

**RANDOM ERROR AND SIMULATION MODELS WITH
AN UNOBSERVED DEPENDENT VARIABLE AS APPLIED TO
THE BENEFITS AND COSTS OF THE CLEAN AIR ACT**

Scott Farrow*

University of Maryland-BC and
The Woods Hole Oceanographic Institution

January 26, 2008

Abstract

Most empirical simulation models used in benefit-cost or risk analysis investigate uncertainty based on variability in parameters and conditioning factors. A pure random error term is frequently omitted. Ex-ante benefit-cost analyses create a particular problem because there are no historically observed values of the dependent variable, such as net present value. An estimator for the error variance is developed based on analysis of variance measures and R^2 . When applied to a model of the net benefits of the Clean Air Act, the probability that the net present value is negative increases from .2 to 4.5 percent.

JEL: H4, C5

Appreciation for comments is extended to Greg M. Duncan, Bob Hahn, David Hassenzahl and Andrew Solow. The author is responsible for any errors.

I. Random Error Estimation and Benefit-Cost Analysis

Many benefit-cost analyses provide point estimates, a few provide ranges or confidence intervals, even fewer provide the distribution of measures such as impacts, benefits, costs, or net present value. Although the information desires of policy makers for decision-making under uncertainty remain ambiguous, information about the distribution of performance measures is believed to be useful from either a theoretical or a descriptive perspective (Eckhout, et al., 2005; Farrow, et al., 2001; Hassenzahl, 2006). However, including a random error term in the typical benefit-cost analysis is complicated by the absence of observations on constructed measures such as net present value.

Government guidance recommends using the expected value for point estimates of benefits and costs but to include information on the statistical distribution of key measures for regulations with particularly large impacts (OMB, 1992, 2003).

Researchers are working more frequently to integrate risk and benefit-cost analysis and to report distributions of key performance measures. However, a random error component is generally not included in such distributions.

The most frequent use of benefit-cost analysis is to forecast, ex-ante, the welfare impact of a governmental action. In contrast to a model that estimates the mean of an observed outcome, say the number of accidents; in benefit-cost analysis a forecast is sought given some random parameters and conditioning factors but observations are lacking for the

dependent variable which exists only in the modeled world. Current practice in constructing a stochastic (simulated) benefit-cost analysis is to build, from the bottom up, an estimate of benefits and costs using distributions for both parameters and conditioning factors. Simulation tools are then used to sample over the distributions of the underlying parameters and factors. An example for one component is multiplying the estimated number of fatalities by a parameter representing the value of a statistical life. This “bottom-up” type of modeling typically excludes unexplained variation in the constructed model and understates the variance or other measures of dispersion of the constructed measure.

II. An estimator for random error with an unobserved dependent variable

A method that incorporates an estimate of the random error in a model is developed based on a linear benefit-cost model and common measures used in the analysis of variance.

Consider that many benefit-cost applications can be modeled as a vector product of parameters, $\underline{\beta}$, and variables, \underline{X} . Benefits are distinguished from costs by the sign of the parameter. Discounting is incorporated into appropriate parameters. Consequently measures such as net present value, Y , can often be expressed as:

$$Y = \underline{X}\underline{\beta} + \varepsilon \quad (1)$$

The random error term, ε , is unknown but if even considered is typically assumed to be normally distributed with mean zero and variance σ^2 . The analyst seeks an estimate of Y and information about its distribution without ever observing values of Y . The usual simulation approach to stochastic benefit-cost analysis utilizes either historical or subjective data to provide distributions for each element of β and \underline{X} to estimate the distribution of Y , including its mean and variance. The typically implicit random error is ignored.

In contrast, in the standard statistical setting values of Y exist and so estimates of ε can be constructed as the deviation between the true value and the estimated value and used to estimate the error variance. Estimators exist for the forecast variance when parameters are stochastic and the \underline{X} are fixed (e.g. Wooldridge, 2001, p. 216) or with stochastic \underline{X} (Feldstein, 1971) where the variance of Y would be determined by simulation methods in the latter case. Wooldridge speculates that in many applications, the random error will be a larger source of variation than uncertainty in the parameters. Given the lack of observed Y in benefit-cost analysis, how might one estimate the error variance?

A Bayesian approach could be utilized to directly estimate the error variance and the distribution based on the analyst's prior beliefs. Such elicitation would have little case specific structure on which to base an estimate.

The estimator developed here uses empirical information from the problem to anchor the size of the error variance, and uses a well known but bounded measure of fit, the R^2 , to

scale up or down the model sum of squares to obtain an estimate of the error variance. Given typical distributional assumptions, this is sufficient to define an estimator for ε in Equation 1. Most analysts are familiar with R^2 as a measure of fit of an equation and information is sometimes available on sub-models or calibration efforts in a particular setting. If the analyst can subjectively provide an estimate of R^2 , supported by whatever evidence is available, an estimator for the error variance can be developed as below with extensions presented at the conclusion.

Consider the standard definition of R^2 :

$$R^2 = 1 - SSE/SST \quad (2)$$

Where, from standard analysis of variance decompositions:

- SST, the total sum of squares=SSE + SSM where SSE is the error or residual sum of squares and SSM is the model or explained sum of squares.
- SSE/N, is the mean square error equal to $\sum(Y - \hat{Y})^2 / N$
- SSM/N, model mean square equal to $(\sum(\hat{Y} - \bar{Y})^2) / N$ noting that this value is observable from a numerical model if the analyst assumes $E(Y) = \bar{Y} = E(\hat{Y})$.
- $\hat{\sigma}^2 = SSE / N$ is a consistent estimator for the error variance

Suppose an analyst can provide a subjective estimate of R^2 , \hat{R}^2 . Substituting the component elements for the total sum of squares and dividing both numerator and denominator by N results in:

$$\hat{\sigma}^2 = \left(\frac{1 - \hat{R}^2}{\hat{R}^2} \right) \frac{SSM}{N} \quad (3)$$

As the model sum of squares (SSM) is calculable from a simulation model based on the error between the predicted and the average value, the right hand side is estimable given an estimate for R^2 . The resulting estimate of the variance can be used as the variance component of the random error term in equation 1.

III. Example: Uncertainty in the net benefits of air pollution control

Models have been developed to estimate the net benefits from controlling air pollutants such as particulate matter and sulphur dioxide (EPA, 2008; Farrow, et al., 2001).

Variability in parameters and conditioning factors has been built into the models. To demonstrate implementation of the estimator and its possible importance, we use an illustrative estimate of the fit of the model of .6 indicating that the model is believed to capture about 60 percent of the variability in net benefits. In actual practice, one may survey experts in the field or assess sub-models and obtain estimates through other means. The possible use of a distribution for this value is discussed later.

The model used for the base analysis is reported in Farrow, et al., (2001). The model was augmented to compute the model sum of squares based on the mean net present value of the net benefits of the Clean Air Act in the year 2010 of \$81,159 billion. Equation 3 was then used as an estimator for the error variance. The resulting estimate of the standard error for the random disturbance term was \$31,633 billion. A final simulation was conducted in which an additive, normal, zero mean, standard error as above was added to

the base net present value. The result increased the standard error of the net present value from \$36,669 billion to \$48,330 billion. The probability that the net benefits were below zero increased from .2 percent to 4.5 percent.

These changes may be viewed as statistically significant, for instance a one tailed test at the 99% level that the net benefits are equal to zero would have been rejected in the first instance but not in the second. However, the policy process incorporates uncertainty in a much more ambiguous manner. Incorporating additional uncertainty through an estimate of the random error may simply help inform the decision-maker that there is a larger probability in the extremes of the distribution that would be estimated by the standard method.

The approach may be extended in several ways. Instead of a point estimate for R^2 , the analyst may specify a distribution. In the simulation the error term may be drawn from a distribution defined by a random draw for R^2 . Alternatively, the analyst may approximate the required R^2 to achieve a given level of significance, although this requires additional distributional assumptions as might justify use of a t test.

Consequently, there appear to be several implementable approaches to incorporating random error into simulation models when the dependent variable is unobserved.

Including random error may influence policy decisions if the result indicates that greater uncertainty may exist about the outcome.

References

1. Eckhout, L., C., Gollier, and H. Schlesinger, Economic and Financial Decisions under Risk, Princeton: Princeton University Press, 2005.
2. Environmental Protection Agency (EPA), “Environmental Benefits Mapping and Analysis Program: BenMAP,” accessible at <http://www.epa.gov/air/benmap/> , 2008.
3. Farrow, S., R. Ponce, E. Wong, et. al. “Facilitating Regulatory Design and Stakeholder Participation” in Improving Regulation: Cases in Environment, Health and Safety, S. Farrow and P. Fischbeck, eds., Resources for the Future Press, Washington, D.C. 2001.
4. Feldstein, M., “The Error of Forecast in Econometric Models when the Forecast-Period Exogenous Variables are Stochastic,” Econometrica, 39(1): 55-60, 1971.
5. Hassenzahl, D. M., "Implications of Excessive Precision for Risk Comparisons: Lessons from the Past Four Decades." Risk Analysis, 26(1):1–12, 2006.
6. Office of Management and Budget (OMB), “Regulatory Analysis,” Circular A-4, Washington, D.C., 2003, accessible at <http://www.whitehouse.gov/omb/circulars/index.html> .
7. _____, “Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs,” Circular A-94, Washington, D.C., 1992, accessible at <http://www.whitehouse.gov/omb/circulars/index.html> .
8. Wooldridge, J., Introductory Econometrics: 3rd Edition, Thompson-Southwestern, Mason, Ohio. 2005.