

# Simulation modeling for cost estimation

BY RICHARD WATERMAN

*Department of Statistics, Wharton School, University of Pennsylvania,*  
*and*

DONALD RUBIN

*Department of Statistics, Harvard University,*  
*and*

NEAL THOMAS

*Datametrics Research Inc.*

*and*

ANDREW GELMAN

*Department of Statistics, Columbia University.*

Draft of June 9th 1999

Prepared for the Seventh Conference on Postal and Delivery Economics: Current  
Directions in Postal Reform. June 23–26, 1999, Sintra, Portugal.

---

The authors acknowledge the contribution of many individuals to both the design and construction of the model. Paul Kleindorfer and Michael Crew were instrumental to the design of the underlying econometric models. Without the inputs from the U.S. Postal Service, provided by Ross Bailey, John Reynolds and their staff, there would be no data on which to run the model. The members of the LINX DQS team turned the Postal data

into formats amenable for input to the economic models. Many discussions with both the Postal Rate Commission and the General Accounting Office helped steer the modeling in relevant directions.

# 1 The simulation model context

## 1.1 Introduction

What do a group of statisticians have to contribute to the subject of good costing practices, and in particular the United States Postal Service (USPS) costing methodology? Credible cost estimates require the collection of high quality information as component inputs. Deciding what information to collect lies in the province of the economist, but exactly *how to collect* that information, *when to collect* it, *how much of it to collect*, as well as a significant part of the overall evaluation of the quality of the information itself, lies in the province of the statistician.

Our involvement with the Postal Service costing dialog began as members of the LINX team on a large scale Data Quality Study (DQS) of Postal Service data inputs to the Postal Service rate making process. This study took place from June 1997 to April 1999 and is fully described in the Summary Report and four supporting Technical Reports [LINX, 1999]<sup>1</sup>. Though the study was broad in its approach, including economic, statistical and an Industry Survey studies, one component involved the construction of a simulation model to investigate a variety of questions including the overall quality of specific marginal cost estimates, as well as an examination of issues and concerns raised by intervenors during various Postal Service rate hearings. This paper describes the rationale for the simulation model, explains the key ideas on which it is founded, and illustrates its use. Furthermore, it expands on some of the insights provided by the model.

In particular, among the benefits of the simulation model approach are that it forces the user to think hard about their assumptions and to focus on what exactly it is that needs to be measured. Thereby it may provide a means of exploring conjectures and their consequences, of different or even opposing viewpoints.

## 1.2 The role of a simulation model within cost accounting systems

Accurate costing of products is an essential activity within any large company with a diverse product mix. It is a key requirement for identifying the organizations ultimate profitability. A diverse and complex product mix is likely to require

---

<sup>1</sup>Available from <http://www.usps.gov/clr/dqs.htm>. A full description of the simulation model, its results and conclusions constitute Technical Report #3.

an involved process to reveal individual product costs. In these circumstances it can be a major achievement simply to arrive at a product level cost estimate. However, there is a second and even more demanding dimension to the cost estimation process: to ask how reliably (described in terms of *precision* and *accuracy*) those costs have been estimated. If we agree that it is important to estimate costs, then it is clearly equally important to quantify the quality of those cost estimates. Cooper and Kaplan [1991, pg. 368] discuss possible reasons for, and the impact of, measurement errors in cost management systems.

The simulation model is one way to approach this second-level question – the question that asks, not simply “how should we estimate costs”, but adds caveats “how well have these costs been estimated”, and “what are the likely consequences of potential errors in the cost estimation process”.

### **1.3 The multi-product multi-driver firm**

Because costs arise from a variety of sources, it is necessary to construct a cost formulation that incorporates a range of cost drivers. In the case of the Postal Service, there are nineteen separate cost segments involved, (for example, Purchased Transportation, Supervisors and Technical Personnel *etc.*), which are further subdivided into 59 cost components.

In addition, the annual Cost and Revenue Analysis (CRA) presents attributable costs from numerous categories of mail and services. Though the CRA costs are based on accounting records, the accounts do not differentiate the costs by class and subclass of mail. In order to provide this breakdown by mail class and subclass, additional sources of information have to be utilized. These sources include large scale multi-stage sample surveys, operating data systems and special purpose econometric studies. Data from these sources most often makes their appearance in (i) the distribution keys used to distribute the attributable cost, and (ii) the elasticities of accrued component cost with respect to the cost driver. Due to the diversity of the inputs to the cost calculations, it is extremely difficult to identify analytically the quality of the resulting cost estimates, despite being a very legitimate question to ask. Further, one of the cost measures of interest to the Postal Service, the marginal cost estimate (Unit Volume Variable Cost in Postal Service parlance) is calculated by combining four multiplicative factors:

1. The accrued cost.
2. The elasticity of the accrued cost with respect to the cost driver.

3. The elasticity of the driver with respect to mail volumes (Distribution Key Share – DKS).
4. The mail volumes.

Bradley et al. [1993] and Panzer<sup>2</sup> provide a detailed description of this product costing procedure.

An estimate of each of the four components must be derived, and it is not at all obvious how the uncertainty in each component relates to the uncertainty in the overall marginal cost estimate. One immediate practical application that results from measuring this uncertainty is that it enables the analyst to begin addressing the following question “if there were one million dollars to spend on improved information collection, where should those dollars be spent – better elasticity estimates, distribution key shares or volume estimates”. Furthermore, the simulation model helps direct the analyst to the specific cost components (for example Delivery, Transportation or Mail Processing) where better component estimates would provide significantly better overall cost estimates.

## **1.4 The Data Quality Study**

Information Technology is often cited as the key driver of current productivity increases. A sometimes overlooked component is the raw material of the IT system itself, that is the data/information that these systems work with. If the IT system is ideally a machine that constructs knowledge, then what the DQS looked at was the sometimes less than glamorous, but clearly essential, raw material inputs to the machine. Simply put, without quality inputs there are unlikely to be quality outputs.

## **1.5 A neutral analytical tool**

Value from an endeavor often arises indirectly, even serendipitously, and that appears to have been the case during the implementation of the simulation model. The reason why value from the model may be gained indirectly, is that the construction of an acceptable cost estimation model requires *dialog* concerning:

- An agreed upon language.
- The definition of terms.

---

<sup>2</sup>Testimony in Docket No. R97-1

- The formulation of an accepted cost–generation process.
- The relevant outputs from the model.

As such, the model may play the role of a *rule book* in a sporting contest. The players should agree on the rules *a priori* and accept the outcome. This is not to make the claim that the rules should be immutable; clearly, over time adaptation of the rules is a necessary consideration, but for any particular game they should be fixed.

Two examples follow that illustrate the manner in which the model lead to potentially useful insights. The project had initially focused entirely on marginal cost estimates for the mail subclasses of interest. After consideration of results from the model, parties to the project began to focus attention also on *relative* marginal costs, which represented a major change in the main outcome measure of interest.

Further, concern had been expressed over the consequences from a recent decrease in data collection resources in the core statistical sampling systems. The model simulation model suggested that this was not of primary concern, because there were other components in the cost estimation process that contributed more to the overall uncertainty of cost estimates. This example illustrates how the simulation model offered the potential to focus attention on those parts of the process that were most influential with respect to the outcome of interest.

## 2 What we want to measure and why

### 2.1 Key objectives

The principle objectives of the simulation model were two-fold

1. To calculate marginal costs of a set of subclass products and to ascertain the precision of those estimates – in colloquial parlance, to determine whether the marginal cost estimates were hard or soft.
2. To investigate and quantify the sensitivity of the marginal cost estimates to potential systematic errors in the inputs, termed *biases*.

The quantification of these factors enables the strong and weak components of a complex system to be isolated. If we consider the entire costing process as a very large black box, with all raw data inputs entering one side of the box and the final

cost estimates emerging from the other, then the simulation model corresponds to taking the lid off the box and systematically, revealing, diagnosing and perturbing its inner workings. In addition it can help identify which components are vital to the smooth running of the machine, and others that may be somewhat peripheral.

This type of information should be useful for purposes other than solely the rate setting environment. This is because the real goal here is to measure both the cost causation relationships as well as how much is known about those relationships. This data quality information addresses the issues of where to collect information to describe with sufficient accuracy these cost causation characteristics of the firm. That is, this information should be valuable to help manage the business as well as to set rates.

## **2.2 Marginal Cost / Unit Volume Variable Cost**

The simulation model was built with the objective of measuring marginal cost at the subclass level, in Postal Service parlance, Unit Volume Variable Cost (UVVC). UVVC has been accepted as a fundamental input for the rate making process<sup>3</sup>. However, though rates are built from marginal cost, their final determination involves an additional set of inputs and judgments, for example the *mark up* process, that are not dealt with by the simulation model. The model only goes so far, and it is important to recognize its scope and limitations.

## **2.3 The structure of the UVVC equation**

Figure 1 displays the structure of the UVVC estimation equation, and indicates the components involved in its calculation. Clearly, within each costpool 4 information elements are required: the costpool dollars, the cost pool elasticity, the costpool distribution key share, and the product volumes. Within any particular costpool, the subclass' UVVC is calculated as the product and ratio of various inputs. The point of Figure 1 is that it shows the complexity of this process – that UVVC's are calculated from a variety of inputs, and the overall quality of the UVVC estimate will depend on a complicated way on the quality of its inputs. The UVVC for each subclass, which is generically denoted as UVVC(i), results from the summation of the subclass' UVVC in each of the costpools. The costpools are again generically indexed by the letter (j).

---

<sup>3</sup>Testimony of Michael D. Bradley in Docket No. R97-1, UPS-T-14

Figure 1: Schematic representation of the Unit Volume Variable Cost estimation equation

$$\boxed{\begin{array}{c} \text{UVVC(i)} \\ \text{\$/Piece} \end{array}} = \sum_{\text{costpools(j)}} \frac{\boxed{\begin{array}{c} \text{Cost(j)} \\ \$ \end{array}} \boxed{\text{Elasticity(j)}} \boxed{\text{DKS(i,j)}}}{\boxed{\begin{array}{c} \text{Volume(i)} \\ \text{Pieces} \end{array}}}$$



## 2.4 The input sources to the UVVC equation

Figure 2 indicates the sources of the data required by the UVVC estimation equation. The costpool dollar quantity is derived from the General Ledger, and is thus the responsibility of the accountants. Costpool elasticities (volume variabilities in Postal Service language) may be derived from econometric studies or by expert opinion and thus are the responsibility of the econometricians. Finally, the distribution key shares and volumes are typically estimated from statistical sampling systems and are the responsibility of statisticians.

Not only are the inputs combined in a complicated fashion, they are produced by the triumvirate of accountants, economist and statisticians. The fact that different professional disciplines have responsibility for the different components only exacerbates the difficulty in evaluating the overall quality of the UVVC estimates.

The simulation model dealt with 8 mail subclasses and 29 cost segments, and accounted for approximately 50% of each subclass' UVVC. It utilized a database containing 8000 separate estimation components. To construct the model not only were estimates required of all input parameters, but precision estimates (cv's) were needed as well.

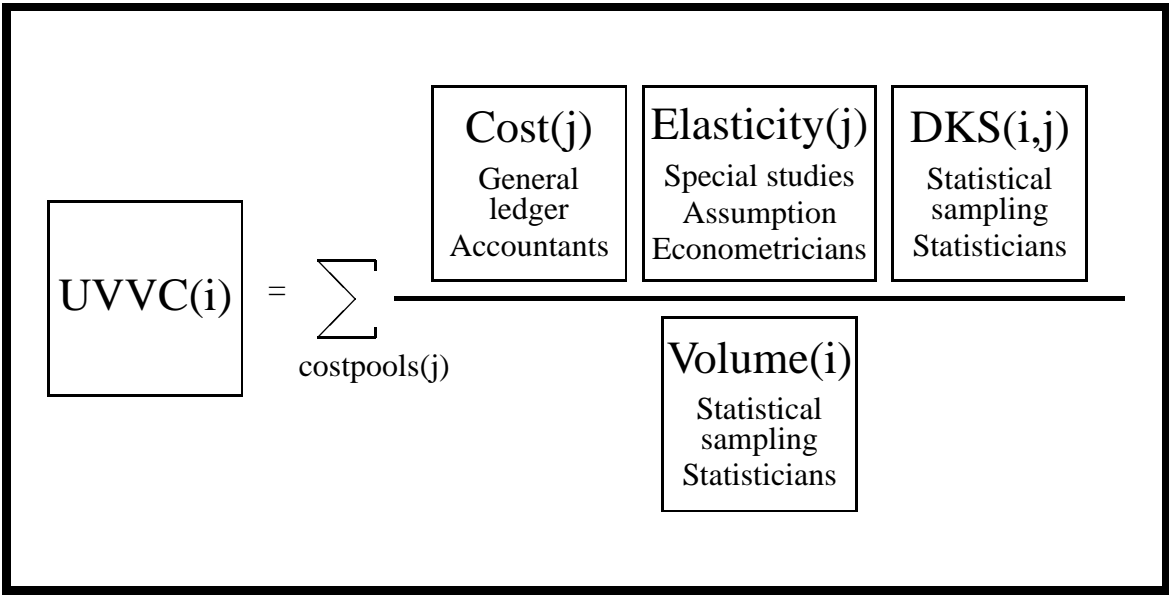
## 3 The necessity for a simulation model

A review of Figure 2 reveals immediately the complexity of the UVVC estimation process. The calculation of the UVVC involves a

- ratio
- products
- interdependencies between its elements because common data inputs may feed into multiple components (for example the same data elements may appear in both the DKS and the allocated dollars to the costpool)
- finally the UVVC in each costpool must be *summed* to obtain the overall UVVC for the subclass of interest

This complexity means that an analytical or formulaic approach is infeasible. It could be argued that an analytic approach is not even desirable, because it would have to be based on a series of simplifying assumptions. The simulation model effectively trades analytical complexity for computational intensity. In this sense

Figure 2: Sources of the inputs to the UVVC estimation equation



the model is similar to a computer chess game, that derives its strength from the ability to explore billions of possible situations.

## 4 Simulation model foundations

### 4.1 The estimation paradox

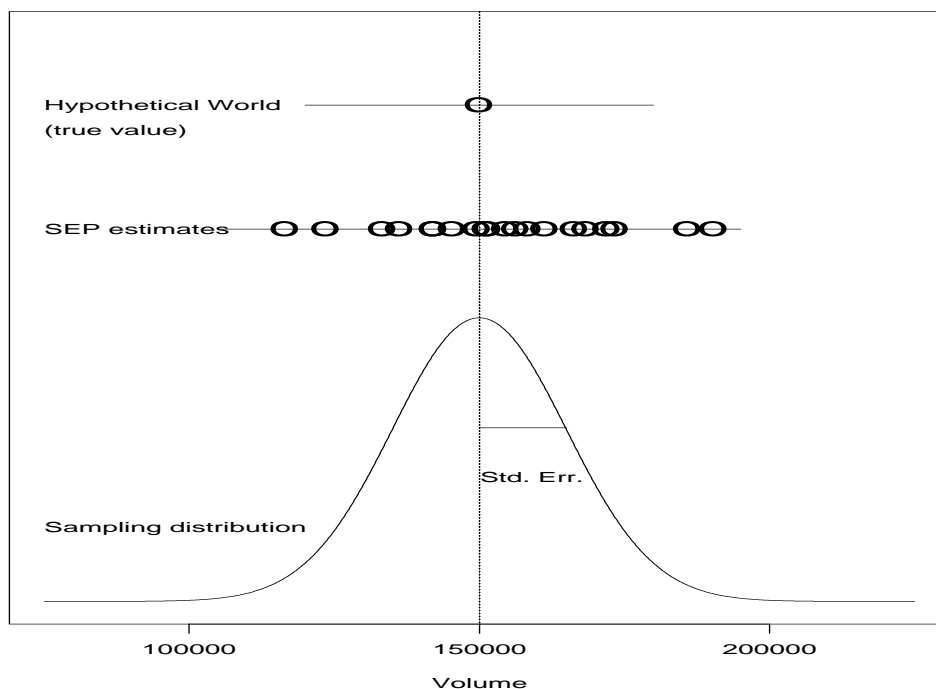
In order to ascertain if a methodology provides accurate estimates, it is necessary to know the value of the quantity being estimated, in this case the true value of a subclass' UVVC. By analogy, to determine if a shot is close to a target, the location of the target must be known. This realization at first appears to be an insurmountable paradox; after all if the true values of the UVVC's were known then there would be no need for a simulation study.

The approach of the simulation model is to assess the accuracy of the cost estimates, by **constructing** realistic values of Postal Service costs, volumes, and input parameter estimates. That is, a world is constructed, termed the *Hypothetical World* in which the true costs are known. The estimation methodology used by the Postal Service in their cost estimation procedures is then applied to this generated Hypothetical World to quantify the accuracy of those methods. These Postal Service estimation methodologies are termed *Survey Estimation Procedures*, or SEPs for short.

By analogy, a set of hypothetical targets is generated and how well the procedures perform against these targets is measured. If the estimation methodologies perform well across a range of targets, then there is much more confidence that they will perform well on the true but unobserved targets (the UVVC's). If a weapon is accurate when aimed at artificial targets, it increases ones confidence that it will be accurate in the field. This approach is analogous to the approach taken by clinical trials that test drugs under experimental conditions on non-random samples of subjects, who are randomized into treatment groups. If a drug is effective for various subsets of subjects with differing characteristics (age, gender etc) enrolled at different clinical sites, it is likely to be effective in actual medical practice.

A potential criticism of this approach is that if the values of the UVVC's and other inputs are unknown, then there is no way to know if the simulated targets are close to the truth. This criticism can be answered from two perspectives. First, targets similar to those observed are constructed. The simulation model achieves this by matching important characteristics such as attributable dollars and

Figure 3: Simulation model for a mail piece volume, under a no bias assumption



elasticities. A second answer to this criticism is that it is, in fact, not important for the simulated targets to be true, because the “truth” itself is dynamic. Of most interest is whether the **differences** between the generated “true” costs in the simulation world and those produced by the estimation procedure are small. Because the focus is on differences, the targets from which those differences are measured are no longer of primary importance. Recall the weapons analogy where the size of the targets are roughly the size of the actual objectives and the errors in hitting the targets are more important than the location of the targets.

The simulation is replicated a large number of times (usually 1000 in the study). Even though the HW is fixed, the simulation does not give identical results for the SEP estimate for each replication because for each replication, the inputs to the SEP are *randomly* drawn from a statistical distribution. This distribution is constructed from information supplied by the Postal Service. As an example, consider a hypothetical product’s volume estimate as would be supplied by the Revenue Pieces and Weight (RPW) system. Assume that the product’s true vol-

ume is 150,000 with a cv of 10%. From these inputs we may construct a *sampling distribution* for the volume estimator. Reasonable considerations suggest a Normal Distribution, which under a no bias assumption would be centered at 150,000. The precision of the estimator is described by its *standard error*, obtained from the cv, and in this case would be 15,000 (10% of 150,000). Figure 3 shows a graphical representation of how the SEP provides a set of estimates for the HW value. Notice that with no bias, the SEP estimates cluster around the true HW value. The variability of the SEP estimates about the true volume parameter is defined as the *sampling error* of the estimator. As the sampling error decreases so the estimate becomes more precise.

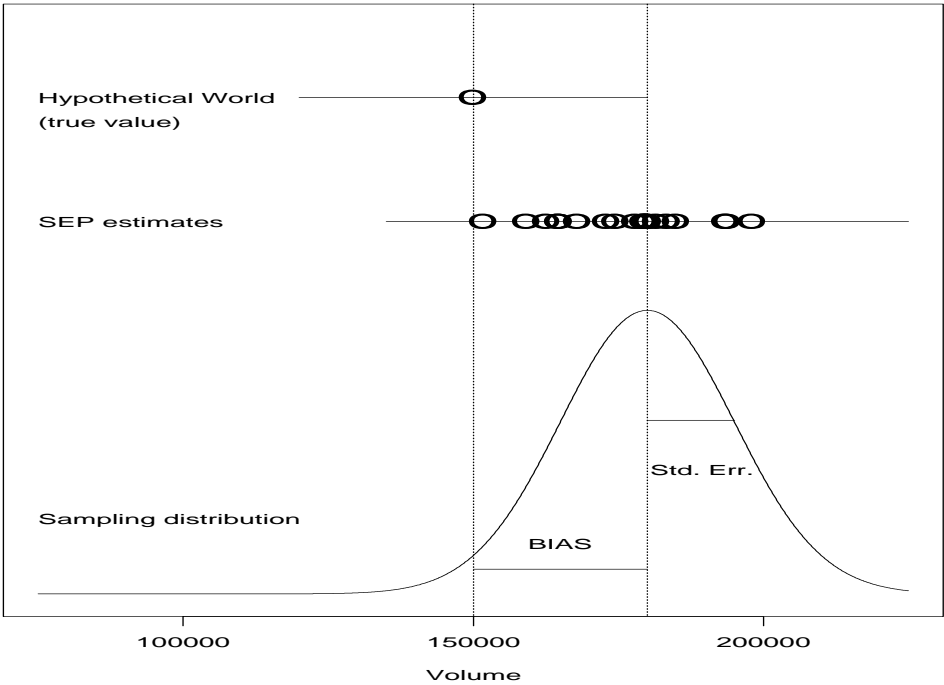
Figure 4 displays the same graphic, except this time the estimates now cluster about the wrong value. The fact that they cluster about the wrong value is termed *non-sampling error*, and their average distance from the true value is called the *bias* of the estimation procedure. A problem such as double counting of mail would give rise to a non-sampling error in a volume estimate.

The power of the simulation model lies in taking thousands of such inputs and combining them according to the methods that the Postal Service uses in its marginal cost estimation procedures.

Even though the simulations produce quantitative results, the exact numerical results are not emphasized in the interpretations. Rather, the numerical results are used as indicators of general areas of strengths and weaknesses. For example, the exact value of a bias, quantified in the simulation model, is not by itself directly relevant since it is computed with respect to a particular simulated world. The relative magnitude of biases resulting from different types of potential errors in the input parameter estimates is, however, indicative of the relative importance of different potential input errors.

Two recent examples of the use of simulation modeling within government agencies include an analysis of the National Center for Health Statistics NHANES survey [Ezzati-Rice et al., 1995] and the more recent evaluation by the Bureau of Labor Statistics of multiple imputation techniques in labor force surveys [Raghunathan and Rubin, 1997]. This latter study is similar to the simulation modeling carried out in this report. It requires the creation of a “Hypothetical World”, but again the accuracy of the Hypothetical World itself is not of primary importance because its role is as a test bed for evaluating estimation procedures. Both of these simulation models have been well received by the government research and academic communities. Considerable current interest is being shown in the latter simulation model as evidenced by a special session devoted to it at the 1998 American Statistical Association annual meeting. Similar acceptance of the sim-

Figure 4: Simulation model for a mail piece volume, under a no bias assumption



ulation study for UVVC's is anticipated. A presentation on the method has been invited for the 1999 American Statistical Association annual meeting.

## **4.2 A language for the simulation environment**

### **4.2.1 Hypothetical Worlds**

The Hypothetical World describes how the economic reality was generated. It contains a full description of cost generation processes, cost drivers, elasticities and distribution key shares.

In the full simulation model seven different HW's were used; in this section we describe the 2 main ones. More than one Hypothetical World is considered, that is more than one truth, because people argue over what exactly the truth is; the simulation model allows for a flexible description of reality.

The first (HW1) assumes that costs are generated according to an identical cost generation process as described by the Postal Service during Fiscal Year 1996 (FY96). Economic assumptions are made, such that if they were to hold true, then they would provide a rationale for the existing cost estimation methodology. In particular, it is assumed that there is a log-linear relationship between costs and a cost driver, and true values for model parameters are provided by FY96 estimates. In particular the cost elasticity for Mail Processing is taken to be 1.0.

A second Hypothetical World (HW2) includes an identical set of assumptions, with the exception of the cost elasticity and DKS assumptions for Mail Processing, which are taken from analyses provided by Bradley and Degen in Docket No. R97-1, USPS-T-14 and USPS-T-12 respectively. The purpose of HW2 is to provide a testbed to investigate the consequences of different approaches to setting/estimating cost elasticities and distribution key shares. In the Bradley-Degen approach, which would be the correct estimation approach if HW2 were true, these are derived from empirical estimates and MODS data. In HW1, these elasticities are set to one and the distribution key shares are derived at a more aggregate level.

These hypothetical worlds could clearly be expanded, and as noted earlier the full simulation model considered 7 different HW's including one in which costs were generated according to drivers sensitive to peak load effects [Crew and Kleindorfer, 1992, Chapter 3].

## 4.2.2 Survey Estimation Procedures

The Survey Estimation Procedure describes the methodology used to measure the parameters of interest in the Hypothetical World.

Three SEPs were used in the full simulation model. For illustrative purposes two are described here, which correspond to the 2 Hypothetical Worlds introduced in §4.2.1.

**SEP1.** Estimation based on FY96 procedures.

This SEP is based on the statistical properties of the procedures underlying the FY96 estimation by the Postal Service.

**SEP2.** Estimation based on FY96 procedures but with alternative mail processing measures.

This SEP is derived from SEP1 except for the mail processing costs, where the Bradley-Degen approach to measuring cost elasticities and distribution key shares is simulated. SEP2 is a step in the direction of more refined empirical estimation of mail processing labor costs based on operating data systems. Comparing the results of using SEP2 in a world where it does not apply (i.e., HW1) with results in a world where it does apply (HW2) would provide valuable information on the likely impact of implementing more detailed empirically-based SEPs.

## 4.2.3 Scenarios and runs

A scenario is defined as the crossing of a Hypothetical World with a SEP. That is, a scenario is generated from 2 elements; first a mechanism for generating a reality, the HW, and then a means for revealing that reality, the SEP. The HW can be thought of as a set of rules for constructing an artificial world in which everything that there is to know about costs, elasticities, drivers, distribution key shares *etc.* is known. The SEP can be interpreted as a lens through which this artificial world is observed. A good lens results in a clear view of the world, whereas a bad one, may distort and mis-represent the truth.

As we have described 2 HW's and 2 SEPs, 4 scenarios are possible and are displayed in Table 1.

In order to draw attention to the difference between the scenarios we concentrate for illustratory purposes on the role of mail processing; scenario one investigates the case when mail processing elasticities are 1, and the Postal Service estimates them as 1. Scenario 2 considers what happens if mail processing elasticities are not one, but the Postal Service estimates them as 1, in other words there is a systematic bias in the elasticity estimates. Scenario 3 looks at the case when



Table 1: The four possible scenarios available from the construction of 2 Hypothetical Worlds, and 2 Survey Estimation Procedures

	<i>Hypothetical World</i>	
<i>SEP</i>	HW1	HW2
SEP 1	Scenario 1	Scenario 2
SEP 2	Scenario 3	Scenario 4

the true elasticities are in fact 1, but the Postal Service estimates them, possibly as numbers less than 1, and finally scenario 4 deals with the case when the true elasticities are not one and the Postal Service estimates them.

Within any particular scenario, there may be a number of what-if questions of interest. As examples one could ask (and indeed these questions were asked within the simulation study in the context of Scenario 1),

- What would the impact be, if data collection resources were reduced by 40%, on the quality of the UVVC estimates?
- What would the impact be on UVVCs if there were a systematic 10% over-reporting of first class mail piece counts?
- What contribution do the special studies used to estimate cost elasticities make to the overall estimates of UVVCs?

These questions are all examples of what are termed *runs* of the simulation model. When a question is put forward that uses the same set of data inputs, but changes their values, this defines a *run* of the simulation model. To summarize, the simulation model terminology involves the use of *Hypothetical Worlds*, *Survey Estimation Procedures*, *scenarios* and *runs*.

## 5 Example: the Intra-BMC Costpool

This section provides an example to illustrate the development and use of the simulation model. It shows how the simulation methodology is applied to calculate the UVVC in the Intra-BMC costpool, which consists of the costs of transporting

mail between Bulk Mail Centers and mail processing facilities within their area. The Intra-BMC costpool is one of the more straight forward costpools considered in the simulation model, but even so the presentation here is somewhat of a simplification of the costpool. There is always a model design decision that has to be made concerning the level of detail to incorporate in the model. Practically, too much detail and the model may never be built, too little detail and the results could be misleading.

The first aspect of formulating the simulation model is to define both the Hypothetical World and the Survey Estimation Procedure, which specifies the methodologies used to estimate the quantities defined in the Hypothetical World. A single Hypothetical World (HW1) and a single Survey Estimation Procedure (SEP1) are presented in this section.

## 5.1 Hypothetical World construction for the Intra-BMC Costpool

Costs in the Hypothetical World (HW) are generated according to the log-linear cost model,

$$C = e^{\eta_0} D^{\eta_1}, \quad (1)$$

where  $C$  denotes the costpool dollars,  $D$  denotes the cost driver,  $\eta_1$  is the elasticity of cost with respect to the driver, and  $\eta_0$  is a multiplicative constant that determines the overall level of costs.

Costpools are dollar costs associated with some activity, here, transportation. A cost driver is a measure of the activity that determines the marginal costs in the costpool. In the Intra-BMC costpool, the costs accrue from transporting mail between Bulk Mail Centers, and the cost driver is cubic-foot miles. In a mail processing costpool, for example, the costs are employee salaries and the cost driver may be Total Pieces Handled, or Total Pieces Handled During a Peak Time Period. Equation (1) specifies the relationship between the total costs and the cost driver. The focus of the methods evaluated by the simulation study is to distribute costs to specific products, termed subclasses. The goal is to compute marginal costs as a function of volume for each subclass. For the  $i^{th}$  subclass, denote the volume (pieces) by  $M_i$ . The cost driver measuring activity associated with the  $i^{th}$  subclass in the costpool is denoted by  $D_i$ , and it is assumed to be proportional to the volume for the  $i^{th}$  subclass,

$$D_i = \alpha_i M_i. \quad (2)$$

The proportionality constants,  $\alpha_i$ , can vary across subclasses. The sum of the  $D_i$  across all subclasses equals  $D$ , the value of the driver. The true UVVC for subclass  $i$  is the derivative of (1) with respect to  $M_i$ . Based on the log-linear cost model in Equation (1), and the proportional relationship between the cost driver and the volume for each subclass in (2), the UVVC is

$$\text{UVVC}_i = \frac{dC}{dM_i} = \frac{C \eta_1 \kappa_i}{M_i}, \quad (3)$$

where  $\kappa_i$  is the distribution key share for subclass  $i$  in the Intra-BMC costpool:  $\kappa_i = D_i/D$ . Note that equation 3 is comprised four components and forms the basis for Figure 1. The formula for the UVVC does not require the explicit specification of the  $\alpha_i$ . This model completes the structural specification of costs in the HW and allows “true” UVVC’s to be calculated for each subclass. The hypothetical targets against which the Survey Estimation Procedure is evaluated are the subclass UVVCs.

## 5.2 Survey Estimation Procedures (SEPs) for the Intra-BMC Costpool

The Survey Estimation Procedure (SEP) describes the methodological approach used to estimate the true but (supposed) unknown costs in the HW. In most of the simulations, the UVVC formula in the SEP has the same form as the true cost-generating model. The values estimated by the SEP differ from the UVVC values generated in the HW because the SEP uses estimates of the inputs to the UVVC formula.

## 5.3 Data Sources for the Intra-BMC Costpool

In the Intra-BMC costpool, the estimates of the inputs to the SEP are based on four data sources:

1. The cost value from the General Ledger account (Postal Service accounting data).
2. The elasticity estimate from a special study.
3. The distribution key share (DKS) from the TRACS purchased transportation sampling system.
4. The volume estimates from the RPW system.

## 5.4 Assigning Input Values for a Hypothetical World for the Intra-BMC Costpool

To compute the UVVC for each subclass in the HW, values must be assigned to each input parameter in Equation (3). There are 9 subclass volumes (8 specific subclasses plus a category for “other”). There are also 9 subclass cubic-foot mile dollar equivalents,  $D_i$ , an elasticity term,  $\eta_1$  and the costpool dollars  $C$ , yielding a total of 21 input values.

The values for the input parameters are in the third column of Table 2. The log-linear UVVC equation (3) for the HW is solved for  $\eta_0$  to make the right hand side equal the allocated cost pool dollars of \$251,149 thousand<sup>4</sup>. Solving for the parameter  $\eta_0$  thus insures that costs generated in the HW will equal the cost reported by the Postal Service:  $\$251,149 = e^{\eta_0} (244,087^{0.9743})$ , and thus,  $\eta_0 = \ln(\$251,149) - 0.9743 \ln(244,087) = 0.3473$ . The sum of the  $D_i$  is  $D = 244,087$ .

Using the inputs in column 3 of Table 2, the  $UVVC_i$  in the HW can be calculated. In this example, the calculation is only shown for the Standard B – parcel post subclass.

The cost and elasticity values in the HW read from Table 2 are \$251,149 and 0.97, respectively. The distribution key share (DKS) for Standard B – parcel post subclass is found by dividing the Standard B – parcel post subclass,  $D_i$ , 71,330, by  $D = 244,087$  to obtain a value of 0.2922. Finally, the Standard B – parcel post subclass volume in Table 2 is 212,828. Putting all these components into the UVVC estimating equation (3) gives a UVVC in the HW of

$$UVVC_i = \frac{dC}{dM_i} = \frac{251,149 \times 0.97 \times 0.2922}{212,828} = 0.3345, \quad (4)$$

so that the UVVC for Standard Parcel Post subclass from the Intra-BMC costpool is roughly 33 cents.

## 5.5 Simulating Input Parameter Estimates for the Intra-BMC Costpool

Each parameter estimate input to the UVVC equations is generated according to its statistical properties determined from the analyses of the major data systems and special studies. The statistical properties are the mean, cv, bias, sampling

---

<sup>4</sup>Data inputs were obtained from Postal Service FY96 values

distribution (e.g., Normal, Gamma etc.), and dependence on other estimates. In this example, the SEP estimators are assumed to have zero bias. For instance, consider the elasticity  $\eta_1$ , which has a true value of 0.97 in the HW. The SEP value of  $\eta_1$  is generated by taking a random draw from a normal distribution with a mean of 0.97 (no bias in this input), and a cv of 1.65%. The value for the cv is based on the Postal Service report for the Fiscal Year 1996. The SEP estimate of the elasticity will **not** be exactly equal to the true value in the HW, but the smaller the cv, the closer it will tend to be to the HW value (see Figure 3). Because the HW value and SEP estimate of the elasticity are not identical (and this is the case for most input estimates), the SEP UVVC estimates will differ from their corresponding HW values.

One of the key aspects of the simulation is that within any single replication of the simulation model, all estimates of volume, revenue, etc., are generated only once, and the same value is used in every costpool calculation. In this way, the complex dependencies and covariances inherent in the UVVC estimates across all the costpools are implicitly accounted for by the simulation model. To simplify the generation of the numerous parameter estimates, they are refined to a level where they can be assumed statistically independent in the SEP. An example of the simulated (SEP) estimates of the input parameters to the UVVC estimation equation, is in the fourth column of Table 2 for a single replication (a typical run of the simulation model has 1000 replications).

Evaluating the UVVC equation with the SEP values for the Standard Parcel Post gives 37.32 cents.

In section §5.4, the UVVC equation evaluated with HW values yielded 0.3345. Thus, the simulation model has provided a target value of 33 cents for the “truth” in the HW, and a value of 37 cents for the estimate by the SEP. To evaluate the SEP, numerous replications of the SEP estimates are generated for each HW. Statistical summaries of the differences between the SEP estimates and the HW true values provide the basis for evaluating the SEP.

## 6 Precision/accuracy and mean squared error

This section describes the statistical summaries used to evaluate the Survey Estimation Procedures. They focus on the statistical properties of the difference between the estimated and true values of an attribute of interest, the UVVC in particular. The difference between the estimated and true values of the attribute is the *estimation error*.

Table 2: Input Parameter Values for the HW and SEP for One Replication of the Intra-BMC Costpool

Description	Input Parameter	HW values	SEP values
General Ledger Dollars ('000)	Accrued dollars	\$251,149	\$251,149
Elasticity	$\eta_0$	0.3473	0.3473
	$\eta_1$	0.97	0.99
subclass Cost Drivers (based on TRACS)	First Class Letters Flats & IPPs (incl. Presorts)	27,380	30,810
	Periodicals – w/n County	0	0
	Periodicals – Regular Rate	19,160	15,860
	Standard – Enhanced Carrier Route	12,850	8,240
	Standard – Regular	46,740	48,350
	Standard – Non-profit	7,930	7,650
	Standard – Parcel Post	71,330	81,460
	Standard – Library	4,010	2,970
	OTHER	54,690	57,720
subclass Volumes in thousands (based on RPW)	First Class Letters Flats & IPPs (incl. Presorts)	93,207,952	93,302,256
	Periodicals – w/n County	877,829	888,834
	Periodicals – Regular Rate	6,984,301	6,995,104
	Standard – Enhanced Carrier Route	29,180,737	29,319,610
	Standard – Regular	30,150,508	30,156,708
	Standard – Non-profit	9,300,466	9,277,951
	Standard – Parcel Post	212,828	216,116
	Standard – Library	30,133	34,399
	OTHER	13,494,719	13,578,647

## 6.1 Coefficient of Variation (cv)

The coefficient of variation (cv) is a summary measure used to report uncertainty in sample survey estimates of positive quantities. It is used to describe the *sampling error* of the estimate. There is uncertainty regarding an estimate from a survey, because if the survey were repeated the estimate would almost certainly change. If a census, rather than a survey were undertaken, then there would be no uncertainty in the estimate and the estimate should equal the true value. The cv is defined as  $\sigma/\mu$ , where  $\sigma$  is the standard error of the estimate, and  $\mu$  its expected value. The cv expresses the standard error of an estimate as a proportion of the mean value of the estimate. In practice, the numbers  $\sigma$  and  $\mu$  are not known so they are replaced by estimates from the survey. The smaller the cv, the more precise the estimate. In the simulation model, these summaries are calculated as follows:

1. Each replication of the simulation provides an estimate of a subclass' Unit Volume Variable Cost from the SEP. A simulation run with 1,000 replications thus produces 1,000 such estimates.
2. After the 1,000 estimates have been computed, two summaries are derived: the mean of the 1,000 values, which is an estimate of the expected value of the estimator; and the standard deviation of the 1,000 values, which is an estimate of the standard error of the estimator.
3. The ratio of the standard error to the mean is then calculated to estimate the cv.
4. The cv's are often reported on a percentage scale computed by multiplying by 100.

The cv's, combined with standard assumptions that lead to the approximate normality of the estimate, can be used to simulate the sampling distribution of an estimator. For example, if the cv is 5%, then standard theory indicates that approximately 95% of values generated from the sampling distribution are within +/- 10% of the mean of the estimator.

## 6.2 Bias

Bias exists when an estimate departs systematically from the parameter it is designed to estimate. The smaller the bias, the more accurate the estimate. For

example, if the true elasticity in a costpool is 0.7 and the SEP always uses an estimate of 1.0, then this estimate would be *biased* by 0.3. This situation describes bias in an input parameter and is an example of a *non-sampling* error. Non-sampling errors can also arise from problems such as double counting, systematic mis-classification of the mail and incomplete sampling frames.

Biases in the inputs will typically lead to biases in the estimates of the output quantities of interest as well, in particular in the UVVC estimates. In this context, a UVVC estimate would be termed biased if on average it was different from the true UVVC value it was attempting to estimate. For example, if the true UVVC of a product were 2c and the estimate had an average value of 3c, then this would constitute a 1c, or 50% bias in the UVVC estimate. It is quite possible for all input quantities to be unbiased but the output quantity of interest still to be biased, if these inputs are combined in a complicated (non-linear) fashion as is illustrated in Figure 1. In practice, it is the qualitative magnitude of the bias that is important, not simply the fact that bias exists. In real world sampling systems, total eradication of all potential bias is not feasible.

Bias may also be described as a proportion of the parameter to be estimated. This facilitates comparing biases across the various subclasses of mail. When the bias is expressed as a proportion in this way, it is termed *relative bias*.

### 6.3 Root Mean Squared Error

Root Mean Squared Error (RMSE) is an alternative way of judging the quality of a SEP. The RMSE provides an enhancement to simply using the bias and cv as descriptors of a SEP, because it incorporates both the cv and the bias. The Mean Squared Error (MSE) is defined as the standard error squared plus the bias squared – so that either a large standard error **or** a large bias makes the MSE large. The RMSE is the square root of the MSE, and provides a summary on the same scale as the estimate. This means that if the cost estimates are in cents then the RMSE is measured in cents providing a directly comparable measure. RMSE provides the foundation for comparisons between sampling and non-sampling errors. As the potential non-sampling error, which is revealed through *bias* in the UVVC estimator, becomes larger it contributes a greater component to the RMSE.



## 7 Illustration of model results

In this section 2 sets of results are summarized from the simulation model. The purpose is to give an idea of the types of information that the model may provide.

### 7.1 UVVC cv's, measuring sampling error

One of the main contributions to the cost estimation debate provided by the simulation model, was that it represented the first time to our knowledge that uncertainty estimates (cv's) have been calculated for the UVVC's. The third column of Table 3, under the heading *SEP cv*, presents these uncertainties for the 8 sub-classes incorporated in the model.

Column 1 of Table 3 indicates the subclass, column 2 contains the average value of the UVVC estimate in cents. Column 3 presents the overall cv for the UVVC estimate as a percentage. Columns 4 through 9 display the cv for the UVVC, but from a run in which the variability due to a particular data system has been removed. For example, the change in cv for *Standard A enhanced carrier route* between column 3 and column 9, where it drops from 8.00% to 1.32% shows that much of the uncertainty in this subclass arises from the Delivery Special Studies (SS). When a system's specific cv's have been set to zero in the model then we term this as "turning the system off".<sup>5</sup>

Each of columns 4 through 9 of Table 3 shows the impact on uncertainty of a particular data system. "ELAS" refers to elasticity data elements and "SS" to carrier special study data elements. One conclusion is that different subclasses are impacted differentially by the various data systems. For example, the UVVC estimate for Standard B parcel post realizes much of its uncertainty from the TRACS system. TRACS measures purchased transportation costs. This is indicated by the fact that the cv falls from 4.59 to 3.75, a drop of about 15% when the TRACS system is turned off. Likewise, Periodicals, within county, receives much of its uncertainty from the IOCS system, which is due to the fact that Periodicals within county is a low volume subclass and thus hard to estimate from the IOCS (In Office Cost System) sample.

In summary, Table 3 helps identify which products are most impacted by which data system.

---

<sup>5</sup>Small increases in the UVVC cv after "turning off" a system are due to simulation errors that occur because only 4000 replications were performed. When this occurs, it indicates that sampling and non-sampling errors from the system have a very small effect on the UVVC's. These instances are denoted in the table by the \* symbol.

Table 3: Summary of the sampling variability in the UVVC estimates for the 8 products used in the simulation model.

Subclass	UVVC mean	SEP cv	Subclass cv when a specific subsystem is “turned off”					
			RPW	IOCS	TRACS	CCS	ELAS	SS
First Class Letters Flats & IPPs	11.60	1.26	1.24	1.19	1.14	*1.30	1.24	0.81
Periodicals within county	3.38	10.09	*10.23	1.30	*10.14	10.06	10.07	10.02
Periodicals regular rate	10.04	2.51	2.51	2.05	1.53	2.51	2.47	*2.52
Standard A enhanced carrier route	3.07	8.00	7.90	*8.14	*8.04	8.00	8.06	1.32
Standard A regular rate	8.52	2.11	2.00	1.88	2.03	2.06	2.05	1.04
Standard A non-profit	6.65	2.70	2.64	2.22	2.57	*2.73	2.67	1.76
Standard B parcel post	146.17	4.59	4.25	4.14	3.75	*4.71	*4.64	3.85
Standard B library	96.33	13.48	8.79	11.46	12.49	13.26	*13.49	13.24

## **7.2 Mail processing elasticity estimates, an example where potential bias overwhelms sampling error**

This example summarizes the potential consequences of a divergence between the Hypothetical World and the SEP. In particular, the HW is taken as HW2 but the SEP used is SEP1. That is, true elasticity values in mail processing are not equal to one, but in the SEP they assumed equal to one. In addition, the SEP over-aggregates the DKS.

The analysis reported in Table 4 reports the UVVC's from only the Mail Processing costpools and shows that changes in elasticity estimates and DKS in Mail Processing will have a large potential impact on subclass UVVC's. The main consequence of using SEP1 when HW2 represents the "truth", is seriously biased estimates of the UVVC's, by between 10 to 40%.

This potential over-estimation of UVVC is due to the use of a higher elasticity in the SEP than the HW. Further, this potential relative bias is an order of magnitude larger than the cv of the UVVC estimate for all but the low volume products, indicating the potential bias would be much more important than the SEP sampling variability. For example, the cv for First Class Letters Flats and IPPs is 0.82%, whereas the relative bias is 31.55%.

## **8 Conclusions**

The simulation approach to cost estimation was shown to be a feasible method with which to estimate UVVCs and to evaluate the impact on these estimates of both the quantity of data collected and various potential sources of systematic error. It allowed for the first time, uncertainties on subclass UVVCs to be calculated. Though the model was limited in its scope, covering only 8 subclasses and 29 costpools there is no fundamental reason why it could not be expanded to capture a greater portion of the UVVC. Such an expansion would entail a sizeable effort as the existing model took 6 months to construct. Further, if the model is to remain useful then its data inputs would have to be updated on a regular basis.

The construction of the model contributes to answering the question "how well are costs estimated"? As such it allows one to direct both dialog and potentially, action in terms of information collection resources, to those areas of the cost generation process that are most influential on the costs themselves. The model has the potential to be the foundation of a methodology that would provide a rational allocation of information collection resources. However, if it were to take on this

Table 4: UVVC Estimates and cv's of UVVC's in the Mail Processing costpools when the Hypothetical World is HW2 but the SEP is SEP1

Subclass	HW UVVC cents	SEP UVVC cents	Relative bias (%)	cv (%)
First Class Letters Flats & IPPs	5.42	7.13	31.55	0.82
Periodicals within county	1.52	1.87	22.37	14.12
Periodicals regular rate	5.83	6.58	12.86	2.05
Standard A enhanced carrier route	0.80	0.91	12.50	2.60
Standard A regular rate	4.51	5.59	23.95	1.40
Standard A non-profit	3.48	4.36	25.29	2.12
Standard B parcel post	64.98	82.78	26.45	3.38
Standard B library	47.07	67.84	44.13	13.31

role then additional information would need to be incorporated, in particular the costs of collecting information.

Though the model was constructed within the context of the Data Quality Study there is no reason why the results from the model could not be used in a wider context than the Postal rate setting environment alone. The underlying questions that motivated the construction of the model, that is the quality of the raw material to an information based decision system, could be posed to any information system, and in particular could include operating data systems.

As a final point, there is nothing intrinsic to the simulation methodology that would prohibit its use outside of the Postal Service environment. Other industries may find benefit in taking a hard look, and attempting to quantify the quality of their data. As management decisions are increasingly information driven, there must come a point when the quality of the raw material inputs to those information systems becomes paramount. The approach taken here may provide a framework from which to formulate data quality metrics in a wider context.

In summary, if it is critically important to measure cost, then one should want to know how well those costs have been measured. The simulation model is a first step in this endeavor.

## References

- M.D. Bradley, J.L. Colvin, and M.A. Smith. Measuring product costs for ratemaking: The United States Postal Service. In M. A. Crew and P. R. Kleindorfer, editors, *Regulation and the Nature of Postal and Delivery Services*. Kluwer Academic Publishers, Boston, MA, 1993.
- R. Cooper and R.S. Kaplan. *The Design of Cost Management Systems*. Prentice Hall, Englewood Cliffs, NJ, 1991.
- M.A. Crew and P.R. Kleindorfer. *The Economics of Postal Service*. Kluwer Academic Publishers, Boston, MA, 1992.
- T.M. Ezzati-Rice, W. Johnson, M. Khare, R.J.A. Little, J.L. Schafer, and D.B. Rubin. A simulation study to evaluate the performance of multiple imputation in the NCHS Health Exam Survey. In *Proceedings of the Bureau of the Census. 11th Annual Research Conference*, pages 257–266, 1995.
- LINX. Data Quality Study, April 1999. Available online at <http://www.usps.gov/clr/dqs.htm>.
- T.E. Raghunathan and D.B. Rubin. Roles for Bayesian techniques in survey sampling. In *Proceedings of the Survey Methodology Section of the Statistical Society of Canada*, pages 51–55, 1997.