

# Quantifying the credibility of energy projections from trends in past data

## The US energy sector

Alexander I. Shlyakhter, Daniel M. Kammen, Claire L. Broido and Richard Wilson

*For energy forecasts to be useful in modelling or in policy efforts, the associated uncertainties must be known reliably. We analyse the actual errors in past forecasts of over 170 energy producing and consuming sectors of the US economy. We find that the often assumed normal distribution fails to model frequency of extreme outcomes (those lying far from the mean) accurately. Triangular distributions perform even worse as they assign zero probability to the outliers. We develop a simple one-parameter model that can be used to estimate a probability distribution for future projections. In addition to energy forecasts, our method can be applied to any field where a history of forecasting is available.*

---

*Keywords:* Energy projections; Forecasting; Uncertainty

Sophisticated modelling systems are generally used to produce the most realistic possible projections. The reliability of the projections is limited, however, because of the uncertainties inherent in any model.<sup>1</sup> The range of uncertainty is usually estimated by running the model first under a set of assumptions

---

The authors are with the Department of Physics and Northeast Regional Center for Global Environmental Change, Harvard University, Cambridge, MA 02138 USA (AIS, CLB, RW) and the Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544 USA (DMK). Correspondence should be addressed to D.M. Kammen at the Woodrow Wilson School, 444 Robertson Hall, Princeton University, Princeton, NJ 08544 USA.

deemed the most realistic (the base or reference case) and then under a few less realistic but still reasonable assumptions. The resulting ensemble of estimates, however, does not constitute a classical statistical sample, and can only be used to obtain a subjective characterization of the true probabilities. The outputs of energy supply and demand models and forecasts are frequently used as input to decision theory models or are directly cited in policy analyses. Decision theory, however, requires that probability values be assigned to each alternative before risks and benefits can be compared.<sup>2</sup>

A lack of formal statistical probability distributions for projections or extrapolations is encountered in a variety of disciplines, and various attempts have been made to surmount the resulting difficulties, including the elicitation of 'subjective confidence intervals' for model parameters from forecasters. Empirical methods of building confidence intervals around point estimates are successfully used in the weather, population, and economic forecasting.<sup>3</sup> They rely on the assumption that an estimate of the reliability of predictions can be derived from an examination of the way in which similar predictions made in the past actually turned out. In engineering applications, the importance of empirical control of experts' probability assessments is also well recognized.<sup>4</sup> With the exception of weather forecasts, however, there is a strong tendency to underestimate uncertainties in predictions,<sup>5</sup> the probability of a 'surprise' (an outcome different from the prediction by much more than the estimated uncertainty) is increased and usefulness of the forecasts is reduced.

The goal of this paper is to present an empirical approach for improved uncertainty estimation and

**Table 1. Explanation of notation**

$L, R, U$	= respectively the low, reference, and high scenario forecasts
$T$	= true value
$x$	= normalized measure of the deviation of the 'old' (projected) values, $R$ , from the true value, $T$ : $x = (T - R)/\Delta$
$\alpha$	= the assumed probability of the forecast range, $[L, U]$ , to cover the true value, $T$ . We use $\alpha = 68\%$ which gives $\Delta = (U-L)/2.0$ for the standard deviation of the equivalent normal distribution
$u$	= the additional parameter describing the relative uncertainty in the forecast range
$\Delta$	= standard deviation corresponding to the estimated uncertainty
$\Delta'$	= standard deviation corresponding to the 'true' uncertainty
$t$	= $\Delta'/\Delta$
$f(t)$	= probability distribution of $t$
$Z$	= inflation factor for the forecast range (see Appendix 3)

to illustrate its use in the analysis of energy projections. This paper therefore divides naturally into two parts: the characterization of uncertainty, particularly for low probability events; and the construction of future forecasts. The value of our method lies in the confidence intervals that are expanded to reflect more accurately the frequency of unsuspected errors. These confidence intervals can then be applied to scenario evaluation and planning, both in the energy sector and more generally in any field where modelling methods are commonly used. The notation used throughout the paper is summarized in Table 1.

### Probability distributions

Uncertainty in energy forecasts is defined less formally than random errors in physical measurements where a Gaussian (normal) distribution is usually implied even though it often underestimates the probability of large deviations caused by unaccounted systematic uncertainties. We will use a convenient normalized measure ( $x$ ) of the deviation of the 'old' (projected) values,  $R$ , from the true value,  $T$  where  $x = (T - R)/\Delta$ . Here  $\Delta$  is uncertainty estimated by the forecaster. In this paper we will determine the actual distribution of  $x$  values from historical energy forecasts and compare with the result predicted by Gaussian and other theoretical models.

A comparison of the empirical frequency of large deviations of the true value from the predicted values with frequency given by the normal distribution provides researchers with an intuitive analogy with the well-understood case of random errors. We grant that the expert forecasters may not necessarily

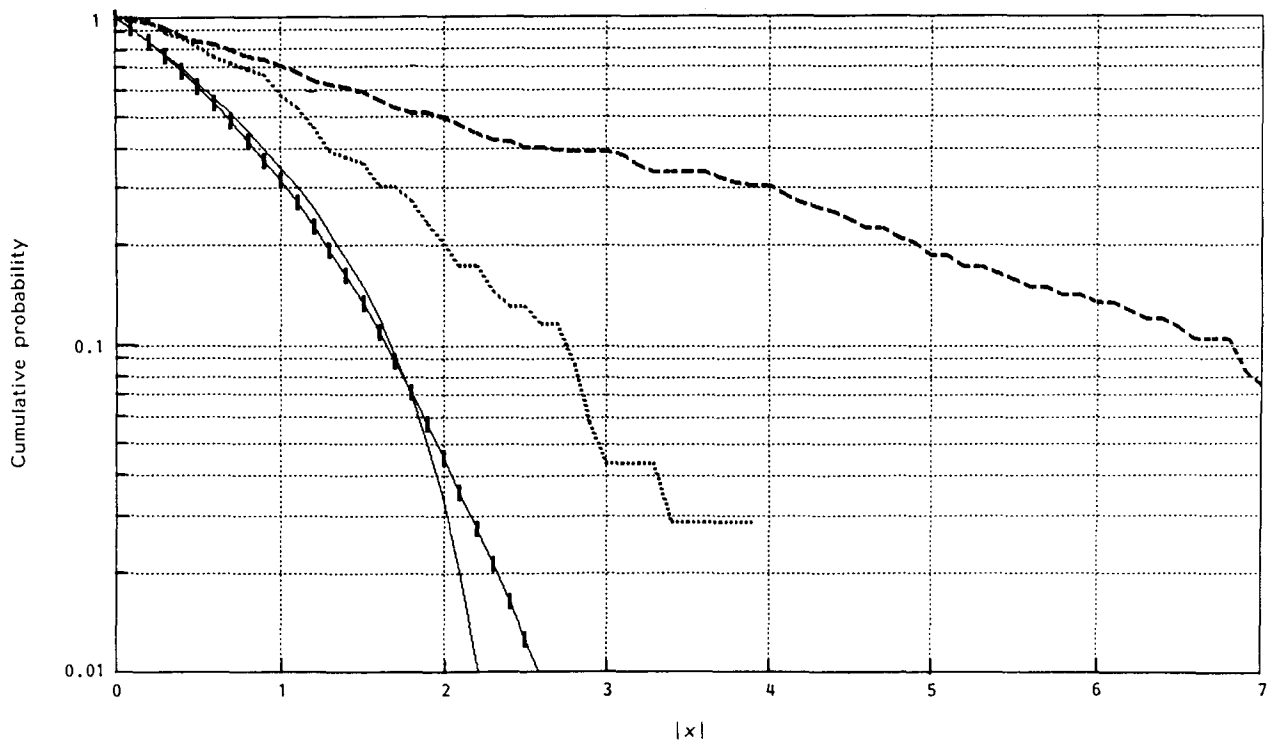
imply that the error terms are normally distributed. However, the users of the results are usually less sophisticated, and, reasonably enough, assume a normal form when they are provided with no explicit alternative. Furthermore, it is common to assume that deviations exceeding the estimated uncertainty by several times are improbable. In our view, a comparison of errors in historical data sets with those predicted by the normal distribution provides a useful measure of the credibility of current uncertainty estimates. We demonstrate that in the data sets we have analysed, the normal distribution is not an appropriate model, particularly far from the mean.

Uncertainty in energy forecasts is usually presented in the form of reference ( $R$ ) lower ( $L$ ) and upper ( $U$ ) estimates that are obtained by running a model with different sets of exogenous parameters (eg annual rate of growth or the size of a carbon emissions tax). The range of scatter around the reference value,  $R$ , does not formally define a Gaussian standard deviation because the fundamental uncertainties involved (eg the rate of future economic growth) are frequently not random. However, it seems reasonable to assume that the range of parameter variation used by a forecaster represents a subjective judgement about the probability that the true value,  $T$ , lies between the lower and upper estimates ( $L \leq T \leq U$ ).<sup>6</sup> Generally, lower and upper bounds present what is believed to be an 'envelope' most likely to bracket the true value. While often  $R$  does not coincide with  $T$ , the goal of the forecaster is generally to closely approximate this condition through the construction of an accurate model. A comparison of the empirical frequency of large deviations from the predicted values with an 'equivalent' normal distribution allows an analogy with the well understood case of stochastic uncertainties.

Upper, reference and lower variants are sometimes perceived as parameters of a triangular distribution. We note that using the bounded distributions (such as triangular) assigns zero probability to the outliers. Historical data presented below, however, suggest that deviations exceeding the expected uncertainty range are not uncommon. Therefore, using a normal (unbounded) distribution as a frame of reference is an assumption that results in a lower bound on the degree of overconfidence.

### Previous studies

As others have noted and grappled with before, forecasters often underestimate their uncertainties,



**Figure 1.** Frequency of unexpected results in US energy projection for the year 2000 (dotted line), and population projections for 133 countries for the year 1985 (dashed line).<sup>a</sup>

<sup>a</sup>The plots depict the cumulative probability,  $S(x) = \int_x^\infty p(t)dt$ , that true values,  $T$ , will be away from the reference values for old projections,  $R$ , by at least  $x$  times the estimated uncertainty range;  $x = (T - R)/\Delta$ . Here  $\Delta$  is the standard error of the mean for energy projections and half of uncertainty range  $(U - L)/2$  for population projections (see Appendix 1 for details). Also shown are the Gaussian (light solid line with vertical ticks) and symmetric triangular distribution with standard deviation  $\Delta = 1$  (light solid line).  $S$  is plotted against the absolute value of  $x$ ,  $|x|$ .

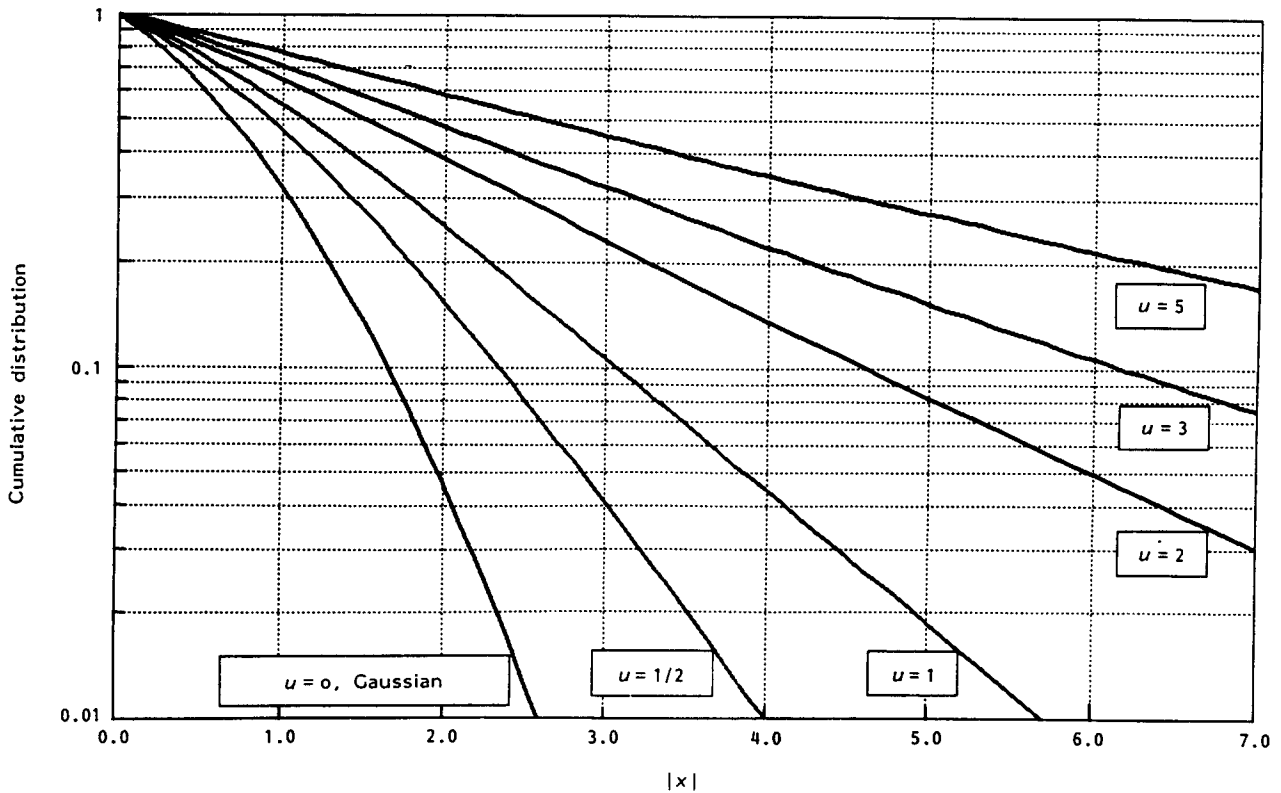
resulting in long-tailed  $x$  distributions.<sup>7</sup> As previously noted, an exception is provided by weather forecasts where daily empirical verification helped to improve the calibration of the probabilities estimated by the experts.<sup>8</sup> We propose that forecasts should be interpreted in accordance with the historical record of  $x$  values characteristic of the particular field. To test this approach, Shlyakhter *et al* recently analysed trends in two separate fields: forecasts made between 1972 and 1990 of the primary energy demand for the USA in the year 2000; and the predictions of population for different countries for the year 1985 made in 1973 (see Appendix 1).<sup>9</sup>

The cumulative probability distributions for  $|x|$  are shown in Figure 1 together with the cumulative Gaussian curve (thin solid line with vertical ticks), and cumulative triangular distribution (thin solid line). Both the Gaussian and triangular distributions significantly underestimate the probability of large deviations. With the Gaussian distribution, about 1% of cases are more than 2.5 standard deviations away from the true value ( $|x| > 1$ ), while triangular distribution excludes cases with  $|x| > \sqrt{6} = 2.45$

altogether. To account for the tendency for the estimates generated by even independent experts to cluster excessively near the mean, Nordhaus and Yohe collected sets of forecasts, and then inflated the standard deviations by 50% to obtain what they felt were more realistic representations of the actual situation.<sup>10</sup> But as can be seen in Figure 1 it is not merely the width of the experimental distribution that is underestimated by the Gaussian, but the shape is fit poorly as well. The fact that the empirical data deviate so dramatically from the normal distribution (and even more dramatically from triangular and rectangular distributions) at large  $|x|$  (a region which is often of concern for policy questions), suggests that we pursue a new parametrization based, in this case, on a compound distribution.

### The compound distribution

The primary goal of this paper is to illustrate the problem of overconfidence in energy projections and to develop methods that can be employed to correct for this in a consistent manner. The database



**Figure 2.** One-parameter family of probability distributions.  $u$  quantifies the uncertainty in the standard deviation of the Gaussian distribution (Equation (5)).<sup>a</sup>

<sup>a</sup>The values of  $u$  are indicated in the figure. The curves demonstrate the continuum of probability distributions: from Gaussian ( $u=0$ ) to exponential ( $u > 1$ ). In accordance with Equation (4), the curve  $u=1$  has the same asymptotic slope as  $\exp(-|x|)$  although they differ by a factor of about 2 that is a consequence of the deviation from exponential behaviour at small  $|x|$ .

examined here suggests a particular model. This model is slightly more complex than the procedure of merely inflating previous uncertainty estimates by a constant factor, but it is still straightforward and soluble. The distribution of large deviations observed in the population and energy data (Figure 1) can be characterized by an asymptotically exponential compound distribution with one additional parameter,  $u$ , the relative uncertainty in the original standard deviation,  $\Delta$  (see Appendix 2).

The cumulative probability functions for a family of curves ( $0 \leq u \leq 5$ ) are plotted in Figure 2. The normal ( $u = 0$ ) and exponential distributions ( $u > 1$ ) are members of a single-parameter family of curves. In this framework the parametric uncertainty can be quantified by analysing the record of prior projections and estimating the value of  $u$ . From Figure 1 we see that  $u \sim 1$  for the set of energy demand forecasts originally compiled by Goldemberg *et al.*, and  $u \sim 3$  for current models of population growth.<sup>11</sup> Thus, while  $u$  is not the same for different types of forecasts, these data exhibit a consistent functional form that can be simply computed from a set of past

projections and subsequently reported true values. Here we apply this approach to the largest available set of US energy forecasts for the year 1990 and estimate the 'credibility intervals' for future projections.

### The database: Department of Energy Annual Energy Outlook

We found that the *Annual Energy Outlook* (AEO) published by the US Department of Energy was the largest coherent set of past domestic projections available.<sup>12</sup> It is produced using an integrated energy modelling system which includes supply modules for oil markets, coal, gas, and electricity, and a set of energy demand models. The supply models determine supply and price for each fuel conditional upon consumption levels, while the demand models determine consumption conditional upon end-use price. The forecasting module solves for market equilibrium for each fuel by balancing supply and demand to produce an energy balance for each forecast year.<sup>13</sup>

The low, reference and high ( $L$ ,  $R$ ,  $U$  respectively) scenario forecasts are aggregated by fuel type within the supply module, and by end use within the demand module. Over 170 separate supply and demand sectors are included in the model.<sup>14</sup> When taken as a group the three scenarios present a range of possible outcomes. To assign a probability interval to  $(U - L)$  we construct a normal distribution with the mean,  $(L + U)/2$ , (generally equal to the reference case  $R$ ) and standard deviation,  $\Delta$ , in such a way that the area between  $L$  and  $U$  is equal to a specified probability value,  $\alpha$ . For  $\alpha=95\%$ ,  $\Delta=(U-L)/3.92$  and for  $\alpha=68\%$ ,  $\Delta=(U-L)/2.0$ . We shall use  $\alpha=68\%$  in this paper and therefore calculate  $x=2\cdot(T-R)/(U-L)$  where  $T$  again is the actual value observed for the year in which  $L$ ,  $R$ ,  $U$  are forecast. This choice of  $\alpha$  corresponds to the common practice of splitting the difference between high and low estimates and using half of this as a surrogate for the standard deviation. If the reference value,  $R$ , does not coincide with the mean value,  $(L+U)/2$ , then  $x$  is defined using the uncertainty range on the same side of  $R$  as  $T$ :  $x=(T-R)/(R-L)$  for  $R > T$  and  $x=(T-R)/(U-R)$  if  $R < T$  with  $L < R < U$  assumed for both cases.

The purpose of introducing the compound distribution model is to provide a parameter parametrization of the observed probabilities of large deviations. Uncertainty ranges are widely perceived as formal confidence intervals of normal distributions and values exceeding several standard deviations are viewed as surprises. We therefore believe that a comparison of the actual data and new models with the normal distribution is in order. To compare the actual uncertainties in energy projections with those expected from a normal distribution we assign a confidence,  $\alpha$ , to the uncertainty range provided by the forecasters. As noted above, the forecast credibility value adopted here is  $\alpha=68\%$  which should be conservative in light of the 90–95% confidence intervals claimed for some models.<sup>15</sup> Larger values of  $\alpha$  produce even larger  $x$  values and deviations from the simple Gaussian model. Triangular and rectangular distributions assign 100% probability to their range and therefore provide even less realistic description of the data than does the Gaussian distribution.

## Results

To determine the appropriate value for parameter  $u$ , we analysed the AEO projections for 1990 made in 1983, 1985, and 1987 which respectively consisted of 182, 185 and 177 energy producing or consuming sectors of the US economy. The variation in the

number of sectors resulted because the low and high projections coincided in some cases, and no corresponding uncertainty range could be derived. In 47, 50 and 47 cases respectively, the  $x$  value exceeded 100. We assumed that the AEO model might not be applicable in those cases and omitted them; had we kept them, derived  $u$  values would be even higher. For all remaining cases the  $x$  values were calculated and the frequency distributions analysed.

Figure 3a demonstrates that the distribution of  $x$  values is approximately symmetric with respect to zero; there is no large systematic bias (eg a gross underestimation of energy consumption in all or many sectors). The correlation structure of the sectors between the 1983–85, 1985–87, and 1983–87 AEO forecasts for 1990 is shown in Figures 3b–3d. The largest linear correlation coefficient,  $r = 0.55$ , is observed between the 1983 and 1985 forecasts. The lack of consistent trends in the scattergrams of  $x$  values is good evidence that the forecasts are generally independent.

Figure 4 shows the cumulative probability distributions of  $|x|$  (the magnitude of  $x$  without regard to sign) for the projections made for 1990 in 1983, 1985, and 1987 together with the Gaussian and exponential distributions. The three empirical distributions are strikingly similar, with  $u$  values in each case of about 3. The similarity could be due in part to the modest correlation between 1983 and 1985 forecasts (Figure 3b) although the lack of any such correlation between either of the two later models (Figure 3c and 3d) suggests that this is a minor effect. Although the absolute error in forecasts made in 1987 for 1990 is somewhat smaller than the error in forecasts made in 1983 for 1990, the range of uncertainty suggested by the forecasters is also smaller so that probability of ‘large’ deviations relative to the stated uncertainty is roughly the same as for the other two years.

Initially we expected that energy forecasts for aggregated sectors of economy would be more reliable than projections for individual sectors. However, we found this not to be the case: distribution of  $x$  values for aggregated forecasts is similar to the distribution for the individual sectors (Figure 4). Overall we find that the Gaussian is not the best model far from the mean while the exponential distribution with  $u$  about 3 fits the data well.

The difference between the exponential and Gaussian models can be striking: for  $u = 3$  there is a 7.5% probability that a value of a parameter predicted by a model will be more than seven standard deviations above or below the true value (see Figures 4 and 5). If, however, the distribution is

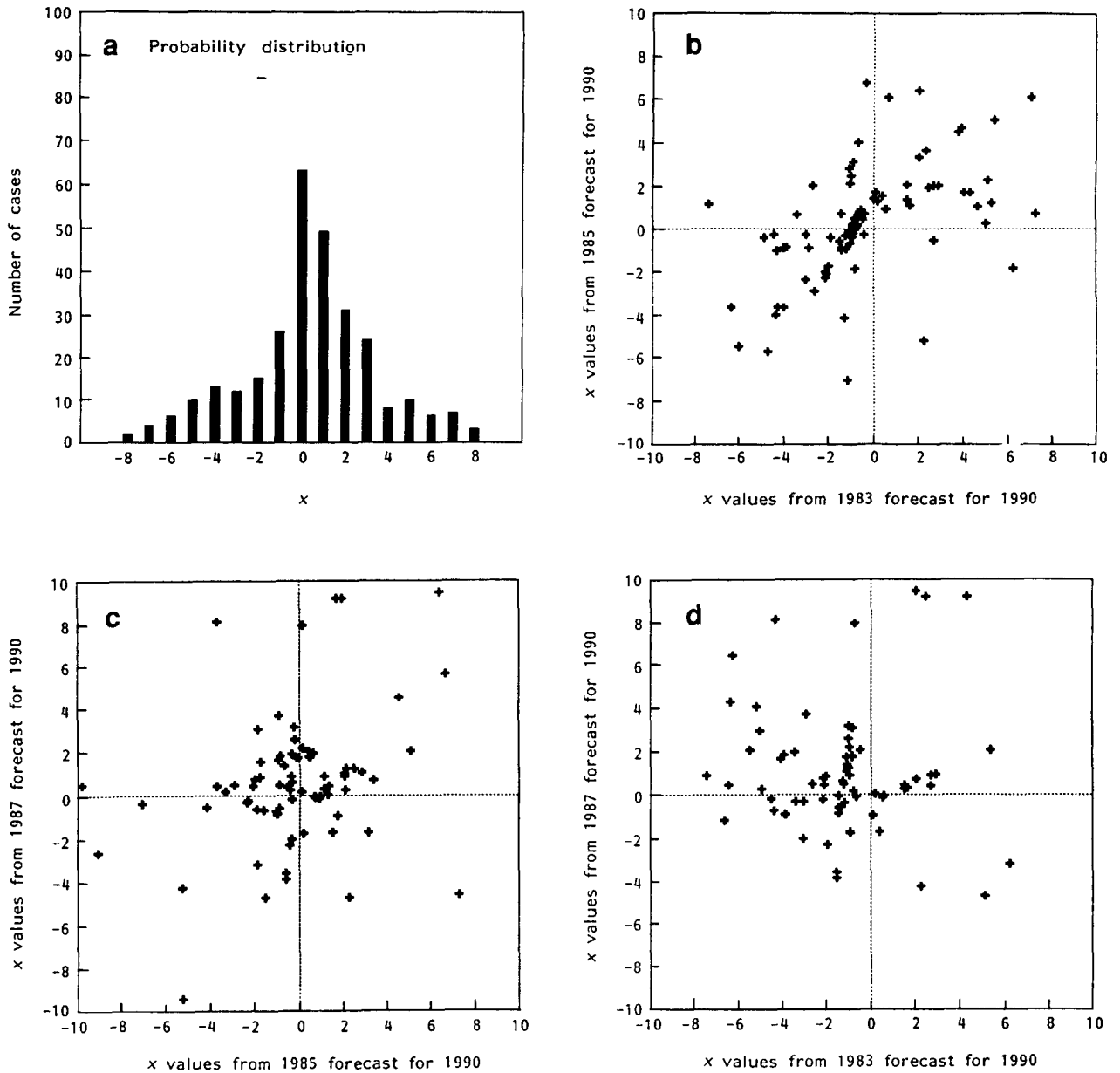


Figure 3. Annual Energy Outlook projections.<sup>a</sup>

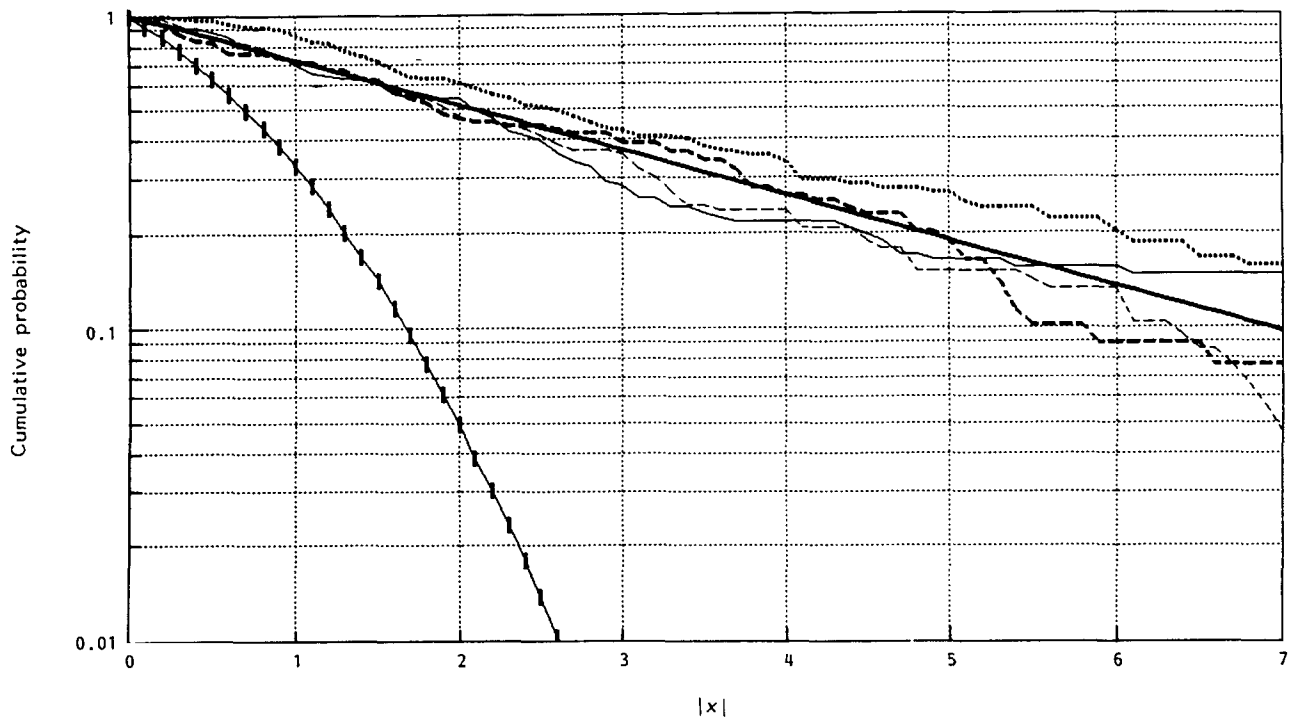
<sup>a</sup>The data in 3a are an accumulation of the 1983, 1985 and 1987 values truncated at  $|x| > 8$ . The distribution shows no statistically significant deviation from a symmetric profile for even large values of  $|x| > 20$ . The data demonstrate overall that there is no significant correlation between the 1983 and 1987, and 1985 and 1987 AEO forecasts. The two earliest AEO models, from 1983 and 1985, are moderately correlated (linear correlation coefficient,  $r = 0.55$ ).

assumed instead to be Gaussian, the probability that the prediction is so far from the true value is negligibly small ( $4 \times 10^{-12}$ ).

### Application to existing forecasts

Our method can be applied to current AEO energy projections by inflating the estimated 90% confidence intervals with  $u = 3$ , corresponding roughly to an inflation by a factor of four (see Appendix 3 for

the details of our procedure). This results in a revised 90% credibility interval. For example: in the current (1992) AEO total US production projected for the 2010 from nuclear power is 6.9 quads (1 quad =  $1.05 \times 10^{18}$  joules) (a reference case with 0.6% annual growth) with  $U$  and  $L$  estimates set at 7.5 and 6.7 quads respectively.<sup>16</sup> Based on  $u=3$  in our compound exponential model we instead forecast the 90% confidence interval to be from  $U = 9.3$  to  $L = 6.3$ , as shown in Figure 5. The actual values tell us



**Figure 4.** Annual Energy Outlook projections.<sup>a</sup>

<sup>a</sup>The presentation is as in Figure 1: 1983 to 1990 (heavy dotted line); 1985 to 1990 (dashed line); 1987 to 1990 (solid line), aggregated forecasts for all three forecast years (heavy dashed line); exponential distribution,  $\exp(-|x|/3.0)$ , approximating, as described in Table 2, the compound distribution with  $u=3.4$  ( $0.7 \cdot 3.4 + 0.6 = 3.0$ ; heavy solid line); Gaussian (thin solid line with vertical ticks).

that far less confidence should have been placed on past AEO forecasts than was claimed by the forecasters. Without significant revision and recalibration it is prudent to apply the same scepticism to current and future AEO forecasts. We suggested that the 1992 AEO revised confidence intervals should apparently be based on  $u = 3$ . These are shown in Figure 5 for three production sectors (crude oil, nuclear power and renewables) and three consumption sectors (liquefied natural gas, coal and residential electricity).

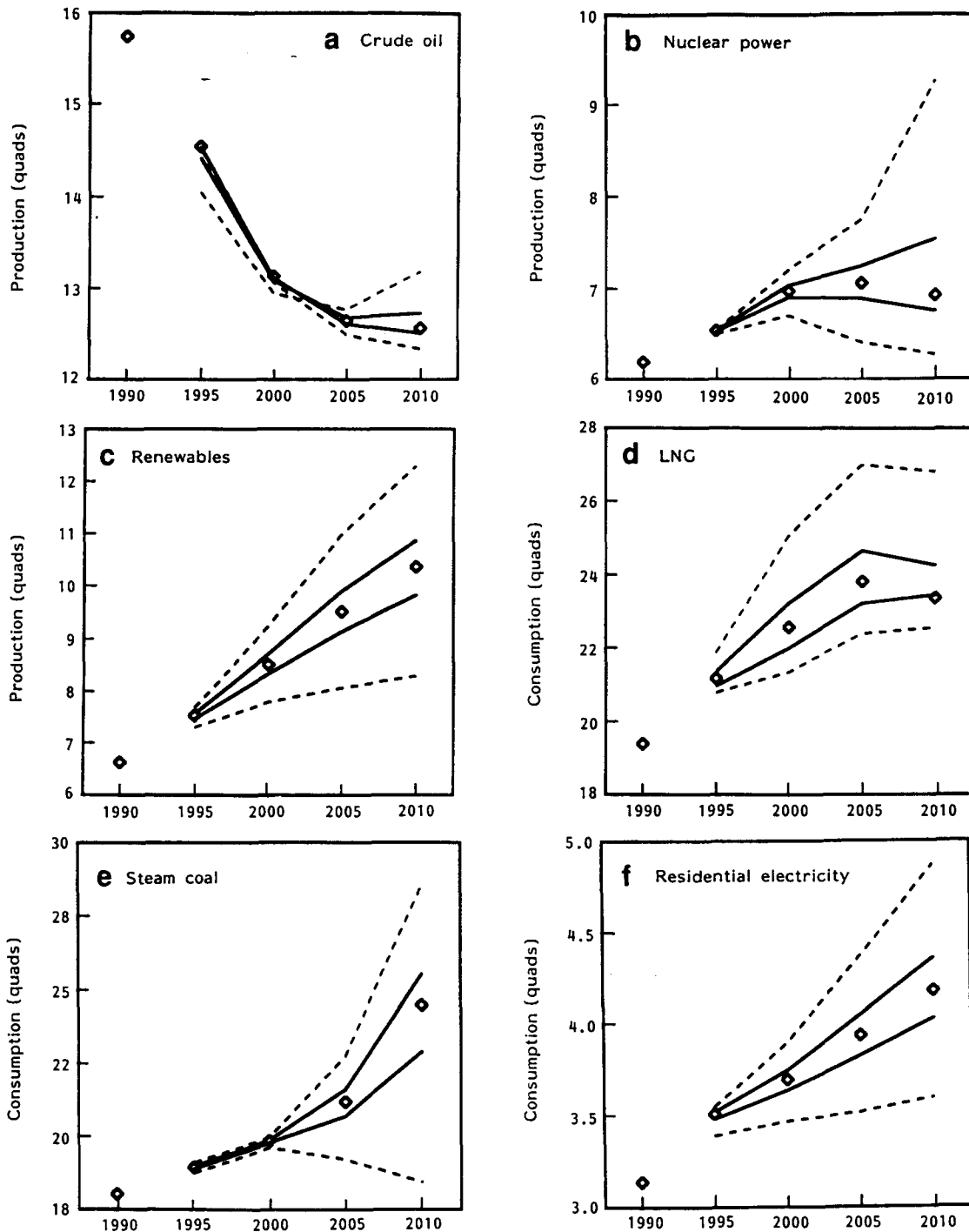
Note that in estimating  $u$  values we assumed  $\alpha=68\%$  for the *old* forecasts but for the *current* projections we assume  $\alpha=90\%$ . In this way we account for the improved reliability of more recent forecasts. Had we assumed  $\alpha=90\%$  for the old forecasts, the derived standard deviations would be smaller and all  $x$  values would be larger. The resulting  $u$  values and the corresponding inflation factors would be also larger than the ones we used.

Particularly dynamic sectors of the energy economy logically exhibit significant uncertainty because even the best models remain out of date. The history of past projections suggests that the expected production from renewable sources (Figure 5c) in 2010 actually lies between 8 and 12.5 quads (5–95%

confidence), compared with a range of 9.8–10.8 under the AEO analysis. The intrinsic variability in energy supply derived from intermittent solar and wind sources adds to the problem and necessitates careful planning when developing plans to reliably integrate renewables into a commercial grid.<sup>17</sup> Even higher  $u$  values than those derived for the energy economy as a whole might be appropriate to quantify the credibility of forecasts in these sectors.

The revised projections for coal consumption (Figure 5e) are interesting in that the AEO forecasts already assume some environmental pressure on the coal industry with a negative effect on consumption. We find that in 1990, with no greenhouse gas regulations in effect,  $x_{\text{coal}} = -2.91$ . This suggests that the latest AEO model does not fully incorporate industry uncertainty over further developing new coal technologies in an uncertain atmosphere.

While domestic crude oil production has declined by almost 20% in the last decade, natural gas and in particular liquefied natural gas production and consumption (Figure 5d) increased sharply. This trend was in part driven by changes in demand and regulatory structure to the point that current production and delivery capacity is in excess of demand.<sup>18</sup> The AEO model did not anticipate the variability in the



**Figure 5.** Confidence intervals for six production and consumption sectors in the AEO database.<sup>a</sup>

<sup>a</sup>For each panel 5a–5f (read as rows from left to right): the diamonds are for the reference case, with the 1990 value that actually reported by the Energy Information Administration (see Ref 12); the low and high confidence limits as reported in the AEO are shown as solid lines; our recalculation of the confidence limits (5–95% using  $u = 3.4$  throughout) is shown as dashed lines. The computation of the confidence intervals is described in Appendix 3. All percentage growth rates shown are for the base case. In 5a the production estimates for crude oil, including other hydrocarbons from drilling, are shown. The AEO projects oil recovery will on average decrease by 1.1% annually until 2010 while we found that  $x_{oil} = -1.4$  based on a composite of the 1983, 1985, and 1987 forecasts for 1990 (Figure 4). Nuclear power generation 5b is expected to grow by only 0.6% per year in the AEO model while in 5c renewables (including both utility and non-utility generating capacity) are projected to grow at 2.3% per year. Consumption projections for several sectors are also shown: 5d liquid natural gas,  $x_{LNG} = 7.3$ , 5e coal utilized in thermal power plants, steam coal,  $x_{coal} = -2.91$ ; and 5f residential electricity demand,  $x_{ec} = 4.05$ .



**Table 2. Flowchart recipe for estimating uncertainty****Ideal procedure**

0. Perform a standard uncertainty analysis and determine the confidence interval
1. Find a data set of prior projections that includes  $N$  ( $L$ ,  $U$ ) pairs with estimates of the standard deviation, (ideally  $N > 100$  individual estimates)
  - 1a. IF: A suitable data set is available, but uncertainty estimates are presented only as a range between lower and upper bounds  
 THEN: Use half of the range as a surrogate for the standard deviation. This corresponds to  $\alpha = 68\%$ ; numerical values for other confidence intervals are given in terms of  $\alpha$  in the section in the text on the AEO database
  - 1b. IF: no suitable data base is available  
 THEN: select a value for the uncertainty parameter based on a data set that most closely resembles the data in question and go to step 5  
 From our previous analysis (see Ref 6 and *op cit*, Ref 10) and that in this paper we can suggest:
 

Data base	$u$
Physical constants	1
Energy projections	1 to 3.5
Population projections	3
2. Calculate  $x = (a - A)/\Delta$  for each estimate
3. Plot the cumulative probability against the absolute magnitude,  $|x|$  (see Figure 4)
4. Estimate  $u$  by comparing this empirical cumulative probability distribution with the curves in Figure 2. For  $u \geq 1$  and  $|x| \geq 3$  one can also use the approximation  $\exp(|x|/(0.7u+0.6))$ .
5. Utilize  $u$  to determine the modified confidence intervals (Appendix 3).

demand for liquefied natural gas (LNG): we found that  $x_{LNG} = 7.3$  for the period 1983–90. The dynamic nature of the LNG sector, as indicated by the large  $x$  value, suggests that in policy decisions concerning gas and LNG in particular, planners should pay careful attention to the large uncertainty in forecasts for LNG availability.

## Discussion and applications

Statistical analysis of past trends in projected values permits a characterization of the uncertainty and overconfidence in model parameters. For energy forecasts and projections of population growth we find that the observed long tails are well fit by simple exponential functions with one additional parameter,  $u$ , which can be interpreted as the ratio of unsuspected systematic errors to the recognized uncertainties.<sup>19</sup> Estimation of  $u$  for specific data sets provides a procedure for inflating confidence intervals that are applicable in scenario planning that requires quantitative probability estimates far from the mean.

An alternative way to account for overconfidence would be to use the uncertainty range provided by the forecaster and to determine the empirical prob-

ability that this range covered the true value derived from the historical data sets. As Figure 4 shows, this probability is 10 to 30% so there is 70 to 90% chance that the true value will fall outside the estimated range. This procedure, however, does not address the low probability cases of large deviations which may have large consequences and be of particular importance to decision making.

Interestingly,  $u$  values for three sets of projections of the US energy consumption for 1990 made in 1983, 1985, and 1987 that encompass a range of different sectors all converge on  $u \sim 3$ . Furthermore, aggregating several sectors together does not improve the precision of estimates. Moreover, although the absolute error in the 1987 to 1990 forecasts is smaller than in 1983 to 1990 forecasts, the range of uncertainty estimated by the analysts is also smaller so that probability of large deviations, when expressed as a multiple of the estimated error (degree of overconfidence) appears roughly the same. As a final note, the purpose of this exercise is not to criticize the AEO; theirs is, in fact, a remarkably useful and sophisticated model. Overconfidence is endemic in model efforts. The goal here is to illustrate the problem and suggest a solution.

An important difference exists between the  $u$  value that we determined previously for a set of 69 forecasts of US energy demand made between 1972 and 1982 for the year 2000<sup>20</sup> ( $u \sim 1$ ) and the value  $u \sim 3$  that is discussed in this paper for the AEO database (see Appendix 1). The earlier work used forecasts taken from Goldemberg *et al* and augmented through our analysis of the energy literature.<sup>21</sup> They were made over an 18 year interval by institutions as ideologically diverse as the Rocky Mountain Institute and the Sierra Club (10 forecasts together), the US government (17 forecasts), oil companies (14 forecasts) and several international energy organizations (28 forecasts). By contrast the AEO forecasts span only seven years and represent only one set of economic and resource models. While the AEO is perhaps the most sophisticated single energy forecast model, the forecasts collected by Goldemberg *et al* represent a greater diversity of assumptions and a longer learning curve. Goldemberg *et al* collected a suite of models that produced a larger spread of projections than did the single vision of future growth captured in the AEO model, which led to a greater confidence that the true result will fall within the predicted range.

Producing forecasts of future change is a separate and more subjective issue than the analysis procedure described above. For the AEO the uniformity of the results suggests that adopting an exponential

distribution with  $u \sim 3$  would improve the reliability of future projections under the same model structure. Two forecasting exercises awaiting further data are a re-analysis of Department of Energy forecasts for 2005 and 2010 and 30 year projections from industry analysis.<sup>22</sup> This might significantly affect projections for the US energy sector and more generally global change models that use energy consumption as an important input variable.<sup>23</sup>

The uncertainty in models can be made of many types, for example: uncertainty of fact; misparametrization; and subjective errors. The methods described here may be summarized in a flowchart (Table 2) designed to serve as a protocol for quantifying the uncertainty in new data sets. It is our hope that researchers in other fields will employ this method to improve the confidence intervals in a variety of models and fields of research.

---

This research was supported by the US Department of Energy through the Northeast Center of the National Institute for Global Environmental Change (NIGEC) and by the Biological Effects Branch of the US Department of Air Force through the contract F33615-92-C-0602. We thank E.A.C. Crouch, J. Hammit, W. Nordhaus, A. Solow, and M. Stoto for valuable discussions and comments.

<sup>1</sup>K. C. Land and S.H. Schneider, eds, *Forecasting in the Social and Natural Sciences*, D. Reidel, Dordrecht, 1987, pp 7–31.

<sup>2</sup>H. Raiffa, *Decision Analysis: Introductory Lectures on Choice Under Uncertainty*, Addison-Wesley, Reading, MA, 1968.

<sup>3</sup>A.H. Murphy and R.L. Winkler, 'Reliability of subjective probability forecasts of precipitation and temperature', *Journal of the Royal Statistical Society, Series C*, Vol 26, 1977, pp 41–47; W.H. Williams and M.L. Goodman, 'A simple method for the construction of empirical confidence limits for economic forecasts', *Journal of the American Statistical Association*, Vol 66, 1971, pp 752–754; M.A. Stoto, 'The accuracy of population projections', *Journal of American Statistical Association*, Vol 78, 1983, pp 13–20; N.W. Keilman, *Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations*, Amsterdam, Swets and Zeitlinger, Publications of the Netherlands Interdisciplinary Demographic Institute (NIDI) and the Population and Family Study Centre (CBGS), Vol 20, 1990; V. Zarnowitz, *Business Cycles: Theory, History, Indicators, and Forecasting*, The University of Chicago Press, 1992.

<sup>4</sup>R.M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, Oxford, 1991.

<sup>5</sup>*Ibid.*; *op cit*, Ref 3, Murphy and Winkler; M.G. Morgan and H. Henrion, *Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York, 1990; D. Kahneman, P. Slovic and A. Tversky, eds, 'Calibration of probabilities: the state of the art to 1980', *Judgement under Uncertainty*, Cambridge University Press, Cambridge, 1982, pp 306–334.

<sup>6</sup>*See op cit*, Ref 3, Murphy and Winkler; Stoto; Keilman; and Zarnowitz; *op cit*, Ref 4; *op cit*, Ref 5, Morgan and Henrion;

Kahneman *et al.*; A.I. Shlyakhter and D.M. Kammen, 'Sea level rise or fall?', *Nature*, Vol 357, 1992, p 25.

<sup>7</sup>*See op cit*, Ref 3, Williams and Goodman; Stoto; Keilman.

<sup>8</sup>*See op cit*, Ref 3, Murphy and Winkler; Keilman.

<sup>9</sup>*See op cit*, Ref 6, Shlyakhter and Kammen; A.I. Shlyakhter, I.A. Shlyakhter, C.L. Broido and R. Wilson, 'Estimating uncertainty in physical measurements and observation studies: lessons from trends in nuclear data', and A.I. Shlyakhter and D.M. Kammen, 'Uncertainties in modeling low probability/high consequence events: application to population projections and models of sea-level rise', both in *Proceedings of ISUMA '93 the Second International Symposium on Uncertainty Modeling and Analysis*, University of Maryland, IEEE Computer Society Press, Los Alamitos, CA, 1993, pp 246–253, 310–317.

<sup>10</sup>W. Nordhaus and G. Yohe, *Changing Climate*, report of the Carbon Dioxide Assessment Committee, National Academy Press, Washington, DC, 1983, pp 87–150.

<sup>11</sup>J. Goldemberg, T.B. Johansson, A.K.N. Reddy and R.H. Williams, *Energy for a Sustainable World*, World Resources Institute, Washington, DC, 1987; *op cit*, Ref 6, Shlyakhter and Kammen; Ref 9, Shlyakhter and Kammen; D.M. Kammen, A.I. Shlyakhter, C.L. Broido and R. Wilson, 'Non-Gaussian uncertainty distributions: historical trends and forecasts of the United States energy sector, 1983–2010', *ibid.*, pp 112–119.

<sup>12</sup>*Annual Energy Outlook with Projections to 2010*, Energy Information Administration, Office of Integrated Analysis and Forecasting, US Department of Energy, Washington, DC 20585, DOE/EIA-0383(92), 1992.

<sup>13</sup>*Ibid.*

<sup>14</sup>*Ibid.*

<sup>15</sup>*See op cit*, Ref 3, Murphy and Winkler; *op cit*, Ref 10.

<sup>16</sup>*Op cit*, Ref 12.

<sup>17</sup>*Ibid.*

<sup>18</sup>*Ibid.*

<sup>19</sup>*Op cit*, Ref 6, Shlyakhter and Kammen; Ref 9 Shlyakhter and Kammen.

<sup>20</sup>*Op cit*, Ref 6, Shlyakhter and Kammen.

<sup>21</sup>*Op cit*, Ref 11, Goldemberg *et al.*

<sup>22</sup>*Op cit*, Ref 12; United States Department of Energy, *National Energy Strategy*, US Government Printing Office, Washington, DC, 1991; C. Starr, M.F. Searl and S. Alpert, 'Energy sources: a realistic outlook', *Science*, Vol 256, 1992, pp 981–987.

<sup>23</sup>W. Nordhaus, 'An optimal transition path for controlling greenhouse gases', *Science*, Vol 258, 1992, pp 1315–1319; H. Dowlatabadi and M.G. Morgan, 'A framework for integrated studies of the climate problem', *Energy Policy*, Vol 21, 1993, pp 209–221.

<sup>24</sup>*Op cit*, Ref 6, Shlyakhter and Kammen.

<sup>25</sup>*Op cit*, Ref 10; Ref 23, Dowlatabadi and Morgan.

<sup>26</sup>*See*, United Nations, *World Population Prospects*, Population Studies Paper No 120, 1991; T. Freyka, 'Long term prospects for world population growth', *Population Development Review*, Vol 7, 1981, pp 489–511; N. Keyfitz, 'The limits of population forecasting', *Population Development Review*, Vol 7, 1981, pp 579–594.

<sup>27</sup>A.I. Shlyakhter and D.M. Kammen, *Estimating the Range of Uncertainty in Future Development from Trends in Physical Constants and Predictions of Global Change*, CSIA discussion Paper 92-06, Kennedy School of Government, Harvard University, July 1992.

<sup>28</sup>*Op cit*, Ref 10.

<sup>29</sup>A. Madansky, *Prescriptions for Working Statisticians*, Springer-Verlag, New York, 1988.

<sup>30</sup>S.S. Shapiro and A.J. Gross, *Statistical Modeling Techniques*, Statistics: textbooks and monographs, Vol 38, Marcel Dekker, New York and Basle, 1982.

# Appendix 1

## Previous case studies

Shlyakhter and Kammen analysed trends in 69 forecasts of the primary energy demand for the USA projected for the year 2000 (thick dotted line) and the predictions of population for 133 countries for the year 1985 made in 1973 (thick dashed line): Figure 1.<sup>24</sup>

The energy projections were from a variety of sources and were made between 1972 and 1990.<sup>25</sup> Estimates that were published as a range, for example 80–100 quads, were taken as two separate values representing the endpoints of a confidence interval. The data were grouped into bins spanning three to five years, each containing at least 11 estimates. A mean and

standard errors of the mean were then calculated for each bin. The 'true' value,  $T$ , for the year 2000 AD was taken to be an average over eight estimates published in 1990.

The population database included projections from 164 nations with population exceeding 100 000 and consisted of the upper ( $U$ ), lower ( $L$ ), and reference ( $R$ ) variants of the UN Population Studies series.<sup>26</sup> The projections were made in 1973 for the year 1985, which is the most recent year that the United Nations has released documented population figures for all of the countries in the database. Data for 31 countries were excluded

due to extreme errors ( $|x| > 100$ ) that resulted from unanticipated international migration (in several cases war refugees between relatively small nations), reliability questions surrounding particular census efforts, and clear cases of politically motivated reporting bias. Data for 114 nations satisfying the criteria  $|x| < 20$  were included in the study. The population forecasts for an additional 19 nations fall in the range  $10 < |x| < 20$ . An important area for future study is a detailed analysis and understanding of the reasons for the individual extreme cases in the population, AEO and other databases.

# Appendix 2

## The compound distribution

Let us assume that the projected reference value,  $R$ , is unbiased but that the estimate of the uncertainty as measured by the standard deviation of the equivalent normal distribution is randomly biased with a distribution  $f(t)$  where  $t = \Delta'/\Delta$ . Here  $\Delta$  is the standard deviation corresponding to the estimated uncertainty and  $\Delta'$  is the standard deviation corresponding to the 'true' uncertainty. In other words we assume that  $x' = (T-R)/\Delta'$  follows the standard normal distribution while  $x$  follows a normal distribution with standard deviation  $t$ :

$$p_t(x) = \frac{1}{\sqrt{2\pi}t} \exp\left(-\frac{x^2}{2t^2}\right) \quad (1)$$

Assuming the normal distribution from Equation (1) for each value of  $t$ , we obtain:

$$p(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{dt}{t} f(t) \exp\left(-\frac{x^2}{2t^2}\right) \quad (2)$$

If  $f(t)$  has a sharp peak near  $t=1$ , Equation (2) reduces to the normal distribution Equation (1). If  $f(t)$  is broad, however, the result is different. Consider the asymptotic behaviour of  $p(x)$  when  $|x| \gg 1$ . In this case, the value of the integral in Equation (2) is determined by the asymptotic behavior of  $f(t)$  as  $t \rightarrow \infty$ , as for small  $t$  the exponent is nearly zero. We assume that for  $t \gg u$   $f(t)$  follows a Gaussian law:

$$f(t) \sim \exp\left(-\frac{t^2}{2u^2}\right) \quad (3)$$

The new parameter,  $u$ , quantifies the width of the distribution  $f(t)$ . At  $x \gg 1$  the main contribution to the integral in Equation (2) comes from the vicinity of the saddle-point where the exponential term in Equation (2) reaches a maximum at  $t=t_{\max}$ :  $t_{\max}^2 = u|x|$ . For  $|x| \gg 1$ , the probability distribution is not Gaussian but exponential:

$$p(x) \sim \exp\left(-\frac{|x|}{u}\right) \quad (4)$$

The preceding analysis provides a convenient parametrization of the surprisingly frequent occurrence of extreme values in many models. In this fashion Gaussian and exponential distributions can be related by a single parameter  $u$ .

In this paper we use a truncated normal distribution for  $f(t)$ , where  $f(t)$  is zero for  $t < 1$  and follows Gaussian distribution with the mean value  $t = 1$  and standard deviation  $u$  for  $t > 1$  (multiplied by a factor of two to maintain the proper normalization of the probability density). We have:

$$f(t) = 0, t < 1$$

$$f(t) = \sqrt{\frac{2}{\pi}} \frac{1}{u} \exp\left(-\frac{(t-1)^2}{u^2}\right) t > 1 \quad (5)$$

This choice of  $f(t)$  is consistent with Equation (3) and produces the exponential asymptotic approximation for the compound distribution at large  $u$ . This definition reflects the fact that  $t < 1$  is highly improbable as it corresponds to underconfidence (estimated standard deviation  $\Delta' < \Delta$ ) and negative values of  $t$  that are impossible. The parametrization chosen here has the advantage over our previous parametrization<sup>27</sup> in that the effect of truncation does not depend on the value of  $u$ . Integrating Equation (2) gives the cumulative probability  $S(x)$  of deviations exceeding  $|x|$ :

$$S(x) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{u} \int_1^\infty \exp\left(-\frac{(t-1)^2}{2u^2}\right) \operatorname{erfc}\left(\frac{|x|}{t\sqrt{2}}\right) dt \quad (6)$$

For  $u = 0$  Equation (6) is reduced to  $S(x) = \operatorname{erfc}(|x|/\sqrt{2})$  and the probability distribution is Gaussian. On a logarithmic scale the exponential curves ( $u > 1$ ) are linear. Using normal distribution at  $u \sim 1$  underestimates the

probability of extreme events ( $x \geq 5$ ) by several orders of magnitude.

The compound probability distribution Equation (6) has the same asymptotic slope as  $\exp(-|x|/u)$  although they differ by a factor of about 1.5–2.0 that stems from the deviation from exponential behavior at small  $|x|$ . For quick estimates for  $u \geq 1$ ,  $x \geq 3$ , one can use the approximation  $\exp(-|x|/(0.7u + 0.6))$ .

Researchers have on occasion attempted to account for unknown errors by inflating the standard deviation of the Gaussian distribution.<sup>28</sup> We compared the inflated Gaussian parametrization with the exponential parametrization using formal statistical tests. Test for normality of the distribution of errors in energy forecasts based on the ratio of the sample range to the sample standard deviation<sup>29</sup> rejects Gaussian parametrization with a wide margin while a test for exponentiality based on Shapiro-Wilk W-statistics<sup>30</sup> cannot reject exponential parametrization. We note however that the particular model is case dependent and not necessarily unique. We use a Gaussian form for  $f(t)$  because it yields a simple exponential behaviour for  $|x| \gg 1$ . Other forms of  $f(t)$  can be employed that also fit the data. The choice of the most appropriate probability distribution for systematic uncertainties deserves further investigation and will almost certainly depend upon the application.

## Appendix 3

### Calculation of the new confidence intervals

Specify the subjective probability  $\alpha$  that the true value will lie between the low ( $L$ ) and high ( $U$ ) estimates. We assumed  $\alpha=68\%$ .

Draw an equivalent normal distribution that would have a specified cumulative probability  $\alpha$  between  $L$  and  $U$ . For  $\alpha=68\%$  the standard deviation of the equivalent normal distribution is  $(U - L)/2$ .

If the reference value ( $R$ ) is not the middle of the ( $L, U$ ) interval use two separate normal distributions truncated at zero: left half for ( $L, R$ ) interval and right half for ( $R, U$ ) interval

each having  $\alpha/2$  as the cumulative probability.

Estimate  $u$  from the historical data and calculate the new low ( $LN$ ) and new high ( $UN$ ) limits as follows:

$$\begin{aligned} LN &= R - Z(R - L) \\ UN &= R + Z(U - R) \end{aligned}$$

where the inflation factor can be read from the curves in Figure 2.

For the two-sided 90% confidence interval (5–95%),  $Z$  is the ratio of  $x(u)/x_0$ . Here  $x(u)$  is the value of  $x$  for

which  $S(x) = 0.1$ . For  $u = 0$ ,  $x_0 = 1.645$ . This gives:  $Z = 3.0/1.645 = 1.8$ ;  $4.6/1.645 = 2.8$ ;  $6.1/1.645 = 3.7$ ; and  $7.5/1.645 = 4.6$  for  $u = 1, 2, 3, 4$  respectively. From Figure 4,  $Z_{AEO} = 7.0/1.645 = 4.2$  corresponding to  $u = 3.4$ .

Note that had we assumed a number higher than 68% for the credibility uncertainty range estimated by the forecasters, then the standard deviation of the equivalent normal distribution would be smaller and  $Z$  values respectively higher than those listed above.