Management Science

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

Assessing Uncertainty from Point Forecasts

Anil Gaba, Dana G. Popescu, Zhi Chen

To cite this article:
Anil Gaba, Dana G. Popescu, Zhi Chen (2019) Assessing Uncertainty from Point Forecasts. Management Science 65(1):90-106.
https://doi.org/10.1287/mnsc.2017.2936

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

MANAGEMENT SCIENCE
Vol. 65, No. 1, January 2019, pp. 90–106
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Assessing Uncertainty from Point Forecasts

Anil Gaba,[a] Dana G. Popescu,[b] Zhi Chen[a]

[a] Decision Sciences, INSEAD, Singapore 138676; [b] Darden School of Business, University of Virginia, Charlottesville, Virginia 22903

Contact: anil.gaba@insead.edu, http://orcid.org/0000-0002-8748-6300 (AG); popescud@darden.virginia.edu, http://orcid.org/0000-0001-6758-1872 (DGP); zhi.chen@insead.edu, http://orcid.org/0000-0001-6663-9550 (ZC)

Received: July 11, 2016
Revised: May 3, 2017; August 4, 2017
Accepted: August 17, 2017
Published Online in Articles in Advance: April 9, 2018

https://doi.org/10.1287/mnsc.2017.2936

Abstract. The paper develops a model for combining point forecasts into a predictive distribution for a variable of interest. Our approach allows for point forecasts to be correlated and admits uncertainty on the distribution parameters given the forecasts. Further, it provides an easy way to compute an augmentation factor needed to equate the dispersion of the point forecasts to that of the predictive distribution, which depends on the correlation between the point forecasts and on the number of forecasts. We show that ignoring dependence or parameter uncertainty can lead to assuming an unrealistically narrow predictive distribution. We further illustrate the implications in a newsvendor context, where our model leads to an order quantity that has higher variance but is biased in the less costly direction, and generates an increase in expected profit relative to other methods.

History: Accepted by Vishal Gaur, operations management.
Funding: The Centre for Decision Making and Risk Analysis at INSEAD provided financial support for this project.
Supplemental Material: The e-companion is available at https://doi.org/10.1287/mnsc.2017.2936.

Keywords: correlated experts • point forecasts • demand forecasting • newsvendor model

1. Introduction

Predicting an unrealized future variable is a ubiquitous endeavor, for it is crucial in shaping many decisions that we must make. For example, an investor or a financial institution might predict a future currency exchange rate for its hedging or trading strategy, or a retailer might predict the demand for a new product to determine the order quantity. Despite the presence of extensive data and models, subjective human judgment remains a key element in predictions across numerous real-life settings (Seifert et al. 2015).

Subjective forecasts for a variable often come in the form of point forecasts—i.e., single-valued predictions for that variable. Assessment of complete probability distributions, or summary measures (such as the expected value and variance), is a difficult cognitive task for even the most well-trained in probabilities, and is practically impossible for people with no such training. In part, this is also due to the fact that complete probability distributions or summary measures are often not easily observable, if at all. In one of the well-known industry practices in operations management, experts (such as sales people, designers, product managers, and the like) provide point forecasts for the demand of upcoming new products. These forecasts then become a key input in inventory decisions for new products (Fisher and Raman 1996). This, however, involves an additional step by the decision maker, to convert the point forecasts for a variable into an estimate of uncertainty (such as a predictive distribution) for that variable, which is the broad focus of this paper.

Consider a decision maker who judges that the demand for a new product is likely to follow a normal distribution, but has little or no information on the parameters of the distribution. The decision maker then obtains point forecasts from k experts. Intuitively, the mean and dispersion of the k point forecasts are indicative of the mean and variance of the demand distribution. Fisher and Raman (1996), for example, use the mean of k forecasts as an estimate of the expected (mean) demand and augment the observed standard deviation of the k forecasts by a factor of 1.75 to create an estimate of the standard deviation of demand. The augmentation factor of 1.75 is justified by a calibration exercise with past data. In a similar setting in one of the widely taught cases in operations management, *Sport Obermeyer* (Hammond and Raman 1994), an augmentation factor of 2 is suggested. Gaur et al. (2007), using historical data on three different data sets, test the hypothesis that the variance of demand is positively correlated with the dispersion of the point forecasts. However, several questions remain. First, it is not conceptually clear how the augmentation factor arises and what it depends on. Second, any past data used to estimate the augmentation factor might not be applicable for a new product, such as a new fashion item or short-life-cycle product. Third, while the point forecasts provide relevant information on the distribution parameters, those are unlikely to completely

eliminate uncertainty on the parameters. In this sense, the heuristics mentioned above are only *certainty equivalence* approaches (Poirier 1995), as if the parameters of the demand distribution are certain once their point estimates are obtained from the point forecasts. This is of course not the case. These are some of the specific questions we attempt to answer in this paper.

In answering these questions, we consider two salient issues in combining forecasts. One, we allow forecasts to be correlated. Two, our approach admits parameter uncertainty given the forecasts. We then show that the augmentation factor mentioned above is contingent on the dependence among the forecasts and on the parameter uncertainty given the forecasts. Our approach is easily tractable, and hence practical. We discuss this in finer detail next.

Despite all efforts to create a group of independent experts, some form of stochastic dependence among their forecasts is inevitable. For example, Winkler (1981) notes pairwise correlations of sportswriters' prediction errors in the range of 0.84–0.97, suggesting that experts might have similar training, experience, and access to the same data, or might use similar models for their predictions. Ashton (1986) observes an average correlation of 0.6 between sales forecasts by managers. Fisher and Raman (2010, p. 64) describe a forecasting deliberation process where "one of the buyers was more articulate and assertive than the others. Often, she swayed her colleagues, so the final decisions represented her preferences rather than the collective wisdom." It is intuitive that if the point forecasts are highly correlated, then their dispersion would tend to underestimate the uncertainty about the variable of interest, perhaps substantially so. A greater degree of dependence among experts should hence entail a larger augmentation factor.

Further, number of experts must have a bearing on the parameter uncertainty. For example, five forecasts instead of 20 implies greater uncertainty on the parameters estimated. Hence, a larger number of forecasts should suggest a smaller augmentation factor. A Bayesian approach, in contrast to a certainty equivalence approach, accounts for such parameter uncertainty.

Bayesian models for combining correlated point forecasts under normality have been developed, for example, in Geisser (1965), Winkler (1981), and Clemen and Winkler (1985), where the latter also illustrates loss of information due to dependence among the information sources. A necessary input in these models is a covariance matrix consisting of all pairwise correlations between the experts (measures of dependence among the experts) and the variances of each point forecast (measures of accuracy of the experts). Often, there might be little or no past data on the covariance matrix. Further, there is frequently only one k-variate observation in the sample, such as k experts providing one forecast each for a new product, which then contains no information content on either the pairwise correlations or the variances of the experts' forecasts. This means, with k experts, $k(k+1)/2$ prior parameters need to be assessed on which the sample contains no information, clearly a daunting task. Hence, practical applications of such models have been limited. Moreover, several papers have raised concerns around robustness of such models (Bunn 1985, Winkler and Clemen 1992, Chhibber and Apostolakis 1993). Clemen (1989), in a comprehensive review of the literature on combining forecasts, points out that such a model in real-life situations has had somewhat mixed results, along with a central finding that simple averaging of point forecasts often outperforms more complex methods and is an easy and fairly robust way in terms of accuracy to predict an unrealized variable (see also a review by Armstrong 2001). Simple averaging is consistent with an equal-weights model for experts—i.e., with equal variances and pairwise correlations. Schmittlein et al. (1990), using simulations, also recommend such a model unless there is plentiful information to distinguish between experts. Bolger and Rowe (2015, p. 5), in a perspective on aggregation of expert judgment, argue that "no significant benefits are likely to accrue from unequal weighting in mathematical aggregation," attributing it primarily to the difficulty, if not the impossibility, of measuring individual differences for unequal weights.

Building on these ideas, we develop a Bayesian approach with *exchangeable* experts (i.e., all experts have equal variances and pairwise correlations). This also reduces the number of prior parameters on which the sample information contains no information to just one (the common correlation). In practice, a way to think about this is as if a decision maker has no discerning information on experts but at the same time feels that experts will be correlated on the basis of some shared information. In this respect, the Bayesian models under normality mentioned earlier have been developed with the covariance matrix either known or unknown, whereas we assume that the covariance matrix is partially known (known correlation but unknown variance). This greatly reduces the complexity of a model with correlated forecasts, without having to assume that the variance is also known and without the serious omission of dependence in the simpler models of independent forecasts. Hence, our approach is practically more appealing.

In Section 2, we develop a normal model, and show that the augmentation factor needed to equate the standard deviation of the predictive distribution to the observed dispersion of the point forecasts depends on the correlation between the forecasts and on the number of forecasts. Given a number of forecasts,

a higher correlation leads to a higher augmentation factor, implying greater predictive uncertainty. On the other hand, given a correlation, a higher number of forecasts leads to lower predictive uncertainty resulting from lower parameter uncertainty, but only up to a limit. In other words, dependence leads to a loss of information that cannot be compensated for by simply increasing the number of forecasts. We compare our approach with other methods used, for example, in the operations management literature, and highlight the underestimation of uncertainty due to ignoring dependence or parameter uncertainty or both.

The assumption of known common correlation is quite strong. In Section 3, we show that if there is uncertainty on the correlation or heterogeneity in the pairwise correlations, then treating the mean correlation as a known common correlation is still a reasonable approach. Further, we extend our approach to the case where the variable of interest has a lognormal distribution, and show that the impact of ignoring dependence or parameter uncertainty only gets exacerbated. This is followed by a discussion on the assessment of correlation. An added advantage of our approach is that it can be used even in the absence of relevant past data, by subjectively assessing the correlation. The existing heuristics in the operations literature do not provide any basis for generating a predictive distribution without relevant past data, while acknowledging that availability of such data is the exception rather than the rule. For example, Fisher and Raman (1996, p. 93) state, "estimating demand densities for fashion products is challenging because no history of previous demand is available."

In Section 4, we illustrate our analysis in a decision-making context of a newsvendor, and show the impact of our approach on the order quantity and expected profit. Our model in comparison with other methods leads to an order quantity that is biased in the less costly direction with a higher variance. At the same time, we show with simulations that our model leads to an increase in expected profits that can at times exceed 20%. Section 5 follows with a summary and discussion.

2. A Model for a Probability Distribution from Point Forecasts

Let a random variable of interest to a decision maker be $\tilde{y}$, which for example could be a future observation of demand for a new product. Suppose the decision maker models the probability distribution of $\tilde{y}$ conditional on a vector of parameters θ with density function $f(\tilde{y} \mid \theta)$, with θ unknown. For example, $f(\tilde{y} \mid \theta)$ might be modeled as a normal density function conditional on $\theta = (\mu, \sigma^2)$, where μ and σ^2 are unknown mean and variance, respectively. Further, let a probability density function $h(\theta)$ reflect the decision maker's prior uncertainty on θ. Now suppose that the decision maker obtains some additional information before realization of $\tilde{y}$. We consider such information in the form of k point forecasts $\mathbf{x} = (x_1, x_2, \ldots, x_k)'$ of $\tilde{y}$ from k different sources. Then, given $\mathbf{x}$, the predictive distribution for $\tilde{y}$ is given by

$$f(\tilde{y} \mid \mathbf{x}) = \int_{\Theta} f(\tilde{y} \mid \theta) h(\theta \mid \mathbf{x}) \, d\theta, \tag{1}$$

where $h(\theta \mid \mathbf{x}) \propto h(\theta) l(\mathbf{x} \mid \theta)$ is the posterior distribution of θ given $\mathbf{x}$, and $l(\mathbf{x} \mid \theta)$ is the likelihood function for $\mathbf{x}$ given θ. In this setup, $\tilde{y}$ and $\mathbf{x}$ are conditionally independent given θ, and any dependence among the point forecasts $x_1, \ldots, x_k$ is included in $l(\mathbf{x} \mid \theta)$. If the decision maker does not have any prior information on θ, then a diffuse (flat) prior on θ can be used. On the other hand, if the decision maker does possess some prior information (such as relevant experience or past data), then that information should be included in $h(\theta)$. The predictive distribution for $\tilde{y}$ in (1) accounts for two types of uncertainty, the uncertainty of $\tilde{y}$ given θ and the uncertainty about the parameter θ given $\mathbf{x}$. This is a typical Bayesian approach for aggregating expert opinions. We extend this approach below under normality.

2.1. A Normal Model

Suppose that $\tilde{y} \mid (\mu, \sigma^2) \sim \mathrm{N}(\mu, \sigma^2)$, with μ and σ^2 unknown. The k experts provide point forecasts $\mathbf{x} = (x_1, \ldots, x_k)'$ for $\tilde{y}$. We assume that $\mathbf{x}$ follows a multivariate normal with a mean vector $\boldsymbol{\mu} = \mu \mathbf{e}$, where $\mathbf{e} = (1, \ldots, 1)'$, and a $k \times k$ positive definite covariance matrix $\boldsymbol{\Sigma}$ with diagonal elements σ^2 and off-diagonal elements $\rho\sigma^2$.

This implies that the experts are exchangeable, which is consistent with treating all experts equally. Our model also implies that the data-generating process for an expert's judgment is the same as that for $\tilde{y} \mid (\mu, \sigma^2)$, as if each expert's forecast is a draw from the distribution of $\tilde{y} \mid (\mu, \sigma^2)$. Such an approach is common with calibrated experts (see, e.g., Gneiting et al. 2007, Winkler 1981) and seems reasonable especially in demand forecasting for new products with no or little relevant past data (Fisher and Raman 1996, p. 93).[1]

The dependence among experts is reflected in the correlation of their predictions—i.e., for $i, j \in \{1, \ldots, k\}, i \neq j$, $\mathrm{Corr}[x_i, x_j \mid \mu, \sigma^2] = \rho_{ij} = \rho$. This is again consistent with an equal-weights model for experts. We further assume ρ to be known. This simplifying assumption greatly reduces the complexity of the model for the decision maker. However, later in Section 3.1, we do verify the robustness of our model with respect to uncertainty on ρ and also heterogeneity in pairwise correlations. So, an alternate view is to think of the known common correlation ρ as $E[\rho]$ or the average pairwise correlation between the experts. Hence, our model is somewhat more flexible than it might

appear, in the sense that it can also be used with uncertain ρ or with different known pairwise correlations of experts' forecasts.

The likelihood function for the k-variate forecast $\mathbf{x}$ is given by

$$l(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \quad (2)$$

Setting $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Sigma}_\rho$, where $\boldsymbol{\Sigma}_\rho$ is a $k\times k$ is a matrix with diagonal elements 1 and off-diagonal elements ρ, and precision $\lambda = 1/\sigma^2$, the likelihood function can be rewritten as

$$l(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = l(\mathbf{x}\,|\,\boldsymbol{\mu},\lambda) \propto \lambda^{k/2}\exp\left(-\frac{\lambda}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}_\rho^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \quad (3)$$

Proposition 1 further simplifies the likelihood function in (3).

Proposition 1. *The likelihood function in (3) can be represented as*

$$l(\mathbf{x}\,|\,\mu,\lambda) \propto \lambda^{k/2}\exp\left(-\frac{\lambda k^*}{2}(\mu-\bar{x})^2\right)\exp\left(-\frac{\lambda}{2}(k-1)s^{*2}\right), \quad (4)$$

with sufficient statistic $\{\bar{x}, k^*, s^{*2}, k\}$, *where* $\bar{x} = (1/k)\cdot\sum_{i=1}^k x_i$, $s^{*2} = s^2/(1-\rho)$ *with* $s^2 = \sum_{i=1}^k(x_i-\bar{x})^2/(k-1)$, *and* $k^* = k/(1+(k-1)\rho)$.

The likelihood function in Proposition 1 is now of the same form as with independently and identically distributed (i.i.d.) observations in a normal process with unknown mean and variance, which is analytically convenient. And, as discussed in Proposition 2, it is intuitively appealing to summarize the k point forecasts as a four-dimensional sufficient statistic $\{\bar{x}, k^*, s^{*2}, k\}$ rather than as $\{\bar{x}, s^2, k\}$ in the i.i.d. case.

We model the decision maker's prior information on μ and λ with a normal-gamma (NG) distribution that is a natural conjugate prior distribution for a normal process with a likelihood function of the form in (4) (see, e.g., Bernardo and Smith 1994, Hoff 2009):

$$\begin{aligned} f(\mu,\lambda) &= \mathrm{NG}(\mu,\lambda\,|\,\mu_0,n_\mu,v_0,n_v) = f(\mu\,|\,\lambda)f(\lambda)\\ &= \mathrm{N}(\mu\,|\,\mu_0,(n_\mu\lambda)^{-1})\mathrm{Gamma}\left(\lambda\,\middle|\,\frac{n_v}{2},\frac{n_v v_0}{2}\right)\\ &\propto \lambda^{1/2}\exp\left(-\frac{n_\mu\lambda(\mu-\mu_0)^2}{2}\right)\\ &\quad\cdot\lambda^{n_v/2-1}\exp\left(-\frac{n_v v_0}{2}\lambda\right). \end{aligned} \quad (5)$$

The NG prior in (5) can reflect a wide variety of information regarding μ and λ. A priori, $\mu\,|\,\lambda \sim \mathrm{N}(\mu_0,(n_\mu\lambda)^{-1})$, where $\mu_0\in\mathbb{R}$ and $n_\mu > 0$. And, the marginal distribution of λ is gamma with $n_v > 0$ and $v_0 > 0$, so that $E[\lambda] = 1/v_0$ and $\mathrm{Var}[\lambda] = 2/(n_v v_0^2)$. With this parametrization, one can say that a priori the decision maker's best guess of μ is μ_0 and n_μ can be viewed as the equivalent sample size for the prior information on μ. And, the decision maker's best guess of $\lambda = 1/\sigma^2$ is $1/v_0$ with n_v as the equivalent sample size for the prior information on λ.

With the likelihood function in (4) and the prior in (5), the posterior distribution for μ and λ given $\mathbf{x}$ is shown in Proposition 2.

Proposition 2. (a) $f(\mu,\lambda\,|\,\mathbf{x})$ *is of the same form as in* (5) *but with updated parameters*:

$$\begin{aligned} f(\mu,\lambda\,|\,\mathbf{x}) &= \mathrm{NG}(\mu,\lambda\,|\,\mu^*,n_\mu^*,v^*,n_v^*) = f(\mu\,|\,\lambda,\mathbf{x})f(\lambda\,|\,\mathbf{x})\\ &= \mathrm{N}(\mu\,|\,\mu^*,(n_\mu^*\lambda)^{-1})\mathrm{Gamma}\left(\lambda\,\middle|\,\frac{n_v^*}{2},\frac{n_v^*v^*}{2}\right), \end{aligned} \quad (6)$$

where $n_\mu^* = n_\mu + k^*$, $\mu^* = (n_\mu\mu_0 + k^*\bar{x})/n_\mu^*$, $n_v^* = n_v + k$ *and* $v^* = (1/n_v^*)(n_v v_0 + (k-1)s^{*2} + (n_\mu k^*/(n_\mu+k^*))(\bar{x}-\mu_0)^2)$, *with* k^* *and* s^{*2} *as defined in Proposition 1.*

(b) *For* $\rho = 0$, $f(\mu,\lambda\,|\,\mathbf{x})$ *is the same as with* k *independent experts.*

(c) *For* $\rho > 0$, $f(\mu\,|\,\lambda,\mathbf{x})$ *is the same as with* k^* *independent experts, where* $k^* \le k$ *is a convex decreasing function of* ρ. *And,* $f(\mu,\lambda\,|\,\mathbf{x})$ *is the same as if a sample variance of* s^{*2} *rather than* s^2 *is observed with* k *independent experts, where* $s^{*2} \ge s^2$ *is a convex increasing function of* ρ, *while at the same time as if a sample mean of* $\bar{x}$ *is observed with only* k^* *of* k *independent experts.*

It is worth highlighting Part (c) of Proposition 2, which shows that for $\rho > 0$, the posterior conditional distribution of μ given λ is the same as one would obtain with k^* independent experts. In that sense, k^* can be viewed as an equivalent independent sample size for inferences about μ given λ. It is easy to see that $k^* = k$ for $\rho = 0$. And, as ρ gets larger than zero, k^* shrinks rapidly, decreasing to 1 for $\rho = 1$. This is consistent with the results in Clemen and Winkler (1985) for the case of known λ, that any positive dependence among the point forecasts reduces the information content of the forecasts for inferences about μ. Further, while there is loss of information on μ, there is no loss of information on λ. The sample variance s^2 is simply adjusted by factor $1/(1-\rho)$ to obtain an unbiased estimator s^{*2} of σ^2. Hence, as mentioned earlier, it is intuitively appealing to summarize the k point forecasts with the sufficient statistic $\{\bar{x}, k^*, s^{*2}, k\}$, where $\bar{x}$ is the sample estimate of μ with effective sample size k^*, and s^{*2} is the sample estimate of σ^2 with effective sample size k.

The decision maker's primary interest is in the predictive distribution for $\tilde{y}$ given $\mathbf{x}$, which is shown in Corollary 1.

Corollary 1. *The predictive distribution for $\tilde{y} \mid \mathbf{x}$ is a t distribution with degrees of freedom n_v^*, location parameter μ^*, and scale parameter $\sqrt{(n_\mu^*+1)v^*/n_\mu^*}$, so that*

$$E[\tilde{y} \mid \mathbf{x}] = \mu^*, \quad \text{for } n_v^* > 1, \quad \text{and} \tag{7}$$

$$\begin{aligned}\mathrm{Var}[\tilde{y} \mid \mathbf{x}] &= \mathrm{Var}[\tilde{y} \mid \mu, \mathbf{x}] + \mathrm{Var}[\mu \mid \mathbf{x}] \\ &= \frac{n_v^*}{n_v^*-2}\left(v^* + \frac{v^*}{n_\mu^*}\right), \quad \text{for } n_v^* > 2. \end{aligned} \tag{8}$$

In (8), $\mathrm{Var}[\tilde{y} \mid \mu, \mathbf{x}]$ corresponds to the sampling uncertainty given μ. And, $\mathrm{Var}[\mu \mid \mathbf{x}]$ reflects the uncertainty about μ itself. The uncertainty about the precision λ is embedded in $\mathrm{Var}[\tilde{y} \mid \mu, \mathbf{x}]$ and $\mathrm{Var}[\mu \mid \mathbf{x}]$. All else unchanged, a higher ρ shifts the distribution of $\lambda \mid \mathbf{x}$ toward lower values of λ (i.e., increases $\mathrm{Var}[\tilde{y} \mid \mu, \mathbf{x}]$ and $\mathrm{Var}[\mu \mid \mathbf{x}]$) and leads to a greater loss of information about μ in the sample (i.e., further increases $\mathrm{Var}[\mu \mid \mathbf{x}]$), resulting in a higher variance for $\tilde{y} \mid \mathbf{x}$. On the other hand, all else held equal with $\rho > 0$, a higher k reduces the parameter uncertainty, although much less for μ than for λ, resulting in a lower variance for $\tilde{y} \mid \mathbf{x}$.

We explore the impact of dependence and parameter uncertainty on the predictive uncertainty about $\tilde{y} \mid \mathbf{x}$ in greater detail in Section 2.1.1.

2.1.1. The Impact of Ignoring Dependence and Parameter Uncertainty.

In this section, we compare our model with some other approaches that might ignore dependence or parameter uncertainty or both, to provide a flavor for the relative seriousness of such omissions.

Some of the frequently used models (such as the heuristic in Fisher and Raman 1996, hereafter *FRH*) do not incorporate any prior information on μ and σ^2. Indeed, often, the decision maker might have very little or no such prior information (although, at the same time, might judge the experts to be correlated on the basis of some shared information). In any case, to focus entirely on how different approaches extract the information content of the forecasts regarding μ and σ^2, we restrict attention in our model to predictive distribution for $\tilde{y} \mid \mathbf{x}$ with a diffuse prior on μ and λ. We compare four methods described below.

Predictive Distribution with Known ρ (PD). This is our model with a diffuse prior for μ and λ. A common choice for a diffuse prior is a NG prior in (5) with parameters $n_\mu = n_v = 0$, which is an improper prior. With this diffuse prior, $f(\tilde{y} \mid \mathbf{x})$ is a t distribution with k degrees of freedom, location $\bar{x}$, and scale $\sqrt{(k^*+1)v^*/k^*}$, which yields $E[\tilde{y} \mid \mathbf{x}] = \bar{x}$ and

$$\mathrm{Var}[\tilde{y} \mid \mathbf{x}] = \frac{k-1}{k-2}\left(\frac{1+\rho}{1-\rho} + \frac{1}{k}\right)s^2. \tag{9}$$

The $\mathrm{Var}[\tilde{y} \mid \mathbf{x}]$ in (9) is larger for a higher ρ given k, reflecting a greater loss of information. On the other hand, it is smaller for a higher k given ρ, reflecting reduced parameter uncertainty. For very large k, $\mathrm{Var}[\tilde{y} \mid \mathbf{x}] \approx (1+\rho)s^2/(1-\rho)$. Further, with also $\rho = 0$, $\mathrm{Var}[\tilde{y} \mid \mathbf{x}] \approx s^2$. In other words, with k large and $\rho = 0$, $f(\tilde{y} \mid \mathbf{x})$ converges to a normal distribution with mean $\bar{x}$ and variance s^2. Gaur et al. (2007, p. 480) test the hypothesis that the variance of demand is positively correlated with dispersion among experts' forecasts. Our model shows that this is indeed the case.

Predictive Distribution with $\rho \stackrel{\text{set}}{=} 0$ (PD_0). This is the same as *PD* above but with ρ assumed to be 0—i.e., a Bayesian model that assumes i.i.d. forecasts. In other words, this approach retains parameter uncertainty but assumes independence among the experts. In this setup, $E[\tilde{y} \mid \mathbf{x}] = \bar{x}$ and

$$\mathrm{Var}[\tilde{y} \mid \mathbf{x}] = \frac{k-1}{k-2}\left(1 + \frac{1}{k}\right)s^2. \tag{10}$$

While ignoring ρ does not impact $E[\tilde{y} \mid \mathbf{x}]$, $\mathrm{Var}_{PD_0}[\tilde{y} \mid \mathbf{x}] < \mathrm{Var}_{PD}[\tilde{y} \mid \mathbf{x}]$ for any $\rho > 0$. And, this difference increases rapidly with ρ. Even in the limit with a very large k, $\mathrm{Var}_{PD}[\tilde{y} \mid \mathbf{x}]$ remains larger by a factor of $(1+\rho)/(1-\rho)$.

Certainty Equivalence Method with ρ Known (CE). Here, it is assumed that $\tilde{y} \mid \mathbf{x} \sim \mathrm{N}(\bar{x}, s^{*2} = s^2/(1-\rho))$. In this approach, $\bar{x}$ is used as a point estimate of μ, and s^{*2} is used as a point estimate for σ^2. While this approach incorporates the correlation between experts, it ignores any parameter uncertainty. In this sense, this is a certainty equivalence model corresponding to *PD*. This of course does not impact $E[\tilde{y} \mid \mathbf{x}]$, but as one would expect, $\mathrm{Var}_{CE}[\tilde{y} \mid \mathbf{x}] < \mathrm{Var}_{PD}[\tilde{y} \mid \mathbf{x}]$ for any $k > 2$ and $\rho \geq 0$, with

$$\frac{\mathrm{Var}_{PD}[\tilde{y} \mid \mathbf{x}]}{\mathrm{Var}_{CE}[\tilde{y} \mid \mathbf{x}]} = \frac{k-1}{k-2}\left[(1+\rho) + \frac{(1-\rho)}{k}\right]. \tag{11}$$

And, given $\rho > 0$, *CE* even with a very large k does not account for the loss of information due to dependence.

Certainty Equivalence Method with $\rho \stackrel{\text{set}}{=} 0$ (CE_0). In this approach, $\tilde{y} \mid \mathbf{x} \sim \mathrm{N}(\bar{x}, s^2)$. As in *CE*, this approach does not include any parameter uncertainty, and further ignores ρ. Note that $\mathrm{Var}_{CE_0}[\tilde{y} \mid \mathbf{x}]$ is the smallest of all the models discussed here, and in that sense includes the most spurious accuracy.

For ease of reading, Table 1 summarizes our model along with the three other methods discussed above, and provides for each the distribution of $\tilde{y} \mid \mathbf{x}$, along with the corresponding variance and a $100(1-\gamma)\%$ prediction interval for $\tilde{y} \mid \mathbf{x}$, where $t_k(\theta_1, \theta_2)$ denotes a t distribution with degrees of freedom k, location θ_1, and scale θ_2; $\mathrm{N}(\theta_1, \theta_2)$ denotes a normal distribution with mean θ_1 and variance θ_2, whereas $t_{1-\gamma/2,k}$ and $z_{1-\gamma/2}$ are the $(1-\gamma/2)$ quantiles of the standard t distribution with k degrees of freedom and the standard

Table 1. A Summary of the Four Methods (PD, PD_0, CE, CE_0) in the Normal Model

Method	$f(\tilde{y}\,\vert\,\mathbf{x})$	$\mathrm{Var}[\tilde{y}\,\vert\,\mathbf{x}]$	$100(1-\gamma)\%$ prediction interval	ρ	Parameter uncertainty
PD	$t_k\left(\bar{x}, \sqrt{\frac{k-1}{k}\left(\frac{1+\rho}{1-\rho}+\frac{1}{k}\right)}s\right)$	$\frac{k-1}{k-2}\left(\frac{1+\rho}{1-\rho}+\frac{1}{k}\right)s^2$	$\bar{x} \pm t_{1-\gamma/2,k}\sqrt{\frac{k-1}{k}\left(\frac{1+\rho}{1-\rho}+\frac{1}{k}\right)}s$	✓	✓
PD_0	$t_k\left(\bar{x}, \sqrt{\frac{k-1}{k}\left(1+\frac{1}{k}\right)}s\right)$	$\frac{k-1}{k-2}\left(1+\frac{1}{k}\right)s^2$	$\bar{x} \pm t_{1-\gamma/2,k}\sqrt{\frac{k-1}{k}\left(1+\frac{1}{k}\right)}s$	✗	✓
CE	$\mathrm{N}\left(\bar{x}, \frac{s^2}{1-\rho}\right)$	$\frac{s^2}{1-\rho}$	$\bar{x} \pm z_{1-\gamma/2}\sqrt{\frac{1}{1-\rho}}s$	✓	✗
CE_0	$\mathrm{N}(\bar{x}, s^2)$	s^2	$\bar{x} \pm z_{1-\gamma/2}s$	✗	✗

normal distribution, respectively. The last two columns of the table indicate whether the method accounts for ρ and parameter uncertainty. As mentioned above, $E[\tilde{y}\,\vert\,\mathbf{x}] = \bar{x}$ in all of the four methods, but the variances and prediction intervals differ depending on whether one accounts for ρ and the parameter uncertainty.

Let $\mathrm{Var}[\tilde{y}\,\vert\,\mathbf{x}]$ under Method i be $\mathrm{Var}[i]$. Then, given $\rho > 0$ and $k > 2$, it is easy to see that $\mathrm{Var}[PD] > \max\{\mathrm{Var}[PD_0], \mathrm{Var}[CE]\} > \mathrm{Var}[CE_0]$. And, $\mathrm{Var}[CE] > (<)\,\mathrm{Var}[PD_0]$ for $\rho > (<)(2k-1)/(k^2-1)$, with equality at $\rho = (2k-1)/(k^2-1)$. Moreover, given $\rho = 0$, $\mathrm{Var}[PD] = \mathrm{Var}[PD_0] > \mathrm{Var}[CE] = \mathrm{Var}[CE_0]$.

Since PD and PD_0 lead to a t distribution while CE and CE_0 assume a normal distribution for $\tilde{y}\,\vert\,\mathbf{x}$, it is useful to compare the four methods in terms of their respective prediction intervals for $\tilde{y}\,\vert\,\mathbf{x}$. Let $\mathrm{IW}[i]$ be the interval width of a $100(1-\gamma)\%$ prediction interval for $\tilde{y}\,\vert\,\mathbf{x}$ under Method i.

Proposition 3. *For $k > 2$, conditional on $\mathbf{x}$, we have*

$$\begin{aligned}\mathrm{IW}[PD] &> \max\{\mathrm{IW}[PD_0], \mathrm{IW}[CE]\}\\ &\geq \min\{\mathrm{IW}[PD_0], \mathrm{IW}[CE]\}\\ &> \mathrm{IW}[CE_0], \quad \textit{for } \rho > 0,\\ \mathrm{IW}[PD] &= \mathrm{IW}[PD_0] > \mathrm{IW}[CE] = \mathrm{IW}[CE_0], \quad \textit{for } \rho = 0.\end{aligned}$$

Figure 1 shows, as an example, the relative widths of 95% prediction intervals for $\tilde{y}\,\vert\,\mathbf{x}$ under the PD_0, CE, and CE_0 methods as a function of ρ for $k = 3$, 7, and 100, rescaled by setting the width under PD for a given k equal to 1. In all cases, the widths are below 1, reflecting percentage underestimation of uncertainty due to ignoring ρ, parameter uncertainty, or both. As expected, CE_0 underestimates the most. The underestimation in all three methods is worse with a higher ρ and a smaller k. Both PD and CE account for dependence. However, while the underestimation in CE is less with a higher k, it can remain substantial even for large k. For example, given $\rho = 0.6$, the underestimation in CE is about 43% with $k = 3$, 30% with $k = 7$, and still 22% with $k = 100$. On the other hand, both PD_0 and CE_0 do not account for dependence. Hence, their underestimation is worse with a higher ρ. And, with a higher k, while they are more similar, their underestimation relative to PD and CE gets exacerbated. This is because information loss due to dependence can be substantial—i.e., k^* is a decreasing convex function of ρ. Hence, with a larger k, the spurious accuracy in PD_0 and CE_0 is larger. A comparison of PD_0 and CE reflects the relative seriousness of ignoring ρ and parameter uncertainty. For low values of ρ and low-to-medium values of k, ignoring the parameter uncertainty is relatively more serious. However, if ρ is even moderately large, ignoring ρ becomes the more serious omission, and remains so even with a very high k. As mentioned earlier, it would not be unusual to expect ρ in the range of 0.6–0.8, so it is reasonable to say that ignoring ρ is likely to be the much more serious error.

The augmentation factor in our PD model that is needed to equate the standard deviation s of

Figure 1. Relative Widths of 95% Prediction Intervals for $\tilde{y}\,\vert\,\mathbf{x}$ Under the Three Methods (PD_0, CE, CE_0), Rescaled with the Width Under PD (for Any Given k) = 1, As a Function of ρ for Selected Values of k

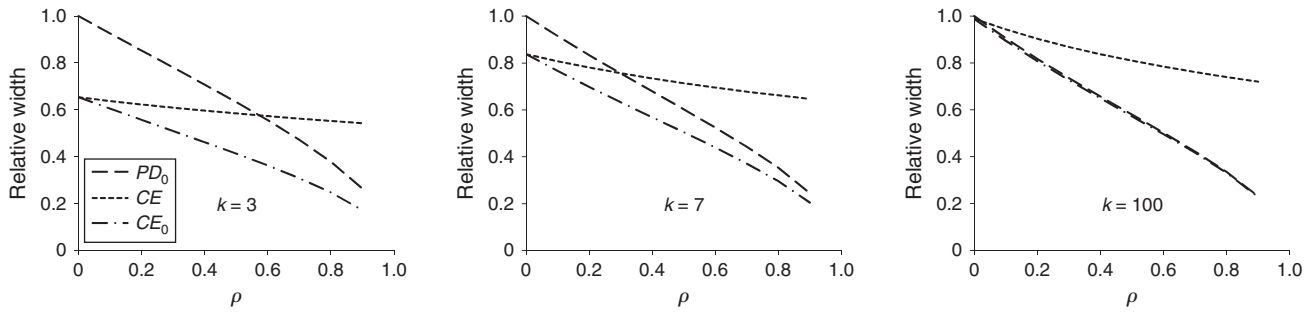


Table 2. Augmentation Factor δ in the Normal Model for Selected Values of k and ρ

	ρ									
k	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	1.63	1.76	1.91	2.09	2.31	2.58	2.94	3.46	4.32	6.22
4	1.37	1.49	1.62	1.78	1.97	2.21	2.52	2.98	3.72	5.37
5	1.26	1.38	1.51	1.66	1.84	2.07	2.37	2.80	3.50	5.06
6	1.21	1.32	1.44	1.59	1.77	1.99	2.28	2.70	3.39	4.89
7	1.17	1.28	1.40	1.55	1.72	1.94	2.23	2.64	3.31	4.79
8	1.15	1.25	1.38	1.52	1.69	1.91	2.19	2.60	3.26	4.72
9	1.13	1.23	1.36	1.50	1.67	1.89	2.17	2.57	3.23	4.67
10	1.11	1.22	1.34	1.48	1.65	1.87	2.15	2.55	3.20	4.64
20	1.05	1.16	1.28	1.42	1.59	1.79	2.07	2.46	3.09	4.48
100	1.01	1.12	1.24	1.37	1.54	1.74	2.01	2.39	3.02	4.38

the k forecasts with the standard deviation of $\tilde{y}\,|\,\mathbf{x}$ is $\delta = \sqrt{((k-1)/(k-2))((1+\rho)/(1-\rho)+1/k)}$, which is an increasing convex function of ρ and a decreasing convex function of k. As an illustration, Table 2 provides δ for different values of ρ and k. Note that, for $\rho > 0$, even with large k, δ is higher than 1, substantially so at high values of ρ. In fact, as shown in (9), in the limit, $\delta = (1+\rho)/(1-\rho)$.

As mentioned before, an added advantage of our approach is that it can be used without past data, by subjectively assessing ρ (which we discuss later in Section 3.3) and then simply applying the corresponding δ. In *FRH*, $\tilde{y}\,|\,\mathbf{x} \sim \mathrm{N}(\bar{x}, (as)^2)$, where a is estimated with past data as the ratio of standard deviation of forecast errors (difference between predicted and actual demand across products in the previous season, where predicted demand is the average forecast by experts) to the average standard deviation of experts' forecasts (across products in the current season). It is assumed that forecast errors follow a similar distribution in past and future seasons, with mean zero, and that the same set of experts are forecasting across seasons with similar dispersion in their forecasts. Besides the issue of applicability of past data for the variable of interest, *FRH* provides no intuition on the drivers for a and hence also no basis for estimating it without past data. Although parameter uncertainty is acknowledged, a is not associated with ρ or with k.[2]

3. Extensions of the Normal Model and Assessment of Dependence

Until now, we have assumed that the pairwise correlations between the experts is the same and known. To investigate the robustness of this assumption, we introduce uncertainty on ρ and heterogeneity in the pairwise correlations, and consider the impact on δ. Further, in this section, we also extend the approach to a lognormal model and discuss the assessment of ρ.

3.1. Uncertainty and Heterogeneity About ρ in the Normal Model

We first explore uncertainty on the common ρ. Recall that in our model, δ depends only on ρ and k. For a given k, we take 10,000 independent draws of ρ from a beta distribution, $f(\rho) \propto \rho^{\alpha-1}(1-\rho)^{\beta-1}$, with $\alpha, \beta > 0$, $E[\rho] = \alpha/(\alpha+\beta)$, and $\mathrm{Var}[\rho] = \alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$, and calculate the average δ. We repeat this for $k = 3, 5$, and 10, and for a variety of values for the parameters of the beta distribution. Keeping $E[\rho]$ constant at $0.1, 0.2, \ldots, 0.9$, we vary $\alpha+\beta$ from 100 (low uncertainty) to 50 (medium uncertainty) to 10 (high uncertainty). Table 3 shows the average δ for different values of k and $E[\rho]$ with varying levels of uncertainty, along with the case of known ρ (no uncertainty).

First note that, given k, δ under no uncertainty is the lowest. It makes sense that any uncertainty about ρ is yet another addition to the parameter uncertainty and should increase δ. Further, for $E[\rho] \le 0.6$, δ does not vary much at all with increasing uncertainty about ρ. Even for $E[\rho] \ge 0.7$, the impact is less than remarkable. For example, with $E[\rho] = 0.7$, the increase in δ from no uncertainty to high uncertainty is 3.46 to 3.84 for $k = 3$, 2.8 to 3.12 for $k = 5$, and 2.55 to 2.83 for $k = 10$. It is only in the extreme case of high uncertainty and high $E[\rho]$ at 0.9 that the jump in δ is substantial.

Next, we explore heterogeneity in the pairwise correlations. For such a case, our model with a common ρ no longer holds. In Appendix B, we develop a more general model, where $\mathbf{x} = (x_1, \ldots, x_k)'$ follows a multivariate normal with mean vector $\boldsymbol{\mu} = \mu\mathbf{e}$, where $\mathbf{e} = (1, \ldots, 1)'$ is a $k \times 1$ column vector, and $k \times k$ positive definite covariance matrix $\boldsymbol{\Sigma}$ with diagonal elements σ^2 and off-diagonal elements $\rho_{ij}\sigma^2$, with ρ_{ij} known. Both $E[\tilde{y}\,|\,\mathbf{x}]$ and $\mathrm{Var}[\tilde{y}\,|\,\mathbf{x}]$ are weighted mixtures of ρ_{ij}s and the observed x. And, hence, the augmentation factor δ depends not only on k and ρ_{ij}s but also on the observed $\mathbf{x}$.

Given k, we generate each pairwise correlation ρ_{ij} independently from a beta distribution as defined before. Here, the mean of the beta distribution is the

Table 3. Impact of Uncertainty About ρ on δ: Average δ for Selected Values of k and $E[\rho]$ with Varying Levels of Uncertainty About ρ

	k and levels of uncertainty about ρ											
	$k=3$				$k=5$				$k=10$			
$E(\rho)$	None	Low	Medium	High	None	Low	Medium	High	None	Low	Medium	High
0.1	1.76	1.76	1.77	1.77	1.38	1.38	1.38	1.38	1.22	1.22	1.22	1.23
0.2	1.91	1.92	1.92	1.94	1.51	1.51	1.51	1.52	1.34	1.34	1.35	1.36
0.3	2.09	2.10	2.10	2.14	1.66	1.66	1.66	1.69	1.48	1.49	1.49	1.52
0.4	2.31	2.32	2.32	2.39	1.84	1.84	1.85	1.89	1.65	1.66	1.66	1.71
0.5	2.58	2.59	2.61	2.71	2.07	2.07	2.08	2.17	1.87	1.88	1.89	1.96
0.6	2.94	2.97	2.98	3.14	2.37	2.38	2.40	2.55	2.15	2.16	2.18	2.30
0.7	3.46	3.50	3.53	3.84	2.80	2.82	2.85	3.12	2.55	2.57	2.60	2.83
0.8	4.32	4.40	4.45	5.24	3.50	3.55	3.63	4.26	3.20	3.26	3.30	3.90
0.9	6.22	6.44	6.69	10.76	5.06	5.26	5.49	8.47	4.64	4.79	4.99	8.13

average ρ_{ij} among the k experts, and a higher variance indicates a higher level of heterogeneity. If a draw results in a covariance matrix that is not positive definite, then that draw is discarded and another is taken until there are 10,000 draws in all. For each draw of a positive definite covariance matrix, we generate a k-variate forecast from a multivariate normal, without loss of generality (WLOG) with mean vector 0 and a covariance matrix based on the drawn observation of a set of ρ_{ij}s and with $\sigma^2 = 1$. We then compute δ as the ratio of $\sqrt{\text{Var}[\tilde{y} \mid \mathbf{x}]}$ with respect to the observed s. The proportion of positive definite covariance matrices is small for large k and for high levels of heterogeneity with high $E(\rho)$. So, we generate simulations for $k = 3, 5$, and 10, and holding *average* ρ_{ij}=$\alpha/(\alpha+\beta)$ constant at 0.1, 0.2,. . . , 0.7, we vary $\alpha+\beta$ (the level of heterogeneity) from 100 (low) to 50 (medium), along with the corresponding case where the average ρ_{ij} is the common ρ (none). For what we label medium, the degree of heterogeneity is still reasonably high. For example, with average $\rho_{ij} = 0.5$, the standard deviation of beta distribution is 0.07 for medium heterogeneity and 0.05 for low heterogeneity. Table 4 shows the average δ for these cases.

Once again, the overall results are similar to the case of known and common ρ (shown in Table 2). Given k, a higher average ρ_{ij} leads to a higher δ. And given average ρ_{ij}, a higher k leads to lower δ. What is more reassuring is that, at least for the results we obtained, heterogeneity has no to low effect on δ.

In sum, the results indicate that the δ in our model with common and known ρ provides a lower bound for cases where ρ is common but uncertain and for cases where ρ equals the average of ρ_{ij}s. The level of uncertainty on the common ρ matters substantially only for extreme cases of high uncertainty and high $E[\rho]$, whereas δ is robust with respect to heterogeneity in pairwise correlations. This provides further support for our model with common and known ρ. And, more broadly, this is consistent with the discussion in Section 1 on equal-weights model for experts.

3.2. A Lognormal Model

In Section 2.1, we assume that the distribution of $\tilde{y}$ given μ and σ^2 is normal. This is often a reasonable approximation for many real-life settings. However, in some instances, the decision maker might consider $f(\tilde{y} \mid \theta)$ to be skewed and/or bounded from below with a long right tail, such as demand for a new product that has a small chance but a large potential for blockbuster sales. As mentioned in Bimpikis and Markakis (2016), the markets for fashion, technology, and creative goods

Table 4. Impact of Heterogeneity in Pairwise Correlations on δ: Average δ for Selected Values of k and Average ρ_{ij} with Varying Levels of Heterogeneity in ρ_{ij}s

	k and levels of heterogeneity in ρ_{ij}s								
	$k=3$			$k=5$			$k=10$		
Average ρ_{ij}	None	Low	Medium	None	Low	Medium	None	Low	Medium
0.1	1.76	1.77	1.76	1.38	1.38	1.38	1.22	1.22	1.22
0.2	1.91	1.92	1.92	1.51	1.51	1.51	1.34	1.34	1.34
0.3	2.09	2.10	2.10	1.66	1.66	1.66	1.48	1.49	1.49
0.4	2.31	2.32	2.32	1.84	1.84	1.85	1.65	1.66	1.66
0.5	2.58	2.59	2.60	2.07	2.07	2.08	1.87	1.88	1.89
0.6	2.94	2.96	2.97	2.37	2.38	2.39	2.15	2.17	2.18
0.7	3.46	3.49	3.52	2.80	2.82	2.85	2.55	2.57	2.58

Table 5. Summary of the Four Methods (PD, PD_0, CE, CE_0) in the Lognormal Model

	$f(\ln \tilde{y} \mid \mathbf{x})$	A $100(1-\gamma)\%$ prediction interval for $\tilde{y} \mid \mathbf{x}$
PD	$t_k\left(\bar{x}_t, \sqrt{\frac{k-1}{k}\left(\frac{1+\rho_L}{1-\rho_L}+\frac{1}{k}\right)}s_t\right)$	$\exp\left(\bar{x}_t \pm t_{1-\gamma/2,k}\sqrt{\frac{k-1}{k}\left(\frac{1+\rho_L}{1-\rho_L}+\frac{1}{k}\right)}s_t\right)$
PD_0	$t_k\left(\bar{x}_t, \sqrt{\frac{k-1}{k}\left(1+\frac{1}{k}\right)}s_t\right)$	$\exp\left(\bar{x}_t \pm t_{1-\gamma/2,k}\sqrt{\frac{k-1}{k}\left(1+\frac{1}{k}\right)}s_t\right)$
CE	$\mathrm{N}\left(\bar{x}_t, \frac{s_t^2}{1-\rho_L}\right)$	$\exp\left(\bar{x}_t \pm z_{1-\gamma/2}\sqrt{\frac{1}{1-\rho_L}}s_t\right)$
CE_0	$\mathrm{N}(\bar{x}_t, s_t^2)$	$\exp(\bar{x}_t \pm z_{1-\gamma/2}s_t)$

typically exhibit heavy-tailed demand distributions (see also Chevalier and Goolsbee 2003, Gaffeo et al. 2008). In such cases, a lognormal distribution might be more appropriate than a normal. In that spirit, we modify our model in Section 2.1 such that $\ln \tilde{y} \sim \mathrm{N}(\mu, \sigma^2)$ and $\ln x_1, \ldots, \ln x_k$ follow a multivariate normal with mean vector $\boldsymbol{\mu} = \mu\mathbf{e}$, where $\mathbf{e} = (1, \ldots, 1)'$ is a $k \times 1$ column vector, and $k \times k$ positive definite covariance matrix $\boldsymbol{\Sigma}$ with diagonal elements σ^2 and off-diagonal elements $\rho_L\sigma^2$, with $\rho_L = \mathrm{Corr}[\ln x_i, \ln x_j \mid \theta]$, $i \neq j$, which implies that $\rho = \mathrm{Corr}[x_i, x_j \mid \theta] = (e^{\rho_L\sigma^2} - 1)/(e^{\sigma^2} - 1)$ (Johnson et al. 2002). Note that ρ_L depends also on σ (and hence the CV of the distribution) and is monotonically increasing in ρ between $\rho_L = \rho = 0$ and $\rho_L = \rho = 1$.

In our normal model, $f(\tilde{y} \mid \mathbf{x}) = f(\tilde{y} \mid \bar{x}, s^2)$ is a t distribution. Using the same results, in our lognormal model, $f(\ln \tilde{y} \mid \mathbf{x}) = f(\ln \tilde{y} \mid \bar{x}_t, s_t^2)$ is a t distribution with the same parameters as in the normal model except for $\bar{x}_t = (1/k)\sum_{i=1}^k \ln x_i$ and $s_t^2 = (1/(k-1))\sum_{i=1}^k (\ln x_i - \bar{x}_t)^2$ replacing $\bar{x}$ and s^2, respectively. The distribution for $\tilde{y} \mid \mathbf{x}$ is then a log t distribution that is very heavy tailed with undefined moments. However, it is still possible to construct prediction intervals for $\tilde{y} \mid \mathbf{x}$ by simply taking the exponential transforms of the corresponding quantiles of the underlying t distribution for $f(\ln \tilde{y} \mid \mathbf{x})$. Table 5 summarizes the lognormal model under the four methods (PD, PD_0, CE, and CE_0) defined in Section 2.1.1, and provides for each of those the distribution of $\ln \tilde{y} \mid \mathbf{x}$ and the corresponding $100(1-\gamma)\%$ prediction interval, $0 \leq \gamma \leq 1$, for $\tilde{y} \mid \mathbf{x}$.

Note that the prediction intervals are stated in terms of $\rho_L = \mathrm{Corr}[\ln x_i, \ln x_j \mid \theta]$. It is easy to see that the rank order of the widths of these prediction intervals under the four methods is the same as in our normal model, with the prediction interval for PD (CE_0) being the widest (narrowest). Furthermore, given $0 \leq \rho \leq 1$, $\rho_L - \rho \geq 0$ and this difference is increasing in σ^2. This then implies that the impact of ignoring parameter uncertainty or the dependence in the normal model is only exacerbated in the lognormal model, even more so as the coefficient of variation (CV) of the lognormal distribution increases. In other words, ignoring dependence or parameter uncertainty would create even greater spurious accuracy in the prediction intervals for $\tilde{y}$ given $\mathbf{x}$ in the lognormal model than in the normal model, and the same would get worse with higher CV of the lognormal. This renders our approach even more crucial in a lognormal model.

3.3. Assessing Dependence

In Section 3.1, we observe that our model is fairly robust with respect to uncertainty about ρ or any heterogeneity in the pairwise correlations. This adds to the attractiveness of the equal-weights model discussed in Section 1. However, there still remains the issue of estimating the common correlation.

If relevant past data exist, or if experts provide several forecasts of the same variable (for example, multiple time-based forecasts such as price of oil for various periods), standard statistical methodologies exist to estimate the pairwise correlations (see, e.g., Meyer and Booker 2001, for a discussion specific to expert judgment), and a mean of such correlations would be a reasonable input in our model. Equally if not more likely is a situation where no relevant past data are available. There might be new experts or a new set of information with a different proportion of common to private information among experts. As mentioned earlier, with the experts providing only one k-variate observation, there is no information content in the sample on the correlation. For example, to obtain a maximum likelihood estimate (MLE) of ρ, $0 \leq \rho \leq 1$, we need more than one observation to avoid the degenerate case with MLE of $\rho = 0$. Hence, some subjective assessment on ρ might be necessary, simply assuming $\rho = 0$ as the default option clearly not being advisable.

For such a scenario, Clemen et al. (2000) provide a review of past research and a comparison of six assessment procedures with a prescriptive orientation. They find that asking for a direct estimate of correlation works best. However, in a further excellent review of related literature, Garthwaite et al. (2005) find this somewhat surprising and point to studies that propose method of assessing *concordance probability* as being

more preferable for eliciting correlations (Gokhale and Press 1982, Kunda and Nisbett 1986). Yet, another way to think about the common correlation between experts is as the proportion of information that is common to all the experts (Lichtendahl et al. 2013a, Winkler 1981). More recent work by Palley and Soll (2016) proposes an elicitation method for estimating the proportion of shared information to private information among forecasters by asking them to provide not just their individual forecasts but also the average forecast that will be provided by the other forecasters, which can then become the basis for the assessment of the common correlation.

Akin to the idea of *wisdom of crowds* and the common practice of plurality of approaches in prescriptive decision analysis (Brown and Lindley 1986, Lindley 1986), Winkler and Clemen (2004) consider combining correlation assessments by simple averaging with multiple experts and multiple methods, and find that there is substantial improvement in accuracy by increasing the number of experts or methods, but with diminishing returns to extra experts or methods. They state (pp. 173–174) that "the gains are much greater from multiple experts than from multiple methods, with a second expert being worth more on average than four methods." Besides averaging, there might be other ways to reconcile different assessments of correlation. For example, the differences in estimates from two different methods could be used in some process of reconciliation to make the experts "think harder and more carefully," by sharing with them that one estimate is higher than the other, and then asking them to readjust their initial estimates until the estimates more or less converge. This, however, involves repeated assessments from the experts, which could be burdensome with no guarantee of reconciliation.

In sum, combining assessments (e.g., by averaging) using at least two methods across multiple experts seems the most reasonable to us. In the e-companion to this paper, we provide a discussion and an example of how the correlation might be assessed, along with a step-by-step demonstration of our *PD* method. However, this remains a worthwhile area for further research.

4. An Illustration for the Newsvendor Problem

Until now, we considered the impact of ignoring dependence and parameter uncertainty on $\tilde{y}\,|\,\mathbf{x}$ in purely inferential terms. However, a decision maker's primary concern might be whether such an impact affects the ultimate consequences of a decision. In this section, we extend the analysis to a frequently encountered decision-making context of a newsvendor. In a typical newsvendor setting, a decision maker must make a one-time ordering decision ahead of the selling season without knowing the demand. The parameters for the newsvendor problem are as follows: $c > 0$ is the unit cost, $p > c$ is unit selling price, and $v < c$ is the unit salvage value. Let $\Pi(y, q)$ be the profit function, where y is the realized demand and q is the order quantity. For a given demand distribution, the solution to the problem is $q^* = \arg\max_q E[\Pi(\tilde{y}\,|\,q)] = F^{-1}_{\tilde{y}}(CR)$, where $F_{\tilde{y}}$ is the cumulative distribution function of demand and $CR = (p-c)/(p-v)$ is the fractile of the demand distribution corresponding to the optimal order quantity.

We show in Section 2.1.1 that ignoring dependence or parameter uncertainty leads to underestimation of uncertainty about the variable of interest. Here, we investigate the resulting implications for the newsvendor quantity and the profit.

4.1. Impact of Ignoring Dependence or Parameter Uncertainty on the Order Quantity

Following Table 1 in Section 2.1.1, it is easily seen that conditional on $\mathbf{x}$ the optimal order quantities under each of the four methods discussed are

$$\begin{aligned} q_{PD} &= \bar{x} + t_{CR,k}\, s \sqrt{\frac{k-1}{k}\left(\frac{1+\rho}{1-\rho} + \frac{1}{k}\right)}; \\ q_{PD_0} &= \bar{x} + t_{CR,k}\, s \sqrt{\frac{k-1}{k}\left(1 + \frac{1}{k}\right)}; \\ q_{CE} &= \bar{x} + z_{CR}\, s \sqrt{\frac{1}{1-\rho}}; \quad q_{CE_0} = \bar{x} + z_{CR}\, s. \end{aligned} \tag{12}$$

Proposition 4 provides a comparison of the four order quantities.

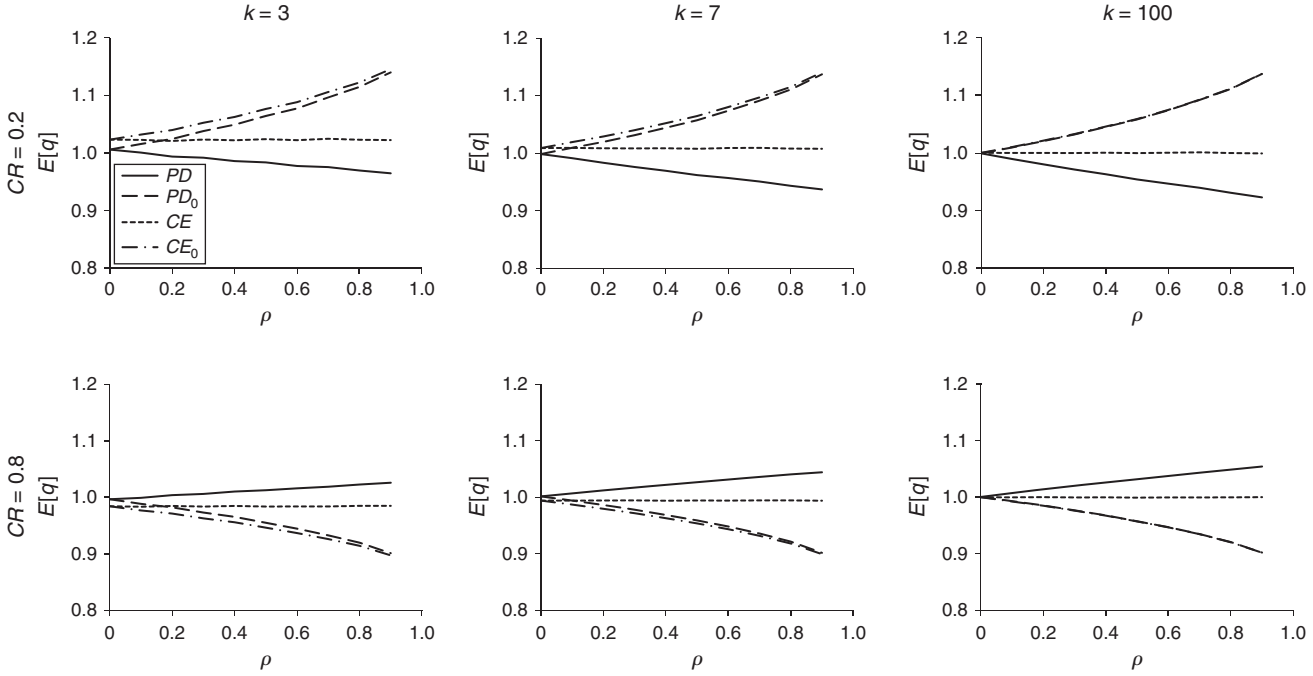
Proposition 4. (a) *For $\rho > 0$ and $k > 2$, conditional on $\mathbf{x}$, we have*

$$\begin{aligned} &q_{PD} < \min\{q_{PD_0}, q_{CE}\} \le \max\{q_{PD_0}, q_{CE}\} < q_{CE_0}, && \text{for } 0 < CR < 0.5, \\ &q_{PD} > \max\{q_{PD_0}, q_{CE}\} \ge \min\{q_{PD_0}, q_{CE}\} > q_{CE_0}, && \text{for } 1 > CR > 0.5, \\ &q_{PD} = q_{PD_0} = q_{CE} = q_{CE_0} = \bar{x}, && \text{for } CR = 0.5. \end{aligned}$$

(b) *Let $\tilde{q}_i$ be the unrealized value of q_i for $i \in \{PD, PD_0, CE, CE_0\}$—i.e., before $\mathbf{x}$ is observed. Then,* $\mathrm{Var}[\tilde{q}_{PD}] \ge \max\{\mathrm{Var}[\tilde{q}_{PD_0}], \mathrm{Var}[\tilde{q}_{CE}]\} \ge \min\{\mathrm{Var}[\tilde{q}_{PD_0}], \mathrm{Var}[\tilde{q}_{CE}]\} \ge \mathrm{Var}[\tilde{q}_{CE_0}]$ *for $\rho > 0$ and $k > 2$.*

For any $\mathbf{x}$, the order quantity in our model (*PD*) is the most conservative of the four, by erring on the side of the less costly consequence: lowest estimation of demand for $CR < 0.5$ and highest estimation of demand

Figure 2. Expected Order Quantities ($E[q]$) Under the Four Methods (PD, PD_0, CE, and CE_0) in the Normal Model, Rescaled with q^*_{PI} (Optimal Order Quantity with μ and σ^2 Known) = 1, for Selected Values of CR and k



for $CR > 0.5$. However, the variance of $\tilde{q}_{PD}$ is also the highest of the four.

It is useful to further compare these quantities with the optimal order quantity q^*_{PI} under perfect information about the demand distribution parameters—i.e., with μ and σ^2 known. From (13), it is easy to verify that $E[\tilde{q}_{PD}]=E[\tilde{q}_{PD_0}]=E[\tilde{q}_{CE}]=E[\tilde{q}_{CE_0}]=q^*_{PI}$ for $CE=0.5$, and that $q^*_{PI} < (>) E[\tilde{q}_{CE}] < (>) E[\tilde{q}_{CE_0}]$ for $CR < (>) 0.5$. Further comparisons are, however, less straightforward, for which we use simulations.

For a given CR, CV, k, and ρ, we generate 100,000 k-variate observations of $\mathbf{x}$ from a multivariate normal and for each observation compute the order quantities under each of the four methods along with q^*_{PI} (which does not depend on $\mathbf{x}$ and hence is the same across all observations). We do this for the cases of low (0.2) and high (0.8) CR, $CV = 0.2$ (assuming WLOG $\mu = 10$ and $\sigma = 2$), $k \in \{3, 7, 100\}$, and $\rho \in \{0, 0.1, \ldots, 0.9\}$. Figure 2 shows the expected order quantities, rescaled with $q^*_{PI} = 1$, under the four methods for the various levels of the parameters. The PD method shows a bias relative to q^*_{PI} in the less costly direction, and this bias increases with ρ. Ignoring parameter uncertainty, CE on average yields an order quantity more consistent with q^*_{PI}, but always with a small bias in the more costly direction. Ignoring both ρ and the parameter uncertainty, CE_0 amplifies this costly bias further. Ignoring ρ but not parameter uncertainty, PD_0 leads to a bias relative to q^*_{PI} in the more costly direction whenever ρ is significantly different from zero or k is large. Given ρ, all that a higher k does is bring the order quantities under PD_0 and CE_0 closer.

It is interesting to contrast our results with some results in the behavioral operations literature. For instance, Ren and Croson (2013) and Ren et al. (2017) show that an overconfident newsvendor will underestimate the variance of the demand distribution (equivalent to ignoring ρ or parameter uncertainty in our model). As a consequence, the order quantity will have a bias relative to q^*_{PI} in the more costly direction, and this bias is linear in the level of overconfidence. One implication from that literature is that correcting for overconfidence will eliminate the decision bias and increase profits. As we show next, in our model, a method that only eliminates the bias need not produce optimal profits.

4.2. Impact of Ignoring Dependence and Parameter Uncertainty on the Newsvendor Profit

Here, we investigate the profit distributions associated with each of the four methods. For analytical tractability, we first analyze profit under the assumption that σ^2 is known. Later, we show numerically that all of the results remain valid also for the case of unknown σ^2.

4.2.1. Unknown μ and Known σ^2. When demand is normally distributed with unknown μ and known σ^2, with a diffuse prior on μ, the predictive distribution of demand $f(\tilde{y} \mid \mathbf{x})$ is normal with mean $\bar{x}$ and variance $(1+1/k^*)\sigma^2$, where $k^* < k$ is the equivalent sample

size as before. The corresponding predictive distribution that ignores ρ is normal with mean $\bar{x}$ and variance $(1+1/k)\sigma^2$. With σ^2 known, the CE and CE_0 are the same and involve simply estimating the unknown parameter μ and result in a demand distribution that is normal with mean $\bar{x}$ and variance σ^2. Hence, ρ does not play any role in these two methods. The optimal order quantities under the different estimation methods are

$$\begin{aligned} q_{PD} &= \bar{x} + z_{CR}\sigma\sqrt{1+\frac{1}{k^*}};\\ q_{PD_0} &= \bar{x} + z_{CR}\sigma\sqrt{1+\frac{1}{k}};\\ q_{CE} &= q_{CE_0} = \bar{x} + z_{CR}\sigma. \end{aligned} \tag{13}$$

Conditional on μ, $\bar{x} \sim \mathrm{N}(\mu, \sigma^2/k^*)$. And, hence, conditional on μ, each of the order quantities in (14) are normally distributed with variance σ^2/k^* and

$$\begin{aligned} E[\tilde{q}_{PD} \mid \mu] &= \mu + z_{CR}\sigma\sqrt{1+\frac{1}{k^*}};\\ E[\tilde{q}_{PD_0} \mid \mu] &= \mu + z_{CR}\sigma\sqrt{1+\frac{1}{k}};\\ E[\tilde{q}_{CE} \mid \mu] &= \mu + z_{CR}\sigma. \end{aligned} \tag{14}$$

Note that conditional on μ, the certainty equivalence model in expectation will lead a decision maker to the optimal order quantity q^*_{PI}. However, as shown next, that expected profit is highest under the PD method. The following theorem states a more general result.

Theorem 1. *Let $\tilde{y} \sim \mathrm{N}(\mu, \sigma^2)$ be the unrealized demand for a newsvendor, and q^*_{PI} the optimal order quantity when μ and σ^2 are known. Let $\tilde{q}_{\epsilon,\tau} \sim \mathrm{N}(q^*_{PI} + \epsilon, \tau^2)$ be a random order quantity, and $E[\Pi(\tilde{y}, \tilde{q}_{\epsilon,\tau})]$ the expected newsvendor profit.*

(a) *$E[\Pi(\tilde{y}, \tilde{q}_{\epsilon,\tau})]$ is decreasing in τ.*

(b) *$E[\Pi(\tilde{y}, \tilde{q}_{\epsilon,\tau})]$ is a concave function of ϵ and attains its maximum at $\epsilon^* = z_{CR}(\sqrt{\sigma^2+\tau^2} - \sigma)$.*

The implication of Theorem 1 for a decision maker is twofold. First, everything else equal, a higher variance in the order quantity will result in a lower expected profit. Second, for a given variance in the order quantity, a demand estimation method that leads to an order quantity that is equal to q^*_{PI} on average will result in a lower expected profit than a model that errs on the side of caution, provided that the bias is not too severe.

In some behavioral research, Schweitzer and Cachon (2000) and Rudi and Drake (2014), for example, show that the order quantity exhibits variability as well as a bias relative to the normative optimal order quantity. In particular, such literature documents the pull-to-center effect by which people tend to order, on average, quantities between the mean demand and the optimal order quantity. Moreover, people also vary their orders from period to period, despite unchanging demand conditions. These systematic deviations are attributed mainly to an anchoring bias and a preference to reduce ex post inventory error. Rudi and Drake (2014) try to disentangle the performance impact of the bias (first moment effect) from that of the variability of the order quantity (second moment effect). Theorem 1 shows that one could compensate for the variability in order quantity by increasing the bias in the less costly direction. Thus, the first and second moment effects together could either amplify the cost or partially cancel each other out depending on the direction of the bias.

Corollary 2. (a) *The following inequalities hold*: $E[\Pi(\tilde{y}, \tilde{q}_{PD})] \geq E[\Pi(\tilde{y}, \tilde{q}_{PD_0})] \geq E[\Pi(\tilde{y}, \tilde{q}_{CE})]$.

(b) *Assume a forecasting method $\mathcal{F}$ produces $\tilde{y} \mid x \sim \mathcal{D}$, where $\mathcal{D}$ is a distribution with mean and location parameter $\bar{x}$ and standard deviation $a\sigma + b$, with $a \in R^+$ and $b \in R$. Then, $E[\Pi(\tilde{y}, \tilde{q}_{PD})] \geq E[\Pi(\tilde{y}, \tilde{q}_{\mathcal{F}})]$ for any such $\mathcal{F}$.*

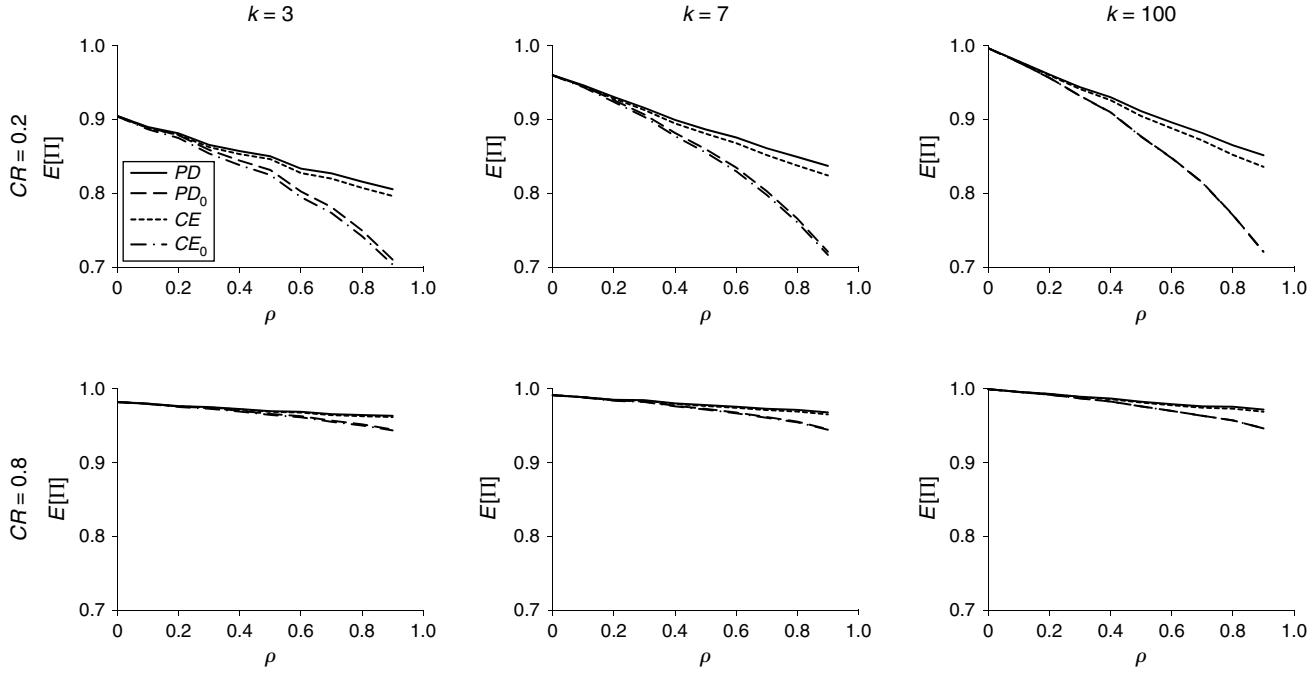
Corollary 2 shows an immediate implication of Theorem 1, that the expected profit is the highest for PD model, and no other model that produces a forecasting distribution with a mean and location parameter equal to the experts' mean forecast and a standard deviation equal to some linear transformation of σ will perform better than the PD model in expectation. This result also makes it easier to compare our method with the one in Fisher and Raman (1996), for example. Recall that in FRH, $\tilde{y} \mid x \sim \mathrm{N}(\bar{x}, (as)^2)$. Corollary 2 shows that such a method cannot do better than PD. Next, we show with simulations that these results also hold for a normal model with unknown μ and σ^2.

4.2.2. Unknown μ and σ^2.

We generate expected profit under the four methods, using the same set of simulation parameters as in Section 4.1. Figure 3 shows the expected profit, rescaled with $E[\Pi(\tilde{y}, q^*_{PI})] = 1$, under the four methods for various levels of the parameters. In all cases, the expected profit is highest under PD, followed by CE, then PD_0, and the expected profit is lowest under CE_0. Given k and $CR \leq 0.5$, the expected profits under PD and CE (methods that account for ρ) are close for low values of ρ but get higher under PD for moderate to large values of ρ. Similarly, the expected profits under PD_0 and CE_0 (methods that ignore ρ) are very close but less than those under PD and CE, and this difference gets larger at higher values of ρ. All that a higher k does is bring PD_0 and CE_0 much closer together. This overall pattern is less pronounced for $CR \geq 0.5$, where expected profits tend to be much closer under the four methods, but still highest under PD.

To summarize, we show that our model in comparison with the other three methods yields the lowest (highest) order quantity of all of the models under consideration for low (high) CRs. In fact, our model is biased in the direction of the less costly decision—i.e.,

Figure 3. Expected Profit ($E[\Pi]$) Under the Four Methods (PD, PD_0, CE, and CE_0) in the Normal Model, Rescaled with $E[\Pi(q^*_{PI})] = 1$, Given $CV = 0.2$, for Selected Values of CR and k



underestimation of demand with low CR and overestimation of demand with high CR. Further, the expected profit is the highest under our model. For low CRs, this increase in expected profit can exceed 20% compared to the other methods discussed. For high CRs, the impact is more muted, and the increase in expected profit under our model compared to the other methods is generally less than 5%. Ignoring ρ appears to be a more costly mistake than ignoring parameter uncertainty, with expected profit being higher under CE than under PD_0. It is worth noting that this is the impact on gross profit, whereas the benefits in terms of net profit can be many-fold higher. Fisher and Raman (1996), for example, state that in some apparel industries, the average net profit is about 3% of sales. In their example, a mere 1.8% increase in gross profit translates into a 60% increase in net profit.

5. Summary and Discussion

We develop a parsimonious and practical approach for generating a predictive distribution for a variable of interest based on point forecasts provided by experts. Our approach allows for the point forecasts to be correlated and admits uncertainty on the distribution parameters given the forecasts. In keeping with the extensive empirical findings on combining forecasts, we use an equal-weights model for experts—i.e., all experts are treated equally in terms of their accuracy and pairwise correlations. The resulting predictive distribution shows that ignoring either parameter uncertainty or dependence can lead to much spurious accuracy in terms of an unrealistically narrow distribution. Further, we provide a rationale for an augmentation factor that is needed to equate the standard deviation of the predictive distribution to the observed dispersion of the point forecasts, as a simple scalar factor. This augmentation factor, which is always greater than 1, depends on the correlation and on the number of forecasts. Given a number of forecasts, a higher correlation leads to a higher augmentation factor, indicating greater predictive uncertainty. On the other hand, given a correlation, a higher number of forecasts reduces the augmentation factor, indicating lower predictive uncertainty. However, loss of information due to dependence cannot be overcome by simply increasing the number of forecasts, even to an extreme. For example, given $\rho = 0.8$, even with 100 point forecasts, the augmentation factor remains at about 3.

Our model assumes all pairwise correlations to be equal and known. However, we show that our model is fairly robust to uncertainty about the correlation and to heterogeneity in the pairwise correlations. In any case, the mean correlation in case of uncertainty or heterogeneity provides at the very least in our model a lower bound on the predictive uncertainty. Another advantage of our approach is that it can be used even in the absence of any past data. Although in that case, a subjective assessment of the correlation is necessary. We discuss some of the existing approaches for such a task. Averaging estimates of correlation across experts and

methods appears to be the most promising. Clearly, simply ignoring dependence is not a good option.

We further extend our model under normality to the case where the variable of interest and the point forecasts have a lognormal distribution. The results under normality only get exacerbated in the lognormal case—i.e., ignoring dependence or parameter uncertainty leads to even a more unrealistically narrower predictive distribution.

Finally, we illustrate the implications in a decision-making context of a newsvendor setting. Our model, when compared to other methods that ignore dependence or parameter uncertainty, leads to an order quantity that on average is smaller (larger) for $CR \leq 0.5$ ($CR \geq 0.5$) but has higher variance. However, our model leads to the highest expected profit. In fact, we show that a method that ignores dependence and parameter uncertainty leads to an order quantity that is on average equal to the optimal quantity under perfect information on the distribution parameters but yields lower expected profit compared to our model. This is because our model errs on the side of caution, whereas a method that ignores dependence and parameter uncertainty retains much spurious accuracy (in terms of a distribution that is tighter than it should be) and hence a greater chance of a costly mistake (in terms of the order quantity).

There remain a number of directions for possible further research. As mentioned earlier in Section 3.3, further development of robust intuitive methods for subjective assessment of dependence among forecasters is a worthwhile direction. This could be useful also in some other domains, such as exploiting the correlation structure in hierarchical forecasting of the type discussed in Kremer et al. (2015). Another direction is the role of dependence when experts might be asked for multiple estimates rather than just a point estimate. While assessing an entire probability distribution for a variable of interest can be a difficult cognitive task, experts might be asked to provide three-point distributions (Keefer and Bodily 1983), quantile or cumulative probability judgments (Wallsten et al. 2016), or prediction intervals (Moder et al. 1995). There are a number of issues to consider in such cases. Since early papers by Alpert and Raiffa (1982) and Lichtenstein et al. (1982), overconfidence in such subjective forecasts is one of the most documented findings. In addition, one might expect that the multiple estimates from an expert might be correlated. And, then there is the issue of dependence among the experts. Jose et al. (2013) and Lichtendahl et al. (2013b) look at trimming and averaging quantiles to reduce overconfidence. Bansal et al. (2017) and Bansal and Palley (2017) consider an optimization approach to estimate the mean and standard deviation from noisy quantile judgments of experts. However, in these studies, the role of dependence among experts is not clear. Gaba et al. (2017) explore a variety of heuristics for combining prediction intervals from experts. Using simulations and data sets with interval forecasts made by professionals in their domain of expertise, they find that the relative performance of heuristics is influenced by the degree of overconfidence in and dependence among the individual forecasts. Dependence impacts the location and hence the calibration of the combined interval, leading to underestimation of uncertainty when a heuristic does not account for the dependence, which is consistent with the results in this paper. From a prescriptive point of view, it is useful to keep in mind that decision makers may be less inclined to use typically more complex methods associated with probability-based forecasts relative to point forecasts because of the more detailed modeling and data collection involved for estimation of model parameters. On the other hand, it appears to us a worthwhile attempt to explore whether multiple probability-based judgments, despite issues of overconfidence and dependence, provide a more reliable assessment of the uncertainty compared to point forecasts.

Acknowledgments

The authors are grateful to Vishal Gaur, the associate editor, and three referees for very constructive comments. They also thank Steve Chick, Kevin McCardle, Serguei Netessine, Ilia Tsetlin, and Bob Winkler for helpful suggestions.

Appendix A. Proofs

Proof of Proposition 1. Completing the squares on μ in (4) and verifying that $e'\boldsymbol{\Sigma}_\rho^{-1}e = k^*$ with $k^* = k/[1 + (k-1)\rho]$, $(e'\boldsymbol{\Sigma}_\rho^{-1}\mathbf{x})/(e'\boldsymbol{\Sigma}_\rho^{-1}e) = \bar{x}$ with $\bar{x} = (1/k)\sum_{i=1}^k x_i$, and $(\mathbf{x}'\boldsymbol{\Sigma}_\rho^{-1}\mathbf{x})/(e'\boldsymbol{\Sigma}_\rho^{-1}e) - ((e'\boldsymbol{\Sigma}_\rho^{-1}\mathbf{x})/(e'\boldsymbol{\Sigma}_\rho^{-1}e))^2 = (k-1)(s^*)^2$ with $(s^*)^2 = s^2/(1-\rho)$ and $s^2 = \sum_{i=1}^k (x_i - \bar{x})^2/(k-1)$ yields the result. □

Proof of Proposition 2. Using (4) and (5), and based on the identity,

$$n_\mu(\mu-\mu_0)^2 + k^*(\mu-\bar{x})^2 = (n_\mu + k^*)(\mu-\mu^*)^2 + \frac{n_\mu k^*(\bar{x}-\mu_0)^2}{n_\mu + k^*},$$

where $\mu^* = (n_\mu\mu_0 + k^*\bar{x})/(n_\mu + k^*)$, we obtain the joint posterior distribution,

$$\begin{aligned} f(\mu,\lambda\,|\,\mathbf{x}) &\propto \lambda^{1/2}\exp\left(-\frac{\lambda}{2}(n_\mu+k^*)(\mu-\mu^*)^2\right)\lambda^{n_v/2+k/2-1} \\ &\quad\cdot\exp\left(-\left(\frac{n_v v_0}{2} + \frac{(k-1)(s^*)^2}{2} + \frac{n_\mu k^*(\bar{x}-\mu_0)^2}{2(n_\mu+k^*)}\right)\lambda\right) \\ &\propto \mathrm{NG}(\mu,\lambda\,|\,\mu^*, n_\mu^*, v^*, n_v^*), \end{aligned}$$

where $n_\mu^* = n_\mu + k^*$, $n_v^* = n_v + k$ and $v^* = (1/n_v^*)(n_v v_0 + (k-1)\cdot(s^*)^2 + (n_\mu k^*/(n_\mu + k^*))(\bar{x}-\mu_0)^2)$. □

Proof of Proposition 3. It is easy to see that $\mathrm{IW}[PD] > (=)\ \mathrm{IW}[PD_0]$ and $\mathrm{IW}[CE] > (=)\mathrm{IW}[CE_0]$ for $\rho > (=)0$. Next, we compare $\mathrm{IW}[PD]$ to $\mathrm{IW}[CE]$ and $\mathrm{IW}[PD_0]$ to $\mathrm{IW}[CE_0]$.

Let $x_k(u)$ be the solution of the equation $T_k(x) = \Phi(u)$, where T_k and Φ are distribution functions of the standard t with k degrees of freedom and the standard normal distributions, respectively. A simple lower bound on $x_k(u)/u$ can be obtained from the Cornish–Fisher expansion of $x_k(u)$ (see, e.g., Fujikoshi and Mukaihata 1993):

$$x_k(u) = u + \frac{1}{4k}(u^3+u) + \frac{1}{96k^2}(5u^5+16u^3+3u) + \cdots$$

For all $u > 0$ and $k > 0$, we have $x_k(u)/u \geq 1 + 1/(4k)$. Similarly, for $u < 0$ and $k > 0$, $x_k(u)/u \leq 1 + 1/(4k)$. But, for $k > 2$ we have $(1 + 1/(4k))\sqrt{((k-1)/k)((1+\rho)/(1-\rho)+1/k)} > \sqrt{1/(1-\rho)}$. It then follows that $\mathrm{IW}[PD] > \mathrm{IW}[CE]$ and $\mathrm{IW}[PD_0] > \mathrm{IW}[CE_0]$ for $\rho \geq 0$. □

Proof of Proposition 4. (a) This follows directly from Proposition 3.

(b) We make use of the result that $\bar{x}$ and s^2 are independently distributed when $\mathbf{x}$ follows a multivariate normal distribution with identical pairwise correlations ρ—i.e., $\mathrm{Cov}[\bar{x}, s] = 0$ (Rao 1973, pp. 196–197). Hence,

$$\mathrm{Var}[\tilde{q}_{PD}] = \mathrm{Var}[\bar{x}] + t^2_{CR,k}\frac{k-1}{k}\left(\frac{1+\rho}{1-\rho}+\frac{1}{k}\right)\mathrm{Var}[s],$$

$$\mathrm{Var}[\tilde{q}_{PD_0}] = \mathrm{Var}[\bar{x}] + t^2_{CR,k}\frac{k-1}{k}\left(1+\frac{1}{k}\right)\mathrm{Var}[s],$$

$$\mathrm{Var}[\tilde{q}_{CE}] = \mathrm{Var}[\bar{x}] + z^2_{CR}\frac{1}{1-\rho}\mathrm{Var}[s],$$

$$\mathrm{Var}[\tilde{q}_{CE_0}] = \mathrm{Var}[\bar{x}] + z^2_{CR}\mathrm{Var}[s].$$

Using the same reasoning as in Part (a), the inequalities follow. □

Proof of Theorem 1. (a) Given an order quantity, the expected profit is given by $E[\Pi(\tilde{y} \mid \tilde{q}_{\epsilon,\tau} = q] = (p-v)\mu - (c-v)q - \sigma L(z)(p-v)$, where $z = (q-\mu)/\sigma$ and $L(z) = \int_z^\infty (x-z)\,\mathrm{d}\Phi(x)$. Then,

$$E[\Pi(\tilde{y},\tilde{q}_{\epsilon,\tau})] = (p-v)\mu - (c-v)(q^*_{PI}+\epsilon) - \sigma(p-v) \cdot \int_{-\infty}^{\infty} L(z)\frac{1}{\tau\sqrt{2\pi}}e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}\,\mathrm{d}q. \quad (A.1)$$

$$\frac{\partial E[\Pi(\tilde{y},\tilde{q}_{\epsilon,\tau})]}{\partial\tau} = \frac{\sigma}{\tau}(p-v)\int_{-\infty}^{\infty} L(z)\frac{e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}}{\tau\sqrt{2\pi}}\,dq - \sigma(p-v)\int_{-\infty}^{\infty} L(z)\frac{(q-q^*_{PI}-\epsilon)^2}{\tau^3} \cdot \frac{e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}}{\tau\sqrt{2\pi}}\,\mathrm{d}q. \quad (A.2)$$

Integrating by parts twice the second term in (A.2), we get:

$$\frac{\partial E[\Pi(\tilde{y},\tilde{q}_{\epsilon,\tau})]}{\partial\tau} = -\tau\sigma(p-v)\int_{-\infty}^{\infty}\frac{\partial^2 L(z)}{\partial q^2}\frac{1}{\tau\sqrt{2\pi}}e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}\,\mathrm{d}q.$$

But $L(z)$ is a convex decreasing function of q and the result follows.

(b) From (A.1), differentiating with respect to ϵ, we get the following first-order condition:

$$\frac{\partial E[\Pi(\tilde{y},\tilde{q}_{\epsilon,\tau})]}{\partial\epsilon} = -(c-v) - \sigma(p-v)\int_{-\infty}^{\infty} L(z)\frac{(q-q^*_{PI}-\epsilon)}{\tau^3\sqrt{2\pi}}e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}\,\mathrm{d}q = 0.$$

Integrating by parts, we get

$$\begin{aligned}\frac{\partial E[\Pi(\tilde{y},\tilde{q}_{\epsilon,\tau})]}{\partial\epsilon} &= -(c-v) - \sigma(p-v)\int_{-\infty}^{\infty}\frac{\partial L(z)}{\partial q}\frac{1}{\tau\sqrt{2\pi}}e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}\,\mathrm{d}q\\ &= -(c-v) + (p-v) - (p-v)\int_{-\infty}^{\infty}\Phi(z)\frac{1}{\tau\sqrt{2\pi}}e^{-(q-q^*_{PI}-\epsilon)^2/2\tau^2}\,\mathrm{d}q\\ &= -(c-v) + (p-v) - (p-v)P(Y\leq 0) = 0,\end{aligned}$$

where Y is a normally distributed random variable with mean $(\mu - q^*_{PI} - \epsilon)/\sigma$ and standard deviation $\sqrt{(\sigma^2+\tau^2)/\sigma^2}$. Equivalently, $P(Y \leq 0) = CR$. Then $0 = (\mu - q^*_{PI} - \epsilon^*)/\sigma + z_{CR}\sqrt{(\sigma^2+\tau^2)/\sigma^2}$, or equivalently $\epsilon^* = z_{CR}(\sqrt{\sigma^2+\tau^2} - \sigma)$. We used the fact that $q^*_{PI} = \mu + z_{CR}\sigma$. Note also that $\partial^2 E[\Pi(\tilde{y}, q(\epsilon))]/\partial\epsilon^2 \leq 0$, so ϵ^* maximizes (A.1). □

Proof of Corollary 2. (a) Conditional on μ, $\bar{x} \sim \mathrm{N}(\mu, \sigma^2/k^*)$. And, hence, conditional on μ, each of the three order quantities are normally distributed with variance σ^2/k^*. Moreover, $E[\tilde{q}_{CE} \mid \mu] = q^*_{PI}$, $E[\tilde{q}_{PD} \mid \mu] = q^*_{PI} + z_{CR}\sigma(\sqrt{1+1/k^*} - 1)$ and $E[\tilde{q}_{PD_0} \mid \mu] = q^*_{PI} + z_{CR}\sigma(\sqrt{1+1/k} - 1)$. Also, $0 \leq |z_{CR} \cdot \sigma(\sqrt{1+1/k} - 1)| \leq |\epsilon^*| = |z_{CR}\sigma(\sqrt{1+1/k^*} - 1)|$, because $k \geq k^*$. Then, from Theorem 1, it follows that $E[\Pi(\tilde{y},\tilde{q}_{PD}) \mid \mu] \geq E[\Pi(\tilde{y},\tilde{q}_{PD_0}) \mid \mu] \geq E[\Pi(\tilde{y},\tilde{q}_{CE}) \mid \mu]$. Hence, for any prior density function on μ, the result follows.

(b) If distribution $\mathscr{D}$ has location parameter and mean $\bar{x}$ and standard deviation $a\sigma + b$, then $\mathscr{D}_{\bar{x}}$ is a distribution with mean zero and standard deviation $a\sigma + b$, where $\mathscr{D}_{\bar{x}}(y) = \mathscr{D}(y - \bar{x})$ does not depend on $\bar{x}$. Then, $q_{\mathscr{F}} = \mathscr{D}^{-1}(CR) = \bar{x} + \mathscr{D}^{-1}_{\bar{x}}(CR)$. It follows that, conditional on μ, $q_{\mathscr{F}}$ is normally distributed with variance σ^2/k^*. From Theorem 1, the inequality follows. □

Appendix B. Normal Model with Heterogeneous ρ_{ij}s

We extend Section 2.1 by allowing heterogeneous ρ_{ij}s in $\boldsymbol{\Sigma}$. In this case, the covariance matrix $\boldsymbol{\Sigma}$ has diagonal elements σ^2 and off-diagonal elements $\rho_{ij}\sigma^2$, with $\rho_{ij} = \mathrm{Corr}[x_i, x_j \mid \theta]$, $i \neq j$. Assume ρ_{ij}s are known.

Setting $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Sigma}_{\rho_{ij}}$ where $\boldsymbol{\Sigma}_{\rho_{ij}}$ is a $k \times k$ is a matrix with diagonal elements 1 and off-diagonal elements ρ_{ij}. With $\lambda = 1/\sigma^2$, the likelihood function can be rewritten as

$$l(\mathbf{x} \mid \mu, \lambda) \propto \lambda^{k/2}\exp\left(-\frac{\lambda k^*}{2}(\mu - \hat{\mu})^2\right)\exp\left(-\frac{\lambda}{2}(k-1)\hat{s}^2\right), \quad (B.1)$$

where $k^* = \mathbf{e}'\boldsymbol{\Sigma}^{-1}_{\rho_{ij}}\mathbf{e}$, $\hat{\mu} = (\mathbf{e}'\boldsymbol{\Sigma}^{-1}_{\rho_{ij}}\mathbf{x})/(\mathbf{e}'\boldsymbol{\Sigma}^{-1}_{\rho_{ij}}\mathbf{e})$ and $\hat{s}^2 = ((\mathbf{x} - \hat{\mu}\mathbf{e})' \cdot \boldsymbol{\Sigma}_{\rho_{ij}}{}^{-1}(\mathbf{x} - \hat{\mu}\mathbf{e}))/(k-1)$.

Using the normal-gamma prior distribution on μ and λ in Section 2.1, the posterior distribution of μ and λ is

$$\begin{aligned}f(\mu, \lambda \mid \mathbf{x}) &= \mathrm{NG}(\mu, \lambda \mid \mu^*, n^*_\mu, v^*, n^*_v)\\ &= \mathrm{N}(\mu \mid \mu^*, (n^*_\mu\lambda)^{-1})\mathrm{Gamma}\left(\lambda \,\middle|\, \frac{n^*_v}{2}, \frac{n^*_v v^*}{2}\right), \quad (B.2)\end{aligned}$$

where $n^*_\mu = n_\mu + k^*$, $\mu^* = (n_\mu\mu_0 + k^*\hat{\mu})/n^*_\mu$, $n^*_v = n_v + k$ and $v^* = (1/n^*_v)(n_v v_0 + (k-1)\hat{s}^2 + (n_\mu k^*/(n_\mu + k^*))(\hat{\mu} - \mu_0)^2)$.

It then follows that $f(\tilde{y} \mid \mathbf{x})$ is a t distribution with degrees of freedom n^*_v, location parameter μ^*, and scale parameter

$\sqrt{(n_\mu^*+1)v^*/n_\mu^*}$, so that $E[\tilde{y}\,|\,\mathbf{x}] = \mu^*$ for $n_v^* > 1$ and, for $n_v^* > 2$, $\mathrm{Var}[\tilde{y}\,|\,\mathbf{x}] = n_v^*/(n_v^*-2)(v^* + v^*/n_\mu^*)$.

With a diffuse prior on μ and σ^2 (i.e., with $n_\mu = n_v = 0$), $f(\tilde{y}\,|\,\mathbf{x})$ is a t distribution with k degrees of freedom, location $\hat{\mu}$, and scale $\sqrt{(k^*+1)v^*/k^*}$, which yields $E(\tilde{y}\,|\,\mathbf{x}) = \hat{\mu}$ and

$$\mathrm{Var}[\tilde{y}\,|\,\mathbf{x}] = \left(1 + \frac{1}{k^*}\right)\frac{(\mathbf{x}-\hat{\mu}\mathbf{e})'\boldsymbol{\Sigma}_{\rho_{ij}}^{-1}(\mathbf{x}-\hat{\mu}\mathbf{e})}{k-2}. \tag{B.3}$$

Endnotes

[1] As an alternative, consider that $\mathbf{x}$ follows a multivariate normal with a mean vector $\boldsymbol{\mu} = (a+\mu)\mathbf{e}$, where $\mathbf{e} = (1,\ldots,1)'$, and a $k \times k$ positive definite covariance matrix $\boldsymbol{\Sigma}$ with diagonal elements $b\sigma^2$ and off-diagonal elements ρ, $b \geq 1$. One way such a model could arise is when $x_i = \tilde{y}_i + \epsilon_i$, where, $\forall\, i$, given the parameters $\theta = (\mu, \sigma^2)$ and $\theta_\epsilon = (a, \sigma_\epsilon^2)$, $\tilde{y}_i \sim \mathrm{N}(\mu, \sigma^2)$, $\epsilon_i \sim \mathrm{N}(a, \sigma_\epsilon^2)$, and $\tilde{y}_i$ and ϵ_i are conditionally independent. Then, $E(x_i\,|\,\theta, \theta_\epsilon) = \mu + a$ and $\mathrm{Var}[x_i\,|\,\theta, \theta_\epsilon] = \sigma^2 + \sigma_\epsilon^2 = b\sigma^2$, with $b \geq 1$. And, for $i, j \in \{1,\ldots,k\}, i \neq j$, $\mathrm{Corr}[x_i, x_j\,|\,\theta, \theta_\epsilon] = \mathrm{Corr}[\tilde{y}_i + \epsilon_i, \tilde{y}_j + \epsilon_j\,|\,\theta, \theta_\epsilon] = \rho_{ij} = \rho$. Here, a is the average bias in experts' forecasts, and b can be thought of as related to the forecasting ability of experts, with σ^2 as the *irreducible uncertainty* of $\tilde{y}$. As the forecasting ability improves, b approaches 1. Such a model is in the same spirit as those considered in, among others, Raju and Roy (2000) and Vives (1984). With a and b known from extensive past data, the experts forecasts can simply be adjusted for the bias a to obtain a $\boldsymbol{\mu}$ mean vector. And, $b > 1$ does not change any of the results in our paper with respect to the role of dependence or parameter uncertainty in inferences about $\tilde{y}\,|\,\mathbf{x}$.

[2] A possible way to view a is as $E[s_{fe}\,|\,\theta, \rho]/E[s\,|\,\theta, \rho]$, where $s_{fe} = \sqrt{\mathrm{Var}(\bar{x} - \tilde{y})}$ and $\theta = (\mu, \lambda)$. It can then be shown that (i) for $k > 2$, $a < \delta$ and $\mathrm{Var}[PD] > \mathrm{Var}[FRH] > \mathrm{Var}[CE]$, and (ii) as $k \to \infty$, $a \approx \delta$ and $\mathrm{Var}[PD] \approx \mathrm{Var}[FRH] > \mathrm{Var}[CE]$. With $\rho > 0$, there is a loss of information about μ that cannot be recovered even with a very large k. On the other hand, uncertainty about λ can be reduced proportionally with a higher k. Hence, in such a setup, *FRH* can be seen as a variation of the *CE* method that accounts for ρ as *CE* does, but in addition also accounts for parameter uncertainty about μ while ignoring that for λ. Therefore, with a very high k, *FRH* gets close to *PD*, whereas *CE* still underestimates predictive uncertainty.

References

Alpert M, Raiffa H (1982) A progress report on the training of probability assessors. Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, UK), 294–305.

Armstrong JS (2001) *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Springer Science and Business Media, New York).

Ashton RH (1986) Combining the judgments of experts: How many and which ones? *Organ. Behav. Human Decision Processes* 38(3):405–414.

Bansal S, Palley AB (2017) Is it better to elicit quantile or probability judgments? A comparison of direct and calibrated procedures for estimating a continuous distribution. Kelley School of Business Research Paper No. 17-44, https://ssrn.com/abstract=2981840.

Bansal S, Gutierrez GJ, Keiser JR (2017) Using experts' judgments to quantify risks: Theory and application to agribusiness. *Oper. Res.* 65(5):1115–1130.

Bernardo JM, Smith AF (1994) *Bayesian Theory* (John Wiley & Sons, New York).

Bimpikis K, Markakis MG (2016) Inventory pooling under heavy-tailed demand. *Management Sci.* 62(6):1800–1813.

Bolger F, Rowe G (2015) The aggregation of expert judgment: Do good things come to those who weight? *Risk Anal.* 35(1):5–11.

Brown RV, Lindley DV (1986) Plural analysis: Multiple approaches to quantitative research. *Theory and Decision* 20(2):133–154.

Bunn DW (1985) Statistical efficiency in the linear combination of forecasts. *Internat. J. Forecasting* 1(2):151–163.

Chevalier J, Goolsbee A (2003) Measuring prices and price competition online: Amazon.com and barnesandnoble.com. *Quant. Marketing Econom.* 1(2):203–222.

Chhibber S, Apostolakis G (1993) Some approximations useful to the use of dependent information sources. *Reliability Engrg. System Safety* 42(1):67–86.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting* 5(4):559–583.

Clemen RT, Winkler RL (1985) Limits for the precision and value of information from dependent sources. *Oper. Res.* 33(2):427–442.

Clemen RT, Fischer GW, Winkler RL (2000) Assessing dependence: Some experimental results. *Management Sci.* 46(8):1100–1115.

Fisher M, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* 44(1):87–99.

Fisher M, Raman A (2010) *The New Science of Retailing: How Analytics Are Transforming the Supply Chain and Improving Performance* (Harvard Business Press, Boston).

Fujikoshi Y, Mukaihata S (1993) Approximations for the quantiles of Student's t and f distributions and their error bounds. *Hiroshima Math. J.* 23(3):557–564.

Gaba A, Tsetlin I, Winkler RL (2017) Combining interval forecasts. *Decision Anal.* 14(1):1–20.

Gaffeo E, Scorcu AE, Vici L (2008) Demand distribution dynamics in creative industries: The market for books in Italy. *Inform. Econom. Policy* 20(3):257–268.

Garthwaite PH, Kadane JB, O'Hagan A (2005) Statistical methods for eliciting probability distributions. *J. Amer. Statist. Assoc.* 100(470):680–701.

Gaur V, Kesavan S, Raman A, Fisher M (2007) Estimating demand uncertainty using judgmental forecasts. *Manufacturing Service Oper. Management* 9(4):480–491.

Geisser S (1965) A Bayes approach for combining correlated estimates. *J. Amer. Statist. Assoc.* 60(310):602–607.

Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* 69(2):243–268.

Gokhale DV, Press SJ (1982) Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *J. Roy. Statist. Soc. Ser. A General* 145(2):237–249.

Hammond JH, Raman A (1994) *Sport Obermeyer, Ltd.* (Harvard Business School, Boston), HBS Case 695-022.

Hoff PD (2009) *A First Course in Bayesian Statistical Methods* (Springer Science and Business Media, New York).

Johnson NL, Kotz S, Balakrishnan N (2002) *Continuous Multivariate Distributions, Volume 1, Models and Applications* (John Wiley & Sons, New York).

Jose VRR, Grushka-Cockayne Y, LLichtendahl KC Jr (2013) Trimmed opinion pools and the crowd's calibration problem. *Management Sci.* 60(2):463–475.

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

Kremer M, Siemsen E, Thomas DJ (2015) The sum and its parts: Judgmental hierarchical forecasting. *Management Sci.* 62(9): 2745–2764.

Kunda Z, Nisbett RE (1986) The psychometrics of everyday life. *Cognitive Psych.* 18(2):195–224.

Lichtendahl KC Jr, Grushka-Cockayne Y, Pfeifer PE (2013a) The wisdom of competitive crowds. *Oper. Res.* 61(6):1383–1398.

Lichtendahl KC Jr, Grushka-Cockayne Y, Winkler RL (2013b) Is it better to average probabilities or quantiles? *Management Sci.* 59(7):1594–1611.

Lichtenstein S, Fischhoff B, Phillips LD (1982) Calibration of probabilities: The state of the art. Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, UK), 306–334.

Lindley DV (1986) The reconciliation of decision analyses. *Oper. Res.* 34(2):289–295.

Meyer MA, Booker JM (2001) *Eliciting and Analyzing Expert Judgment: A Practical Guide* (Society for Industrial and Applied Mathematics, Philadelphia).

Moder JJ, Phillips CR, Davis EW (1995) *Project Management With CPM, PERT and Precedence Diagramming* (Blitz Publishing Company, Middleton, WI).

Palley AB, Soll JB (2016) Eliciting and aggregating forecast when information is shared. Working paper, Duke University, Durham, NC.

Poirier DJ (1995) *Intermediate Statistics and Econometrics: A Comparative Approach* (MIT Press, Cambridge, MA).

Raju JS, Roy A (2000) Market information and firm performance. *Management Sci.* 46(8):1075–1084.

Rao CR (1973) *Linear Statistical Inference and its Applications* (John Wiley & Sons, New York).

Ren Y, Croson R (2013) Overconfidence in newsvendor orders: An experimental study. *Management Sci.* 59(11):2502–2517.

Ren Y, Croson D, Croson R (2017) The overconfident newsvendor. *J. Oper. Res. Soc.* 68(5):496–506.

Rudi N, Drake D (2014) Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Sci.* 60(5):1334–1345.

Schmittlein DC, Kim J, Morrison DG (1990) Combining forecasts: Operational adjustments to theoretically optimal rules. *Management Sci.* 36(9):1044–1056.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Seifert M, Siemsen E, Hadida AL, Eisingerich AB (2015) Effective judgmental forecasting in the context of fashion products. *J. Oper. Management* 36:33–45.

Vives X (1984) Duopoly information equilibrium: Cournot and Bertrand. *J. Econom. Theory* 34(1):71–94.

Wallsten TS, Shlomi Y, Nataf C, Tomlinson T (2016) Efficiently encoding and modeling subjective probability distributions for quantitative variables. *Decision* 3(3):169–189.

Winkler RL (1981) Combining probability distributions from dependent information sources. *Management Sci.* 27(4):479–488.

Winkler RL, Clemen RT (1992) Sensitivity of weights in combining forecasts. *Oper. Res.* 40(3):609–614.

Winkler RL, Clemen RT (2004) Multiple experts vs. multiple methods: Combining correlation assessments. *Decision Anal.* 1(3):167–176.