CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

The Value of Sample Information for Water Quality Management

Authors

Hwansoo Sung
Ph.D Candidate of Agricultural and Environmental Economics
Department of Agricultural Economics and Rural Sociology
308 Armsby Building
Pennsylvania State University
University Park PA 16802
Phone: 814-863-8248

James Shortle
Professor of Agricultural and Environmental Economics
Department of Agricultural Economics and Rural Sociology
112 Armsby Building
Pennsylvania State University
University Park PA 16802
Phone: 814-865-7657

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005

Copyright 2005 by Hwansoo Sung and James Shortle. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract

There is considerable interest in watershed-based pollution water quality protection but the approach can be highly information intensive (USEPA 2004, NRC 2000). This study examines the value of different types and levels of information for water quality management in the Conestoga watershed. For this estimation, a Monte Carlo procedure is used to construct the posterior expected value. Then, an Evolutionary Optimization Strategy with Covariance Matrix Adaptation (CMA-ES) is used to compute the expected value of optimized resources allocations given posterior information structures for specific sample sizes. This posterior optimization is nested within a second Monte Carlo simulation that computes the preposterior expectation (a nested Monte Carlo procedure). Thus, this paper provides some insight about the relative values of these alternative types of information for controlling water pollution from agriculture, and the gains from more intensive sampling.

I. Introduction

The U.S. Environmental Protection Agency's (EPA) Total Maximum Daily Load (TMDL) initiative requires states to develop and implement watershed-based plans for surface waters that do not meet in-stream water quality standards even after point sources of pollution have installed the minimum required levels of pollution control technology (Ribaudo 2001, USEPA 2004a). The states must identify the maximum total pollution load consistent with satisfying the water quality standards and allocate the loads among point and nonpoint sources. While there is much to be said in favor of this comprehensive, watershed-based approaches to water quality protection, it is also clear that implementation requires much more information about pollution sources, water quality conditions, relationships between land uses and pollution loads, and pollution loads and water quality conditions than does the traditional approach (NRC 1999, 2000). And, as highlighted by the recent National Research Council (NRC) report on the TMDL approach, essential information is often lacking (NRC 2000). The report emphasizes the essential role of information acquisition by water quality managers for improving the effectiveness and efficiency of water quality management. However, given that information acquisition is costly, to make good use of scarce resources for water quality management calls for attention to the benefits and costs of information collection.

This paper examines the sample value of various types of information for water quality management. Value of information studies often focus on the value of perfect information. A recent example, relevant to this study is Borisova et al (2004), which estimates the value of perfect information about the benefits and costs of water quality

protection under alternative water quality policy regimes. However, because perfect information is an unrealistic goal, a more meaningful measure is the expected value of sample information (EVSI) that reduces but does not eliminate uncertainty. The expected value of sample information (EVSI) that reduces but does not eliminate uncertainty is a widely cited but little used measure of the contribution of information to decision making (Yokota and Thompson 2004).

Calculating EVSI requires preposterior knowledge of how newly added information is used to update the posterior density functions of uncertain parameters. Then, using this preposterior knowledge, EVSI is generally defined as the difference between the expected value of optimal action selected with the updated posterior probability of parameters, and the expected value of optimal decision selected only with the prior information about the parameters.

We analyze the EVSI of various types of economic and biophysical parameters in the context of nitrogen pollution control from agricultural nonpoint and municipal point sources in the Conestoga watershed of Pennsylvania. The Conestoga is a major source of nutrients entering the Susquehanna River and in turn the Chesapeake Bay. We take the objective of water quality management in the Conestoga to be maximization of the expected benefits less the expected costs of nitrogen pollution control regulations. Uncertainty is modeled from the perspective of a social planner seeking to develop an ex ante efficient allocation of resources for water quality protection. The planner is uncertain of:(1) the private costs of changes in resource allocation for nitrogen pollution control, (2) the relationships between land use practices and nitrogen load at the mouths of the watersheds, (3) the transport of pollution loads from the watersheds to the Bay, and

(4) the economic benefits of reduced pollution loads. In this context, sample information about the information components, has value when its acquisition and use increases the posterior net benefits.

The analysis is performed using a model that couples economic and biophysical components to simulate nitrogen delivery from point and nonpoint sources to the Bay from the Conestoga (non-point water pollution). The unknown parameters are treated as random variables with known distributions from the planner's perspective. The EVSI is computed for individual parameters, sets of parameters, and alternative sample sizes to learn how different types of information contribute to the water quality management, and to learn how the value of information changes with the extent of information acquisition.

A nested Monte Carlo procedure is used in combination with the Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) to compute the expected value of optimized resources allocations under alternative information collection strategies. The CMA-ES is an evolutionary (search) algorithm for highly nonlinear optimization problems. The CMA-ES is typically applied to unconstrained or bounded constraint continuous optimization problems, and search space dimensions between three and a hundred. The CMA-ES is used to compute agricultural practices and point source abatement levels that maximize expected net benefits given posterior information structures for specific sample sizes. A Monte Carlo procedure is used to construct the posterior expected value. This posterior optimization is nested within a second Monte Carlo simulation that computes the preposterior expectation.

II. Expected Value of Sample Information

Information is valuable when reducing uncertainty leads to better decisions. Value of information (VOI) analysis provides a quantitative means to assess the gains from improving information. Policymakers can use this analysis to determine which uncertainties, if reduced, would change their decisions and thus give them a better idea of where resources should be devoted to research. Thus, the VOI framework can provide helpful insights for determining the appropriate balance between taking action and waiting for more information.

To examine how EVSI is calculated, consider the following maximization problem

(1) $\underset{X}{Max}\, f(X, \theta_1, \theta_2)$

where X is a control variable, θ_1 and θ_2 are unknown parameters, f(.) is a concave objective function. In this study, θ_1 and θ_2 can be bio-environmental resource factors. So, it is most impossible to get true values of these factors. Instead, what decision maker (social planner) only knows is information of prior distribution s to these factors. With EVSI analysis, these input distributions are simulated as a solution technique. The simulation approach tends to yield estimates of the distributions closer to those of the perfect information by simply increasing the number of trials (Monte Carlo simulation). Then, the newly added information on site-specific observed data is combined with prior information on parameter distributions so that posterior probability distributions of the random parameters are derived. Bayesian inference affirms that these posterior distributions will contain less uncertainty than the prior distributions.

Assume that only information of θ_1 is updated by the sampling. That is, set $\theta_2 = \theta_2^B$ (θ_2^B is a baseline value; no additional information). Then, with sample size m, the updated information can be reflected to posterior probability of parameter θ_1, $P(\theta_{1i} \mid \theta_{11}, \theta_{12}, ..., \theta_{1m})$ where $\theta_{11}, \theta_{12}, ..., \theta_{1m}$ are observed data on θ_1 through m times of sampling. On the other hand, social planer tries to optimize the objective function with realized value of θ_1, θ_{1i} at each sample i (i = 1,…,m) such as

(2) $$L(\theta_{1i}, \theta_2^B) = \underset{X}{Max}\, f(X, \theta_{1i}, \theta_2^B) \text{ for i = 1,…,m}$$

In the next step, with posterior posterior probabilities of parameter θ_1, $P(\theta_{1i} \mid \theta_{11}, \theta_{12}, ..., \theta_{1m})$ for all i, a Monte Carlo procedure (Rubinstein (1981), Borisova et al (2005)) is used to estimate expected optimal value of objective function such as

(3) $$J(\theta_{11}, \theta_{12}, ..., \theta_{1m}, \theta_2^B) = \sum_{i=1}^{m} P(\theta_{1i} \mid \theta_{11}, \theta_{12}, ..., \theta_{1m}) \times L(\theta_{1i}, \theta_2^B)$$

Finally, if the baseline optimal value (V_B) with the baseline information of $\theta_1 = \theta_1^B$ and $\theta_2 = \theta_2^B$, is derived such as $V_B(\theta_1^B, \theta_2^B) = \underset{X}{Max}\, f(X, \theta_1^B, \theta_2^B)$, then, the expected value of sample information of sample size, m (EVSI(m)) can be expressed as

(4) $$EVSI(m) = J(\theta_{11}, ..., \theta_{1m}, \theta_2^B) - V_B(\theta_1^B, \theta_2^B)$$

In general, the data collection costs should be compared to the EVSI to determine the optimal data collection strategy as follows. If there is a data collecting cost C(m) of sample size m, the optimal sample size (m*) can be earned by solving the second optimization,

(5) $$\underset{m}{Max}\, EVSI(m) - C(m)$$

The first order condition of (7) is,

$$(6) \qquad \frac{\partial EVSI(m)}{\partial m} = \frac{\partial C(m)}{\partial m} = 0$$

From (6), the optimal sample size m* is earned (Figure 1).

III. Conestoga Watershed Model

In the following, for the Conestoga watershed in the Pennsylvania portion of the Susquehanna River Basin (SRB), a hypothetical planner is assumed to maximize the expected net social benefit of water quality protection from agricultural nonpoint sources, and point sources of pollution. For the analysis, agricultural land use, associated nitrogen loadings and point source emissions are the instrument variables that are targeted by the planner.

The Conestoga watershed model consists of an economic model of agricultural production and pollution control decisions, point source pollution control costs, a model that quantifies nutrient transport, and the economic costs of nutrients entering the Chesapeake Bay from the watershed.

1. Economic model for agricultural nonpoint sources

Corn production in the watershed is a function of farmer decisions involving the use of land (L) and nitrogen fertilizer (N). The corresponding profit equation associated with these input choices is defined generally as $\pi(\mathrm{N}, \mathrm{L})$. Agricultural land is scarce and earns economic rents. Accordingly, agricultural land rents are considered when defining social surpluses from agricultural production. The net benefits (NB) to nonpoint sources in the watershed in is expressed as

(7) NB = π(N, L) + R(L)

where land supply and demand are same in the equilibrium.

2. Nonpoint nutrient loadings model

Following Borisova et al. (2005), the expected annual load to the mouth of the Conestoga watershed b as a function of nitrogen concentration in runoff N_c, agricultural land area, and mean annual precipitation Z:

(8) $$b = \left[\varphi_1 Z\mu(1-u)N + \frac{\varphi_2 (Z\mu(1-u)N)^2}{L} + \varphi_3 Z \right]$$

where φ_1, φ_2, and φ_3 are coefficients. Here, nitrogen concentration N_c is defined as the ration of nitrogen runoff mass ((1-u) N) and water volume (Z×L):

$$N_c = \mu \frac{(1-u)\cdot(N/L)}{Z}$$

where μ is a calibration coefficient, and u is the share of applied nitrogen which is taken (utilized) by plants. Thus, nonpoint loadings function for the watershed is defined as l(N, L, Z) such as $\frac{\partial l}{\partial N} > 0, \frac{\partial l}{\partial L} < 0$. Precipitation Z is stochastic in the simulation.

3. Point source model

The total cost of abatement depends on the present level of abatement $a_p = e_0 - e_p$ where the abatement level are physically bounded at its upper bound, e_0. Reflecting this bound e_0, total abatement cost is expressed as

(9) $$C(MC_0, e_0, e_I) = MC_0 \times (e_p - e_0) \times \ln\left(\frac{e_I - e_0}{e_p - e_0}\right)$$

where e_I is an emission level of non-control case and MC_0 is marginal cost of baseline information.

4. Nutrient Delivery

Only a fraction of Nonpoint source loadings and point source emissions are combined to build an ambient concentration of nitrogen in the Bay. The proportion of the load that is delivered is modeled with constant delivery coefficient ϑ (Horan et al., 2001), so that total delivered nitrogen load from the Conestoga watershed to the Bay (a) is

(10) $\quad a = \vartheta b + \vartheta e_p$

The transport coefficient ϑ is imperfectly known, and is modeled as random variables with a mean and variance.

5. Economic Damages from pollution

Following Borisova et al. (2005), the mean annual damage from the Conestoga nitrogen loads to the Chesapeake Bay is modeled as a convex increasing function of the total nitrogen load, a to the Bay such as

(11) $\quad D(a) = \rho a^q$

where D is economic damage, ρ is a coefficient, q is elasticity of damage function, and $\partial D/\partial a > 0, \partial^2 D/\partial a^2 > 0$. To reflect the social planner's imperfect knowledge about environmental damage, both parameters ρ and q are random variables.

6. Social Net Benefit (SNB)

Combining the above equations ((7)-(11)), a 'Social Net Benefit (SNB)' function is constructed, which represents net economic returns in the consideration of negative externality of nitrogen residuals in the Bay such as

$$(12)\ SocialNetBenefit = \pi(N,L) + R(L) - MC_0 \times (e_p - e_0) \times \ln\left(\frac{e_I - e_0}{e_p - e_0}\right) - \rho(\vartheta b + \vartheta e_p)^q$$

Social planner tries to maximize SNB (12) with respect to L, N and e_p. At last, planner has the following uncertainty about the values of six parameters: 1) the planner has imperfect information on the pollution transport parameters (Z, ϑ), 2) the planner has imperfect information about substitution elasticity between land and nitrogen fertilizer (σ), 3) the planner has imperfect information about abatement cost parameters (MC_0) and 4) the planner has imperfect information about the damage cost parameters (ρ and q).

Among six random parameters, for the concavity of objective function SNB baseline point source marginal abatement cost is assumed to be 100,000 after several candidate values are examined by authors. The information of distributions of rest of uncertain parameters is earned from Horan et al (2002) and Borisova et al (2005). The random and fixed values are used in the analysis (Table 1).

In Figure 2, a flow chart for estimation of for the value sample information is presented. At each iteration i of sample size M, random numbers of five uncertain parameters are generated. Then, we do optimize the objective function, SNB_i and get ith optimal value of social net benefit (J_i). After all the M optimizations are performed, these optimal

values are summed up and averaged to earn the average optimization value ($\hat{J}_M^s$) from sample size M. To test the robustness of the sample mean of optimal values with sample size M, the expected social net benefit calculation are repeated S times in the outer circle of the flow chart and acquire the representative social net benefit of sample size M ($\hat{\hat{J}}_M^S$) (Nested Monte Carlo Simulation). Here, to get the ith optimization value J_i, varying variables, L, N and e_p, we should extract Kuhn-Tucker condition. However, since J_i itself has a nonlinear form, we cannot get the optimal condition directly. Thus, among heuristic methods, the Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) is used to get ith optimization value J_i (i = 1,…., M). The model is computed using the CMA-ES by Matlab 7.0.

IV. The CMA Evolution Strategy for Noisy and Global Optimization

The CMA-ES (Evolution Strategy with Covariance Matrix Adaptation) is an evolutionary (search) algorithm for difficult optimization problems. The CMA-ES is typically applied to unconstrained or bounded constraint continuous optimization problems, and search space dimensions between three and a hundred. The method should be applied, if derivative based methods, e.g. conjugate gradient, fail due to a rugged search landscape (e.g. discontinuities, sharp bends, noise, local optima, outliers).

Originally designed for small population sizes, the CMA-ES was interpreted as a robust local search strategy. In Hansen and Kern (2004), the CMA-ES was expanded by the so-called rank-μ-update. The rank- μ-update exploits the information contained in large

populations more effectively, because the algorithm selects only μ best individuals in the next generation. It can reduce the time complexity of the strategy from quadratic to linear. Similar to quasi-Newton methods the CMA-ES estimates the inverse Hessian matrix (here: the covariance matrix) within an iterative procedure. In the end, any convex-quadratic (ellipsoid) objective function such as social benefit function in this study is transformed into the spherical function (Hansen, 2005). This can improve the performance on ill-conditioned problems. In addition, the CMA-ES has several invariance properties. Two of them are (i) invariance against order preserving (i.e. strictly monotonic) transformations of the objective function value, and (ii) invariance against angle preserving transformations of the search space (including rotation, reflection, and translation), if the initial search point is transformed accordingly. These invariances are highly desirable, because they imply uniform behavior on classes of functions and therefore allow for generalization of empirical results. The complete algorithm is presented in Hansen and Kern (2004), and Hansen (2005).

V. CMA-ES Application to Simulations

In this section, we apply CMA-ES to optimization problem of water quality management in Conestoga, Pennsylvania. The non-linear objective function of social net benefit is described in chapter III. Here, five randomly generated parameters are implemented into the optimization problem: substitution elasticity between land and nitrogen fertilizer (σ), damage exponent (q), damage coefficient (ρ), transport coefficient for watershed in Conestoga river (ϑ), and regional precipitation (z).

This problem is entered into CMA-ES Matlab code (Hansen, 2005). We collect the information of all five randomly generated parameters, simultaneously in every sample trial. Then, we perform the optimization procedure at each sample in the same manner with the perfect information case. After that, we sum up all the optimized values of social net benefit and divide the sum by the number of samples (m). This final value is the optimized social net benefit of the sample size, m. For each sample size, we iterate 100 times so that a representative optimized values of social net benefit can be earned at each sample size. In the same manner, ex post[1] mean quantities of three control variables (N, L and e_p) are calculated such that every ex post optimal quantity of each control variable is summed up and averaged.

VI. Conclusion

As water quality protection has become one of critical issues in the regional systems in the U.S., information acquisition on the water quality management has been crucial topic in the environmental science. Given that collecting information is costly and imperfect, strategic information acquisition is essential to improving performance of water quality management. Onto this necessity, the concept of expected value of sample information offers a tool for evaluating alternative types and amounts of information. As a contribution, this paper provides some insight about the relative values of these alternative types of information for controlling water pollution from agriculture in the Conestoga watershed, and the gains from more intensive sampling. From a water quality management model EVSI analysis is performed using Monte Carlo method. In the

[1] Ex post quantity is defined as the quantity level that social planner actually chooses in response to the realized values of uncertain parameters at each replication.

analysis, it is shown that incorporating new data to the decision framework leads to the updated information state with reduced uncertainty from which better decision may follow.

EVSI is a measure of the value of the reduction in uncertainty that may result from the collection of new information. For the maximization of SNB in Conestoga watershed, EVSI involving larger number of data can be expected to increase up close to the expected value of perfect information. The EVSI can be used as an upper bound on what should be spent on data collection. If the cost of data collection can be estimated, these costs can be compared to the EVSI to determine the optimal data collection strategy (equation (6)).

As an analytical challenge in EVSI, the correlation in input distributions and dependence in information collected can be examined. For example, precipitation in the region can affect the productivity of corn production. Generally, climate factors such as monthly average temperature and precipitation are expected to have bio-chemical effects on the water pollution level such that there are differences in level of pollution by the various regional climate factor conditions even if emissions of pollutants are recorded at the same level. Accordingly, parametric assumptions on precipitation and usage rate of nitrogen fertilizer taken up by the plant, etc., can influence the EVSI through social planner's decisions. Over the correlations, a sensitivity analysis can be conducted to examine the effect of parametric assumptions on the optimal decision and the EVSI. Since plenty of rain can deliver most of nitrogen fertilizer residuals to the Bay, the optimal decision is

expected to be sensitive to the distribution of precipitations in the watershed. Similarly, the optimal decision and the EVSI are expected to be sensitive to the change of usage rate of fertilizer such that higher usage rate makes a control of nitrogen fertilizer more flexible.

References

Abler, D, J. Shortle, J. Carmichael, and R. Horan (2001) Climate change, agriculture, and water quality in the Chesapeake bay region. *Climatic Change, 55(3),* 339-359.

Borisova, T., J. Shortle, R.D. Horan and D. Abler (2005) The value of information for water quality management. *Water Resources Research*, (In Press).

Daskins, M.E., J.E. Toll, M.J. Small and K.P. Brand (1996) Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis*, 16(1).

Dilks, D.W., R.P. Canale and P.G. Meier (1992) Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling*, 62, 149-162.

Hansen, N. (2005) The CMA Evolution Strategy: A Tutorial.

Hansen, N. and S. Kern (2004) Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Eighth International Conference on Parallel Problem Solving from Nature PPSN VIII, Proceedings*, pp. 282-291, Berlin: Springer.

Horan, R., J. Shortle, and D. Abler (2002) Point-nonpoint nutrient trading in the

Susquehanna river basin. *Water Resources Research, 38(5)*.

Horan, R. and J. Shortle (2001) Environmental Instruments for Agriculture. In: *Environmental Policies for Agricultural Pollution Control* / Shortle J. S. and D. Abler (eds.) CABI Publishing, Oxon (UK), NY (USA).

National Research Council (NRC) (2000) *Assessing the TMDL Approach to Water Quality Management*. National Academy Press, Washington, DC.

National Research Council (NRC) (1999) *New Strategies for America's Watersheds*. Washington, DC: National Academy Press.

Ribaudo M. (2001) Non-Point Source Pollution Control Policies in the USA. in J.S. Shortle and D.G. Abler (Ed), *Environmental Policies for Agricultural Pollution Control*. CABI Publishing, Oxon, UK and NY, USA.

Ribaudo, M.O., R.D. Horan, and M.E. Smith (1999) *Economics of Water Quality Protection from Nonpoint Sources: Theory and Practice*. Agricultural Economic Report Number 782. USDA. http://www.ers.usda.gov/publications/aer782/ (last accessed September 19, 2003)

Ribaudo, M.O., and J.S. Shortle (2001) Estimating Benefits and Costs of Pollution Control Policies, in J.S. Shortle and D.G. Abler (Ed), *Environmental Policies for Agricultural Pollution Control*. CABI Publishing, Oxon, UK and NY, USA.

Rubinstein, R. Y. (1981) *Simulation and the Monte Carlo Method*. John Wiley and Sons, Inc., New York, NY.

Shortle, J. S. and R. D. Horan (2001) The Economics of Non-Point Pollution Control. *Journal of Economic Surveys*, *15(3)*, 255-289.

Varian, H.R. (1992), *Microeconomic Analysis*. WW Norton and Company, NY

Yokota, F. and K.M. Thompson (2004) Value of Information Analysis in Environmental Health Risk Management Decisions: Past, Present, and Future. *Risk Analysis, 24(3)*, 635-650.

Table 1. Model Parameters

Variable	Notation	Distribution Characteristics
Substitution elasticity between land and nitrogen fertilizer	σ^a	Uniform, mean =1.25, variance = 0.025
Damage exponent	q^b	Uniform, mean = 2, variance = 0.1089
Damage coefficient	ρ^b	Uniform, mean = 1.2×10^{-4}, variance = 4.41×10^{-18}
Transport coefficient for watershed 2	$\vartheta_2{}^a$	Gamma, mean = 0.731, variance = 0.114
Load regression coefficient φ_1	$\varphi_1{}^b$	Deterministic, mean = 646×10^{-5}
Load regression coefficient φ_2	$\varphi_2{}^b$	Deterministic, mean = 8602×10^{-11}
Load regression coefficient φ_3	$\varphi_3{}^b$	Deterministic, mean = 136×10^{4}
Calibration coefficient μ	μ^b	10^5
Precipitation, millimeters	$z_i{}^a$	Gamma, mean = 40.19, variance = 7.943
Proportion of nitrogen taken up by the plant	u^b	*0.7*
Baseline point source marginal abatement cost	***MC₀***	***100,000***

Sources for values: a = Horan et al (2002), b = Borisova et al (2005)

Figure 1. The optimal sample size m*

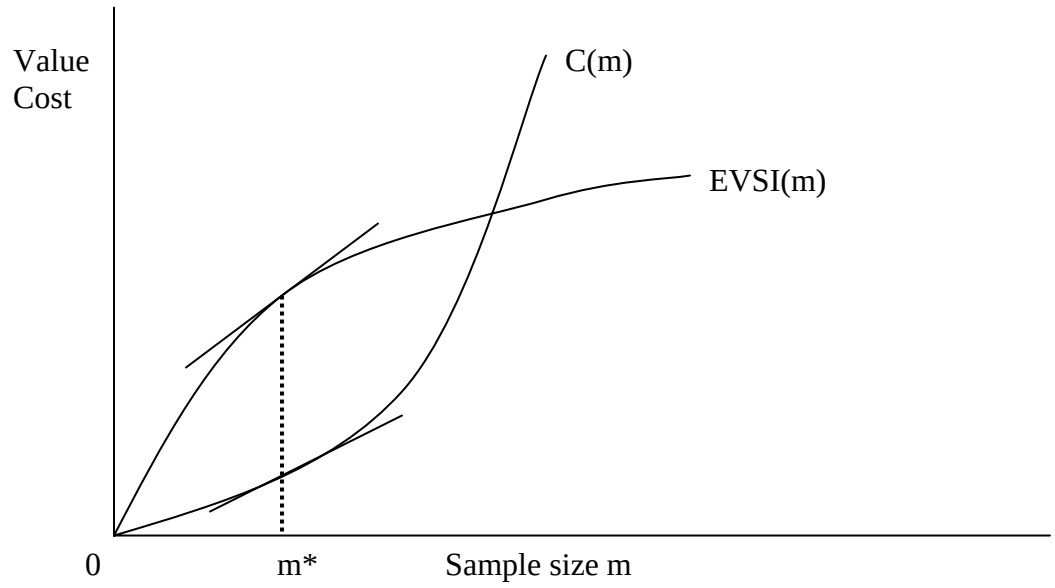

Figure 2. Flow Chart of Simulation for Sample Information Analysis