Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Using Experts' Noisy Quantile Judgments to Quantify Risks: Theory and Application to Agribusiness

http://orcid.org/0000-0001-5443-4144Saurabh Bansal, Genaro J. Gutierrez, John R. Keiser

informs
http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Articles in Advance, pp. 1–16
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Using Experts' Noisy Quantile Judgments to Quantify Risks: Theory and Application to Agribusiness

Saurabh Bansal,[a] **Genaro J. Gutierrez,**[b] **John R. Keiser**[c]

[a] The Pennsylvania State University, University Park, Pennsylvania 16802; [b] The University of Texas at Austin, Austin, Texas 78712; [c] Dow AgroSciences, Marshalltown, Iowa 50158

Contact: sub32@psu.edu, http://orcid.org/0000-0001-5443-4144 (SB); genaro@austin.utexas.edu (GJG); JRKeiser@dow.com (JRK)

Received: July 31, 2015
Revised: September 17, 2015; July 6, 2016; November 2, 2016
Accepted: December 15, 2016
Published Online in Articles in Advance: July 24, 2017

Subject Classifications: decision analysis: risk; forecasting applications; industries: agriculture/food
Area of Review: OR Practice

https://doi.org/10.1287/opre.2017.1627

Abstract. Motivated by a unique agribusiness setting, this paper develops an optimization-based approach to estimate the mean and standard deviation of probability distributions from noisy quantile judgments provided by experts. The approach estimates the mean and standard deviations 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 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 and enables combining the expert's judgments with those of other experts. It also shows analytically that the weights for the mean add up to one and the weights for the standard deviation add up to zero—these properties have been observed numerically in the literature in the last 30 years, but without a systematic explanation. The theory has been in use at Dow AgroSciences for two years for making an annual decision worth $800 million. The use of the approach has resulted in the following monetary benefits: (i) firm's annual production investment has reduced by 6%–7% and (ii) profit has increased by 2%–3%. We discuss the implementation at the firm, and provide practical guidelines for using expert judgment for operational uncertainties in industrial settings.

Funding: This research was supported, in part, by grants from the Center for Supply Chain Research at Penn State, and the Supply Chain Management Center of the McCombs School of Business at The University of Texas at Austin.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2017.1627.

Keywords: expert judgments • quantile judgments • estimating distributions • bootstrap • yield uncertainty

1. Introduction and Industry Motivation

1.1. Problem Context

Understanding and quantifying production-related uncertainties is critical for decision making in businesses. The probability distributions for these uncertainties are usually estimated using historical data obtained during repetitive manufacturing. But these data may not be available when firms frequently launch new products in the market, e.g., at firms in the semiconductor and in the agribusiness industry. In the absence of historical data, these firms turn to domain experts for obtaining subjective probability distributions (e.g., Baker and Solak 2014).

Prior literature (e.g., O'Hagan and Oakley 2004) advises that in these situations, one should avoid obtaining direct estimates of the standard deviation from domain experts as this quantity is not intuitive to estimate. This literature recommends obtaining experts' input in the form of judgments for specific discrete points on distributions, for example, judgments for specific quantiles, but also cautions that these judgments are subject to judgmental errors (Ravinder et al. 1988). However, a systematic approach that uses these judgments to deduce the mean and standard deviation of probability distributions, *while explicitly modeling and accounting for experts' judgmental errors*, is not yet established. In this paper, we accomplish this task. Specifically, we develop an approach to deduce the mean and standard deviation using judgments provided by one or multiple experts for distribution quantiles (or fractiles). This approach is analytically tractable, provides the flexibility of using judgments for any set of quantiles that an expert is willing to provide, and is amenable to combining an expert's quantile judgments with those of other experts. The approach also establishes a novel equivalence between the quality of an expert's judgments and the size of an experimental sample that is equally informative about the distribution.

This approach was developed to manage a dynamic new product development situation at Dow AgroSciences (DAS) for an annual decision worth $800 million. Analysis on the firm's historical decisions shows that the use of the approach has resulted in the following monetary benefits: (i) firm's annual production investment has reduced by 6%–7% and (ii) profit has increased by 2%–3%. We also discuss the implementation of the approach developed at the firm, and practical guidelines for seeking expert judgment for operational uncertainties in industrial situations.

The rest of this paper is organized as follows. Section 2 provides an overview of our approach and the contributions to the existing literature. Sections 3 and 4 discuss a model for deducing mean and standard deviation from quantile judgments, derive the solution and its structural properties. Section 5 discusses an equivalence of expertise with randomly collected data and results for combining judgments from multiple experts. Section 6 describes implementation of the approach at DAS and quantification of benefits of using the approach. Section 7 concludes with insights for practice.

2. Overview of Approach and Our Contributions to the Existing Literature

2.1. Overview

We develop an optimization-based approach to estimate the mean and standard deviation from quantile judgments provided by an expert; specifically, we obtain the minimum variance estimates of the mean and standard deviation of yield distributions as weighted averages of the quantile judgments provided by the expert, subject to the constraint that the estimates are unbiased. The model has two inputs. The first input is an identification of the quantiles for which the expert will provide judgments. For example, at DAS, the expert chose to provide judgments for the 10th, 50th, and 75th quantiles because his software was set to show these quantiles during data analysis, and he was accustomed to thinking about these quantiles. The second input is a quantification of the noise present in the expert's judgments for the quantiles. This quantification is done separately by comparing the expert's judgments for the quantiles with the true values for a number of distributions constructed using historical data. We discuss this empirical estimation at DAS in Section 6.

The solution to the optimization model assigns two sets of weights to the quantile judgments (e.g., for the 10th, 50th, and 75th quantiles). The first set of weights is for estimating the mean as a weighted average of quantile judgments. The second set of weights is used similarly to estimate the standard deviation. The weights are specific to the noise quantified in the second input discussed above for the expert's quantile judgments.

2.2. Contributions to Literature

A large body of literature considers situations in which experts provide their assessments for events with binary outcomes (e.g., Ayvaci et al. 2017). In contrast, we focus on situations where continuous distributions need to be specified over the outcomes, and expert judgment is sought to estimate these probability distributions. Two streams of literature are relevant to our focus: (i) models of judgmental errors and (ii) practice-driven literature on the use of expert judgments.

2.2.1. Models on Model-Driven Theory on Judgments.

The first stream of related literature is on model-driven theory of judgmental errors. The existing literature on expert judgments acknowledges the potential severity of judgmental errors and focuses on developing elicitation guidelines for reducing judgmental errors (e.g., Koehler et al. 2002). In contrast, articles on moment estimation from quantile judgments have explored the problem of deducing moments from the median and two additional symmetric quantiles (typically, the 5th and 95th) or four additional symmetric quantiles (typically, the 5th, 25th, 75th, and 95th), *numerically* with a key assumption: *no judgmental errors are present*. Pearson and Tukey (1965) and Keefer and Bodily (1983) follow this paradigm. But no prior articles consider the problem where subjective quantile judgments from multiple experts need to be combined to deduce the mean and standard deviation or the case where an expert provides judgments for an arbitrary set of quantiles that are different from the standard ones mentioned above. We contribute to this literature by developing a tractable solution approach to this problem. A salient feature of this approach is that one can use *any* set of quantile judgments that an expert can provide (over the 5th, 25th, 75th, and 95th as discussed in the prior literature) to estimate the mean and standard deviation. This feature is useful for practice since an expert may not be willing to provide quantile judgments for specific symmetric quantiles. For example, the expert at DAS was habituated to seeing the 10th, 50th, and 75th quantiles for historical data on his software and was willing to estimate only these quantiles.

The approach also provides the following structural insights. First, regardless of the magnitude of an expert's judgmental errors and the quantiles elicited, the variance-minimizing 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), among others, who all assume that judgmental errors do not exist. Second, 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 the expert's judgments for quantiles. This equivalence enables an objective comparison of experts. Finally, in our approach, the optimal weights provide point estimates and variability in the estimates for the moments. This quantification of variability of moment estimates enables us to combine quantile judgments from multiple experts in a rational and consistent manner.

Prior literature, e.g., O'Hagan (1998), Stevens and O'Hagan (2002), discusses the role of expert judgments in the absence of data for constructing prior distributions; when data become available, posterior distributions for parameters are obtained using Bayesian updating. The literature discusses two cases. When conjugate priors are used, the posterior distributions are obtained in closed form. When conjugate prior cannot be used, numerical approaches must be used to obtain posterior distributions. We make two contributions to this literature. First, we develop a novel approach to obtain the prior distributions on the parameters for the mean and standard deviations of distributions using expert judgments for quantiles. Second, we show that the joint prior distributions are correlated and are not conjugate priors, and we develop a Copula-based approach to obtain the posterior distributions.

2.2.2. Practice-Driven Tools and Insights. Our contributions to practice are as follows. We provide a step-by-step approach to quantifying an expert's judgmental errors and then discuss some practical issues observed during this quantification at DAS. Specifically, we discuss a bootstrapping approach to separate judgmental errors from sampling errors during the error quantification process. Then, we show that the information provided by the expert is equivalent to five to six years of data collection at DAS using our approach. Such quantification has not been reported in practice literature before. We also report that the expert at DAS was reluctant to provide judgments for extreme quantiles because of his inability to distinguish between random variations and systematic reasons as causes of extreme outcomes. This observation suggests that contextual reasons and experts' preferences can lead to elicitation of quantiles that are different from the standard values (median and two or four symmetric quantiles); our approach is especially useful in such situations.

3. Analytical Model

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 $\boldsymbol{\theta} = [\theta_1, \theta_2]^{\mathrm{T}}$ are the parameters of the PDF, and μ_1, μ_2 denote the mean and standard deviation, respectively. Similar to Lindley (1987), O'Hagan (2006), we assume that the distribution family is known from the application context, but the parameters are not known. This framework is especially relevant to a number of operations contexts in which the parametric family of probability distributions is known from the historical data available or from formal models. The cumulative distribution function (CDF) of X is denoted as Φ. A source of information such as an expert provides quantile judgments $\hat{x}_i$ corresponding to probability CDF values p_i for $i = 1, 2, \ldots, m$. 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 $\mathbf{p} = [p_1, \ldots, p_m]^{\mathrm{T}}$.

We seek to develop an approach to deduce μ_1, μ_2 from the quantile judgments $\hat{\mathbf{x}}$. From theoretical and application perspective, it is desirable that the approach's formulation provides a unique solution to the problem, preferably in closed form, and is amenable to sensitivity analysis. Prior literature in this domain (e.g., Keefer and Bodily 1983, Johnson 1998) also suggests that for an ease of implementation, the approach should be consistent with moment matching and with other probability discretization practices in use, e.g., program evaluation and review technique (PERT) for project management. Our approach accomplishes these objectives and additionally provides a quantification of the quality of expert's judgments into an equivalent sample size.

3.1. Preliminaries for Expert Judgments

We assume that the quantile judgments are obtained using an underlying process or mental model (we discuss the mental model used by the expert at DAS, in Figure 1, Section 6.1), which is error prone but is used consistently for generating quantile judgments. This assumption means that the expert's judgmental errors are stable during elicitation. We further assume an additive error structure that is used frequently in the literature (e.g., Ravinder et al. 1988): the quantile judgment $\hat{x}_i$ is composed of a true value x_i and an additive error e_i:

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

In vector notation, the error model is $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{e}$. Consistent with this literature, we assume that the error e_i has two parts: a systematic component or bias δ_i and a random component or noise 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 judgments for quantile i from the true value. The noise ϵ_i captures the spread in the error due to random variations. In vector notation, the bias and residual variation are denoted as $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$, respectively. The noise $\boldsymbol{\epsilon}$ is quantified in variance-covariance matrix Ω. The diagonal elements of this matrix $\omega_{ii} = \mathrm{Var}(\epsilon_i)$ denote the variance in the unbiased judgment of quantile i. The off-diagonal elements are covariances of unbiased judgments $\omega_{ij} = \mathrm{Cov}(\epsilon_i, \epsilon_j)$. We discuss the empirical estimation of $\boldsymbol{\delta}$ and Ω separately in Section 6, and assume for now that these quantities are available.

From the biased judgments $\hat{\mathbf{x}}$, the unbiased judgments $\hat{\mathbf{q}}$ are obtained by removing bias as $\hat{\mathbf{q}} = \hat{\mathbf{x}} - \boldsymbol{\delta}$. Substituting this relationship into $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{e} = \mathbf{x} + \boldsymbol{\delta} + \epsilon$, we obtain

$$\hat{\mathbf{q}} = \mathbf{x} + \boldsymbol{\epsilon}. \tag{2}$$

The matrix Ω for $\boldsymbol{\epsilon}$ is used as an input in the optimization model discussed next.

3.2. Optimization Problem

We seek to obtain the estimates of the mean $\hat{\mu}_1$ and standard deviation $\hat{\mu}_2$ as pooled or weighted linear functions of the debiased quantile judgments as $\hat{\mu}_k = \mathbf{w}_k^\intercal \hat{\mathbf{q}}$; $k = 1, 2$ with the weights $\mathbf{w}_1 \equiv [w_{11}, w_{12}, \ldots, w_{1m}]^\intercal$ and $\mathbf{w}_2 \equiv [w_{21}, w_{22}, \ldots, w_{2m}]^\intercal$. Since the unbiased judgments $\hat{\mathbf{q}}$ are subject to noise ϵ, the estimates $\hat{\mu}_k$; $k = 1, 2$ have variances $\mathrm{Var}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}]$. Smaller values of the variances of these estimates are desirable as it would imply that the estimates are more precise. To this end, it is desirable to select weights $\mathbf{w}_k$; $k = 1, 2$ that lead to a small variance in the estimates $\hat{\mu}_k$. We first restate the variance of estimates $\hat{\mu}_k$ in terms of the weights as

$$\begin{aligned}\mathrm{Var}[\hat{\mu}_k] = \mathrm{Var}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] &= \mathrm{E}[(\mathbf{w}_k^\intercal \hat{\mathbf{q}} - \mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}])^2] \qquad (3)\\ &= \mathrm{E}[(\mathbf{w}_k^\intercal (\mathbf{x}+\boldsymbol{\epsilon}) - \mathrm{E}[\mathbf{w}_k^\intercal (\mathbf{x}+\boldsymbol{\epsilon})])^2]\\ &= \mathrm{E}[(\mathbf{w}_k^\intercal (\mathbf{x}+\boldsymbol{\epsilon}) - \mathbf{w}_k^\intercal \mathbf{x})^2]\\ &= \mathrm{E}[\mathbf{w}_k^\intercal \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\intercal \mathbf{w}_k] = \mathbf{w}_k^\intercal \Omega \mathbf{w}_k. \qquad (4)\end{aligned}$$

Prior literature (e.g., Bates and Granger 1969, Granger 1980) shows that only minimizing this variance is not informative as it is minimized by setting $\mathbf{w}_k = \mathbf{0}$ and the resultant weighted linear estimate is always equal to $\hat{\mu}_k = 0$ for all judgments $\hat{\mathbf{q}}$. This literature suggests adding constraints to make statistical estimates responsive to forecasts or judgments. Our focus will be on a specific class of such constraints. We seek variance-minimizing weights such that the obtained estimates $\mathbf{w}_k^\intercal \hat{\mathbf{q}}$ are unbiased, i.e., $\mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] = \mu_k$, leading to the following optimization formulations for $k = 1, 2$:

$$\begin{aligned}\min_{\mathbf{w}_k} \quad & \mathrm{Var}[\hat{\mu}_k] = \mathbf{w}_k^\intercal \Omega \mathbf{w}_k\\ \text{s.t.} \quad & \mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] = \mu_k\end{aligned} \qquad (5)$$

Problem (5) consists of finding the weights $\mathbf{w}_k$ that lead to the minimum variance unbiased estimates of μ_k. In the next section, we determine these weights for location-scale distributions using structural properties of these distributions. The focus on location-scale distributions is motivated by their widespread application in numerous operations management contexts (see Kelton and Law 2006, for a list of these applications) as well as their specific application at DAS, where the in-house statistics team has shown using existing data that yields are normally distributed. This analysis is in Section 6.2.

4. Solution: Weights for Quantile Judgments

In Section 4.1 we specialize the problem (5) for distributions of a location-scale family, and obtain the optimal weights for quantile judgments and the weights' structural properties in Section 4.2.

4.1. Reformulation and Solution for Distributions of a Location-Scale Family

We now assume that X is a location-scale random variable with location and scale parameters $\theta_1 \in \mathbb{R}$ and $\theta_2 \in \mathbb{R}_{++}$, respectively, and transform the constraint $\mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] = \mu_k$ in formulation (5) using two properties of location-scale distributions. The first property enables us to rewrite the left-hand side (LHS) of this constraint as a function of θ_1, θ_2. If X is a location-scale random variable with PDF $\phi(\cdot; \boldsymbol{\theta})$, then a specific value x that corresponds to probability p can be expressed as $x = \theta_1 + \theta_2 z$, where z denotes the value of *standardized* random variable with the *standardized* PDF $\phi(\cdot; [0,1]^\intercal)$ for probability p (Casella and Berger 2002, p. 116). We write this expression in vector form as

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\theta}, \qquad (6)$$

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. Substituting (6) into the LHS of the error model in (2), $\hat{\mathbf{q}} = \mathbf{x} + \boldsymbol{\epsilon}$, it follows that

$$\mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] = \mathrm{E}[\mathbf{w}_k^\intercal (\mathbf{x}+\boldsymbol{\epsilon})] = \mathbf{w}_k^\intercal \mathrm{E}[\mathbf{x}+\boldsymbol{\epsilon}] = \mathbf{w}_k^\intercal \mathbf{Z}\boldsymbol{\theta}. \qquad (7)$$

The second property, formalized in Lemma 1 below, enables us to rewrite the right-hand side (RHS) of the unbiasedness constraint $\mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] = \mu_k$ as a function of θ_1, θ_2.

Lemma 1 (Characterization of Location-Scale Moments). *If X is a location-scale random variable with parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]^\intercal$ with finite jth moments for $j = 1, 2, \ldots$, then,*

(a) *the raw moments* $\mathrm{E}[X^j]$ *are given by* $\mathrm{E}[X^j] = \sum_{i=0}^{j} \binom{j}{i} \theta_1^j \theta_2^{j-i} \kappa_{j-i}$ *and*

(b) *the central moments are given by* $E[(X - \mu_1)^j] = \theta_2^j \sum_{i=0}^{j} (-\kappa_1)^i \kappa_{j-i}$,

where the constants κ_i are $\kappa_0 = 1$ and $\kappa_j = \mathrm{E}[Z^j]$ for $j = 1, 2, \ldots$.

The proof is in Appendix A1. The values of κ_j are documented in the literature for location-scale distributions (see, e.g., Johnson et al. 1994). For example, for a normal distribution, we have $(\kappa_0, \kappa_1, \kappa_2) = (1, 0, 1)$. It follows from part (a) of Lemma 1 that $\mu_1 = E[X] = [1, \kappa_1]\boldsymbol{\theta}$. It follows from part (b) of the lemma that variance of X is equal to $E[(X - \mu_1)^2] = \theta_2^2(\kappa_2 - \kappa_1^2)$, and therefore the standard deviation is equal to $\mu_2 = \sqrt{E[(X-\mu_1)^2]} = \sqrt{\theta_2^2(\kappa_2 - \kappa_1^2)} = [0, \sqrt{\kappa_2 - \kappa_1^2}]\boldsymbol{\theta}$. We write both relationships in vector notation as

$$\mu_k = \mathbf{a}_k^\intercal \boldsymbol{\theta} \qquad (8)$$

with $\mathbf{a}_1^\intercal = [1, \kappa_1]$ and $\mathbf{a}_2^\intercal = [0, \sqrt{\kappa_2 - \kappa_1^2}]$.

Substituting (7) and (8) into the LHS and RHS of the constraint $\mathrm{E}[\mathbf{w}_k^\intercal \hat{\mathbf{q}}] = \mu_k$, respectively, we obtain the following condition on the weights $\mathbf{w}_k$ for the estimate $\mathbf{w}_k^\intercal \hat{\mathbf{q}}$ to be unbiased. This condition will help us solve the problem in a tractable form.

Proposition 1. *If X is a random variable with a location-scale distribution, the weighted linear estimator $\mathbf{w}_k^{\mathrm{T}}\hat{\mathbf{q}}$ is unbiased for μ_k, if and only if the weights $\mathbf{w}_k$ satisfy*

$$\mathbf{Z}^{\mathrm{T}}\mathbf{w}_k = \mathbf{a}_k; \quad k = 1, 2. \tag{9}$$

Proof. By definition, the estimate $\mathbf{w}_k^{\mathrm{T}}\hat{\mathbf{q}}$ is unbiased if and only if $\mathrm{E}[\mathbf{w}_k^{\mathrm{T}}\hat{\mathbf{q}}] = \mu_k$. Substituting (7) and (8) into the LHS and RHS of the constraint, it follows that the estimate is unbiased if and only if $\mathbf{w}_k^{\mathrm{T}}\mathbf{Z}\boldsymbol{\theta} = \mathbf{a}_k^{\mathrm{T}}\boldsymbol{\theta}$ for all values of θ_1 and θ_2. It follows that the estimator is unbiased if and only if $\mathbf{w}_k^{\mathrm{T}}\mathbf{Z} = \mathbf{a}_k^{\mathrm{T}}$, i.e., $\mathbf{Z}^{\mathrm{T}}\mathbf{w}_k = \mathbf{a}_k$; $k = 1, 2$. □

The implication of the iff in Proposition 1 is that we can replace the constraint $\mathrm{E}[\mathbf{w}_k^{\mathrm{T}}\hat{\mathbf{q}}] = \mu_k$ in formulation (5) with the condition on weights $\mathbf{Z}^{\mathrm{T}}\mathbf{w}_k = \mathbf{a}_k$; $k = 1, 2$. After this substitution, we obtain the formulation for $k = 1, 2$ as

$$\begin{aligned} \min_{\mathbf{w}_k} \quad & \mathbf{w}_k^{\mathrm{T}}\Omega\,\mathbf{w}_k \\ \text{s.t.} \quad & \mathbf{Z}^{\mathrm{T}}\mathbf{w}_k = \mathbf{a}_k. \end{aligned} \tag{10}$$

The matrix Ω is a covariance matrix, and therefore it must be positive semidefinite. It follows that the problem (10) is a quadratic convex problem, and its solution is obtained by solving a Lagrange formulation of the problem. The next result establishes this unique solution.

Theorem 1. *The weights that solve problem* (10) *are given by* $\mathbf{w}_k^* = \Omega^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a}_k$.

The proof is in Appendix A2. The conspicuous feature of the optimal weights $\mathbf{w}_k^*$ is that they are explicit functions of the expert's precision encoded in Ω. Therefore, a change in an expert's precision in providing quantile judgments will modify Ω, which, in turn, will change the optimal weights $\mathbf{w}_k^*$ for the quantile judgments. Finally, we note that the variance of the estimates $\mathrm{Var}[\hat{\mu}_k]$ at the optimal weights is equal to $\mathrm{Var}[\hat{\mu}_k] = \mathbf{w}_k^{*\mathrm{T}}\Omega\,\mathbf{w}_k^*$, which simplifies to $\mathrm{Var}[\hat{\mu}_k] = \mathbf{a}_k^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a}_k$, establishing a direct link between the variance in the estimates $\hat{\mu}_k$ to the expert-specific Ω.

4.2. Structural Properties and Generalization of Results Available in Literature

The development thus far provides new generalizations and insights to the existing literature. First, in our approach, 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 5th, 25th, 50th, 75th, and 95th quantiles as specified in extant literature such as Lau et al. (1996). This flexibility is useful since we no longer need to convince an expert to provide judgments for these 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 existing literature that the weights add up to a constant as follows.

Proposition 2. *The optimal weights for quantiles add up to constants. Specifically,* $\sum_{i=1}^{m} w_{1i}^* = 1$ *and* $\sum_{i=1}^{m} w_{2i}^* = 0$.

The proof is in Appendix A3. This result is true regardless of the numerical values of Ω; therefore, it holds true even when the judgmental errors are arbitrarily small, e.g., when $\Omega = \lim_{\lambda \to 0} \lambda I$. A number of prior articles (Pearson and Tukey 1965, Perry and Greig 1975, Keefer and Bodily 1983, Johnson 1998) numerically discuss the limiting case when the errors are absent. They select specific numerical 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 result in the smallest squared deviations between the true and the estimated means over all cases. A similar analysis provides the weights to obtain the standard deviation. The weights recommended in this literature add up to 1 and 0 for the mean and standard deviation, respectively. Proposition 2 establishes that the additivity properties observed numerically in these articles are structural properties of probability distributions, and hold true for any magnitude of judgmental errors.

Third, these additivity properties are also shared by the weights assigned in the project management technique PERT to the estimates for the optimistic, pessimistic, and most likely scenarios. The weights for the mean are (1/6, 4/6, 1/6), respectively, for the estimation of the mean adding up to one, and (−1/6, 0, 1/6) for the estimation of the standard deviation adding up to zero. Fourth, one can show using straightforward algebra that our approach automatically assigns lower weight to a quantile judgment that has large noise. In situations where an expert provides a number of quantile judgments, this feature is useful in identifying which quantile judgments have large noise and are therefore not useful for the estimation of the moments; the weights for these quantile judgments will be negligible.

5. Data Equivalence, Multiple Experts, and Other Relationships

In Section 5.1, we determine the size of a randomly drawn sample that is equivalent in terms of precision to the expert's judgments. In Section 5.2 we discuss combining judgments of one expert with the judgments from other experts. In Section 5.3, we discuss the consistency of the approach developed with least squares and moment matching.

5.1. 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, provide point-estimates $\hat{\mu}_k$ and the variances in these estimates $\mathrm{Var}[\hat{\mu}_k]$. We can compare this variance with the variance of the mean and standard deviation obtained from a sample of random observations for X if data collection is possible. Specifically, it is well known that a sample mean has a variance of σ^2/N_1, where N_1 is the sample size. In our approach, the variance in the estimate of the mean is equal to $\mathrm{Var}[\hat{\mu}_1] = \mathbf{w}_1^{*\mathrm{T}}\Omega\mathbf{w}_1^*$ (see Appendix A4 for proof), which can be simplified to $[1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}$. 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. A similar analysis provides the equivalent sample size for the standard deviation. The next result provides expressions of these equivalent sample sizes.

Proposition 3. *The precision of the estimates $\hat{\mu}_k$ 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_k, where*

$$N_1 = \frac{\mu_2^2}{[1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}}} \quad \textit{and}$$

$$N_2 \approx \frac{\mu_2^2\left(\dfrac{\sum_{j=0}^{4}(-\kappa_1)^j\kappa_{4-j}}{(\kappa_2-\kappa_1^2)^2} - \dfrac{(\sum_{j=0}^{2}(-\kappa_1)^j\kappa_{2-j})^2}{(\kappa_2-\kappa_1^2)^2}\right)}{4[0,\sqrt{\kappa_2-\kappa_1^2}](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}[0,\sqrt{\kappa_2-\kappa_1^2}]^{\mathrm{T}}}.$$

The proof is in Appendix A5. This result has two profound implications. First, using this result, multiple experts can be compared objectively based on their judgmental errors quantified in Ω. More specifically, for two experts A and B with matrices Ω_A and Ω_B, the ratios of equivalent sample sizes are given as $N_1^A/N_1^B = ([1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega_B^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}})/([1,\kappa_1](\mathbf{Z}^{\mathrm{T}}\Omega_A^{-1}\mathbf{Z})^{-1}[1,\kappa_1]^{\mathrm{T}})$ and $N_2^A/N_2^B = ([0,\sqrt{\kappa_2-\kappa_1^2}](\mathbf{Z}^{\mathrm{T}}\Omega_B^{-1}\mathbf{Z})^{-1}[0,\sqrt{\kappa_2-\kappa_1^2}]^{\mathrm{T}})/([0,\sqrt{\kappa_2-\kappa_1^2}](\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})_A^{-1}[0,\sqrt{\kappa_2-\kappa_1^2}]^{\mathrm{T}})$, and they are independent of the true value of μ_1,μ_2. For example, if $N_k^A/N_k^B = 2$, then the estimates of μ_k obtained from expert A are two times as reliable as the estimates obtained from expert B. This benefit from using expert A over expert B is equal to the benefit from doubling the sample size of experimental or field data for the purposes of estimating μ_k. Second, some recent literature (e.g., Akcay et al. 2011) quantifies the marginal benefit of improving estimates of probability distributions by collecting more data before making decisions under uncertainty. Proposition 3 provides a natural connection to these results by quantifying the economic benefits of improving the precision in judgments.

5.2. Combining Estimates from Multiple Experts

The technical development extends to multiple experts $j = 1,2,\ldots,n$ as follows. We first construct the combined matrix

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1n} \\ \Omega_{12} & \Omega_{22} & \cdots & \Omega_{2n} \\ \vdots & & & \\ \Omega_{1n} & \Omega_{2n} & \cdots & \Omega_{nn} \end{bmatrix}$$

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. Then, the matrix Ω is used in Theorem 1 along with matrix $\mathbf{Z}^{\mathrm{T}}$ of size $2\times mn$, $\mathbf{Z}^{\mathrm{T}} = [\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_2^{\mathrm{T}},\ldots,\mathbf{Z}_n^{\mathrm{T}}]$, where each $\mathbf{Z}_j = [\mathbf{1},\mathbf{z}_j]$, $\mathbf{z}_j$ is the column vector of standardized quantiles corresponding to the probabilities $\mathbf{p}_j$ that expert j has chosen to provide judgments for, and $\mathbf{1}$ is a column vector of ones. The use of Theorem 1 provides mn weights; the first m weights for the first expert, the next m weights for the second expert, and so on.

We discuss a special case of interest here. Suppose n experts $j = 1,2,\ldots,n$ provide judgments for the same set of quantiles, i.e., $\mathbf{Z}^{\mathrm{T}} = [\mathbf{Z}_0^{\mathrm{T}}\mathbf{Z}_0^{\mathrm{T}},\ldots,\mathbf{Z}_0^{\mathrm{T}}]$, and the covariance matrix of each expert j is given by $\Omega_{jj} = r_j\Omega_0\ \forall\, j$ (i.e., the judgmental error structure of one expert is a scaled version of another expert) and further assume that the errors of any two experts are mutually independent, i.e., all elements of Ω_{ij}, $i\neq j$ are equal to 0. The weights for the mn quantile judgments obtained using Theorem 1 are denoted as $\mathbf{w}_k^*$ with elements w_{kt}^* for $t = 1,2,\ldots,mn$ and $k = 1,2$. The first m weights are for expert 1, the next m weights are for expert 2, and so on. We can write these weights as $\mathbf{w}_k^* = [\mathbf{w}_k^1,\ldots,\mathbf{w}_k^n]$, where $\mathbf{w}_k^j$ is the vector of weights of expert j. We can also decompose the weights $\mathbf{w}_k^*$ as the product of constant α_j for expert j and a common weight vector of m weights $\mathbf{w}_k^{\mathrm{c}}$ that would be obtained if each expert was the only one available, i.e., $\mathbf{w}_k^* = [\alpha_1\mathbf{w}_k^{\mathrm{c}},\ldots,\alpha_n\mathbf{w}_k^{\mathrm{c}}]$. The values of α_j and the relationships between $\mathbf{w}_k^j$ and $\mathbf{w}_k^{\mathrm{c}}$ are as follows.

Proposition 4. *Consider experts $j = 1,2,\ldots,n$, 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, the optimal weights for his unbiased judgments would be $\mathbf{w}_k^{\mathrm{c}\,\mathrm{T}} = \mathbf{a}_k^{\mathrm{T}}(\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}\mathbf{Z}_0)^{-1}\mathbf{Z}_0{}^{\mathrm{T}}\Omega_0^{-1}$ 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 $\mathbf{w}_k^j = \alpha_j\mathbf{w}_k^{\mathrm{c}}$ with $\alpha_j = (1/r_j)/R$, where $R = \sum_{j=1}^{n}(1/r_j)$.*

The proof is in Appendix A6. As an illustration, suppose that we have two experts with $r_1 = 1$ and $r_2 = 2$, i.e., expert 2 is half as precise as expert 1. 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 are obtained as $\mathbf{w}_1^{c\mathrm{T}} = [-0.167\ \ 1.484\ \ -0.317]$ and $\mathbf{w}_2^{c\mathrm{T}} = [-0.576\ \ 0.190\ \ 0.386]$ 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 $\mathbf{w}_k^c$ with the expert-specific marginal weight α_j as $\mathbf{w}_k^j = \alpha_j \mathbf{w}_k^c$. It follows from part (ii) of the proposition that the expert-specific constants are $\alpha_1 = (1/r_1)/(3/2) = 2/3$ and $\alpha_2 = (1/r_2)/(3/2) = 1/3$. The weights for the mean, for example, are obtained as $\mathbf{w}_1^1 = (2/3) \times (-0.167, 1.484, -0.317) = (-0.111, 0.989, -0.211)$ and $\mathbf{w}_1^2 = (1/3) \times (-0.167, 1.484, -0.317) = (-0.055, 0.495, -0.105)$ for experts 1 and 2, 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 $\mathbf{w}_1^* = [-0.111, 0.989, -0.211, -0.055, 0.495, -0.105]$.

5.3. Relationship with Classical Least Squares Regression and Moment Matching

We now discuss how our model and its solution is consistent with (i) the classical least square minimization-based regression framework and (ii) with moment matching. In the classical regression framework, the variance-covariance matrix is $\Omega = K\Omega'$, where $K > 0$ is a scalar, the diagonals elements of Ω' are equal to 1, and the off-diagonal elements are equal to 0, i.e., $\Omega' = I$. In our context, this would be the noninformative case when the expert is equally good at estimating all quantiles and his judgmental errors are mutually independent. We showed in Theorem 1 that the optimal weights are equal to $\mathbf{w}_k^* = \Omega^{-1}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\Omega^{-1}\mathbf{Z})^{-1}\mathbf{a}_k$. Now, substituting $\Omega = KI$, we obtain the weights as $\mathbf{w}_k^* = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{a}_k$, or alternately, as the familiar kernel of the ordinary least squares in the big parentheses: $\mathbf{w}_k^{*\mathrm{T}} = \mathbf{a}_k^{\mathrm{T}}((\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{-1}\mathbf{Z}^{\mathrm{T}})$.

In the moment matching framework, we would seek to minimize the squared deviations of the debiased quantile judgments $\hat{q}_i$ obtained from the expert for probability p_i from the unobserved values of mean and standard deviation, i.e., we would seek to solve: $\min_{\mu_1,\mu_2}\{\sum_{i=1}^m (\Phi^{-1}(p_i;\mu_1,\mu_2) - \hat{q}_i)^2\}$. This approach is codified in many commercial software (e.g., @RISK) and has been used in prior academic literature (e.g., Wallsten et al. 2013). For location-scale distributions, $\Phi^{-1}(p_i;\mu_1,\mu_2) = \theta_1 + z_i\theta_2$. Using the properties that $\theta_1 = \mu_1 - (\kappa_1/\sqrt{\kappa_2-\kappa_1^2})\mu_2$ and $\theta_2 = (1/\sqrt{\kappa_2-\kappa_1^2})\mu_2$, we can rewrite this problem as

$$\min_{\mu_1,\mu_2} \sum_{i=1}^m \left(\mu_1 - \frac{\kappa_1}{\sqrt{\kappa_2-\kappa_1^2}}\mu_2 + z_i\frac{\mu_2}{\sqrt{\kappa_2-\kappa_1^2}} - \hat{q}_i\right)^2.$$

The next result establishes the classical least squares and moment matching as a special case of our approach.

Proposition 5. *Consider the original optimization problem*: $\min_{\mathbf{w}} \mathbf{w}_k^{\mathrm{T}}\Omega\mathbf{w}_k$ *subject to* $\mathrm{E}[\mathbf{w}_k^{\mathrm{T}}\hat{\mathbf{q}}] = \mu_k$.

(1) *This problem reduces to ordinary least squares solution when* $\Omega = K\Omega'$, *where* $\Omega' = I$.

(2) *Consider the moment matching problem in the form* $\min_{\mu_1,\mu_2}\{\sum_{i=1}^m (\Phi^{-1}(p_i;\mu_1,\mu_2) - \hat{q}_i)^2\}$. *It's solution is* $\hat{a} = \mathbf{w}_a^T\hat{\mathbf{q}}$ *for* $a \in \{\mu_1,\mu_2\}$ *and it is identical to the solution obtained from the original problem for* $\Omega = K\Omega'$, *where* $\Omega' = I$.

The proof is in Appendix A7. The second part of the proposition implies that given quantile judgments and no information for the noise in the judgments, the best estimates of the mean and standard deviation (under quadratic penalty) for location-scale distributions are linear functions of the quantile judgments. And these estimates coincide with solution obtained in our approach for the noninformative case of $\Omega = I$. Our approach extends the moment matching model to account for expert's judgmental errors as captured in Ω as $\min_{\mathbf{w}} \mathbf{w}_k^{\mathrm{T}}\Omega\mathbf{w}_k$ subject to $\mathrm{E}[\mathbf{w}_k^{\mathrm{T}}\hat{\mathbf{q}}] = \mu_k$ for the case when information for the expert's judgmental errors is available, i.e., when $\Omega \neq KI$. Finally, we note that our approach is amenable to Bayesian updating using Markov chain Monte Carlo methods, and we omit details for sake of brevity. We next discuss the implementation of the approached developed at DAS.

6. Implementation Details and Benefits at DAS

6.1. Industry Context: Estimating Production Yield Distributions for Hybrid Seeds

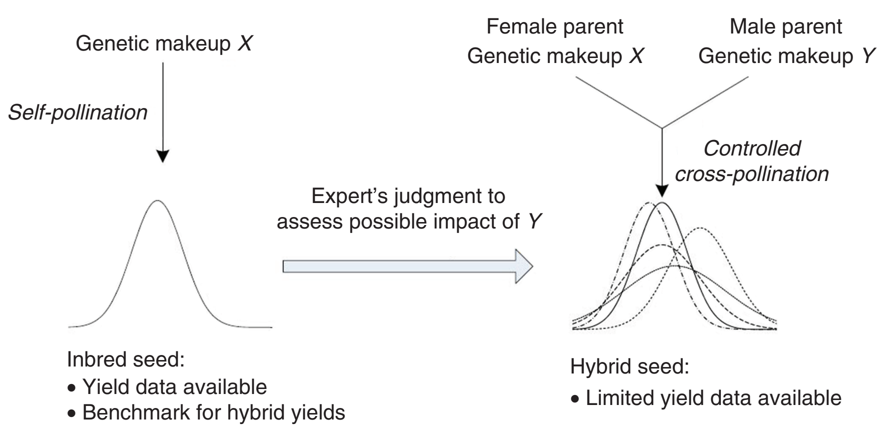
DAS produces seeds for various crops such as corn and soybean, and sells these seeds to farmers. Our focus is on the production of hybrid seed corn. DAS decides annually how many acres of land to use to produce hybrid seed corn. The yield, or amount of hybrid seed corn obtained per acre of land by DAS, is uncertain. Under this yield uncertainty, producing hybrid seed corn on a large plot of land may result in a surplus with a large up-front production cost if the realized yield is high; using a small plot of land 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 that DAS should use, but the historical yield data are not available for obtaining a statistical distribution. The unique industry-specific reasons for this lack of historical data are discussed next, but before discussing these reasons, we note an important characteristic of our focus. Our focus is on the production yield realized by DAS when it produces hybrid seeds, and this is the context in which the term yield will be used in the remainder of the paper.

6.1.1. Biological Context for Expert Judgments. DAS has a pool of approximately 125 types of parent or purebred seed corn; each type has a unique genetic structure, and these purebred varieties are used to produce hybrid varieties of seed corn. To replenish the stock of a specific parent seed, DAS plants this seed in a field. Self-pollination among the plants produces seeds of the same type, this is why the parent seed are purebred seeds. This inbreeding is carried out regularly to maintain inventories of parent seeds, and statistical distributions of the yields obtained during this inbreeding process are available from historical data. But these seeds are not sold to farmers; rather, the seeds sold to farmers are hybrid seeds that are obtained by cross-mating pairs of parent seeds. This cross-mating occurs when two different parent seeds are planted in the field. Corn plants have male and female reproductive parts. Plants growing from the parent seeds of one type; say, X, are treated chemically and physically (in a process called detasseling, e.g., see http://ntrdetassel.com/detasseling/) to make them act as female, and the plants growing from the parent seeds of the other type; say, Y, are made to act as male. The cross-pollination between these parents produces the hybrid seeds. DAS offers more than 200 varieties of hybrid seed corn every year in the market targeted to diverse soil and climate zones of the continental United States. Each variety is obtained from a different set of parents.

Due to the rapid pace of innovation in this industry, the average life of hybrid varieties is short. DAS produces and sells most hybrid varieties only three or four times before replacing them with new hybrids. Therefore, sufficient historical yield data necessary to obtain statistical distributions are not available for most hybrid seeds. In the absence of these data, DAS relies on a yield expert to estimate the yield distributions for producing the seeds. Before describing the process the expert uses, we note that a set of hybrid seeds has also been produced and sold repetitively. The historical yield data of these hybrids serve an important purpose in our estimation approach.

6.1.2. Expert's Mental Model. The yield expert at DAS uses a mental model for estimating the yield distribution for the production of a hybrid seed without historical data. This model is illustrated in Figure 1 for the hybrid seed obtained by crossing varieties X and Y. Female parent plants (type X) provide the body on which the hybrid seed grows; the male parent plants (type Y) 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 (left part of Figure 1) for the hybrid seed. The male parent affects this distribution during cross-pollination through its pollinating power and other genetic characteristics, leading to var-

Figure 1. (Color online) During Cross-Pollination, the Male Y Changes the Inbred Yield Distribution of X Shown on the Left



Notes. The expert's mental model involves judgments about changes in the location and/or spread of the distribution due to Y. Possible distributions after crossbreeding are shown in dotted lines on the right.

ious likely distributions as shown in dotted lines in the right part of the figure. This may include a shift in the median and/or changes in the spread of the distribution. The expert's contextual knowledge for the biology of both parents provides him with insights into how the distribution might change during cross-pollination.

6.1.3. Practice at DAS Before New Approach. In the past, the yield expert has adjusted the median of the inbreeding female distribution higher or lower to provide an estimate of the median yield for the production of hybrid seed. Thus the estimate of the median seed production yield has been based on indirect data and is judgmental in nature. This median yield was used for production planning decisions as follows. Managers would first calculate the area needed as area = demand/median yield judgment and then increase it by 20% or 30% to account for high profit margins. In our interactions, managers articulated the need for a rigorous approach to estimate the spread in the uncertain yields, which could then be used to determine the number of acres for each hybrid using optimization models. Furthermore, since the yield expert is required to provide judgments for almost 200 hybrid seeds within a span of two weeks every seed production season, it was necessary to develop an approach that could be implemented within this time window. Our analytical approach accomplishes these tasks.

In Section 6.2, we discuss how our approach continues to use the expert's judgments for the median (that he has estimated in the past) to exploit the mental model that he has developed and used over years as well as two additional quantiles selected by him. DAS makes the production planning decision once a year, typically during January–February. Our approach was first used in 2014, and has been in use since then. In the first step of our implementation, we determined the bias vector $\boldsymbol{\delta}$, the matrix Ω of judgmental errors, and the matrix $\mathbf{Z}$ corresponding to the quantiles for which the expert will provide judgments. This was done using historical yield data for a set of hybrid seeds that have been produced repetitively in the past. Details of this step are in Section 6.2. From these quantities, we obtained the optimal weights $\mathbf{w}_k^*$; $k = 1, 2$. Details of this determination are in Section 6.3. Finally, we quantified the benefit from using our approach using the data from the 2014 production planning decisions. This analysis is presented in Section 6.4. In Section 6.5, we discuss the integration of our approach into DAS's operational decision making.

6.2. Implementation: Data Collection at DAS and Calibration of Judgmental Errors

The first task during the implementation was to collect data from the firm to determine the appropriate distribution to use to model yield uncertainty, and to calibrate the expert. We first describe this data and the statistical tests performed to determine the parametric family of yield distributions. Then, we describe the process used for calibrating the expert at DAS.

Table 1. Tests to Accept/Reject Normality of Historical Yield Data for a Subset of Seeds

Test: H_0: Data are normal	p value for seed 1	p value for seed 2	p value for seed 3
Kolmogorov-Smirnov test	>0.15	>0.15	>0.15
Anderson-Darling test	0.90	0.51	0.51
Lilliefors-van Soest test	>0.20	>0.20	>0.20
Cramer-von Mises test	0.92	0.56	0.59
Ryan-Joiner test	>0.10	>0.10	>0.10

We asked DAS to identify a set of hybrid seeds that have been produced repetitively in the last few years. Overall, DAS found $L = 22$ such hybrid seeds indexed by $l = 1, 2, \ldots, 22$ and provided us with the historical yield data for these seeds. Using this data and other sources, we sought to determine the appropriate parametric family to model yield distribution. First, we analyzed the available yield data and ran a battery of tests, including the Kolmogorov-Smirnov test, Anderson-Darling test, and Lilliefors-van Soest test and found that they all failed to reject the hypothesis that the data was normally distributed. Table 1 shows the results for three such seeds. These tests only confirm that normality cannot be ruled out. We then ran a second test to see if normality provided the best fit with the data. In this test, we determined the parametric family with the best fit with the data using the chi-square test and Anderson-Darling test. The candidate distributions were the normal distribution, the gamma distribution, the uniform distribution, the log-normal distribution, the beta distribution, the Gumbel distribution, the exponential distribution, the Weibull distribution, the logistic distribution, and the inverse normal distribution. The normal distribution was the best fit on the Anderson-Darling test for all hybrid seeds. On the chi-square test, the normal distribution provided the best fit for a majority of the seeds.

In addition to this statistical proof for the hybrid seeds, DAS has extensive data for inbred seeds that supports normality. Since the biological factors at play during plant growth are the same in hybrid seeds, the yields of the hybrid seed corn also would be normal. Recently, Comhaire and Papier (2015) also provided statistical evidence for normality of yields during seed corn production. After identifying the normal distribution to be appropriate, we determined the quantiles that the yield expert at DAS was comfortable estimating (to obtain $\mathbf{Z}$ for the normal distribution), as well as determined his error structure ($\boldsymbol{\delta}$ and Ω), as described next.

6.2.1. Step 1: Selection of Quantiles for Elicitation and Determination of Z. For each of the hybrid seeds $l = 1, 2, \ldots, L$, we asked the expert to select three quantiles to estimate, without looking at the historical yield data of these hybrids. 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 an agricultural-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. 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 then 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 75th quantiles for several reasons. First, he has developed familiarity with these quantiles in the last few years: his statistical software (JMP) 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 "estimate one-tail quantile and calculate the other symmetric quantile using the properties of the normal distribution." This will be equivalent to estimating only one quantile instead of two.

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 tornados and floods (that we seek to exclude since the yield expert cannot predict these events) that lead to extreme outcomes.

We then determined the matrix **Z** for the 10th, 50th, and 75th quantiles. For the normal distribution, this matrix is calculated as

$$\mathbf{Z} = \begin{bmatrix} 1 & -1.28 \\ 1 & 0 \\ 1 & 0.67 \end{bmatrix},$$

where the value of -1.28 is equal to the inverse of the standard normal distribution at the probability 0.1 and so on.

6.2.2. Step 2: Elicitation Sequence and Consistency Check. For each distribution l, we obtained the three quantile judgments $\hat{x}_{il}(p_i)$; $i = 1, 2, 3$; $l = 1, 2, \ldots, 22$; $p_i = 0.1, 0.5, 0.75$ from the expert; the expert did not have access to the historical yield data for these hybrids during this estimation. We obtained the expert's 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 parent, looking at the properties of the male parent and providing his judgment for the median (see Figure 1). We then asked the expert to provide his quantile judgments for the 10th and 75th quantiles, in that order. 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 deductions about the third one, and then asked the expert to validate these deductions. If the expert did not concur with the deductions, we provided him an opportunity to fine tune the quantile judgments.

As an example, consider a specific seed for which the expert provided values of 15, 70, and 100 for the 10th, 50th, and 75th quantiles, respectively. The stated values of the 10th and 50th quantiles imply a mean yield of 70 and standard deviation of 42.92 for normally distributed yields. These two values imply that there is a 50% chance that the yield will be between 41 and 99 (the implied 25th and 75th quantile). We asked the yield expert the following question: "Your estimate of the 10th quantile implies that there is a 50% 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. Throughout this process, we emphasized that the objective of this fine tuning was to help him reflect on his estimates carefully without leading him to any specific set of numbers. Analysis showed that after this feedback, the standard deviations reduced by 33% for the tail quantiles in round 2, confirming that the feedback was indeed helpful to the expert in improving the quality of his estimates.

6.2.3. Step 3: Separation of Sampling Errors Using Bootstrapping. After elicitation was complete, we quantified the judgmental errors by comparing the expert's stated values for the quantiles with the values obtained from the historical data. For our analysis in Section 2, we assumed that the true values of the quantiles x_i were available. However, since the number of data points for each seed at DAS was limited (the largest sample size was 53), the quantile values obtained from the data were subject to sampling variations that needed to 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:

$$\hat{x}_i - \tilde{x}_i = (\hat{x}_i - x_i) + (x_i - \tilde{x}_i) \tag{11}$$

$$\text{Total Error} = \text{Judgmental Error} + \text{Sampling Error}. \tag{12}$$

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.

Writing (11) 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

$$\begin{gathered} E[\hat{\mathbf{x}} - \tilde{\mathbf{x}}] = E[(\hat{\mathbf{x}} - \mathbf{x})] + E[(\mathbf{x} - \tilde{\mathbf{x}})] \\ \boldsymbol{\delta}^t = \boldsymbol{\delta} + \boldsymbol{\delta}^s, \end{gathered} \tag{13}$$

where $\boldsymbol{\delta}^t$ is the *t*otal bias, and $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^s$ are the expert's judgmental bias and the *s*ampling 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

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

We can write this equation in matrix notation as

$$\Omega^t = \Omega + \Omega^s, \tag{14}$$

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 $\Omega = \Omega^t - \Omega^s$. 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 DAS'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 DAS'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 $\hat{e}^t_{il} = \hat{x}_{il}(p_i) - \tilde{x}_{il}(p_i)$. The average error $\hat{\delta}^t_i = \sum_{l=1}^{L} \hat{e}_{il}/L$ provided the total bias for each quantile. The vector of biases $\hat{\delta}^t_i$ constituted $\hat{\boldsymbol{\delta}}^t$. We then obtained unbiased errors as $\hat{e}^u_{il} = \hat{e}^t_{il} - \hat{\delta}^t_i$; using these, we estimated the 3×3 variance-covariance matrix $\hat{\Omega}^t$. As discussed earlier, 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 30%). The covariance matrix $\hat{\Omega}^t$ and the bias $\hat{\boldsymbol{\delta}}^t$ obtained after the second round are shown in Table 2.

The sampling bias $\boldsymbol{\delta}^s$ and the variance-covariance matrix Ω^s were estimated by bootstrapping as follows. We had data $y_{1l}, y_{2l}, \ldots, y_{n_l l}$ 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 $y_{1l}, y_{2l}, \ldots, y_{n_l l}$ and obtained the quantiles for *this* bootstrapping sample, $\tilde{x}_{ilp}$. We repeated the process for $p = 1, 2, \ldots, P$ 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 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. To ensure a stable variance-covariance matrix $\hat{\Omega}^{sl}$, we used a large value of P, $P = 1{,}000{,}000$. 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 DAS's data, the values of $\hat{\Omega}^s$ and the bias vector $\hat{\boldsymbol{\delta}}^s$ are shown in Table 2. The estimated judgmental bias $\hat{\boldsymbol{\delta}}$ was obtained as $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}^t - \hat{\boldsymbol{\delta}}^s$,

Table 2. Variance-Covariance Matrix and Biases After Bootstrap Adjustment

$$\hat{\Omega}^t = \begin{bmatrix} 113.41 & 50.09 & 46.83 \\ 50.09 & 42.92 & 51.46 \\ 46.82 & 51.46 & 93.37 \end{bmatrix} \qquad \hat{\Omega}^s = \begin{bmatrix} 34.42 & 20.71 & 13.50 \\ 20.71 & 21.00 & 21.16 \\ 13.49 & 21.16 & 25.20 \end{bmatrix} \qquad \hat{\Omega} = \hat{\Omega}^t - \hat{\Omega}^s = \begin{bmatrix} 78.99 & 29.38 & 33.33 \\ 29.38 & 21.92 & 30.30 \\ 33.33 & 30.30 & 68.17 \end{bmatrix}$$

$$\hat{\boldsymbol{\delta}}^t = \begin{bmatrix} 9.43 & 0.94 & -2.48 \end{bmatrix} \qquad \hat{\boldsymbol{\delta}}^s = \begin{bmatrix} -1.05 & 0.00 & 0.55 \end{bmatrix} \qquad \hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}^t - \hat{\boldsymbol{\delta}}^s = \begin{bmatrix} 10.48 & 0.94 & -3.03 \end{bmatrix}$$

and the estimated matrix of judgmental errors $\hat{\Omega}$ was obtained as $\hat{\Omega} = \hat{\Omega}^s$, and are shown in Table 2.

6.3. Implementation: Determination of Weights

For the variance-covariance matrix $\hat{\Omega}$ in Table 2 and the matrix

$$\mathbf{Z} = \begin{bmatrix} 1 & -1.28 \\ 1 & 0 \\ 1 & 0.67 \end{bmatrix},$$

Theorem 1 provides the weights $\mathbf{w}_1^* = [-0.18, 1.51, -0.33]$ for estimating the mean and $\mathbf{w}_2^* = [-0.58, 0.20, 0.38]$ for estimating the standard deviation to be used on the expert's judgments for the 10th, 50th, and 75th quantiles for yield distributions for hybrid seeds without historical data. For these results, the following regime was used at DAS in 2014 for estimating the production yield distributions of each of more than 100 hybrid varieties that did not have historical yield data. First, the expert estimated 10th, 50th, and 75th quantiles $\hat{\mathbf{x}}$ for the yield distribution of that hybrid seed. He provided judgments for these quantiles using the same mental model that he used during calibration, i.e., he looked at the historical statistical distribution of the production yield of the female parent on his computer, considered the pollinating power and other biological factors of the male parent, and then provided the quantile judgments for the hybrid.

From this information, the debiased estimates were obtained as $\hat{\mathbf{q}} = \hat{\mathbf{x}} - \hat{\boldsymbol{\delta}}$ by subtracting the biases $\hat{\delta}_1 = 10.48$, $\hat{\delta}_2 = 0.94$, $\hat{\delta}_3 = -3.03$. Next, the mean and standard deviation were obtained using the weights above on the debiased estimates, $\hat{\mu}_1 = \mathbf{w}_1^{*\mathrm{T}}\hat{\mathbf{q}}$ and $\hat{\mu}_2 = \mathbf{w}_2^{*\mathrm{T}}\hat{\mathbf{q}}$. Finally, these estimates were used in an optimization framework that an in-house team was developing in parallel to determine the optimal area of land to produce each hybrid. Since 2014, this approach has formed the basis of decisions worth \$800 million annually. Equally important, since the approach leveraged the expert's experience and intuition, which he had been using for a few years, the decision to implement the approach at DAS was reached quickly.

6.4. Estimation of Monetary Benefits Using Managerial Decisions

6.4.1. *Status Quo* Approach for Comparison. Before adopting our approach for estimating the mean and standard deviation for yield distributions, DAS used the following method to determine the area of land to use to grow each hybrid. The expert provided his estimate of the median $\hat{x}_2$. The production manager used this point estimate to determine the area to use as $Q^h = (D/\hat{x}_2)f$, where D was the demand of the seed, $\hat{x}_2$ was the median value provided by the expert and f was the risk adjustment factor of 1.2 or 1.3 based on a subjective high/low perceived uncertainty in yield. This framework provides a benchmark for quantifying the benefit of using our approach.

6.4.2. Measures for Quantifying Benefits from Our Approach Over *Status Quo*. The benchmark *status quo* approach affected the firm's finances systematically in three ways. First, the profit margins of the seed did not influence the acreage decision at all even though they clearly should affect the decision. Second, only two values of the factor f did not completely capture the complete range of yield standard deviations that were present in the portfolio. After our approach was implemented to estimate the mean and standard deviation, the firm used them as inputs to a stochastic optimization problem for expected profit maximization, i.e., the firm determine $Q^* = \arg\max_Q \{-cQ + pE[h(Q, D)]\}$, where h is the revenue function, c is the per acre cost, and p is the selling price per bag. This process change was a direct consequence of the availability of the standard deviation. One could then calculate the optimal ratio $f^* = Q^*\hat{x}_2/D$. At Dow, these ratios varied from 1 to 1.4, suggesting that the use of only 1.2 or 1.3 was not optimal. The dollar capital investment in a seed is equal to: Capital Investment = \$4,500 × *Area*, as the per acre cost of growing seed corn is approximately \$4,500 (the number is modified to preserve confidentiality). A reduction in the area used for growing hybrids directly translates into a reduction in initial capital investment, with the savings being equal to $\sum 4{,}500 \times (Q^h - Q^*)$, where the summation is over all 200 seeds. Over the complete portfolio, the cost savings were significant, as we discuss shortly. This reduction in the cost is the first measure for quantifying the benefit of our approach.

Third, as we documented in earlier sections, the yield expert's judgments for the median $\hat{x}_2$ has judgmental error. When using the *status quo* approach, this judgmental error leads to an error in the calculation of the Unadjusted Area = demand$/\hat{x}_2$. This error was further amplified by the use of the scaling factor $f > 1$ during the calculation of the adjusted area using $Q^h = (\text{demand}/\hat{x}_2)f$. For some hybrids, this error in the calculation of adjusted area can be very large and may result in a substantial suboptimal decision with a substantial loss in profit. This loss of profit is the second measure for quantifying the benefits of our approach. The benefit on the two measures was quantified using historical decisions made at DAS, as discussed below.

6.4.3. Analysis for Quantifying the Benefits. Due to confidentiality concerns, we do not provide here the specific numerical values for all 200 seeds, and instead, focus on the process used and the benefits observed. Our approach was first used in 2014 to make the production planning decision. For a number of seeds involved in this decision, we documented the area used for the annual crop plan in two ways: (a) *status quo* approach and (b) using yield distributions estimated using our approach. In approach (a), we determined the area used as $Q^h = (D/\hat{x}_2)f$ at $f = 1.2, 1.3$ for the median estimates $\hat{x}_2$ provided by the expert. In approach (b), we estimated the mean and standard deviation of the yield distribution from the expert's quantile judgments using our approach and then determined the optimal area using a profit-maximization formulation Q^* (discussed in Bansal and Nagarajan 2017), which needs the specification of yield distributions to determine the optimal area. The acreage decisions obtained from our approach were implemented at DAS along with a record of the decisions made using the *status quo* approach that would have been made in the absence of our approach.

The benefit of using our approach was estimated using the sale data available at the end of the season. Specifically, an in-house business analytics team compared the cost of using the area that our approach recommended with the cost of using the *status quo* approach. These results showed that the annual production investment decreased by 6%–7% using our approach. Equally important, DAS did not see a drop in the service levels of the seeds after the adoption of this new approach for estimating yield distributions.

Subsequently, an analysis was performed on the profit. For this analysis, the key item was that the demand, yield, revenue, and profit for each hybrid had been observed by the end of the year. For each hybrid, these quantities provided the revenue if the area in the *status quo* approach had been used. From this revenue, the cost was subtracted to obtain the profit. Comparing the profit from this *status quo* approach with actual profit suggested that our approach led to between 2% and 3% improvement in profit. These documented benefits have led to a continuous use of our approach for estimating yield distributions at the firm. We next discuss how this approach has been integrated into DAS's operations, but first, we discuss some nonmonetary benefits accrued.

6.4.4. Nonmonetary Benefits. Several features of our approach were perceived to be of managerial importance during the implementation. First, it provided a unique quantification of the quality of the expert's judgments. This quantification was important for the firm in understanding the benefit of identifying and training experts in other seed businesses (soybean, cotton, etc.) for which new varieties are being developed. Specifically, at DAS, the yield distributions have a variance of $\mu_2^2 \approx 400$ on average. At $\hat{\Omega}$ shown in Table 2, the variance $\mathbf{w}_1^{*\mathrm{T}}\Omega\mathbf{w}_1^* = 18$. Using Proposition 3, it follows that our approach extracts information from the expert's quantile judgments that is equivalent to the information provided by $400/\mathrm{Var}(\hat{\mu}_1) = 400/18 \approx 22$ data points. We were told that this is equivalent to approximately five to six years of test data at DAS. Second, the approach provides a rational effort to estimate the variability in production yields, enabling the yield expert to support his estimates for yield distributions with scientific tools.

6.5. Integration into Firm's Operations

After the initial implementation in 2014, DAS recognized the value of formal statistical modeling and analysis for yield forecasting and production planning decisions. The firm created a new business analytics group, and two members of this group were tasked with developing optimization protocols to inform DAS's operations. The team was composed of trained statisticians with experience in biostatistics. This niche skill set was considered necessary since the yield distributions and other properties of seeds are driven by biology, and an understanding of plant biology as well as statistics would enable the team to develop context-informed models.

For the annual production planning decision, the team implements the approach in the following manner. The production planning decision is made every year a few weeks before the advent of spring. In the weeks preceding this decision, the team obtains a list of hybrid seeds from the seed business manager that are under consideration for being offered to the market. The portfolio of hybrid seeds offered changes annually and this information is necessary for the team to estimate yield distributions to support the production planning decision. The team then sends this list to the yield expert who is located at a different geographical location. This expert does travel back and forth to the team's location, nevertheless, DAS has emphasized the development of computer-based tools that can be accessed from anywhere. The yield expert obtains this list and provides his judgments for yield distributions. The team of statisticians processes these quantile judgments using the process described earlier to deduce means and standard deviations. A list of these values is then sent back to the business analytics team that is responsible for making the production planning decision.

The business analytics team then uses an optimization framework to determine the number of acres that should be used to grow each hybrid seed. Yield distributions constitute the major source of stochasticity in this model. Under this uncertainty, the model seeks to balance the trade-off between using a very large or a very small area. The per acre tilling and land lease cost is high and using a large area of land needs up-front high investment and could lead to a surplus inventory of hybrid seeds. The use of a small area of land requires less up-front capital investment in the production, but can lead to shortages. Estimating the yield distributions enables the firm to optimize this trade-off in a mathematical fashion, in addition to providing a quantitative decision support.

7. Discussion and Future Research

7.1. Summary of Approach

In changing environments, historical data do not exist to provide probability distributions of various uncertainties. In such environments, judgments are sought from experts. But expert judgments are prone to judgmental errors. In this paper, we develop an analytical approach for deducing the parameters of 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-building perspective, the optimization approach proposed is consistent with moment matching, has a unique analytically tractable solution, and is amenable for comparative static analysis. The approach also provides an analytical foundation for results documented numerically in the prior literature. From a practice perspective, 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. The approach also establishes 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 prior literature that the weights for the mean and the standard deviation add up to 1 and 0, respectively. Equally important, it provides for a linear pooling of quantile judgments from multiple experts, thereby providing a practical toolkit for combining judgments in practice.

From an implementation perspective, the approach has several features that make it viable for an easy adoption by firms. First, it uses judgments for any three or more quantiles that an expert is comfortable providing. In a specific application at DAS, we used the yield expert's judgments for the 10th, 50th, and 75th quantiles to deduce the mean and standard deviations of a large number of yield uncertainties. The expert chose to estimate these quantiles based on his experience with obtaining and using these quantiles in his data analysis responsibilities. Second, the final outcome of the approach is a set of weights that are used to estimate means 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 DAS. 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 parameters (and the decision based on this estimated distribution) will vary as the quality of the expert's judgmental errors improve or deteriorate. More specifically, in using Theorem 1, one can analytically determine how the weights $\mathbf{w}$ change when the variance-covariance matrix Ω changes.

7.2. Other Potential Approaches

In this section, we discuss three other potential approaches to obtain mean and standard deviation from quantile judgments: parameter estimation through entropy minimization, by minimizing sum of absolute errors, and by nonparametric approaches.

In relative entropy methods, the entropy of the distribution obtained from iid randomly sampled data relative to a benchmark distribution is computed to evaluate the similarity of two distributions. In our problem, only three imperfect quantile judgments are available from the expert. Therefore the conventional theory available for comparing distributions with iid randomly sampled data using entropy-based measures is not directly applicable. Motivated by the weighted linear approach suggested by moment matching (in Proposition 5), one possibility is to estimate moments from quantile judgments as $\hat{\mu}_j = \sum_{i=1}^m w_{ji}\hat{q}_i$; $j = 1, 2$, where the quantile judgments $\hat{q}_i$ correspond to probabilities p_i. For the normal distribution, the cross-entropy or the Kullback-Leibler (KL) distance of the estimates $\hat{\mu}_j$ from true values μ_j is given as (Duchi 2007)

$$KL = \log\frac{\hat{\mu}_2}{\mu_2} + \frac{\mu_2^2 + (\mu_1 - \hat{\mu}_1)^2}{2\hat{\mu}_2^2} - \frac{1}{2}.$$

For each debiased quantile judgment, $\hat{q}_i = \mu_1 + z_i\mu_2 + \epsilon_i$, where the term ϵ_i is the noise in the judgment, and therefore $E[\epsilon_i] = 0$; then it follows that

$$\mu_1 = E\left[\sum_{i=1}^m \hat{q}_i w_{1i}\right] = \mu_1\sum_{i=1}^m w_{1i} + \mu_2\sum_{i=1}^m z_i w_{1i} + E\left[\sum_{i=1}^m \epsilon_i w_{1i}\right],$$

and since $E[\sum_{i=1}^m \epsilon_i w_{1i}] = 0$, this implies that (i) $\sum_{i=1}^m w_{1i} = 1$ and (ii) $\sum_{i=1}^m z_i w_{1i} = 0$. Similarly, since $\mu_2 = E[\sum_{i=1}^m \hat{q}_i w_{2i}]$, it follows that (iii) $\sum_{i=1}^m w_{2i} = 0$ and (iv) $\sum_{i=1}^m z_i w_{1i} = 1$.

Using the properties (i)–(iv), the KL distance can be expressed as

$$KL = \log\frac{\mu_2 + \sum_{i=1}^m w_{2i}\epsilon_i}{\mu_2} + \frac{\mu_2^2 + (\sum_{i=1}^m w_{1i}\epsilon_i)^2}{2(\mu_2 + \sum_{i=1}^m w_{2i}\epsilon_i)^2} - \frac{1}{2}. \quad (15)$$

This KL distance is a random variable, which is a function of the estimation errors ϵ_i, thus a plausible approach would be to select the weights w_{ji} that minimize the expected value of the KL distance, $E[KL]$. The limiting behavior of $E[KL]$ provides a point of comparison between this approach and the one developed earlier in this paper. As the expert becomes increasingly more reliable, we have on the limit $E[KL] \to 0$ as $\mathrm{Var}(\epsilon_i) \to 0$, for any values of w_{1i} and w_{2i} that satisfy conditions (i)–(iv). Since the value of KL is nonnegative by construction, on the limit any such (w_{1i}, w_{2i}) minimize $E[KL]$. Moreover, since in the optimization, we can select $2m$ weights and we have only four constraints, anytime we elicit more than two quantiles, in general, we may have an infinite number of optimal weight combinations. The unique weights obtained by our approach automatically satisfy conditions (i)–(iv), hence they also optimize $E[KL]$ on the limit.

With respect to the general case of this approach (minimizing (15)), we make three observations:

1. Notice from Equation (15), that $E[KL]$ is a nontrivial function of the entire error covariance matrix Ω, and obtaining the $E[KL]$-minimizing weights will require numerical optimization.
2. The above definition of $E[KL]$ requires knowledge of μ_2, which we do not have.
3. The uniqueness of the weights is not guaranteed.

Comparing this estimation approach with the one proposed and implemented at Dow, we can appreciate an important difference. Both approaches would require us to estimate the covariance matrix Ω from the calibration data set. But the $E[KL]$ minimization approach also requires knowledge of the parameter μ_2, which Dow did not have. The estimation approach developed in Sections 3–5 does not require this knowledge. These challenges associated with the $E[KL]$ minimization approach will need to be addressed by future research before the approach can be used in practice.

The problem of estimating distribution parameters by minimizing the sum of absolute errors (instead of the sum of squared errors) is stated as $\min_{w_{ik}} \sum_j |\sum_i w_{ik} \cdot \hat{q}_{ji} - \hat{\mu}_{jk}|$, where $\hat{\mu}_{jk}$ is the mean ($k = 1$) and standard deviation ($k = 2$) of the calibration distribution j. The optimal weights for this model are not obtainable in closed form, rather this problem must be solved numerically using a linear programming formulation, and it not always has a unique solution (Harter 1977, Bassett and Koenker 1978, Chen et al. 2008). Furthermore, there is no direct relationship between the sum of squared errors and sum of absolute errors for the data. Due to these two issues, the equivalent sample size for an expert, akin to the result in Proposition 3, cannot be determined.

Nonparametric methods explore various functional forms to fit data, while minimizing the squared distances between the fitted and true values. The Spline fitting approach fits one or more splines of various degrees to the data. The recommended functional form for the predictive model tends to be sensitive to the data (Härdle et al. 2012) and, in our context, may change with the inclusion/exclusion of even one probability distribution in the calibration set. Similarly, the additive kernel model can be sensitive to tuning parameters, which need to be selected subjectively (Härdle et al. 2012). This sensitivity and subjectivity in model recommendation implies, in our context, that the nonparametric model for new seeds may have to be modified for every season, which could be undesirable when a firm seeks to develop a stable and transparent model for a repetitive use. Finally, a direct least squares analysis provides a strong basis for using the linear functional form used in the paper. Proposition 5 shows that the conventional least squares formulation to deduce mean and standard deviation from quantile judgments for location-scale distribution results in the estimation of mean and standard deviation as linear combinations of the quantile judgments. Our approach exploits this result and develops it further in the form of tractable and ease to use results discussed in various propositions.

7.3. Future Research

A scant but important stream of literature has quantified the benefit of a more reliable estimation of operational uncertainties. Akcay et al. (2011), in collaboration with SmartOps Corporation, show that using the demand information computed from 20 data points over 10 data points for inventory decision making reduces the operating cost by 10% (Tables 2—4, p. 307). Our quantification provides a new addition to this literature, especially when the information for an uncertainty is obtained from an expert. In the future, this quantification should be sharpened using Monte Carlo simulation studies for the seed industry as well as other industries. Future research should also explore tighter connections between the yield of hybrid seed production and the genomes of both parents crossed. This industry is making significant investments in genetic research, and a large amount of genomic information for some corn varieties is becoming available. Unfortunately, currently, this task is daunting as corn has one of the most complex plant genomes with some mapped varieties showing sequences of more than two billion genes (Dolgin 2009); this is in stark contrast with the sparsity of the yield data available. Finally, an important requirement for the approach developed for a use, in practice, is that we calibrate the experts by comparing their quantile judgments with the true values for some distributions that are specific to the context, and for which historical data is available at the firm. However, this data may not be available in all businesses. The future research should explore whether it is possible to calibrate experts on almanac events, and then use this information for estimating probability distributions specific to the business.

Acknowledgments

The authors gratefully acknowledge the suggestions made by three anonymous reviewers, associate editor, and area editor Andres Weintraub, which resulted in a much improved paper. The authors thank Dow AgroSciences, especially Sue Gentry and J. D. Williams, for their support in this collaboration. The first version of the paper was developed when the first author was visiting the Department of Supply Chain and Operations at University of Minnesota. The Laboratory for Economics, Management and Auctions (LEMA) at Penn State provided laboratory settings to test the theory developed in the paper before its field deployment. The authors also thank Mike Blanco, Marilyn Blanco, Murali Haran, and Dennis Lin at Penn State for their help during a revision.

References

Akcay A, Biller B, Tayur S (2011) Improved inventory targets in the presence of limited historical demand data. *Manufacturing Service Oper. Management* 13(3):297–309.

Ayvaci MUS, Ahsen ME, Raghunathan S, Gharibi Z (2017) Timing the use of breast cancer risk information in biopsy decision making. *Production Oper. Management*. Forthcoming.

Baker E, Solak S (2014) Management of energy technology for sustainability: How to fund energy technology research and development. *Production Oper. Management* 23(3):348–365.

Bansal S, Nagarajan M (2017) Product portfolio management with production flexibility in agribusiness. *Oper. Res.* 65(4):914–930.

Bassett G Jr, Koenker R (1978) Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* 73(363):618–622.

Bates JM, Granger CWJ (1969) The combination of forecasts. *Oper. Res. Quart.* 451–468.

Casella G, Berger RL (2002) *Statistical Inference*, 2n ed. (Duxbury Press, Pacific Grove, CA).

Chen K, Ying Z, Zhang H, Zhao L (2008) Analysis of least absolute deviation. *Biometrika* 95(1):107–122.

Comhaire P, Papier F (2015) Syngenta uses a cover optimizer to determine production volumes for its European seed supply chain. *Interfaces* 45(6):501–513.

Dolgin E (2009) Maize genome mapped. *Nature News* 1098.

Duchi J (2007) *Derivations for Linear Algebra and Optimization*. Working paper, University of California, Berkeley, Berkeley, CA.

Garthwaite PH, Dickey JM (1985) Double- and single-bisection methods for subjective probability assessment in a location-scale family. *J. Econometrics* 29(1–2):149–163.

Granger CWJ (1980) *Forecasting in Business and Economics* (Academic Press).

Härdle WK, Müller M, Sperlich S, Werwatz A (2012) *Nonparametric and Semiparametric Models* (Springer , New York).

Harter HL (1977) Nonuniqueness of least absolute values regression. *Comm. Statist.-Theory and Methods* 6(9):829–838.

Johnson D (1998) The robustness of mean and variance approximations in risk analysis. *J. Oper. Res. Soc.* 49(3):253–262.

Johnson NL, Kotz S, Balakrishnan N (1994) *Continuous Univariate Distributions*, Vol. 1, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (Wiley, New York).

Keefer DL, Bodily SE (1983) Three-point approximations for continuous random variables. *Management Sci.* 29(5):595–609.

Kelton WD, Law AM (2006) *Simulation Modeling and Analysis*, 4th ed. (McGraw Hill, New York).

Koehler DJ, Brenner L, Griffin D (2002) The calibration of expert judgment: Heuristics and biases beyond the laboratory. *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge University Press, New York).

Lau HS, Lau AHL (1998) An improved PERT-type formula for standard deviation. *IIE Trans.* 30(3):273–275.

Lau HS, Lau AHL, Ho CJ (1998) Improved moment-estimation formulas using more than three subjective fractiles. *Management Sci.* 44(3):346–351.

Lau HS, Lau AHL, Kottas JF (1999) Using Tocher's curve to convert subjective quantile-estimates into a probability distribution function. *IIE Trans.* 31(3):245–254.

Lau AHL, Lau HS, Zhang Y (1996) A simple and logical alternative for making PERT time estimates. *IIE Trans.* 28(3):183–192.

Lindley DV (1987) Using expert advice on a skew judgmental distribution. *Oper. Res.* 35(5):716–721.

O'Hagan A (1998) Eliciting expert beliefs in substantial practical applications. *J. Roy. Statist. Soc.: Ser. D* (*The Statistician*) 47(1): 21–35.

O'Hagan A (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*, Vol. 35 (John Wiley & Sons, Chichester, UK).

O'Hagan A, Oakley JE (2004) Probability is perfect, but we can't elicit it perfectly. *Reliability Engrg. System Safety* 85(1–3):239–248.

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

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

Ravinder HV, Kleinmuntz DN, Dyer JS (1988) The reliability of subjective probabilities obtained through decomposition. *Management Sci.* 34(2):186–199.

Stevens JW, O'Hagan A (2002) Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. *Internat. J. Tech. Assessment in Health Care* 18(04):782–790.

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

Saurabh Bansal is an assistant professor of supply chain management and information systems, and a faculty member of operations research at the Pennsylvania State University. His research focuses on developing mathematical models, algorithms, and protocols to estimate business risks and optimize business operations under risks.

Genaro J. Gutierrez is an associate professor of information risk and operations management at McCombs School of Business, The University of Texas at Austin, where he teaches operations management and supply chain analytics. He is the Director of the Executive MBA Program that McCombs School offers in Mexico City.His current research interests include, in general, the incorporation of data analytics in the supply chain management domain. Specific research projects include: combination of statistical and judgmental approaches for estimating demand, data-driven models to optimize the supply chain for digital advertising, procurement of traded commodities, and reliability models for forecasting and procurement of high cost spare parts. Recent publications of Professor Gutierrez have appeared in *Management Science, Operations Research, IIE Transactions*, and *European Journal of Operations Research*.

John R. Keiser is the global technical expert for corn seed production research, and is responsible for providing guidance and coordination between all corn production research programs globally, as well as technical oversight for the NA Production Research program. He earned a PhD in Crop Production and Physiology from Iowa State University.