Data-informed calibration of experts in Bayesian framework

C. H. Shirazi & A. Mosleh
Center for Risk and Reliability, 0151-F Martin Hall, University of Maryland, College Park, MD, 21044, USA

ABSTRACT: Expert judgment is a major source of estimating the parameters of risk and reliability models. However, expert opinion is subjective and contains some degree of uncertainty, often expressed as error. The negative impact of expert's error on vital tasks forces the decision makers to assess the quality of elicited opinions and perhaps make some adjustments to estimates with the hope of improving their accuracy. The primary focus of this paper is to reduce the inaccuracy of expert's point estimates using formulated likelihood functions from relative errors of experts' judgment. The results of the study reveal an overall improvement in the accuracy of estimates, applying likelihood distributions in the Bayesian framework for homogenous and nonhomogenous empirical data.

1 INTRODUCTION

1.1 *Expert judgment*

Historically, decision-makers have used expert opinion to supplement insufficient data. Relatively cheap and virtually inexhaustible (Cook, 1991), experts greatly influence decisions in key subject matters such as political, financial, legal, and social issues. Additionally, a major source of information in estimating the parameters of risk and reliability models is expert judgment. Cases involving new process or product design, very rare events, and proceedings that are beyond our direct experience, call for the use of expert opinion as a surrogate information source. The evaluation of expert judgment quality starts with a clear definition of 'expert(s)'. For the purpose of this paper, an 'expert' is broadly defined as a field professional that will assist the analysts or decision makers to determine the unknown quantity of interest. Besides expressing their subjective judgments directly, experts can use prototypes, models, simulations, destructive and nondestructive tests (among other tools) to gather information, acquire data, gain pertinent practical knowledge, and to carry out a specified set of tasks proficiently. The association of expert's attributes to the quality of expert judgments is acknowledged in most of the studies. Attributes are the characteristics and qualities relating to an individual. Specific attributes are used to differentiate between experts and novices. However, the identification and selection of experts is subjective; a unique perception of the qualifications is reported in different studies to name an individual the 'expert'. According to Booker and Meyer (1996), expert opinion is used in two ways:

[1] Structuring of technical problems including the determination of relevant information for analyses such as key input and output variables as well as proper assumptions and evaluation techniques.

[2] Direct qualitatively or quantitatively estimates of the unknown of interest, characterize uncertainty, and determine weighting factors for combining data sources

This study focuses on the use of experts' quantitative estimate of the unknown. The issues surrounding the use of expert opinion fall into two broad categories:

a. Elicitation of opinions: selection of experts, number of experts, elicitation process, etc. and,
b. Use of elicited opinions: how to use and aggregate data provided by the expert.

This paper considers only the use of elicited experts' opinions reported as point estimates. An opinion is not a fact or verified by an experiment, but a person's assessment of a subject or judgment towards something, and therefore, contains some degree of uncertainty often expressed as error. The potential negative impact of expert error on vital tasks forces the decision makers to take into consideration the quality level of the judgment.

The poor quality of expert judgment can be broadly classified as those shortcomings associated with attributes defining an expert, the estimation procedure, the elicitation process (formal vs. informal), technical calibration and aggregation (performance measures of experts and expertise), and available information about the unknown. The issue being considered in this research is the feasibility and value of empirically-based calibration of experts within the Bayesian formalism. The main problem in applying the Bayesian technique is the complications associated with the development of a suitable likelihood function. This paper presents the development and use of likelihood functions based on relative errors of experts' estimates. Overall, an investigation of several practical questions about experts' opinion in the Bayesian framework is conducted:

[1] Empirical assessment of expert error in different disciplines

[2] To uncover whether the use of formulated likelihood functions would reduce future prediction errors for homogenous and nonhomogenous data

1.2 *Bayesian formalism*

Conceptually, the application of the Bayesian method to utilize the expert opinion is simple. The expert's estimate is treated as a piece of evidence about the unknown quantity of interest. This estimation is used to update the analyst's own (prior) knowledge through Bayes' theorem. Prior distributions are used to describe the uncertainty surrounding the unknown. After observing the data (in this case, the expert opinion), the posterior distribution provides a coherent post data summary of the remaining uncertainties. Mathematically speaking:

$$\pi(u|u') = \frac{L(u'|u)\pi_o(u)}{\int L(u'|u)\pi_o(u)du} \qquad (1)$$

Here, u is the unknown quantity of interest, u' is the set of the experts' opinions, $\pi(u|u')$ is posterior distribution expressing the analyst's updated knowledge about the unknown, $L(u'|u)$ is the likelihood of the estimate given the true value of the unknown quantity, and $\pi_0(u)$ is the analyst's prior knowledge about the unknown (prior to obtaining the opinion of the experts). The first formal frame of the Bayesian method for use of expert opinion was presented by Morris (1974, 1977). His work fully establishes the foundations for the Bayesian paradigm in the analysis of expert judgment. Building on Morris's method, Mosleh and Apostolakis (1986) proposed the use of 'Additive' and 'Multiplicative' error models for constructing the likelihood functions, expressing the experts' assessments as the sum (or ratio) of the true value of unknown quantity and an error term. Still, the main problem in applying the Bayesian technique remains as complications associated with the development of a suitable likelihood distribution. This function is a probabilistic model for data and must capture the interrelationships among estimates and the quantity of interest. Particularly, it must account for the bias of the individual estimates and be able to model dependencies among experts.

1.3 Collection and Characterization of Data

Data refers to a collection of organized information, usually the result of experience, observation or experiment. Data can be in the form of nominal, ordinal, interval and ratio, depending upon objective of data collection and analysis. Generally in the assessment of uncertainty, subjective data can come in the form of expert's estimate, historical knowledge of the unknown as well as evidence on the credibility, applicability and relevance of the judgment. Experts provide qualitative information or quantitative estimates in form of probability distribution, point estimate, range or limitation, statement or partial evidence of true values. The search for the accuracy of expert opinion began with a general survey of the literature, publications, books and referred sources. The wide literature search included databases such as Econpapers, Elsvier, IEEE Digital Library, and UMCP Library. To conduct this study, over 1500 publications analyzing or reporting the use of expert opinion since 1930s are examined in a three-year time frame, accumulating over 1900 experts' point estimates in 27 different disciplines. In addition, TU Delft University Excalibur Database is used, which reports the assessment of over 800 experts on more than 4000 variables, representing about 80,000 elicited questions. These reported estimates were made in various areas, such as nuclear applications, the chemical and gas industries, toxicity of chemicals, external effects (pollution, waste disposal sites, inundation, volcano eruptions), aerospace sector and aviation sector, the occupational sector, the health sector, and the banking sector. Many discarded sources contained no report of actual data such as expert estimates, true values, or an indication of formal elicitation process. The collected data were either the estimates directly provided by experts or the information supplied by forecasting models, assuming expert's contribution and classified into homogenous and nonhomogenous groups. In a homogenous data set, there are a series of expert's estimates for a true value. An example is a study conducted by National Human Exposure Assessment Survey (NHEXAS) using the estimates of seven experts to obtain exposure assessment in residential ambient, residential indoor and personal air benzene concentrations in U.S. EPA's Region V, experienced by the nonsmoking, non-occupationally exposed population. These experts were selected by a peer nomination process. Individually elicited judgments were gathered from the experts during a 2-day workshop. In a nonhomogenous case, there is an expert's estimate for a true value. An example is found in a weather precipitation research study among expert meteorologists at UMCP. The study involved four experts who were asked to make 48-hour precipitation forecasts projections. In the field of meteorology, a 48-hour forecast of precipitation is considered moderately difficult, and requires specialized skills. The forecast were conducted on three different days for cities of Orlando, Seattle, San Francisco, New Orleans and Detroit. In utilized case studies, there is a wide range of reasons explaining the errors of estimates. This list includes, but not limited to, study topic or subject matter, expert's characteristics, career affiliation, academic degree, field of expertise, or years of experience. Hoffmann, et al. (2006) show that variability in best estimates differs by professional background and discipline. Respondents who identify government as their primary career setting have tighter ranges than those whose careers have been primarily in academia, industry, or multiple sectors. Those with significant career experience in multiple sectors have the largest ranges, followed by those in industry and academia. This study also reports those with master's degrees have the least confidence in their best estimates, and veterinarians have the most. The field of expertise is also reported as a significant matter, in such as way that relative to microbiology (bacteriology, food science microbiology, microbiology, and area pathology), those who identified their field as "public health" (public health, public health epidemiology, and epidemiology) have ranges that are larger than the microbiology group and veterinary medicine had ranges that are smaller than the microbiology group. For the forecasts obtained by model, in addition to model inputs and assumptions, there is a series of reasons listed to explain the error of forecasts such as model types, forecast period and projection horizon, forecast accuracy measures used, additional information that becomes available later, the size of the error, seasonal and geographical errors, etc. Although the impact and consequences of factors affecting the estimates such as experts' expertise, calibration, dependencies, reliability, gender, or ethics were duly noted, variations in accumulated data and heterogeneity among experts were accepted as an inherent variability and a means of quantifying and comparing the uncertainties about parameters of interest.

1.4 *Selection of forecast accuracy measure*

According to Armstrong and Fildes (1995), the objective of a forecast accuracy measure is to provide an informative and clear understanding of the error distribution. Theoretically, when the forecast errors are randomly structured, the form of the forecasts is independent of the selected accuracy measure. Otherwise, it is generally accepted that there is no single best accuracy measure, and deciding on the assessment method is essentially subjective. In this study, a simple form of relative error (*E*) is selected as the forecast accuracy measure, since it offers a number of desirable properties:

$$E = u'/u \quad , u>0 \qquad (2)$$

u: the quantity of interest, u': experts' opinions

[1] Scale-Independence: Relative error is independent of scale or level of series and gives an indication of how good a measurement is relative to the size of the thing being measured.

[2] Outliner-Independence: Absolute performance measures may produce very big numbers due to outliers, which can make the comparison of different estimates not really feasible. In contrast, relative accuracy measures eliminate the bias introduced by possible trends, seasonal components and outliers.

a. series outliner independence: measure is not affected by large errors associated with outliner observation

b. error outliner independence: measure is not affected by large errors associated with outliner errors

[3] Sensitivity: measure is responsive to small changes in error

[4] Stability (Reliability): measure is repeatable;

[5] Correlation Validity (Interpretability): measure is related to specific issue of decision-making

[6] Simplicity: measure is easy to use.

[7] Typicality: measure is representative of its underlying distribution. It has been shown by Chen and Yang (2004) that Mean Square Error (MSE) is the optimal selection when the errors are normally distributed.

2 CONSTRUCTION OF THE LIKELIHOOD AND POSTERIOR

2.1 *Homogenous Data*

In the case of homogenous data type, the available information regarding the quantity of interest (u) is comprised of experts' estimates ($u'_1...u'_n$) and evidence or relative error of estimates ($E_1...E_n$). The error distribution can be characterized in terms of finite set of parameters, i.e., if ln$E$ is normally distributed, E is a lognormal distribution:

$\underline{\theta} = (E_{50}, \sigma_E)$

E_{50} : is the mean of the error distribution,

σ_E : is the standard deviation of the distribution

$$f(E) = \frac{1}{\sqrt{2\pi}\sigma_E E} e^{-\frac{1}{2}\left(\frac{\ln E - \ln E_{50}}{\sigma_E}\right)^2} \quad (3)$$

The error distribution ($E_1...E_n$) represents the likelihood function of errors given the parameters:

$$L(E_1...E_n | E_{50}, \sigma_E) = \frac{1}{\sqrt{2\pi}\sigma_E E} e^{-\frac{1}{2}\left(\frac{\ln E - \ln E_{50}}{\sigma_E}\right)^2} \quad (4)$$

The posterior distribution of the set of likelihood parameters is:

$$\pi(E_{50}, \sigma_E | E_1...E_n) = \frac{L(E_1...E_n | E_{50}, \sigma_E)\pi_0(E_{50}, \sigma_E)}{\int_{\mu_E}\int_{\sigma_E} L(E_1...E_n | E_{50}, \sigma_E)\pi_0(E_{50}, \sigma_E) dE_{50} d\sigma_E} \quad (5)$$

In constructing the likelihood function in terms of relative errors, the relation between the distribution of relative errors, $f(E)$, and the distribution of estimates, $f(u')$, must be established:

$$L(u'|u) = \frac{1}{u} L\left(\frac{u'}{u}|u\right) = \frac{1}{u} L(E|u) \quad (6)$$

By substituting in Equation 4:

$$L(u'|u, \underline{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_E u'} e^{-\frac{1}{2}\left(\frac{\ln u' - \ln u - \ln E_{50}}{\sigma_E}\right)^2} \quad (7)$$

By de-conditioning this distribution from E_{50} and σ_E, the likelihood function for new estimates can be obtained (this is likelihood averaging process, another approach can also be made by posterior averaging):

$$L(u'|u) = \int_{E_{50}}\int_{\sigma_E} L(u'|u, \underline{\theta})\pi(E_{50}, \sigma_E | E_1...E_n) dE_{50} d\sigma_E \quad (8)$$

The new expert's estimates can now be updated using general Bayes' model. The mean of this posterior (μ), as the distribution marker, is compared with the true value (μ/u), in order to determine if and how much the formulated likelihood function has been able to reduced the error of estimates.

2.2 *Nonhomogenous Data*

In case of nonhomogenous data set, there is one expert estimate for each true value. The available information regarding true values of ($u_1...u_n$) is comprised of experts' estimates ($u'_1...u'_n$) and evidence or relative error of estimates ($E_1...E_n$). The error distribution can be marginalized in terms of a finite set of parameters ($\underline{\theta}$), which by itself is a variable characterized by a population variability distribution of g($\underline{\theta}$). This distribution is symbolized by a 'hyper-parameter' ($\underline{\omega}$):

$$\underline{\omega} = (\omega_1...\omega_n) \Rightarrow g(\underline{\theta}) = g(\underline{\theta}|\underline{\omega}) \quad (9)$$

$$L(E|\underline{\omega}) = \int_{\underline{\theta}} L(E|\underline{\theta})g(\underline{\theta}|\underline{\omega})d\underline{\theta} \quad (10)$$

$$\pi(\underline{\omega}|E) = \frac{\left(\int_{\underline{\theta}} L(E|\underline{\theta})g(\underline{\theta}|\underline{\omega})d\underline{\theta}\right)\pi_0(\underline{\omega})}{\int_{\underline{\omega}}\left(\int_{\underline{\theta}} L(E|\underline{\theta})g(\underline{\theta}|\underline{\omega})d\underline{\theta}\right)\pi_0(\underline{\omega})d\underline{\omega}} \quad (11)$$

The posterior expected distribution is:

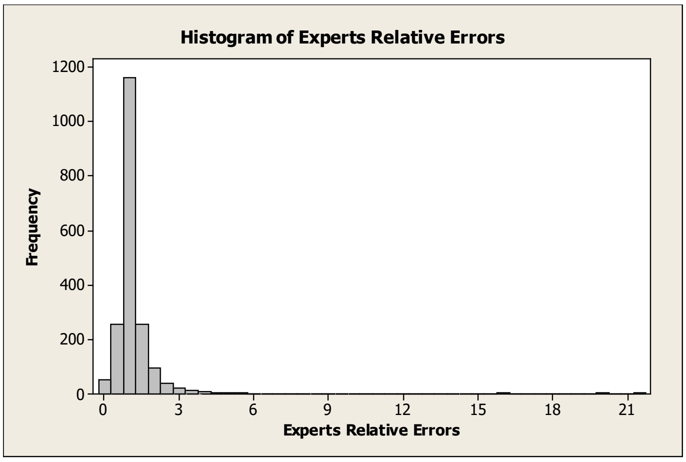
$$\bar{g}(\underline{\theta}|E) = \int_{\underline{\omega}} g(\underline{\theta}|\underline{\omega})\pi(\underline{\omega}|E)d\underline{\omega} \quad (12)$$

$$\pi(u|u') = \frac{\int_{\underline{\theta}} \bar{g}(\underline{\theta}|E)L(u'|u, \underline{\theta})\pi_0(u)}{\int_u\left(\int_{\underline{\theta}} \bar{g}(\underline{\theta}|E)L(u'|u, \underline{\theta})\right)\pi_0(u)du} \quad (13)$$

The mean of this posterior, (μ), as the distribution marker, is compared with the true value, (μ/u_n), in order to determine if and how much the formulated likelihood function has been able to reduced the error of estimates.

2.3 *Empirical Assessment of Experts' Errors in Different Disciplines*

The histogram of experts' relative errors shows that over 45% of relative errors are equal or close to one (expert estimate ~ true value), about 45% of data points are falling between (1 – 2) and 5% falling in the range of (2 – 3). The average relative error is 1.2 and only 5% among all empirical relative errors data are greater than 3. Table 1 shows the best fitted probability distributions for relative errors of experts' estimates. Considering the producer risk of 5% (α=0.05), lognormal is found to be among the top fits. The distribution fitting tests point to Wakeby and Cauchy distribution as the first best fits. This seems logical since, i.e. Cauchy is a ratio distribution (E is the ratio distribution constructed given two stochastic variables u' and u). Additionally, a ratio or proportion distribution is often heavy-tailed, as it is in this analysis. The random variable associated with this distribution comes about as the ratio of two Gaussian distributed variables with zero mean, therefore, the Cauchy distribution is also called the normal ratio distribution. The other best fits are Log-Logistic, Burr, and Dagum distributions, which are continuous probability distributions for a nonnegative random variable. The Pearson distribution is also a fit since it can visibly contain skewed observations. Finally, the lognormal distribution fitting arises when independent random variables are combined in a multiplicative fashion, as expected in the application of relative errors.



Descriptive Statistics: Relative Error (Minitab®)

Variable	N	Mean	StDev	Min	Median	Max
Relative Error	1923	1.2322	1.5146	0.0003	1.0008	21.2766

Figure 1. Distribution of Empirical Relative Errors

Table 1. Best Fitted Distribution for Experts' Relative Errors

Best Fitted Distribution (MathWave-EasyFit)	Kolmogorov Smirnov Rank	Anderson Darling Rank	Chi-Squared Rank
Wakeby	1	1	1
Cauchy	2	2	2
Dagum (4P)	3	5	5
Log-Logistic (3P)	4	4	4
Burr (4P)	5	3	3
Burr	6	7	7
Dagum	7	6	6
Pearson 6 (4P)	8	8	8
Lognormal (3P)	9	9	9

2.4 *Assessment of Likelihood Functions*

In the Bayesian analysis, the likelihood function for homogenous data (Equation 8) is used in general Bayesian formalism (Equation 1) to develop the posterior (the analyst's updated knowledge about the unknown) and calculate the posterior mean. For the nonhomogenous data, the application of formulated likelihood distribution leads to Equation 13. Normally, in a general Bayesian approach, posterior is developed independent of the form of priors. In fact, this approach provides a basis for defining expertise of information sources (in the matter of estimating true value) relative to the decision maker. Additionally, if the decision maker believes, as would normally be the case in consulting experts, that the prior information should have little or no impact on the posterior distribution, a flat prior would be a proper modeling choice (Edwards, 1963). Flat or non-informative priors in Bayesian analyses are very important for two reasons. First, they may reflect the true state of current knowledge. Second, it may be required to construct reference replica against models containing subjective information. In both cases, the goal is to let the data drive the analyses, which can be evaluated by comparing results with outcomes of the assessment. To compare the results of the Baysian calibration, the mean as the posterior marker is compared with the observed or true values to assess the error of updated estimates. According to Christensen and Huffman (1985), the most often used posterior markers have been the mean, median, and mode of the posterior, with no consensus among experts on which is the most appropriate. Barnett (1982) states that there would seem to be no other useful criterion for choosing a single value to estimate true value than to use the most likely value, unless we incorporate further information on the consequences of incorrect choice of true value. Berger (1980) states the mean and median are often better estimates than the mode. Cox and Hinkley (1974) state that if it is required to summarize the posterior distribution in a single quantity, then the mean is frequently the most sensible. In particular, if the prior density is exactly or approximately constant, the use of the mean of the likelihood with respect to the parameter is indicated. Results in Table 2 reveal 285% overall average improvements with 77% of estimates improved, applying the likelihood function developed by relative errors in all homogenous (H) and nonhomogenous (NH) cases. To calculate the average improvements, the mean of the posterior is compared with the true value. The error of Bayesian updating is then weighed against expert's error. The amount of reduction in error is presented as percentage, which reflects the average of improvements using Bayesian updating in each case. The graphical presentation can be found in Figure 2. The results obtained confirm that experts' errors of estimates are reduced by application of formulated likelihood distribution.

Table 2. Bayesian Treatment

Case # H/NH	Average Improvement	Estimates Improved
1 H	368%	71%
2 NH	335%	71%
3 NH	208%	100%
4 NH	72%	100%
5 NH	91%	71%

6 NH	-75%	83%
7 NH	120%	67%
8 NH	63%	67%
9 NH	78%	67%
10 NH	54%	71%
11 NH	316%	100%
12 NH	1989%	57%
13 NH	555%	71%
14 NH	509%	86%
15 NH	220%	86%
16 NH	1171%	100%
17 NH	89%	57%
18 NH	72%	86%
19 H	61%	80%
20 NH	264%	57%
21 NH	98%	86%
22 NH	524%	57%
23 NH	83%	86%
24 NH	87%	86%
25 NH	44%	57%
26 NH	96%	100%
27 NH	237%	57%
28 H*	*243%*	*79%*
Average	*285%*	*77%*
Minimum	*-75%*	*57%*
Maximum	*1989%*	*100%*

**Data provided by R. M. Cooke from TU Delft University*

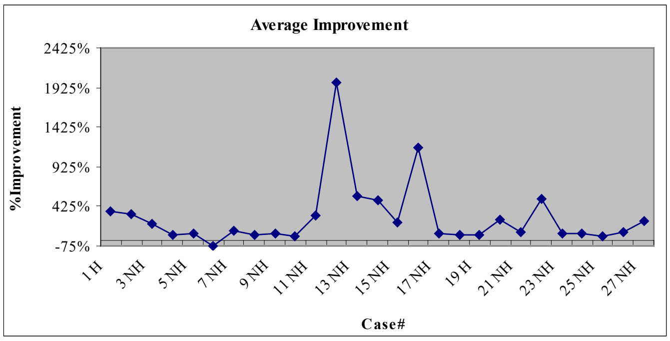


Figure 2. Bayesian Treatment of Empirical Data

For the sake of moderation in the assessment of results, cases with improvement of over 500% are considered improvements with low probability in occurrence and eliminated from calculations. Percentages are recalculated as shown in Table 3 and displayed with Figure 3. The positive results obtained after these modifications still confirm that experts' error of estimates are reduced by application of formulated likelihood distribution function.

Table 3. Bayesian Treatment – modified data

Average Improvement of Error	%Estimates Improved
368%	71%
335%	71%
208%	100%
72%	100%
91%	71%
-75%	83%
120%	67%
63%	67%
78%	67%
54%	71%
316%	100%
220%	86%
89%	57%
72%	86%
61%	80%
264%	57%
98%	86%
83%	86%
87%	86%
44%	57%
96%	100%
237%	57%
243%	*79%*
Average = 140%	*Average = 78%*
Min = -75%	*Min = 57%*
Max = 368%	*Max = 100%*

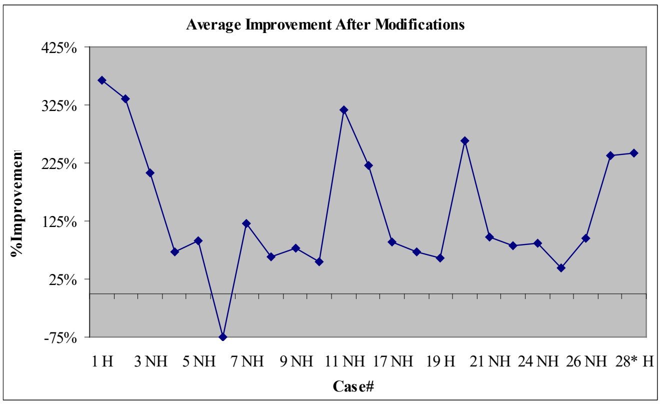


Figure 3. Bayesian Treatment –after Modifications

2.5 *Conclusion*

The empirical assessment of experts' relative error of estimates revealed that over 45% of errors were close to one (expert estimate ~ true value). Additionally, lognormal was identified as one of the best fitted distributions, considering the selection of relative error as the forecast accuracy measure. The study also showed 285% average improvements in experts' estimates with 77% of estimates improved, applying the likelihood function developed by relative errors for homogenous and nonhomogenous cases.

3 *REFERENCES*

Abdullah Khan, H. T. A Comparative Analysis of the Accuracy of the United Nations' Population Projections for Six Southeast Asian Countries. Interim Report, IR-03-015.

Ahlburg, D. A. 1992. Error measures and the choice of a forecast method. *International Journal of Forecasting*, 8: 99-111. D.

Armstrong, S. & R. Fildes 1995. On the selection of error measures for comparisons among forecasting methods. *Journal of Forecasting*, 14: 67-71.

BaFail, A. O 2004. Applying Data Mining Techniques to Forecast Number of Airline Passengers in Saudi Arabia (Domestic and International Travels). King Abdul Aziz University.

Barnett, V. 1982. *Comparative statistical inference* (2nd ed.), New York: John Wiley.

Batz, M. B. et al. 2004. Identifying the most significant Microbiological Food-borne Hazards to Public Health: A New Risk Ranking Model, Food Safety Research Consortium. Discussion Paper Series Number (1), September 2004 - FIRRM Food Attribution Percentages for Illnesses from Food borne Campylobacter and Listeria monocytogenes.

Berger, J. O. 1980. *Statistical decision theory foundations, concepts and methods*. NY: Springer-Verlag.

Booker, J. M. & Meyer, M. 1996. Elicitation and analysis of expert judgment. Los Alamos National Laboratory, LA-UR-99-1659.

Campbell P. R. 2002. Evaluating Forecast Error in State Population Projections Using Census 2000 Counts. U.S. Bureau of Census, Population Division Working Paper Series No. 57.

Chen, Z. & Yang, Y 2004. Assessing forecast accuracy measures.

Christensen, R. & Huffman, M. D. 1985. Bayesian point estimation using the predictive distribution. *The American Statistician*, 39, 4, Part 1: 319-321.

Clifford A. Cathers, C. A. & Thompson, G. D. 2006. Forecasting Short-Term Electricity Load Profiles. Sierra Southwest Cooperative Services, Inc. The University of Arizona, Cardon Research Papers, August 2006.

Cooke, R. M. 1991. *Experts in uncertainty: opinion and subjective probability in science*. NY: Oxford University.

Cox, D. R. & Hinkley, D. V. 1974. *Theoretical statistics*, London: Chapman & Hall.

Droguett, E. L. & Mosleh, A. Assessment of model uncertainty: use of model performance data.

Droguett, E. L. & Mosleh, A. Framework for integrated treatment of model and parameter uncertainties.

Dumler, T. J. 2003. Rainfall and Farm Income. Risk and Profit Conference.

Edwards, W., Lindman, H. & Savage, L. J. 1963. Bayesian statistical inference for psychological research. *Psychological Review*, 70: 193-242.

Ehrman, C. M. & Shugan, S. M. 1995. The Forcaster's Dilemma. *Marketing Science*, 14(2): 123-127, Springer.

Forrester, Y. 2005. The quality of expert judgment: an interdisciplinary investigation.

Fullerton Jr., H. N. Evaluating the 1995 BLS Labor Force Projections.

Gilliland, M. (2002). Is Forecasting a Waste of Time? *Supply Chain Management Review*, July/August, 2002.

Hoffmann, et al. 2006. Eliciting Information on Uncertainty from Heterogeneous Expert Panels: Attributing U.S. Foodborne Pathogen Illness to Food Consumption. RFF DP6-17.

Mc Laren, C. H. & Mc Laren, B. J. 2003. Electric Bill Data. *Journal of Statistics*, Edition [Online], (11) 1.

McIntosh, C. S. & Bessler, D. A. 1988. Forecasting Agricultural Prices Using a Bayesian Composite Approach. *Southern Journal of Agricultural Economics.*

Morris, P. A. 1974. Decision analysis expert use. *Management Science*, 20 : 1233-1241.

Morris, P. A. 1977. Combining expert judgments: a Bayesian approach. *Management Science*, 23: 679-693.

Mosleh, A. & Apostolakis, G. 1986. The Assessment of probability distributions from expert opinions with an application to Seismic Fragility Curves. *Risk Analysis*, 6, No. 4: 447-461.

Ouchi, F. 2004. A literature review on the use of expert opinion in probabilistic risk analysis. World Bank Policy Research working paper 3201.

Quigley, M. A. Chandramohan, D. & Rodriguez, L. C. 1999. Diagnosis accuracy of physician review expert algorithms and data-derived algorithms in adult verbal autopsies. International epidemiological Association, 1999. *International Journal of Epidemiology*, 28: 1081-1087.

Ramachandran, G. Banerjee, S. & Vincent, J. H. 2003. Expert Judgment and Occupational Hygiene: Application to Aerosol Speciation in the Nickel Primary Production Industry. Division of Environmental and Occupational Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455; 2 Division of Biostatistics, School of Public Health, University of Minnesota, Department of Environmental Health Sciences, School of Public Health, University of Michigan, 2003. Means and standard deviations of the actual measurements, combined expert priors (for the three experts C, G and J), and posterior distributions.

Rodier, C. J. 2005. Test of Model Error in Travel Forecast, Verify the accuracy of land use model used in transportation and air quality planning: a case study in Sacramento, California region. MTI Report 05-02.

Schwartz, Z. Monitoring the Accuracy of Multiple Occupancy Forecasts.

Sevket Gunter, S. & Aksu, C. Conservative and liberal error measures for evaluating the combination forecasts.

Spiteri, A. et al. 2004. A comparison of clinical paediatric murmur assessment with echocardiography. *Malta Medical Journal*, 16, 04.

Stekler, H. O. & Thomas, R. 2000. Evaluating BLS Labor Force, Employment, and Occupation Projection for 2000.

Walker, K. Catalano, P. Hammitt, J. & Evans, J. 2003. Use of expert judgment in exposure assessment: Part 2. calibration of expert judgments about personal exposures to benzene. *Journal of Exposure Analysis and Environmental Epidemiology,* 13: 1-16.