisit his input and the accompanying feedback question any number of times before moving to the next feedback question for the judgment for the 75th quantile (using the deduced 35th and 85th quantile values obtained from his judgments for the 50th and 75th quantiles). After finishing this feedback, he moved to the next seed.

6.4. Step 4: Separation of Sampling Errors Using Bootstrapping [2]

After the elicitation was complete, to obtain errors we compared the expert's stated values for the quantiles with the values obtained from historical data. Recall that we assumed in Section 2 that the true values of the quantiles x_i are available. However, since the number of data points for each seed at Dow were limited (the largest sample size was 53), the quantile values obtained from the data were subject to sampling variations that must be explicitly accounted for. Specifically, let $\tilde{x}_i$ denote the value of quantile i for the empirical distribution. Then, for the true value x_i and the expert's estimate $\hat{x}_i$, we have the following decomposition of errors:

$$\hat{x}_i - \tilde{x}_i = (\hat{x}_i - x_i) + (x_i - \tilde{x}_i) \tag{14}$$

$$\text{Total Error} = \text{Judgmental Error} + \text{Sampling Error}$$

The comparison of the expert's assessment $\hat{x}_i$ with the empirical value $\tilde{x}_i$ has two sources of errors: the expert's judgmental error and the sampling error. The judgmental error is the difference between the quantile judgment and the true quantile $(\hat{x}_i - x_i)$. The sampling error $(x_i - \tilde{x}_i)$ captures the data variability that is present because the empirical distribution is based on a random sample of limited size from the population. The expert did not see the historical data, therefore both sources of errors can be considered to be mutually independent. Note that if the sample size is large, then the quantile value obtained from the empirical distribution will be close to the true value; consequently, the difference $(x_i - \tilde{x}_i)$ is negligible and the expert's judgmental error completely captures the uncertainty in the quantile value.

[2] We thank an anonymous reviewer for suggesting this procedure.

Writing (14) in a vector form, we have $\hat{\mathbf{x}} - \tilde{\mathbf{x}} = (\hat{\mathbf{x}} - \mathbf{x}) + (\mathbf{x} - \tilde{\mathbf{x}})$. It follows that the total bias is equal to

$$E[\hat{\mathbf{x}} - \tilde{\mathbf{x}}] = E[(\hat{\mathbf{x}} - \mathbf{x})] + E[(\mathbf{x} - \tilde{\mathbf{x}})] \tag{15}$$

$$\boldsymbol{\delta}^t = \boldsymbol{\delta} + \boldsymbol{\delta}^s$$

where $\boldsymbol{\delta}^t$ is the total bias, and $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^s$ are the expert's judgmental bias and the sampling bias respectively. The expert's judgmental bias is computed as: $\boldsymbol{\delta} = \boldsymbol{\delta}^t - \boldsymbol{\delta}^s$.

Similarly, the variance in the estimates of quantiles, assuming independence of the data-specific sampling error and the expert-specific judgmental error, is

$$Var[\hat{\mathbf{x}} - \tilde{\mathbf{x}}] = Var[(\hat{\mathbf{x}} - \mathbf{x})] + Var[(\mathbf{x} - \tilde{\mathbf{x}})]$$

We can write this equation in matrix notation as

$$\Omega^t = \Omega + \Omega^s \tag{16}$$

where Ω is the matrix of covariances of judgmental errors and needs to be estimated for use in our analytical development described earlier. This matrix is estimated as $\Omega = \Omega^t - \Omega^s$. Note that the matrix Ω must be checked for positive definiteness to be able to take an inverse to obtain the weights using Theorem 1. We next discuss the estimation of $\boldsymbol{\delta}^t$ and Ω^t using Dow's data, and the estimation of $\boldsymbol{\delta}^s$ and Ω^s using bootstrapping. Note that with a large number of historical observations, $\Omega \simeq \Omega^t, \boldsymbol{\delta} \simeq \boldsymbol{\delta}^t$, and the bootstrapping approach is not required.

For Dow's data, the total bias $\boldsymbol{\delta}^t$ and matrix Ω^t were determined using the expert's assessments as follows: In each of the two rounds of elicitation, the expert's quantile judgments $\hat{x}_{il}(p_i); i = 1, 2, 3$ for hybrid l were compared to the quantiles of the empirical distribution, $\tilde{x}_{il}(p_i)$. The differences provided the total errors $\hat{e}_{il} = \hat{x}_{il}(p_i) - \tilde{x}_{il}(p_i)$. The average error $\hat{\delta}_i^t = \sum_{l=1}^{L} \hat{e}_{il}/L$ provided the total bias for each quantile. The vector of biases $\hat{\delta}_i^t$ constituted $\hat{\boldsymbol{\delta}}^t$. We then obtained unbiased errors as $\hat{e}_{il}^u = \hat{e}_{il} - \hat{\delta}_i^t$; using these, we estimated the 3 x 3 variance-covariance matrix $\hat{\Omega}^t$. A comparison of $\hat{\Omega}^t$ from the first round without feedback and the second round with feedback showed that the feedback reduced the spread of the errors significantly (by 33%). The covariance matrix $\hat{\Omega}^t$ and the bias $\hat{\boldsymbol{\delta}}^t$ obtained after the second round are shown in Table 1.

$\Omega^t = \begin{bmatrix} 113.41 & 50.09 & 46.83 \\ 50.09 & 42.92 & 51.46 \\ 46.82 & 51.46 & 93.37 \end{bmatrix}$	$\Omega^s = \begin{bmatrix} 34.42 & 20.71 & 13.50 \\ 20.71 & 21.00 & 21.16 \\ 13.49 & 21.16 & 25.20 \end{bmatrix}$	$\Omega = \Omega^t - \Omega^s = \begin{bmatrix} 78.99 & 29.38 & 33.33 \\ 29.38 & 21.92 & 30.30 \\ 33.33 & 30.30 & 68.17 \end{bmatrix}$
$\boldsymbol{\delta}^t = [9.43 \ 0.94 \ -2.48]$	$\boldsymbol{\delta}^s = [-1.05 \ 0.00 \ 0.55]$	$\boldsymbol{\delta} = \boldsymbol{\delta}^t - \boldsymbol{\delta}^s = [10.48 \ 0.94 \ -3.03]$

Table 1 variance–covariance matrix and biases after bootstrap adjustment

The sampling bias $\boldsymbol{\delta}^s$ and the variance-covariance matrix Ω^s were determined by bootstrapping, as follows. We had data $y_{1l}, y_{2l}, ..., y_{n_l l}$ for seed l and corresponding quantiles $\tilde{x}_{il}$ estimated using these data. For each distribution l, we drew a sample indexed p of size n_l with replacement from the data $y_{1l}, y_{2l}, ..., y_{n_l l}$ and obtained the quantiles for *this* bootstrapping sample, $\tilde{x}_{ilp}$. We repeated the process for *p=1,2,...,P* times. Then we obtained the differences $\Delta_{ilp} = (\tilde{x}_{ilp} - \tilde{x}_{il})$, determined the average difference $\bar{\Delta}_{il} = \sum_p \Delta_{ilp}/P$ and then calculated the unbiased differences $\Delta^u_{ilp} = \Delta_{ilp} - \bar{\Delta}_{il}$. From these $3 \times P$ unbiased differences, we obtained the covariance matrix $\hat{\Omega}^{sl}$ for seed l. Extant literature suggests that a size of P=100 is usually sufficient to ensure a stable variance–covariance matrix $\hat{\Omega}^{sl}$. Since the bootstrapping can be done very efficiently on today's computers, we used P = 1,000,000. Finally, Ω^s was estimated as $\hat{\Omega}^s = \sum \hat{\Omega}^{sl}/L$, implying that each covariance matrix $\hat{\Omega}^{sl}$ is equally likely to be present for each elicitation in the future. The sampling bias for quantile i was estimated as $\hat{\delta}^s_i = \sum_l \bar{\Delta}_{il}/L$. The vector of these biases constituted $\hat{\boldsymbol{\delta}}^s$. For Dow's data, the values of $\hat{\Omega}^s$ and the bias vector $\hat{\boldsymbol{\delta}}^s$ are shown in Table 1.

The estimated judgmental bias $\hat{\boldsymbol{\delta}}$ was obtained as $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}^t - \hat{\boldsymbol{\delta}}^s$ using (15), and the estimated matrix of judgmental errors $\hat{\Omega}$ was obtained as $\hat{\Omega} = \hat{\Omega}^t - \hat{\Omega}^s$ using (16), and are shown in Table 1.

6.5. Benefits of Incorporating Judgmental Errors at Dow

For the variance–covariance matrix Ω in Table 1 and the matrix $\mathbf{Z} = \begin{bmatrix} 1 & -1.28 \\ 1 & 0 \\ 1 & 0.67 \end{bmatrix}$, Theorem 1 provides the weights $\mathbf{w}^*_\mu$=$[-0.18, 1.51, -0.33]$ and $\mathbf{w}^*_\sigma$=$[-0.58, 0.20, 0.38]$ for the 10th, 50th, and 75th quantiles. For these results, the following regime is useful for Dow: First, obtain the estimates of the 10th, 50th, and 75th quantiles $\hat{\mathbf{x}}$ from the expert. Obtain the de-biased estimates $\hat{\mathbf{q}} = \hat{\mathbf{x}} - \hat{\boldsymbol{\delta}}$ by subtracting the biases $\hat{\delta}_1 = 10.48, \hat{\delta}_2 = 0.94, \hat{\delta}_3 = -3.03$. Then, use the weights above on the de-biased estimates to obtain the mean and standard deviation as $\hat{\mu} = \mathbf{w}^{*\mathrm{T}}_\mu \hat{\mathbf{q}}$ and $\hat{\sigma} = \mathbf{w}^{*\mathrm{T}}_\sigma \hat{\mathbf{q}}$.

Using our approach for Ω in Table 1, the variance in the estimate $\hat{\mu}$ is $Var(\hat{\mu}) = 18$ (bags/acre)^2 approximately. This variance more than doubles $Var(\hat{\mu}^I) = 38$ (bags/acre)^2 if we ignore the judgmental errors and use the weights for the error ignoring Benchmark I discussed in Section 4.4, $\mathbf{w}_{\mu}^{I*}$=[0.22,0.36,0.42]. The variance of Benchmark II's estimate is $Var(\hat{\mu}^{II}) = 53$ (bags/acre)^2 assuming elicitation of the 10th and 75th quantiles. Similar trends are present for the standard deviation.

This reduction in variance in the estimates $\hat{\mu}, \hat{\sigma}$ for yield distributions has two measurable impacts. First, at Dow the yield distributions have a variance of $\sigma^2 \approx 400$ on the average. Using Proposition 3, it follows that when the expert's judgmental errors are accounted for using our approach, his quantile judgments can be used to extract information (for the distribution parameters) that is equivalent to the information provided by $\sigma^2/Var(\hat{\mu}) = 400/18 \approx 22$ data points. In contrast, Benchmarks I and II ignore his judgmental errors and extract information that is equivalent to $\sigma^2/Var(\hat{\mu}^I) = 400/38 \approx 11$ and $\sigma^2/Var(\hat{\mu}^{II}) = 400/53 \approx 7$ data points, respectively. For Dow's practice of test-growing a seed in 3–4 test-fields every year, our approach extracts information (from quantile judgments) that is equivalent to approximately 5–6 years of test-data while Benchmarks I and II extract information that is equivalent to test-data of 2–3 years and 1–2 years, respectively.

Second, the reduction in variance of $\hat{\mu}, \hat{\sigma}$ has monetary benefits. Akcay et al. (2011) show that the average cost of operation for a firm that procures products under uncertain demand decreases with the number of data points available for determining the distribution of uncertainty. Using data from SmartOps Corporation, they establish that an increase in the availability of data points from 10 data points to 20 data points reduces the operating cost between 10% to 20% (Tables 2,3,4 on page 307). In our context this finding implies a substantial benefit for Dow's annual \$800 million decision from using our approach that is equivalent to 22 data points as opposed to Benchmark I that is equivalent to 11 data points. Equally importantly, as the cost of implementing the proposed approach is minimal, the decision to implement it at Dow was easy to make.

An additional advantage of the proposed approach is that it is now possible to investigate whether the expert should be asked to provide judgments for specific quantiles. Dow's expert provided us with his judgments for the 10th quantile because he was accustomed to thinking about this quantile.

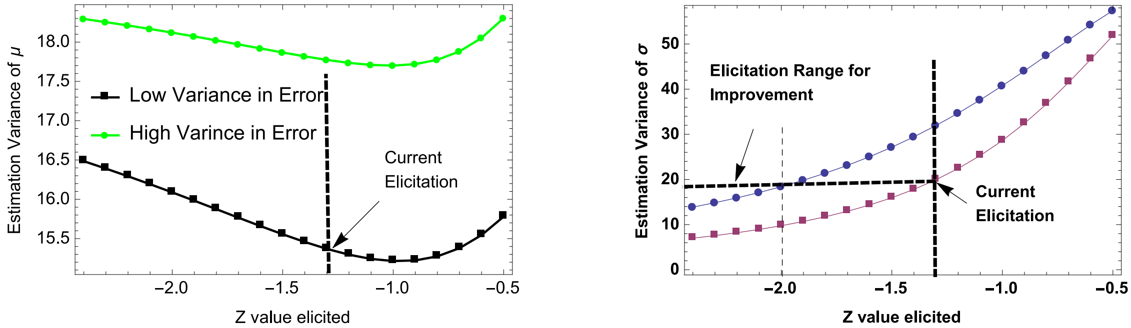


Figure 5 Differences between the variance in $\hat{\mu}$ and $\hat{\sigma}$ at different z values elicited.

If we had asked him to provide his judgments for a more extreme fractile, it is very likely that the variance of the judgmental error for this specific quantile would have been larger. Given this increase in error, which quantiles are worth asking for? Figure 5 illustrates this analysis. For the current matrix Ω of judgmental errors, $Var(\hat{\mu})$ is minimum at $z \approx -1.0$ (line with square markers in left panel). In contrast, $Var(\hat{\sigma})$ in the right panel decreases as we elicit quantiles further out in the tail of the distribution. Now suppose the variance of the judgmental error doubles when the expert is asked to estimate a different quantile *in lieu* of the 10th quantile. The panel on the left shows that regardless of which quantile the expert now provides (line with round markers), the estimated mean will have a larger variance than before. But for the estimation of the standard deviation, the right panel shows that increasing the variance in judgmental errors may still be worthwhile, but only when the expert can provide estimates of the 5th quantile or lower (for z value of -2 or lower). Given this analysis, there does not appear to be a significant benefit in asking the expert to estimate a different quantile instead of the 10th quantile.

7. Summary and Concluding Remarks

In changing environments, historical data does not exist to provide probability distributions of various uncertainties. In such environments, experts' judgments are sought. But experts' judgments are prone to judgmental errors. In this paper we develop an analytical approach to deducing probability distributions from a set of quantile judgments provided by an expert while explicitly taking the expert's judgmental errors into account.

From a theory development perspective, the two-step optimization approach developed, its analytical solution, and its use for developing a number of structural properties and providing an analytical foundation for results documented numerically in the decision analysis literature is a contribution to this literature. A salient feature of the approach is that an expert is no longer required to provide judgments for the median and specific symmetric quantiles studied in the literature but can provide his judgments for any set of quantiles. We also develop a novel equivalence between an expert's quantile judgments and a sample size of randomly collected data; this equivalence is useful for ranking and comparing experts objectively. Finally, the modeling framework explains a consistent numerical finding in the decision analysis literature that the weights for the mean and the standard deviation add up to 1 and 0, respectively.

Equally importantly, the approach establishes new connections between this problem and other streams of literature. It provides for a linear pooling of quantile judgments from multiple experts, thereby extending the existing literature on combining multiple point-estimates to combining quantile judgments from multiple experts. The estimates obtained for distribution parameters in our approach can also be updated using data in a Bayesian fashion with a Copula based procedure.

From a practice perspective, our approach has features that make it viable for implementation: First, our approach uses judgments for any three or more quantiles that an expert is comfortable providing. In a specific application at Dow AgroSciences, we used the yield-expert's judgments for the 50th, 10th, and 75th quantiles to deduce the mean and standard deviations of a large number of yield uncertainties. The expert chose to estimate these quantile based on his experience with seeing and using these quantiles in his data analysis responsibilities at the firm. Second, the final outcome of the approach is a set of weights that are used to estimate mean and standard deviations as weighted linear functions of quantile judgments. The implementation of this procedure requires simple mathematical operations that can be performed in a spreadsheet environment, and it has led to an expedited adoption at Dow. Third, the weights are specific to the expert and capture how good he is at providing estimates of various quantiles. This explicit incorporation of an expert's judgmental errors is useful since we can then determine how the estimated probability distribution (and the decision based on this estimated distribution) will vary as the quality of the expert's judgmental errors improve or deteriorate.

References

Abbas, A.E., D.V. Budescu, H.T. Yu, R. Haggerty. 2008. A comparison of two probability encoding methods: Fixed probability vs. fixed variable values. *Decision Anal* **5**(4) 190–202.

Akcay, Alp, Bahar Biller, Sridhar Tayur. 2011. Improved inventory targets in the presence of limited historical demand data. *Manufacturing & Service Operations Management* **13**(3) 297–309.

Baker, Erin, Olaitan Olaleye. 2013. Combining experts: Decomposition and aggregation order. *Risk Analysis* **33**(6) 1116–1127.

Bates, John M, Clive WJ Granger. 1969. The combination of forecasts. *Operations Research Quarterly* 451–468.

Blattberg, Robert C, Stephen J Hoch. 1990. Database models and managerial intuition: 50% model+ 50% manager. *Management Science* **36**(8) 887–899.

Budescu, David V, Ido Erev, Thomas S Wallsten. 1997. On the importance of random error in the study of probability judgment. part i: New theoretical developments. *Journal of Behavioral Decision Making* **10**(3) 157–171.

Casella, G., R.L. Berger. 2002. *Statistical Inference*. 2nd ed. Duxbury Press.

Clemen, Robert T., Canan Ulu. 2008. Interior additivity and subjective probability assessment of continuous variables. *Management Science* **54**(4) 835–851.

Clemen, Robert T, Robert L Winkler. 1999. Combining probability distributions from experts in risk analysis. *Risk analysis* **19**(2) 187–203.

Dickinson, JP. 1975. Some comments on the combination of forecasts. *Operational Research Quarterly* 205–210.

Garthwaite, Paul H., James M. Dickey. 1985. Double- and single-bisection methods for subjective probability assessment in a location-scale family. *Journal of Econometrics* **29**(12) 149 – 163.

Garthwaite, Paul H, Joseph B Kadane, Anthony O'hagan. 2005. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* **100**(470) 680–701.

Gelman, Andrew, John B Carlin, Hal S Stern, Donald B Rubin. 2003. Bayesian data analysis, (chapman & hall/crc texts in statistical science) .

Johnson, D. 1998. The robustness of mean and variance approximations in risk analysis. *Journal of the Operational Research Society* **49**(3) 253–262.

Johnson, D. 2002. Triangular approximations for continuous random variables in risk analysis. *Journal of the Operational Research Society* **53**(4) 457–467.

Johnson, N. L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* **36**(1/2) pp. 149–176.

Johnson, N.L., S. Kotz, N. Balakrishnan. 1994. *Continuous univariate distributions, vol. 1 of Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York,.

Keefer, D.L. 1994. Certainty equivalents for three-point discrete-distribution approximations. *Management science* **40**(6) 760–773.

Keefer, D.L., S.E. Bodily. 1983. Three-point approximations for continuous random variables. *Management Science* **29**(5) 595–609.

Keefer, D.L., W.A. Verdini. 1993. Better estimation of PERT activity time parameters. *Management Science* **39**(9) 1086–1091.

Keelin, Thomas W., Bradford W. Powley. 2011. Quantile-parameterized distributions. *Decision Analysis* **8**(3) 206–219.

Koehler, Derek J, Lyle Brenner, Dale Griffin. 2002. The calibration of expert judgment: Heuristics and biases beyond the laboratory. *Heuristics and biases: The psychology of intuitive judgment,Cambridge University Press* .

Lau, A.H.L., H.S. Lau, Y. Zhang. 1996. A simple and logical alternative for making PERT time estimates. *IIE Transactions* **28**(3) 183–192.

Lau, H.S., A.H.L. Lau. 1998. An improved PERT-type formula for standard deviation. *IIE transactions* **30**(3) 273–275.

Lau, H.S., A.H.L. Lau, C.J. Ho. 1998. Improved moment-estimation formulas using more than three subjective fractiles. *Management Science* **44**(3) 346–351.

Lau, H.S., A.H.L. Lau, J.F. Kottas. 1999. Using Tocher's curve to convert subjective quantile-estimates into a probability distribution function. *IIE transactions* **31**(3) 245–254.

Lindley, Dennis V. 1987. Using expert advice on a skew judgmental distribution. *Operations research* **35**(5) 716–721.

O'Hagan, A. 2006. *Uncertain judgements: Eliciting experts' probabilities*, vol. 35. John Wiley & Sons Inc.

Pearson, ES, JW Tukey. 1965. Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika* **52**(3-4) 533.

Perry, C., ID Greig. 1975. Estimating the mean and variance of subjective distributions in pert and decision analysis. *Management Science* **21**(12) 1477–1480.

Peterson, Cameron, Alan Miller. 1964. Mode, median, and mean as optimal strategies. *Journal of Experimental Psychology* **68**(4) 363.

Ravinder, H.V., D.N. Kleinmuntz, J.S. Dyer. 1988. The reliability of subjective probabilities obtained through decomposition. *Management Science* **34**(2) 186–199.

Stuart, Alan, J Keith Ord. 1994. Kendalls advanced theory of statistics. vol. i. distribution theory. *Arnold, London* .

Wallsten, T S., C. Nataf, Y. Shlomi, T Tomlinson. 2013. Forecasting values of quantitative variables. *Paper presented at SPUDM24, Barcelona, Spain, August 20, 2013.* .

Wallsten, Thomas S, David V Budescu. 1983. State of the art encoding subjective probabilities: A psychological and psychometric review. *Management Science* **29**(2) 151–173.

Wang, Tianyang, James S Dyer. 2012. A copulas-based approach to modeling dependence in decision trees. *Operations Research* **60**(1) 225–242.

Zapata-Vázquez, Rita Esther, Anthony O'Hagan, Leonardo Soares Bastos. 2014. Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics* **41**(9) 1919–1933.

Appendix

1. Proof of Lemma 1

We first prove an intermediate result in Lemma 3 below.

Lemma 3 (Characterization of location–scale Moments) *Let X be a random variable with finite j^{th} moments for $j = 1, 2, \ldots, k$, then the moments $\mu_k(\boldsymbol{\theta})$ have the functional form:*

$$\mu_k(\boldsymbol{\theta}) = \mu_k(\theta_1, \theta_2) = \sum_{i=0}^{k} \binom{k}{i} \theta_1^i \theta_2^{k-i} \kappa_{k-i}, \tag{17}$$

where $\kappa_i = \mu_i(0,1)$ and $\kappa_0 = 1$, if and only if X is a random variable with a distribution with a location–scale family. Further, the central moments are obtained as $\bar{\mu}_k(\boldsymbol{\theta}) = E[(x-\mu)^k] = \theta_2^k \sum_{i=0}^{k} (-\kappa_1)^i \kappa_{k-i}$.

Clearly $\mu_0(\boldsymbol{\theta}) = 1$ for all random variables. To prove the "if" claim ($\Leftarrow$) denote as $\phi_x(\cdot;\theta_1,\theta_2)$ the pdf of X, and define a random variable Z with pdf $\phi(z) = \phi_x(x;0,1)$. It then follows from the definition of distributions with a location–scale family that (17) can be written as

$$\mu_k(\theta_1,\theta_2) = \frac{1}{\theta_2}\int(\theta_1+\theta_2 z)^k\, d\Phi(z) \quad \text{for } k = 1,\, 2,\, \ldots \tag{18}$$

The claim follows by expanding $(\theta_1+\theta_2 z)^k$ and expressing the individual integrals as moments of Z. To establish the converse claim ($\Rightarrow$) we assume, contrary to fact, that X is not a random variable with a distribution with a location–scale family and $\mu_1(\theta_1,\theta_2) = \theta_1 + \kappa_1\theta_2$. It then follows that there is a transformation $x = g(z)$ with

$$\mu_1(\theta_1,\theta_2) = \int x\phi_x(x;\theta_1,\theta_2)\, dx = \int g(z)\phi_x(g(z);\theta_1,\theta_2)g'(z)\, dz = \theta_1 + \kappa_1\theta_2. \tag{19}$$

Since the expectation is a linear operator and for any $x = g(z)$ we have $\phi_x(x;\theta_1,\theta_2)\, dx = \phi_x(g(z);\theta_1,\theta_2)g'(z)\, dz$, we must also have $g(z) = \theta_1+\theta_2 z = x$ and $\phi_x(x;\theta_1,\theta_2) = (1/\theta_2)\phi\big((x-\theta_1)/\theta_2\big)$. But this means that X has a distribution with a location-scale family, which is a contradiction. Hence X must be a random variable with a distribution with a location-scale family. The expressions in Lemma 1 are obtained by expanding (17) for *k=1,2.*

To show that $\bar{\mu}_k(\boldsymbol{\theta}) = E[(x-\mu)^k] = \theta_2^k\sum_{i=0}^{k}(-\kappa_1)^i\kappa_{k-i}$, we substitute $x = \theta_1 + z\theta_2$ and $\mu = \theta_1+\kappa_1\theta_2$ and obtain $\bar{\mu}_k(\boldsymbol{\theta}) = E[(x-\mu)^k] = E[(\theta_1+z\theta_2-\theta_1-\kappa_1\theta_2)^k] = \theta_2^k E[(z-\kappa_1)^k]$. On binomial expansion of the last term, and using $E[z^{k-i}] = \kappa_{k-i}$, we obtain $\bar{\mu}_k(\boldsymbol{\theta}) = \theta_2^k E[(z-\kappa_1)^k] = \theta_2^k\sum_{i=0}^{k}\binom{k}{i}(-\kappa_1)^i\kappa_{k-i}$. $\square$

2. Proof of Lemma 2

The result follows because Ω is a covariance matrix, hence it is positive definite and the objective function (8) is convex.

3. Proof of Theorem 1

The Lagrangian of the problem is

$$L = \mathbf{w}^{\mathrm{T}}\Omega\mathbf{w} - [\lambda_1,\lambda_2][\mathbf{w}^{\mathrm{T}}\mathbf{Z} - \mathbf{a}^{\mathrm{T}}] \tag{20}$$

This can be written as

$$L = \mathbf{w}^{\mathrm{T}}\Omega\mathbf{w} - \lambda(\mathbf{w}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta} - \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}) \tag{21}$$

when the constraint $\mathbf{w}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta} = \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta}$ must be valid for all vectors $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1,\theta_2) \in R\times R_{++}$. Then the first order optimality conditions are given by the following $n+1$ equations:

$$\nabla_w L^{\mathrm{T}} = 2\mathbf{w}^{\mathrm{T}}\Omega - \lambda\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}} = 0, \tag{22}$$

$$\frac{\partial}{\partial \lambda} L = -\mathbf{w}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta} + \mathbf{a}^{\mathrm{T}}\boldsymbol{\theta} = 0. \tag{23}$$

Post-multiplying the first equation by $\Omega^{-1}\mathbf{Z}$, we obtain $2\mathbf{w}^{\mathrm{T}}\mathbf{Z} = \lambda\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z}$. Since (23) must be satisfied by all valid $\boldsymbol{\theta}$, we must have $\mathbf{w}^{\mathrm{T}}\mathbf{Z} = \mathbf{a}^{\mathrm{T}}$, and we can then solve for $\lambda\boldsymbol{\theta}^{\mathrm{T}}$ as

$$\lambda\boldsymbol{\theta}^{\mathrm{T}} = 2\mathbf{a}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1} \tag{24}$$

The inverse of $\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z}$ exists because $\mathbf{Z}$ has a full column rank. Substituting (24) in (22) and post-multiplying by Ω^{-1} we obtain the optimal weights $\mathbf{w}^{*}$.

4. Proof of Proposition 1

We showed above that

$$\mathbf{w}^{\mathrm{T}}\mathbf{Z} = \mathbf{a}^{\mathrm{T}}. \tag{25}$$

For the estimation of mean μ, we have $\mathbf{a} = [1, \kappa_1]$, which then implies that $\sum w^{*}_{\mu i} = 1$. Similarly, for the estimation of standard deviation $\mathbf{a} = \left[0, \sqrt{\kappa_2 - \kappa_1^2}\right]$, which implies that $\sum w^{*}_{\sigma i} = 0$. $\square$

5. Proof of Proposition 2

(i) After expanding the solution $\mathbf{w}^{*\mathrm{T}}_{\mathbf{a}} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}$ and substituting $z_2 = 0$, we obtain the following expression of $w^{*}_{\mu_2}$:

$$w^{*}_{\mu_2} = \frac{z_3^2 Var(\hat{q}_1) + z_1^2 Var(\hat{q}_3) - 2z_1 z_3 \rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)} - \sqrt{Var(\hat{q}_2)}(z_1 - z_3)\left(z_1\rho_{23}\sqrt{Var(\hat{q}_3)} - z_3\rho_{12}\sqrt{Var(\hat{q}_1)}\right)}{Var(\hat{q}_2)(z_1 - z_3)^2 + z_3^2 Var(\hat{q}_1) + z_1^2 Var(\hat{q}_3) - 2z_1 z_3\rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)} - 2\sqrt{Var(\hat{q}_2)}(z_1 - z_3)\left(z_1\rho_{23}\sqrt{Var(\hat{q}_3)} - z_3\rho_{12}\sqrt{Var(\hat{q}_1)}\right)}$$

which we can write as $w^{*}_{\mu_2} = \frac{c_1 - \sqrt{Var(\hat{q}_2)}(z_1 - z_3)c_2}{c_3 Var(\hat{q}_2) - 2\sqrt{Var(\hat{q}_2)}(z_1 - z_3)c_2 + c_1}$ where $c_1 = z_3^2 Var(\hat{q}_1) + z_1^2 Var(\hat{q}_3) - 2z_1 z_3\rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)}$, $c_2 = z_1\rho_{23}\sqrt{Var(\hat{q}_3)} - z_3\rho_{12}\sqrt{Var(\hat{q}_1)}$, $c_3 = (z_1 - z_3)^2$. We can also factorize the squares in the denominator and write it as $Var(\hat{q}_2)(z_1 - z_3)^2 + Var(z_3\hat{q}_1 - z_1\hat{q}_3) - 2\sqrt{Var(\hat{q}_2)}(z_1 - z_3)\left(z_1\rho_{23}\sqrt{Var(\hat{q}_3)} - z_3\rho_{12}\sqrt{Var(\hat{q}_1)}\right)$. Since $(z_1 - z_3) < 0$, the denominator is strictly positive when $z_1\rho_{23}\sqrt{Var(\hat{q}_3)} \geq z_3\rho_{12}\sqrt{Var(\hat{q}_1)}$). We will denote the denominator as D. For ease of analysis, we will focus on $1 - w^{*}_{\mu_2}$. On simplification, the expression for this quantity is obtained as:

$$1 - w^{*}_{\mu_2} = \frac{\sqrt{Var(\hat{q}_2)}(z_1 - z_3)\left(\sqrt{Var(\hat{q}_2)}(z_1 - z_3) + z_3\rho_{12}\sqrt{Var(\hat{q}_1)} - z_1\rho_{23}\sqrt{Var(\hat{q}_3)}\right)}{D} \tag{26}$$

We can further complete the squares in the numerator and write it as $N = \left(\sqrt{Var(\hat{q}_2)}(z_1 - z_3) + (c_5/2)\right)^2 - (c_5/2)^2$ where $c_5 = z_3\rho_{12}\sqrt{Var(\hat{q}_1)} - z_1\rho_{23}\sqrt{Var(\hat{q}_3)}$. In the numerator expression of N, $(z_1 - z_3) < 0$. Therefore, the numerator is greater than zero when $c_5 < 0$. That is $1 - w^{*}_{\mu_2} > 0$, or $w^{*}_{\mu_2} < 1$ when $c_5 < 0$ or when $z_3\rho_{12}\sqrt{Var(\hat{q}_1)} < z_1\rho_{23}\sqrt{Var(\hat{q}_3)}$. It follows that when $z_3\rho_{12}\sqrt{Var(\hat{q}_1)} < z_1\rho_{23}\sqrt{Var(\hat{q}_3)}$, the weight $w^{*}_{\mu_2} > 1$ and $w^{*}_{\mu_2} < 1$ otherwise. The limit of $w^{*}_{\mu_2} = 1$ at $\sqrt{Var(\hat{q}_2)} = 0$ is obtained by substituting this value in the expression of $w^{*}_{\mu_2}$.

(ii) The analysis for the standard deviation follows similarly with the expression of the weight: $w^*_{\sigma_2} = \frac{z_3Var(\hat{q}_1)+z_1Var(\hat{q}_3)+(z_1+z_3)\rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)}-\sqrt{Var(\hat{q}_2)}(z_1-z_3)\left(\rho_{23}\sqrt{Var(\hat{q}_3)}-\rho_{12}\sqrt{Var(\hat{q}_1)}\right)}{Var(\hat{q}_2)(z_1-z_3)^2+z_3^2Var(\hat{q}_1)+z_1^2Var(\hat{q}_3)-2z_1z_3\rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)}-2\sqrt{Var(\hat{q}_2)}(z_1-z_3)\left(z_1\rho_{23}\sqrt{Var(\hat{q}_3)}-z_3\rho_{12}\sqrt{Var(\hat{q}_1)}\right)}$. Or, we can write it as $w^*_{\sigma 2} = \frac{c_4-\sqrt{Var(\hat{q}_2)}(z_1-z_3)c_2}{c_3Var(\hat{q}_2)-2\sqrt{Var(\hat{q}_2)}(z_1-z_3)c_2+c_1}$ where $c_4 = z_3Var(\hat{q}_1) + z_1Var(\hat{q}_3) + (z_1 + z_3)\rho_{13}\sqrt{Var(\hat{q}_1)Var(\hat{q}_3)}$.

$\square$

6. Proof of the Variance and Covariance of Estimates

We are interested in the variance of $\mathbf{w}_a{}^{\mathrm{T}}\hat{\mathbf{q}}$ for $a = \mu, \sigma$. Since $\mathbf{w}_a{}^{\mathrm{T}}\hat{\mathbf{q}} = \mathbf{w}_a{}^{\mathrm{T}}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon}) = \mathbf{w}_a{}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta} + \mathbf{w}_a{}^{\mathrm{T}}\boldsymbol{\epsilon}$ and the term $\mathbf{w}_a{}^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta}$ is a constant, it follows that the variance of $\mathbf{w}_a{}^{\mathrm{T}}\hat{\mathbf{q}}$ is equal to the variance of $\mathbf{w}_a{}^{\mathrm{T}}\boldsymbol{\epsilon}$ which is given by $\mathbf{w}_a{}^{\mathrm{T}}\Omega\mathbf{w}_a$ for $a = \mu, \sigma$. Replacing the expressions for $\mathbf{w}_a$ obtained in Theorem 1 we obtain the variance as

$$\mathbf{w}_a{}^{\mathrm{T}}\Omega\mathbf{w}_a = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\Omega\Omega^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a}. \tag{27}$$

Note that $\mathbf{a} = [1, \kappa_1]^{\mathrm{T}}$ for estimating μ, $\mathbf{a}= \left[0, \sqrt{\kappa_2-\kappa_1^2}\right]^{\mathrm{T}}$ for estimating σ, $\mathbf{a} = [1, 0]^{\mathrm{T}}$ for estimating θ_1, and $\mathbf{a} = [0, 1]^{\mathrm{T}}$ for estimating θ_2. The covariance $Cov[\hat{\theta}_1, \hat{\theta}_2]$ is calculated as $Cov[\hat{\theta}_1, \hat{\theta}_2] = E[\hat{\theta}_1\hat{\theta}_2] - E[\hat{\theta}_1]E[\hat{\theta}_2]$. Now $\hat{\theta}_1 = [1, 0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon})$ and $\hat{\theta}_2 = [0, 1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon})$. Therefore, the value of the covariance $Cov[\hat{\theta}_1, \hat{\theta}_2]$ is determined as

$$Cov[\hat{\theta}_1, \hat{\theta}_2] = E\left[\left[[1, 0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon})\right]\left[[0, 1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}(\mathbf{Z}\boldsymbol{\theta}+\boldsymbol{\epsilon})\right]\right] - E[\hat{\theta}_1]E[\hat{\theta}_2] \tag{28}$$

On subsequent simplification and substitution $E[\boldsymbol{\epsilon}] = \mathbf{0}$ we obtain

$$Cov[\hat{\theta}_1, \hat{\theta}_2] = [1, 0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0, 1]^{\mathrm{T}} \tag{29}$$

$\square$

7. Proof of Proposition 3

(a) For the mean, we note that variance of the distribution of a sample mean of size N_μ is equal to σ^2/N_μ where σ^2 is the population variation. Equating this variance with the variance $[1, \kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1, \kappa_1]^{\mathrm{T}}$ obtained above in (27), we obtain $N_\mu = \frac{\sigma^2}{[1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}}$.

(b) Stuart and Ord (1994) show on p 352 that the variance of sample standard deviation can be approximated as

$$Var(S) \approx \frac{\bar{\mu}_4 - \bar{\mu}_2^2}{4N_\sigma\sigma^2} \tag{30}$$

where $\bar{\mu}_i$ denotes the i^{th} central moment of the distribution of the random variable X. We know from Lemma 3 that

$$\bar{\mu}_i = \theta_2^i \sum_{j=0}^{i} (-\kappa_1)^j \kappa_{i-j} \tag{31}$$

From Lemma 1, we know that $\theta_2 = \sigma/\sqrt{\kappa_2-\kappa_1^2}$. Substituting this in (31), we obtain $\bar{\mu}_i = \sigma^i \sum_{j=0}^{i}(-\kappa_1)^j\kappa_{i-j}/\left(\sqrt{\kappa_2-\kappa_1^2}\right)^i$ Substituting this in (30), we obtain

$$Var(S) \approx \frac{\sigma^2}{4N_\sigma}\left(\frac{\sum_{j=0}^{4}(-\kappa_1)^j\kappa_{4-j}}{(\kappa_2-\kappa_1^2)^2} - \frac{\left(\sum_{j=0}^{2}(-\kappa_1)^j\kappa_{2-j}\right)^2}{(\kappa_2-\kappa_1^2)^2}\right) \tag{32}$$

Equating this with the variance in (27) $[0,\sqrt{\kappa_2-\kappa_1^2}](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0,\sqrt{\kappa_2-\kappa_1^2}]^{\mathrm{T}}$, we obtain $N_\sigma \approx \frac{\sigma^2\left(\frac{\sum_{j=1}^{4}(-\kappa_1)^j\kappa_{4-j}}{(\kappa_2-\kappa_1^2)^2} - \frac{\left(\sum_{j=1}^{2}(-\kappa_1)^j\kappa_{2-j}\right)^2}{(\kappa_2-\kappa_1^2)^2}\right)}{4[0,\sqrt{\kappa_2-\kappa_1^2}](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0,\sqrt{\kappa_2-\kappa_1^2}]^{\mathrm{T}}}$. □

8. Proof of Proposition 4

(a) We start by noting the general result in Theorem 1: $\mathbf{w}_a^{*\mathrm{T}} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}$. Next, we define each of these components when *j=1,2,...,n* experts provide quantile judgments. The matrix $\mathbf{Z}^{\mathrm{T}} = [\mathbf{Z}_0{}^{\mathrm{T}}, \mathbf{Z}_0{}^{\mathrm{T}}, ..., \mathbf{Z}_0{}^{\mathrm{T}}]$ where $\mathbf{Z}_0{}^{\mathrm{T}}$ appears n times, once for each expert. The subscript 0 simply suggests that this is a constant matrix since all experts provide judgments for the same set of quantiles.. The matrix Ω is a $mn \times mn$ block diagonal matrix with diagonal blocks $r_j\Omega_0$ where Ω_0 is of size $m \times m$. Then,

1. $(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z}) = \mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}\mathbf{Z}_0[\sum_{j=1}^{N}(1/r_j)] = \mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}\mathbf{Z}_0 R$ where $R = [\sum_{j=1}^{N}(1/r_j)]$
2. $\mathbf{Z}^{\mathrm{T}}\Omega^{-1} = [\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}(1/r_1), \mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}(1/r_2), ..., \mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}(1/r_N)]$
3. It follows from points 1 and 2 above that $\mathbf{w}_a^{*\mathrm{T}} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}\mathbf{Z}_0)^{-1}\mathbf{Z}_0^{\mathrm{T}}\Omega_0^{-1}[(1/r_1)/R, (1/r_2)/R, ..., (1/r_N)/R]$.

Now we can write this expression as $\mathbf{w}_a^{*\mathrm{T}} = \mathbf{w}_a^{c\mathrm{T}}[(1/r_1)/R, (1/r_2)/R, ..., (1/r_N)/R]$ where $\mathbf{w}_a^{c\mathrm{T}} = \mathbf{a}^{\mathrm{T}}(\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}\mathbf{Z}_0)^{-1}\mathbf{Z}_0^{\mathrm{T}}\Omega_0^{-1}$. Further, the vector $\mathbf{w}_a^{*\mathrm{T}}$ is composed of the m weights for each expert j: $\mathbf{w}_a^{*\mathrm{T}} = [\mathbf{w}_a^{1\mathrm{T}}, \mathbf{w}_a^{2\mathrm{T}}, ..., \mathbf{w}_a^{n\mathrm{T}}]$. It follows from these two relations that $\mathbf{w}_a^j = \alpha_j\mathbf{w}_a^c$ where the expert j's marginal weight is equal to $\alpha_j = (1/r_j)/R$.

(b) Consider the case when expert j is the only expert available with matrix Ω_0 for eliciting quantiles $\mathbf{Z}_0$. Then $R = (1/r_j)$. Substituting this expression in point 3 above, it follows that the weight for this expert is equal to $(1/r_j)/R = 1$, and the weights for his quantile judgments are equal to his independent weights.

□

9. Details on Bayesian Updating We performed the analysis for a distribution with true values of $\mu = 103, \sigma = 20$, and assumed that the expert provided unbiased judgments $\hat{q}_1 = 74, \hat{q}_2 = 105, \hat{q}_3 = 115$ for the 10th, 50th, and the 75th quantiles. To generate the prior distributions for θ_1, θ_2, we used the estimated matrix $\Omega = \begin{bmatrix} 78.99 & 29.38 & 33.33 \\ 29.38 & 21.92 & 30.30 \\ 33.33 & 30.30 & 68.17 \end{bmatrix}$. Since the expert provided quantile judgments for the 10th, 50th, and 75th quantiles, the matrix $\mathbf{Z}$ was $\mathbf{Z} = \begin{bmatrix} 1 & -1.28 \\ 1 & 0 \\ 1 & 0.67 \end{bmatrix}$. The mean, variance and the covariance between the distributions were obtained using the expressions $E[\hat{\theta}_1] = [1,0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}}$, $E[\hat{\theta}_2] = [0,1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\hat{\mathbf{q}}$, $\mathrm{Var}\left[\hat{\theta}_1\right] = [1,0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,0]^{\mathrm{T}}$, $\mathrm{Var}\left[\hat{\theta}_2\right] = [0,1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0,1]^{\mathrm{T}}$, and covariance $Cov[\hat{\theta}_1, \hat{\theta}_2] = [1,0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0,1]^{\mathrm{T}}$

as provided in Section 6 of the Appendix. To determine the family of the distributions for the prior distributions of $\hat{\theta}_1, \hat{\theta}_2$, we used the data generated during the error quantification process discussed in Section 2.1. For each distribution in the bench marking process *l=1,2,...,L* with true values θ_{1l}, θ_{2l} we determined the unbiased errors in the estimation of θ_1, θ_2 as $e_{1l} = \theta_{1l} - \mathbf{w}^{*}_{\theta_1}{}^{\mathrm{T}}\hat{\mathbf{q}}$ and $e_{2l} = \theta_{2l} - \mathbf{w}^{*}_{\theta_2}{}^{\mathrm{T}}\hat{\mathbf{q}}$. Then we determined the parametric families that best fit these two streams of errors. We focused on the distributions with a positive domain and found the Gamma distribution to be acceptable representations for the error streams for both parameters. We selected a grid (and increased its resolution until the posterior distribution became stable) and then used the Copula based simulation discussed in Section 4.3 to generate realizations of values $\theta_1^r, \theta_2^r; r = 1, 2, ..., 100000$, and estimated the probability mass for the prior distribution.

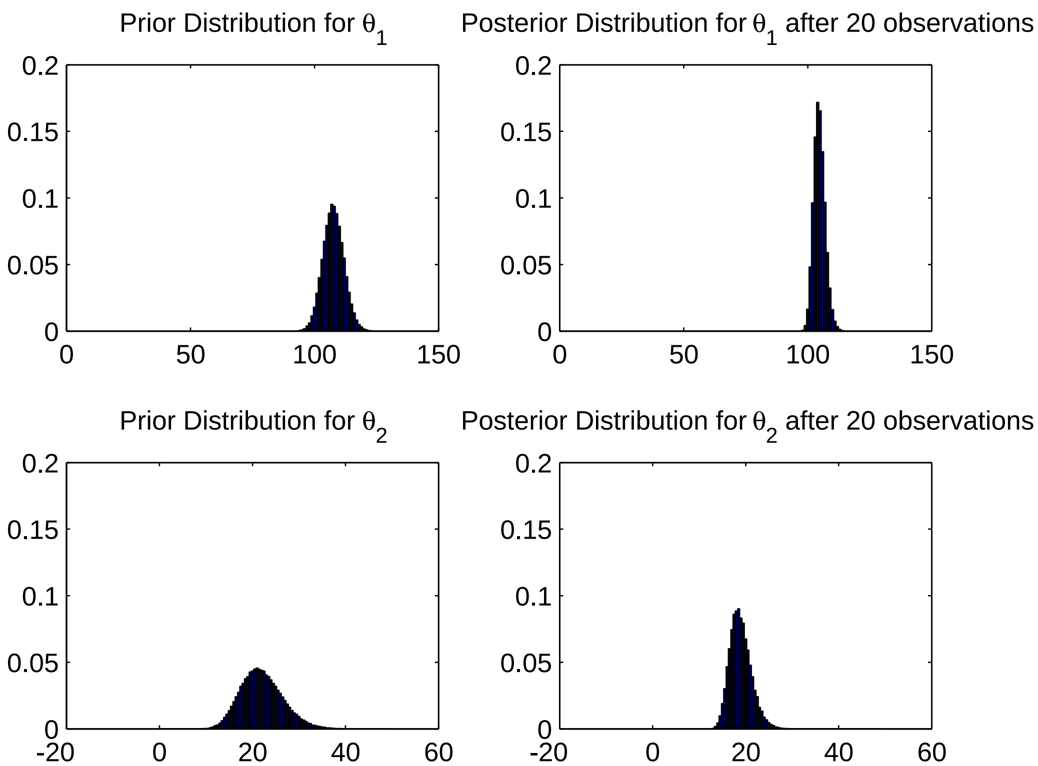


Figure 6 Bayesian updating of θ_1, θ_2 for Ω obtained at Dow and explained in Section 6, and 20 observations.

For updating the distributions for θ_1, θ_2 we followed the practice at Dow of modeling Normally distributed yields, and considered the sample of 20 points with values $119.2, 88.17, 92.83, 102.75, 105.69, 148.34, 90.65, 120.11, 100.46, 119.28, 92.30, 113.97, 119.84, 112.69, 73.53, 88.54, 115.61, 111.77, 71.52, 100$. These observations have a mean of 104.36 and the standard deviation of 18.14. The updated distributions of $\hat{\theta}_1, \hat{\theta}_2$ are shown in Figure 6. For $\hat{\theta}_1$, the mean of the prior distribution was 107.11

and the mean of the data was 104.36. The mean of the posterior distribution was 104.73 with a much smaller spread as compared to the prior as evident in Panel 2 of the figure. For $\hat{\theta}_2$, the mean of the prior distribution was 21.76 and the mean of the data was 18.14. The mean of the posterior distribution was 18.90 with a much smaller spread as compare to the prior as evident in Panel 4 of the figure.

10. Proof of Proposition 5

We need three intermediate results. Consider the matrix $\Omega = K\Omega'$ where $K > 0$ is a scalar, the diagonals elements of Ω' are equal to 1, and the off-diagonal elements are equal to the correlation value ρ. We can ignore K for the analysis below. With this simplification, $\Omega = \Omega'$.

(a) $\Omega^{-1} = \frac{1}{(1-\rho)\sigma^2}\left(I - \frac{\rho M_1}{1+(m-1)\rho}\right)$. To establish this claim define M_1 as an $(m \times m)$ matrix of ones and verify that

$$\begin{aligned}\Omega\Omega^{-1} &= \frac{1}{1-\rho}\Big((1-\rho)I + \rho M_1\Big)\Big(I - \frac{\rho}{1+(m-1)\rho}M_1\Big),\\ &= \frac{1}{1-\rho}\left[(1-\rho)I + \rho M_1 - \frac{\rho(1-\rho)}{1+(m-1)\rho}M_1 - \frac{\rho^2 m}{1+(m-1)\rho}M_1\right] = I.\end{aligned}$$

(b)

$$\mathbf{Z}^{\mathrm{T}}\Omega^{-1} = \frac{1}{\sigma^2(1+(m-1)\rho)}\begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \frac{(1+(m-1)\rho)z_1 - S_1\rho}{1-\rho} & \cdots & \frac{(1+(m-1)\rho)z_i - S_1\rho}{1-\rho} & \cdots & \frac{(1+(m-1)\rho)z_m - S_1\rho}{1-\rho}\end{pmatrix}.$$

To establish this claim we use (a) to obtain

$$\begin{aligned}\mathbf{Z}^{\mathrm{T}}\Omega^{-1} &= \frac{1}{(1-\rho)\sigma^2}\begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \\ z_1 & \cdots & z_i & \cdots & z_m\end{pmatrix}\Big(I - \frac{\rho}{1+(m-1)\rho}M_1\Big)\\ &= \frac{1}{\sigma^2(1+(m-1)\rho)}\begin{pmatrix} 1 & \cdots & 1 \\ \frac{(1+(m-1)\rho)z_1 - S_1\rho}{1-\rho} & \cdots & \frac{(1+(m-1)\rho)z_m - S_1\rho}{1-\rho}\end{pmatrix}.\end{aligned}$$

(c) Denote $S_1 \equiv \sum_{i=1}^m z_i$ and $S_2 \equiv \sum_{i=1}^m z_i^2$, then the inverse of (2×2) matrix $\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z}$ is obtained as

$$(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1} = \frac{\sigma^2}{mS_2 - S_1^2}\begin{pmatrix} S_2(1+(m-1)\rho) - S_1^2\rho & -S_1(1-\rho) \\ -S_1(1-\rho) & m(1-\rho)\end{pmatrix}.$$

Therefore, the weights for the mean are obtained as $\mathbf{w}^*_{\mu_i} = [1,0](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}$, which on simplification reduce to $w_i = \frac{(S_2 - S_1 z_i)}{mS_2 - S_1^2}$. The weights for the standard deviation are obtained as $\mathbf{w}^*_{\sigma_i} = [0,1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\Omega^{-1}$, which on simplification reduce to $w^*_{\sigma_i} = \frac{S_1(S_2 - S_1 z_i)}{S_2(mS_2 - S_1^2)} - \frac{z_i}{S_2}$.

We now show that these weights coincide with the weights obtained for the minimization of the least squares $F = \sum(\theta_1 - z_i\theta_2 - \hat{q}_i)^2$. The summation is over indices *i=1,2,...,m*. Taking the first order derivatives, we obtain:

$$\frac{\partial F}{\partial \theta_1} = 2\sum(\theta_1 - z_i\theta_2 - \hat{q}_i) = 0 \tag{33}$$

$$\frac{\partial F}{\partial \theta_2} = 2\sum(\theta_1 - z_i\theta_2 - \hat{q}_i)z_i = 0 \tag{34}$$

Now, we can write (33) as $n\theta_1 - \theta_2\sum z_i - \sum \hat{q}_i = 0$, or, using $\sum z_i = S_1$ equivalently,

$$\theta_1 = \frac{\theta_2 S_1 + \sum \hat{q}_i}{n} \tag{35}$$

Next, we can simplify (34) as $\theta_1 \sum z_i - \theta_2 \sum z_i^2 - \sum \hat{q}_i z_i = 0$, or, using $\sum z_i^2 = S_2$ equivalently as,

$$\theta_2 = \frac{\theta_1 S_1 - \sum \hat{q}_i z_i}{S_2} \tag{36}$$

Now, substituting (36) into (35), and simplification, we obtain

$$\theta_1 = \frac{\sum \hat{q}_i (S_2 - S_1 z_i)}{m S_2 - S_1^2} \tag{37}$$

This relationship implies the weights of $w_i = \frac{(S_2 - S_1 z_i)}{m S_2 - S_1^2}$ for the estimation of θ_1. The weights for θ_2 are obtained similarly. □

11. Proof of Proposition 6

(i) The weights for benchmarks I and II both are obtained by a problem mis-specification of ignoring errors. Therefore, the optimal (minimum) value of the objective function will always exceed the true optimal value.

(ii) We prove the result for the estimation of the mean. The result for the estimation of standard deviation follows a similar proof. We first note from Proposition 3 that these sizes can be written as $N_\mu = \frac{\sigma^2}{Var(\hat{\mu})}$, $N_\mu^I = \frac{\sigma^2}{Var(\hat{\mu}^I)}$, and $N_\mu^{II} = \frac{\sigma^2}{Var(\hat{\mu}^{II})}$. It follows from part (i) of this proposition that $Var(\hat{a}^I) \geq Var(\hat{a})$ and $Var(\hat{a}^{II}) \geq Var(\hat{a})$ for $\hat{a} \in \{\hat{\mu}, \hat{\sigma}\}$. It follows that $N_\mu \geq N_\mu^I, N_\mu \geq N_\mu^{II}$. □