Working Papers

Industrial and Construction Statistics

Manufacturing and Construction Division

Bureau of the Census

BUREAU OF THE CENSUS
MANUFACTURING AND CONSTRUCTION DIVISION

MCD Working Paper Number:
Census/MCD/WP-95/03

SAMPLE ALLOCATION FOR THE 1994 ANNUAL SURVEY OF MANUFACTURES

by

Douglas Bond
Robert Struble
Lynn Imel

Manufacturing and Construction Division
Bureau of the Census

September 1995

SAMPLE ALLOCATION FOR THE 1994 ANNUAL SURVEY OF MANUFACTURES

Douglas Bond, Robert Struble, and Lynn Imel

ABSTRACT

We used Chromy's algorithm, a convex programming method, to optimally allocate a sample of about 58,000 establishments for the 1994 Annual Survey of Manufactures (ASM). This algorithm, followed by a minor adjustment procedure, satisfied 2,230 variance constraints for industries and product classes simultaneously. Chromy's algorithm saved about 5,000 units in comparison with the method that was previously used to allocate the ASM sample. We reduced processing time by 70 percent to 3.4 cpu hours on a VAX 9000 computer, by rewriting the original SAS program for Chromy's algorithm in SAS/IML (matrix language).

KEY WORDS: Chromy's algorithm, Nonlinear programming, Optimum allocation, Poisson sampling

1. INTRODUCTION

The U.S. Bureau of the Census conducts the Annual Survey of Manufactures (ASM) to derive estimates of U.S. manufacturing activity between the censuses of manufactures. The Census Bureau has been conducting this mandatory survey since 1949. Estimates are published by state and industry group and at the U.S. level. These estimates include employment, payroll, value of shipments, capital expenditures, and inventories. Value of shipments is also estimated by product class. The ASM is the only source of these detailed data, which are needed by government and industry for analysis and planning. The Census Bureau uses ASM results to benchmark some of its other surveys, such as the monthly Manufacturers' Shipments, Inventories, and Orders Survey, a principal Federal economic indicator. Export statistics are derived from ASM results. The ASM is an important source of longitudinal establishment-level data that are used for making public policy decisions.

The Census Bureau selects a new ASM sample every 5 years, using establishments in the most recent census as the sampling frame. (The Census Bureau conducts the census every 5 years to cover years ending in "2" and "7.") The sample is regularly updated for births and deaths of establishments. Zayatz and Sigman (1994) recommended that the Census Bureau use a different method, Chromy's algorithm (Chromy 1987), to allocate the new sample for 1994. This paper describes Chromy's algorithm and our experiences with it for allocating the 1994 ASM sample.

In Section 2, we describe the design of the ASM. We state the optimum allocation problem in Section 3 and outline the steps of Chromy's algorithm for the ASM in Section 4. Section 5 presents a simple example of optimum allocation in the ASM. Section 6 describes how we used Chromy's algorithm for allocating the 1994 sample, and how it compared with the approach that was used for the previous (1989) sample. We draw conclusions and make recommendations in Section 7.

2. ASM DESIGN

The sampling unit for the ASM is the establishment, one physical location where manufacturing is performed. Each establishment is classified in one 4-digit Standard Industrial Classification (SIC) industry (Office of Management and Budget 1987), based on the primary types of products it ships. The Census Bureau groups products into product classes (5-digit codes), and an establishment may ship products in more than one product class. When data are summarized, an establishment can contribute to estimates for only one 4-digit SIC code, but it may contribute to the estimates for more than one product class. For allocation of the 1994 ASM sample, there were 457 4-digit SIC codes and 1,773 product classes, a total of 2,230 estimation cells.

The Census Bureau uses Poisson sampling (Hajek 1964) to select establishments for the ASM sample (Ogus and Clark 1971). The units have independent, and generally unequal, probabilities of selection. The major advantage of Poisson sampling for the ASM is that smaller units can be rotated out when a new sample is selected without biasing the estimator of the total. Also, variances can be easily computed.

The major disadvantage of Poisson sampling is that the sample size is a random variable. That is why we discuss the expected sample size in this paper. The expected sample size is

$$E(n) = \sum_{h=1}^{N} p_h ,$$

where p_h is the probability of selecting unit h, and N is the total number of units in the population or subpopulation of interest (for example, a state, 4-digit SIC code, or product class). This and other formulas in this paper can be used to derive estimates for subpopulations, with appropriate subscripting. The selected sample size is usually different from the expected sample size, and it is possible to draw a sample of all or none of the units in a population or subpopulation. In our experience, the selected sample size for the entire ASM is usually within 1 percent of the expected size, so Poisson sampling does not pose a problem for planning to meet budgetary constraints. Another disadvantage of

Poisson sampling is that the estimator of the total has a larger variance than with a sample design that has a fixed sample size.

Under Poisson sampling, a total Y is estimated by the unbiased "reciprocal" estimator:

$$\hat{Y}_{RECIP} = \sum_{h=1}^{n} \frac{Y_h}{p_h} ,$$

where Y_h is the survey value of Y for unit h, and n is the number of sample units selected from the population or subpopulation. The Census Bureau incorporates the reciprocal estimator into a difference estimator to derive most ASM totals:

$$\hat{Y}_{DIFF} = \hat{D}_{RECIP} + X = (\hat{Y}_{RECIP} - \hat{X}_{RECIP}) + X ,$$

where X is the total from the latest census, and $\hat{D}_{RECIP}$ is the sample estimate of the change since that census. $\hat{Y}_{RECIP}$ and $\hat{X}_{RECIP}$ are the reciprocal estimates for the current ASM and the ASM portion of the latest census, respectively. The variance of the estimator is

$$Var(\hat{Y}_{DIFF}) = Var(\hat{D}_{RECIP}) = \sum_{h=1}^{N} \left(\frac{1}{p_h} - 1\right) D_h^2 ,$$

where D_h is the change in survey values since the latest census for unit h. This formula is used to define constraints for optimum allocation, which is discussed in the next section.

3. OPTIMUM ALLOCATION OF THE ASM SAMPLE

Our goal in optimally allocating the ASM sample is to control variances of estimated value of shipments by 4-digit SIC code and product class. Although the Census Bureau also publishes estimates for states, we do not try to control variances at that level because it would put unnecessary constraints on sample allocation. The optimum allocation problem can be stated as: assign p_h values to units so that the cost function is minimized, subject to variance constraints. The cost function is

$$C = C_0 + \sum_{h=1}^{N} c_h p_h ,$$

where C_0 is a fixed overhead cost, c_h is the cost per sample unit, and N is the total number of eligible sampling units. The constraints are of the form

$$Var(\hat{Y}_{DIFF,i}) \leq V_i^* \text{ , } i \in S$$

where S is the set of 2,230 4-digit SIC codes and product classes. This means that variances by 4-digit SIC code and product class must not exceed target values V_i^*. Additional constraints are that all p_h values must be at most 1 and at least some value such that sample weights ($1/p_h$) are not too large.

For the 1989 ASM, the Census Bureau used an allocation based on a measure of size: each unit's sum of <u>predicted</u> squared D_h values (differences in value of shipments since the latest census) for its product classes. D_h values were predicted for each product class of each unit, using regression equations. The predictor variable was 1987 census value of shipments. Eligible sampling units were arrayed in descending order of measure of size, and a cutoff point selected, above which all units were selected with certainty (p_h = 1). The Census Bureau assigned a probability of selection to the remaining sampling units with the following formula:

$$p_h = \frac{E(n)\sqrt{\sum_{j\in S_h} \hat{D}_{hj}^2}}{\sum_{h=1}^{N}\sqrt{\sum_{j\in S_h} \hat{D}_{hj}^2}} \text{ ,}$$

where N is the number of remaining sampling units, $\hat{D}_{hj}$ is the predicted difference in value of shipments since the 1987 census for product class j of unit h, and S_h is the set of all product classes in unit h. This sample allocation did not satisfy all 4-digit SIC code and product class variance constraints, so the Census Bureau supplemented the expected sample size (by several thousand units) by increasing the measure of size for units so that more constraints would be satisfied. Even then, some constraints were not met. For more details of this method, see Waite and Cole (1980) and Dee and Dorinski (1993).

Zayatz and Sigman recommended a different approach, Chromy's algorithm, which minimizes the cost function, subject to the product class <u>and</u> 4-digit SIC code constraints, all at once. The following is a description of this approach. Assume that the costs for all units are equal, and note that the cost minimization problem will have the same solution if C_0 is removed from the cost function. Then this function can be simplified to

$$C = \sum_{h=1}^{N} p_h$$

With the transformation $x_h = 1/p_h$, the problem becomes: minimize

$$f(x_1, x_2, \ldots, x_N) = \sum_{h=1}^{N} \frac{1}{x_h}$$

subject to

$$\sum_{h=1}^{N_i} (x_h - 1)\hat{D}_{hi}^2 \leq V_i^* \ , \ i \in S$$

where N_i is the number of eligible sampling units in the ith subpopulation (4-digit SIC code or product class). This is a nonlinear programming problem, where the objective function, f, is convex, and the constraints are concave linear functions of the x_h values. Chromy's algorithm is a convex programming method that iteratively seeks the optimum point $x^* = (x_1, x_2, \ldots, x_N)$ that satisfies all the constraints, i.e., the unique point that minimizes f. The constraints on the p_h values are satisfied after each iteration by forcing each x_h value to be in the range

$$1 \leq x_h \leq W \ , \ h = 1, 2, \ldots, N \ ,$$

where W is the maximum desired sample weight. The next section describes Chromy's algorithm for the ASM.

4. CHROMY'S ALGORITHM FOR THE ASM

This is a description of the <u>modified</u> Chromy's algorithm, in which the Lagrange multipliers λ are computed as the product of a "univariate λ" and a "scaling factor." This causes a more rapid movement towards the optimum solution than when the λ values are computed directly, as in the original version of the algorithm. The steps of the algorithm are:

1. Compute the univariate λ_i, denoted a_i, for each 4-digit SIC code and product class:

$$a_i = \left(\frac{\sum_{h=1}^{N_i} \hat{D}_{hi}}{V_i^* + \sum_{h=1}^{N_i} \hat{D}_{hi}^2} \right)^2$$

Values of $\hat{D}_{hi}$ are predicted with regression equations using value of shipments from the latest census. Recall that $\hat{D}_{hi}$ values were also computed this way for the 1989 sample.

2. Compute λ_i:

(a) If this is the first iteration, initialize the scaling factor b_i to 1. Then

$$\lambda_i = a_i b_i = a_i$$

Go to step 3.

(b) If this is the second or later iteration, compute the following value for each 4-digit SIC code and product class:

$$c_i = \frac{\sum_{h=1}^{N_i} \frac{\hat{D}_{hi}^2}{p_h}}{V_i^* + \sum_{h=1}^{N_i} \hat{D}_{hi}^2}$$

Note that $c_i \leq 1$ is equivalent to $Var(\hat{Y}_{DIFF,i}) \leq V_i^*$.

Compute updated scaling factors b_i'' (compute factors b_i' in an intermediate step), using b_i values from the previous iteration:

$$b_i' = \begin{cases} b_i c_i^2 & b_i \neq 0 \\ 1 & b_i = 0 \text{ and } c_i > 1 \\ b_i & \text{otherwise} \end{cases}$$

$$b_i'' = \begin{cases} 0 & b_i' < \epsilon \text{ and } c_i \leq 1 \\ b_i' & \text{otherwise} \end{cases}$$

We used $\epsilon = 0.001$. Then

$$\lambda_i = a_i b_i''$$

3. Compute each unit's selection probability:

$$p_h = \sqrt{\sum_{j \in S_h} \lambda_j \hat{D}_{hj}^2} \ ,$$

and force p_h into the interval [1/W,1].

Repeat steps 2 and 3 until the solution is "near" convergence. The criterion for nearness is that

$$\sum_{i \in S} \lambda_i |Var(\hat{Y}_{DIFF,i}) - V_i^*| \leq K$$

This is a summation over all 4-digit SIC codes and product classes. When this criterion is met, the distance of $f(x)$ from $f(x^*)$ is no more than approximately K (this is based on a result from Causey (1983)). We set K at a level that is small enough to ensure that most variance constraints are satisfied, but large enough that computer time does not become excessive. Section 6 describes results of using K = 5 and K = 50.

Zayatz and Sigman also derived a "stopping rule" as an alternative criterion for nearness to convergence. This stopping rule assumes that expected sample sizes converge to the minimum size as a negative exponential function of the iteration number. We were always able to obtain a solution using the nearness criterion of the previous paragraph, so we did not need to use the stopping rule.

Some variance constraints will not be satisfied even though the nearness criterion is met. To ensure that all are satisfied, compute adjusted probabilities p_h':

$$p_h' = \frac{\max_{j \in S_h} r_j}{\frac{1}{p_h} - 1 + \max_{j \in S_h} r_j} \ ,$$

where

$$r_j = \frac{Var(\hat{Y}_{DIFF,j})}{V_j^*} \ , \ j \in S$$

r_j is the ratio of the variance to the target variance, by 4-digit SIC code and product class. Zayatz and Sigman derived the adjustment formula.

5. EXAMPLE

This is a simple example of the optimum allocation problem for the ASM. Assume there are only two units eligible for sampling. We developed regression equations from 1989-91 ASM data, by 4-digit SIC code and product class. Then we computed predicted differences of value of shipments, using 1992 census values as predictors. The predicted differences are shown in Table 1.

Table 1: Predicted Differences (Thousand $) for 2-Unit Example

Unit	4-Digit SIC #1	4-Digit SIC #2	Product Class #1	Product Class #2	Product Class #3	Product Class #4
1	-	427	-	86	141	43
2	530	-	581	-	-	24

The problem is to minimize

$$f(x_1, x_2) = \sum_{h=1}^{2} \frac{1}{x_h} = \frac{1}{x_1} + \frac{1}{x_2} ,$$

subject to six variance constraints (two 4-digit SIC codes plus four product classes) and constraints on two sample weights. Using the target variance from the 1994 ASM sample selection (see Section 6 for more details), the first variance constraint is

$$(x_2 - 1)(530)^2 \le 4.55 \times 10^{10}, \text{ or}$$

(1) $x_2 \le 162{,}000$ for 4-digit SIC #1

The other variance constraints are

(2) $x_1 \le 232{,}000$ for 4-digit SIC #2
(3) $x_2 \le 3{,}564$ for product class #1
(4) $x_1 \le 46{,}745$ for product class #2
(5) $x_1 \le 61{,}658$ for product class #3
(6) $1.83x_1 + 0.57x_2 \le 91{,}967{,}500$ for product class #4

For this example, we specified an unusually small lower bound for p_h: 0.00001. This is equivalent to constraints on both sample weights of $1 \le x_h \le 100{,}000$. We set the lower bound very low to avoid deriving a trivial solution (corresponding to the lower bound on p_h) to the optimum allocation problem in our example. To actually select the 1994 ASM sample, we set the lower bound on p_h much higher (0.02); see Section 6 for details.

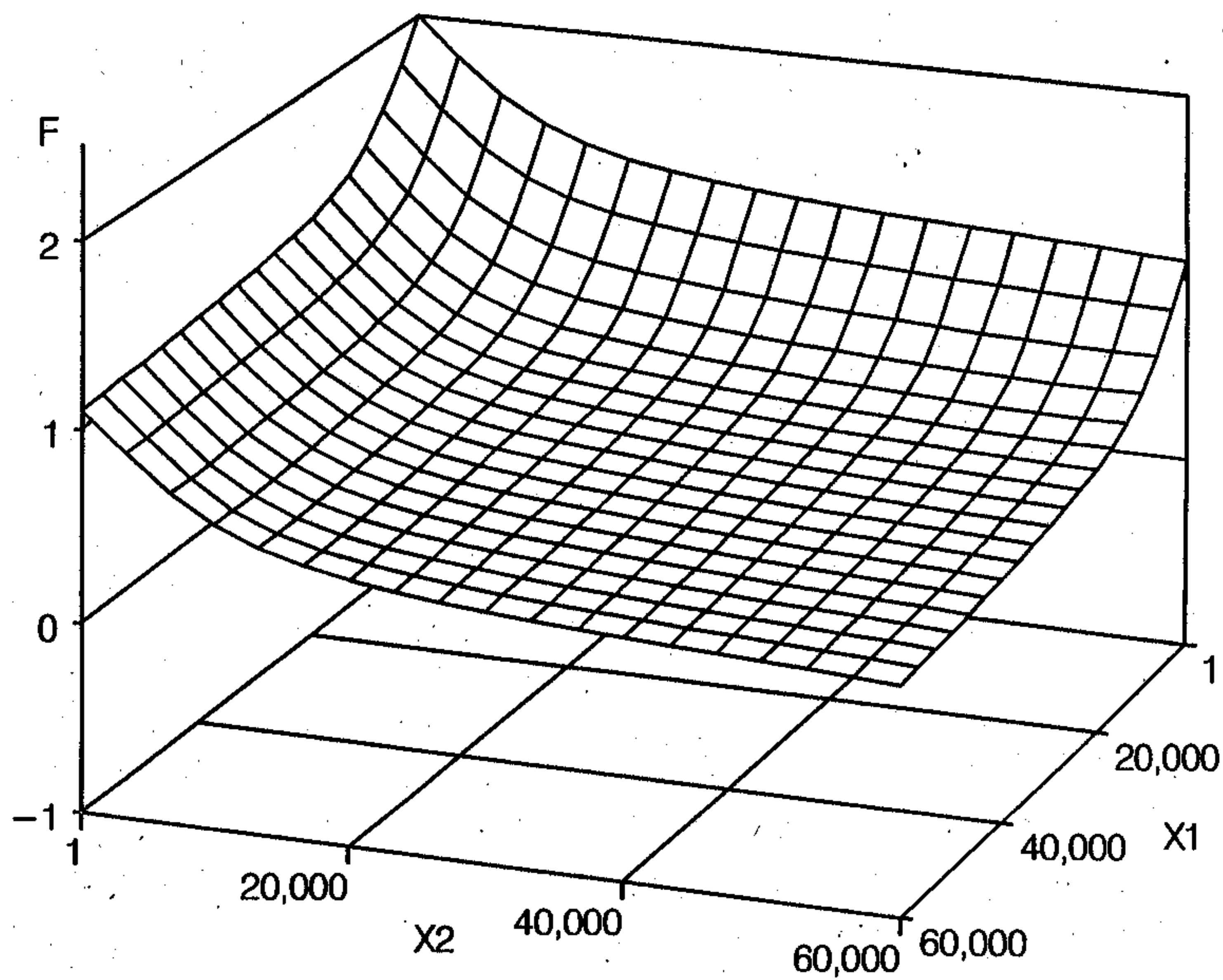
The graph of f is shown in Figure 1. We can find the optimum solution by the following reasoning, without using Chromy's algorithm. The constraint $x_h \in [1,100000]$ limits the region in the x_1x_2 plane over which f is defined. Some of the other constraints further limit this region; constraint 4 most limits the values of x_1 and constraint 3 most limits the values of x_2. Because f decreases monotonically for increasing x_1 and x_2, the minimum value of f is at (46745,3564) (equivalently, at (p_1,p_2) = (0.0000214,0.0002806)). The expected sample size is 0.0000214 + 0.0002806 = 0.000302.

We ran Chromy's algorithm for this 2-unit problem, using a nearness criterion of K = 0. It met this criterion after 15 iterations, converging to the optimum solution found above.

We also computed values of p_h using the measure-of-size approach of 1989 ASM sample selection. We generalized this approach by defining each unit's measure of size as the sum of predicted squared differences for its product classes <u>and</u> its 4-digit SIC code. Using E(n) = 0.000302 in the formula of Section 3 for p_h, we obtained (p_1,p_2) = (0.000111,0.000191), or (x_1,x_2) = (8979,5247). This illustrates a problem with this method: the expected sample size is the same minimum found by Chromy's algorithm, but the solution is outside the area in the x_1x_2 plane defined by the most limiting constraints, i.e., the constraints are not all met.

The problem was much more complex for allocating the 1994 ASM sample. Table 1 would expand to about 201,000 rows and 2,230 columns. The next section provides details.

Figure 1: 2–Establishment Example of Cost Function F
(Scale Exaggerated to Show Shape of Surface)



6. ALLOCATION OF THE 1994 SAMPLE

The budgeted sample size of the 1994 ASM was about 58,000 establishments. There were 371,000 establishments in the 1992 Census of Manufactures. The sampling frame for the 1994 ASM consisted of 231,000 of these establishments. The other 140,000 establishments were excluded from the frame because of their small size; their ASM data came from administrative records of other Federal agencies. The Census Bureau designated about 25,000 establishments as certainties ($p_h = 1$): all establishments of very large companies, establishments with 250 or more employees, plants under construction, and manufacturers of certain computer products. The Census Bureau also set aside nearly 2,000 idle establishments for special sampling, and identified over 3,000 deaths and other deletions since the census. The remaining approximately 201,000 "noncertainty units" were then eligible for drawing a sample of about 33,000 units.

To specify our variance constraints, we began by dividing the 4-digit SIC codes and product classes into deciles, based on 1992 census value of shipments. The resulting groups are shown in Table 2. We set target coefficients of variation (CVs) to achieve the greatest precision in groups with larger values of shipments. We transformed the CV targets to variance constraints with the relationship

$$Variance = (CV)^2 (Value\ of\ Shipments)^2$$

We ran Chromy's algorithm several times, and adjusted the target CVs until the expected sample size from the noncertainties was a little under 33,000. Table 2 shows the final targets.

We also specified that selection probabilities could be no smaller than 0.02, to put a cap on sample weights (1/0.02 = 50) and minimize the effect of individual units on subpopulation totals. For the 1989 ASM, we used minimum probabilities of 0.005. We found that increasing the minimum to 0.02 only increased the expected sample size by about 500 units for the 1994 ASM.

We ran a series of SAS programs to: prepare data, including computation of predicted differences by 4-digit SIC code and product class, based on 1992 census value of shipments data; perform the iterations of Chromy's algorithm; check for constraints that were not met; and adjust probabilities so that all constraints were satisfied. Chromy's algorithm needed 18 iterations to satisfy the nearness criterion with K = 50 (the approximate distance of $f(x)$ from $f(x^*)$ was 45). The expected noncertainty sample size was 31,258. The SAS program that ran the algorithm took by far the most time of the series of programs: 12.0 hours cpu time on a VAX 9000 computer. We rewrote this program in SAS/IML (matrix language) and reduced the cpu time to 3.4 hours.

Table 2: Final Target CVs by 4-Digit SIC Code and Product Class

4-Digit SIC Code		Product Class	
Value of Shipments (Million $)	Target CV (Percent)	Value of Shipments (Million $)	Target CV (Percent)
0-662	17	0-94	15
662-1,111	14	94-189	13
1,111-1,828	11	189-309	10
1,828-2,420	9	309-452	8
2,420-3,110	7	452-626	6
3,110-4,016	6	626-868	4.5
4,016-5,468	5	868-1,237	3.75
5,468-7,635	5	1,237-1,841	2.75
7,635-13,586	3	1,841-3,608	1.75
13,586+	2	3,608+	1

Forty-seven of the 1,773 product class constraints were not met before adjustment (see Table 3). That is, the ratio of the variance of the estimator to the target variance was 1.01 or greater for these product classes. Only two ratios exceeded 1.50, and the largest ratio was 2.11. Many product class constraints were more than satisfied. For example, 288 ratios were 0.9 or less, and 77 ratios were 0.1 or less. All 4-digit SIC code constraints were met (the ratio was 1.00 or less); 448 of the 457 constraints were more than satisfied. For example, the ratio was 0.1 or less for 190 4-digit SIC codes. When we adjusted the selection probabilities to meet all constraints, the expected sample size increased by only 60 units to 31,318.

The faster matrix version of our program enabled us to re-run Chromy's algorithm with a smaller nearness criterion, after we had selected the 1994 sample. We tried K = 5. This required 41 iterations (7.8 cpu hours), and yielded an expected noncertainty sample size of 31,302. This met all 4-digit SIC code constraints, and all but 11 product class constraints. The adjustment to meet all constraints required an increase of seven units to 31,309, virtually the same total as for K = 50.

Table 3: Satisfaction of Variance Constraints by Three Methods, Measured by the Ratio of Variance to Target Variance

Result	Method		
	Chromy $p_h \geq 0.02$	Chromy $p_h \geq 0.000001$	Meas-of-size $p_h \geq 0.000001$
No. of Product Classes:			
Ratio ≥ 1.01	47	101	383
Ratio > 1.50	2	9	309
Ratio > 10	0	3	40
No. of 4-Digit SIC Codes:			
Ratio ≥ 1.01	0	5	0
Exp. Sample Size			
Before adjust	31,258	30,858	30,858
After adjust	31,318	31,893	37,058
Increase	60	1,035	6,200

We compared Chromy's algorithm with the measure-of-size approach that was used for 1989 sample selection. For the comparison, we lowered the minimum probability constraint to $p_h = 0.000001$ for both methods. If we had kept the constraint at $p_h = 0.02$, the two methods would have yielded different expected sample sizes (before adjustment to meet all constraints), making comparisons difficult. We ran Chromy's algorithm with the same nearness criterion as before, K = 50. This criterion was not met after 30 iterations, so we chose the selection probabilities for which the distance was nearest 50. This occurred on iteration 20, when the distance was 52.9 and the expected sample size was 30,858 (see Table 3). Variance constraints were not met for 101 product classes at this point. The ratio of population variance to target variance exceeded 1.50 for nine product classes. Three constraints were badly missed: the ratio exceeded 10 for them. Five 4-digit SIC code constraints were not satisfied, including one 4-digit SIC code for which the ratio exceeded 10. The adjustment to meet all constraints required an increase of 1,035 units to 31,893.

Then we computed selection probabilities with a measure-of-size approach with the same expected sample size as above (30,858). We did not exactly follow the 1989 method, because mainframe computer programs were no longer available and because some unrepeatable judgement was required to select a cutoff for certainties and to select a supplemental sample. As in the 2-unit example of Section 5, we defined each unit's measure of size as the sum of predicted squared differences for its product classes and for its 4-digit SIC code. We derived values of p_h with the formula in Section 3. Some values were 1 or more. We set p_h to 1 for these units, and computed p_h for the remaining noncertainty units, again using the formula in Section 3. We repeated this process several times, until all probabilities were no more than 1. The measure-of-size approach was inferior to Chromy's algorithm, because 383 product class constraints were unsatisfied, and the variance ratio exceeded 10 for 40 product classes (see Table 3). However, all 4-digit SIC code constraints were met. We adjusted probabilities to meet all constraints, using the same method that we used after running Chromy's algorithm. The adjustment required a larger increase than with Chromy's algorithm: 6,200 units to a total of 37,058.

One strategy that the Census Bureau employed with the 1989 and earlier ASM samples was to compute probabilities with a reduced expected sample size (for example, 5,000 less), determine which constraints were badly missed, and supplement the expected sample size to meet as many constraints as possible. We tried a similar approach. We reduced the expected sample size in the measure-of-size formula for p_h by 5,000 (to 25,858). Our subsequent adjustment to meet all constraints required an increase of over 11,000 units to 37,001. This was only a slight improvement over our initial allocation by the measure-of-size approach.

7. CONCLUSIONS AND RECOMMENDATIONS

Chromy's algorithm, followed by an adjustment procedure, enabled us to objectively assign selection probabilities to units so that all variance constraints were satisfied. It yielded an expected sample size that was over 5,000 units smaller than when the measure-of-size approach was used under the same conditions, a savings of tens of thousands of dollars in data collection costs to obtain comparable precision. We required less staff time and we could more easily study alternative constraints (minimum probabilities and target CVs). We recommend the continued use of Chromy's algorithm for allocating the ASM sample.

Computer time is no longer a serious limitation for using Chromy's algorithm, since we reduced cpu time by 70 percent by rewriting the original SAS program in SAS/IML. Now we can complete a run of the algorithm during the day, or set the nearness criterion much smaller. However, we only reduced the sample size slightly by reducing the criterion from K = 50 to 5.

We found that an alternative stopping rule for Chromy's algorithm, developed by Zayatz and Sigman, was unnecessary because we were always able to obtain results by using the nearness criterion.

Predicted squared differences in value of shipments are used throughout the ASM allocation procedure. If these predictions are not accurate, Chromy's algorithm will not be as efficient as it could be. Therefore, research should be conducted to see how well the regression models predict squared differences, and other predictors should be investigated. Methods for dealing with outliers should also be studied. We have begun some of this work, by considering other predictors that make sense according to economic theory, and by investigating "resistant regression" and other methods for handling outliers. We expect to publish initial results later this year.

Before we ran Chromy's algorithm, we designated about 25,000 units in the frame as certainties. This may not be the best approach. Different methods for choosing certainties should be studied, including letting Chromy's algorithm select all the certainties.

REFERENCES

CAUSEY, B.D. (1983), "Computational Aspects of Optimal Allocation in Multivariate Stratified Sampling," SIAM Journal of Scientific and Statistical Computing, 4, 322-329.

CHROMY, J.R. (1987), "Design Optimization with Multiple Objectives," Proceedings of the Section on Survey Research Methods, American Statistical Association.

DEE, T. and DORINSKI, S. (1993), "1994 ASM Sample Design Research," unpublished manuscript, U.S. Bureau of the Census.

HAJEK, J. (1964), "Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population," Annals of Mathematical Statistics, 35, 1431-1523.

OFFICE OF MANAGEMENT AND BUDGET (1987), Standard Industrial Classification Manual, Springfield, VA: National Technical Information Service.

OGUS, J.L. and CLARK, D.F. (1971), "The Annual Survey of Manufactures: A Report on Methodology," Technical Paper 24, U.S. Bureau of the Census.

WAITE, P.J. and COLE, S.J. (1980), "Selection of a New Sample Panel for the Annual Survey of Manufactures," Proceedings of the Social Statistics Section, American Statistical Association.

ZAYATZ, L. and SIGMAN, R. (1994), "Feasibility Study of the Use of Chromy's Algorithm in Poisson-Sample Selection for the Annual Survey of Manufactures," Proceedings of the Section on Survey Research Methods, American Statistical Association.