Contract SRS-9424161
MPR Reference 8239-210

SAMPLING ERRORS FOR SESTAT AND ITS COMPONENT SURVEYS:

1993

January 3, 1997

Authors:

Brenda G. Cox
Donsig Jang
David Edson

Submitted to:

National Science Foundation
Division of Science Resource Studies
4201 Wilson Blvd., Room 965
Arlington, VA 22230

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
(202) 484-9220

Project Officer:
Linda P. Hardy

Project Director:
David Edson

TABLE OF CONTENTS

I. INTRODUCTION

SESTAT combines information from three 1993 NSF-sponsored surveys:

- The National Survey of College Graduates (NSCG),
- The Survey of Doctorate Recipients (SDR), and
- The National Survey of Recent College Graduates (NSRCG).

The combined SESTAT database represents U.S. residents with a baccalaureate degree or higher who were trained as or working as scientists or engineers as of the reference period of April 1993.

This report provides information about sampling errors for SESTAT and its survey components. We discuss how sampling errors were developed and how users can calculate standard errors for estimates derived from SESTAT or one of its component surveys.

II. UNDERSTANDING SAMPLING ERRORS

This chapter contains a brief discussion of the types of survey errors and the accuracy of estimates derived from surveys.

A. Types of Survey Errors

Estimates derived from sample surveys are subject to two types of errors--sampling errors and nonsampling errors. Nonsampling errors[1] can be attributed to many sources, such as response differences, definitional difficulties, differing respondent interpretations, and respondent inability to recall information.

Sampling errors (the focus of this presentation) occur when estimates are derived from a sample rather than a census of the population. The sample used for a particular survey is only one of a large number of possible samples of the same size and design that could have been selected. Even if the same questionnaire and instructions were used, the estimates from each sample would differ from the others. This difference, termed sampling error, occurs by chance, and its variability is measured by the standard error associated with a particular survey estimate. Estimates of the characteristics of scientists and engineers obtained using SESTAT and its survey components are based on sample surveys and are thus subject to sampling errors. The standard errors and related information provided here indicate the general magnitude of the sampling errors for the three surveys when analyzed separately and when combined to create the full SESTAT database.

B. Assessing the Accuracy of Estimates

Having estimated a population quantity such as a mean or total, it is desirable to assess the accuracy of the estimate. The customary approach is to construct a *confidence interval* within which one is sufficiently sure the true population value lies. The *standard error* of a survey estimate measures the

[1]For a general discussion of nonsampling errors, see *Nonsampling Errors in Surveys* by Judith T. Lessler and William D. Kalsbeek (New York: John Wiley & Sons, 1992). While the full extent of nonsampling errors is usually unknown, a variety of related research has been conducted for the SESTAT surveys. Some of the information from this research has been summarized in the technical notes associated with the SESTAT data elements, accessible through the SESTAT Home Page.

precision with which an estimate from one sample approximates the true population value, and thus can be used to construct a confidence interval for a survey parameter to assess the accuracy of the estimate[2]. Let $\hat{\theta}$ be an estimator of a parameter of interest θ with a standard error $SE(\hat{\theta})$. If the sample size is large, then an approximate (1- α)100% confidence interval for θ is:

$$\left\{\hat{\theta} - z_{\alpha/2} SE(\hat{\theta}),\ \hat{\theta} + z_{\alpha/2} SE(\hat{\theta})\right\}$$

where $z_{\alpha/2}$ is the upper α /2 percentage point of the normal distribution with mean zero and variance one.

If the process of selecting a sample from the population were repeated many times and an estimate and its standard error calculated for each sample, then:

a. Approximately 90% (α = 0.10) of the intervals from 1.645 (= $Z_{.05}$) standard error below the estimate to 1.645 standard error above the estimate will include the true population value.

b. Approximately 95% (α = 0.05) of the intervals from 1.96 (= $Z_{.025}$) standard errors below the estimate to 1.96 standard errors above the estimate will include the true population value.

c. Approximately 99% (α = 0.01) of the intervals from 2.575 (= $Z_{.005}$) standard errors below the estimate to 2.575 standard errors above the estimate will include the true population value.

With an estimate of the standard error and the factors above (1.645, 1.96, or 2.575), a data user may construct a confidence interval, or range of values, that includes the true population value with the given probability α (=0.10, 0.05, or 0.01).

[2] Another related term is the *variance,* which is the square of the standard error and is sometimes used in standard error calculations.

III. GENERALIZED VARIANCE FUNCTIONS (GVFs): A METHODOLOGY FOR ESTIMATING STANDARD ERRORS

A *generalized variance function* (GVF) is a mathematical model that describes the relationship between a statistic (such as a population total) and its corresponding variance. GVF models are used to approximate standard errors for a wide variety of estimates of characteristics of the target population[3].

A. GVF Modeling

GVF modeling consisted of two steps:

- calculating population totals and their variances directly for a small subset of the survey items, and
- modeling the relationship between the survey-derived totals and their associated variances.

Step 1 – Calculate Population Totals and Their Variances

For direct calculation of the variance (Step 1), a successive differences method or a resampling method such as random groups, balanced repeated replication, or jackknife replication might be used. Direct variance estimation techniques used in each survey are described in Chapter IV.

Step 2 – Model Relationships Between Survey-derived Totals and Sampling Errors

GVF models (Step 2) use regression modeling techniques and hence are subject to the same limitations of model specification, fit, and estimation as any other model. The principal advantage of the GVF method is that approximations of sampling errors are simplified for the large amount of estimates that are normally generated from a demographic survey with a large number of variables. For the SESTAT and its three component surveys, GVF models are available for the total population and for domains of interest. Analysts can use these models to predict the variance for a particular statistic by inserting the value of the statistic into the model for the appropriate domain and survey component. The models developed for SESTAT and each of its component surveys are also described in Chapter IV.

---

[3]For more information on GVFs, see Chapter 5 of *Introduction to Variance Estimation*, by Kirk Wolter (New York: Springer-Verlag, 1985).

B. A Methodology Overview

Let $\hat{Y}$ denote an estimator of the population total Y. GVF models are usually created for the *relative variance* of the estimated total, or

$$RelVar(\hat{Y}) = \frac{Var(\hat{Y})}{Y^2},$$

where $Var(\hat{Y})$ is the variance of $\hat{Y}$. The modeling typically begins by assuming that the relative variance of an estimated total is a linear function of the inverse of the total Y being estimated, or

$$RelVar(\hat{Y}) = \beta_0 + \frac{\beta_1}{Y}$$

The parameters of the GVF model, β_0 and β_1, are unknown and estimated from a subset of all possible survey-derived totals and their variances by some form of least squares regression estimation.

The relative variance of an estimated total $\hat{Y}$ can be predicted by evaluating the appropriate GVF model using the estimated values for Y, β_0 and β_1. Thus, using the GVF model, the standard error of a specific estimated total can be predicted by inserting the value of the estimated total into the following computational equivalent:

$$SE(\hat{Y}) = \sqrt{\hat{\beta}_0 \hat{Y}^2 + \hat{\beta}_1 \hat{Y}} \qquad \text{(III.1)}$$

where: $SE(\hat{Y})$ is the predicted standard error of the estimated total $\hat{Y}$, and

$\hat{\beta}_0$, $\hat{\beta}_1$ are estimates of the regression parameters β_0 and β_1 .

The GVF model can also be adapted to estimate the standard error of a percentage. Using the same parameters, the standard error for a percentage can be predicted with this formula:

$$SE(\hat{P}) = \sqrt{\frac{\hat{\beta}_1}{\hat{Y}} \hat{P}(100 - \hat{P})} \qquad \text{(III.2)}$$

where: $SE(P)$ is the predicted standard error for a specific estimated percentage $\hat{P}$, and

$\hat{Y}$ is the estimated number of persons in the base of the percentage.

IV. CALCULATING STANDARD ERRORS

This chapter describes how users can calculate standard errors based on GVFs. There is a separate section for each of the component surveys as well as SESTAT as a whole. Each section describes:

(1) the GVF model used,
(2) the method used to obtain directly calculated variance estimators, and
(3) the resulting estimated GVF parameters, displayed as parameter tables.

Parameter tables enable the user to calculate standard errors for a wide range of population totals and percentages. Instead of displaying standard errors, these tables provide parameters that the user inserts into formulas (provided in this section) to calculate standard errors.

Basic Steps to Approximating Standard Errors

The following steps may be followed to approximate the standard error of an estimated total or percentage:

1. Determine the appropriate survey and domain for a characteristic of interest
2. Obtain the estimated total or percentage from the survey;
3. Determine the most appropriate domain for the estimate from parameter table(s);
4. Refer to the parameter table to get the parameter estimates for this domain; and
5. Compute the approximate standard error using the equations provided.

Examples showing how to use these tables for SESTAT and each component survey are also provided.

A. The Scientists and Engineers Statistical Data System (SESTAT)

1. GVF Model

The 1993 Scientists and Engineers Statistical Data System (SESTAT) formed the GVF model for the variance of the estimate as a quadratic function of the total, or:

$$Var(\hat{Y}) = \beta_0 Y^2 + \beta_1 Y$$

where: Y is the population total being estimated

$VAR(\hat{Y})$ is the variance of the estimated total $\hat{Y}$

β_0 and β_1 are parameters of the model.

For the SESTAT data, GVF models were specified for the overall population and for separate subgroups such as gender, race/ethnicity group, field of highest degree, occupation, and combinations of these characteristics.

To fit the model, 60 population totals were estimated for each domain. Direct estimates of the variances for these domain totals were generated using the method of random groups. Ordinary Least Squares Regression was used to derive estimates of β_1 and β_0 with the estimated domain totals and their directly calculated variances as inputs. The results are presented as a table of generalized variance model parameters which can be used to estimate standard errors.

2. Direct Variance Estimates – The Method of Random Groups

The random group technique is appropriate when the sampling structure(s) of the survey(s) is sufficiently complex that analytically-derived variance estimation formulas become unmanageable. In general, variance estimation using the *method of random groups* consists of drawing multiple samples from a target population (or subpopulation) of interest and then constructing separate estimates for each sample. The dispersion of the different population estimates provide the basis for the variance measure.

For the SESTAT variance measures, the survey sample was divided into random subsamples, chosen to mimic the sample design procedures for the total sample and weighted appropriately. From the three

SESTAT component surveys, the observations within each stratum were randomized and twenty separate random group samples were systematically selected without replacement:

1993 NSCG Sampled cases were assigned to twenty random groups within each sampling strata.

1993 SDR Respondent cases were assigned to twenty random groups within each sampling strata.

1993 NSRCG Because of the two step sample design of the NSRCG, two sets of random groups were selected:

- Responding students from certainty institutions were assigned to twenty random groups within each sampling strata.
- Non-certainty institutions were assigned to random groups. All sampled students from the institution were then assigned to that institution's random group.

The sets of random groups for each survey were combined to create twenty SESTAT random groups, with each group representing a valid sample of the combined SESTAT target population.

3. Examples: How Data Users Can Calculate Standard Errors for the 1993 SESTAT

We offer two methods for obtaining standard errors for SESTAT estimates. Method 1 is easiest but is limited in its application to standard errors for estimated totals. Method 2 can be used to predict standard errors for estimated percentages and totals.

Method 1. Obtaining Standard Errors from the Look-up Tables

The look-up tables provide approximate standard errors for estimated counts of scientists and engineers for the total population and for different segments of the population.

Assume the estimate of the number of scientists and engineers employed in S&E occupations was approximately 8 million people. The total column in the table labeled *Scientists and Engineers in 1993 (Total Population): Approximate Standard Errors for Specified Demographic Groups* shows a standard error estimate of 57,450 associated with an estimated count of 8 million. Then the 95% confidence interval is 1.96 (the factor for the 95% confidence interval) times the standard error from the table (57,450). Or

1.96 x 57,450 = 112,602. Thus, the 95% confidence interval for the true value is the interval between 7,887,398 and 8,112,602 (8,000,000 +/- 112,602).

There are several versions of the Look-up Tables. As a general rule, use the table that is most specific to the domain you are studying and the database being analyzed. Thus, the "total" category is used when more than one degree level is included. In many cases, the exact estimate will not be included in the Look-up Tables. For these standard errors, you may use linear interpolation for intermediate values or you may wish to use *Method 2*.

Method 2. Using the Parameter Tables

The parameter tables provide a method for approximating standard errors for estimated counts and percents of scientists and engineers for the total SESTAT population and for different segments of the population.

Example for Totals

Suppose SESTAT data are used to estimate the total population size of individuals employed in science or engineering occupations. As the domain for this population is the total science and engineering population, we look in *Scientists and Engineers in 1993 (Total Population): Listing of* β_0 *and* β_1 *Parameters for Specified Demographic Groups* and determine the values for $\hat{\beta}_0$ (= 0.000029) and $\hat{\beta}_1$ (= 176.694901). Using equation III.1, we estimate the standard error as:

$$SE(\hat{Y}) = \sqrt{(0.000029\,\hat{Y}^2) + (176.694901\,\hat{Y})}$$

We substitute the value of $\hat{Y}$ = 8,311,787 and obtain $SE(\hat{Y})$ = 58,925. Thus, a 95% confidence interval for the true value for the total number of individuals employed in science or engineering occupations would be 8,311,787 +/- 115,493, where 115,493 represents 1.96 times the standard error.

Example for Percentages

To illustrate the use of the formula for determining standard errors for percentages, suppose that we use SESTAT estimates to determine that 82% of female scientists and engineers were participating in the labor force as of the reference week. The base for this percentage is the number of female scientists and engineers, estimated at 3,867,887.

Obtaining the value for $\hat{\beta}_1$ from the appropriate parameter table and using equation III.2:

$$SE(\hat{P}) = \sqrt{\frac{270.328}{\hat{Y}} \hat{P}(100-\hat{P})}$$

Substituting for the values of the base $\hat{Y}$ (3,867,887) and the percentage $\hat{P}$ (82), we obtain an estimated standard error of 0.067. The 95% confidence interval for the labor force participation rate for females is 82% +/- 0.131% (where 0.131 equals 1.96 times the estimated standard error).

B. 1993 National Survey of College Graduates (NSCG)

1. GVF Model

The 1993 National Survey of College Graduates (NSCG) modified the basic GVF model (see Chapter III) to ensure that negative variances were not predicted. Let T_d represent the population total for a domain *d* for which a GVF model was desired for the NSCG. The 1993 NSCG assumed that:

$$\beta_0 = -\frac{\beta_1}{T_d}$$

and fit the following GVF model for totals derived from domain *d*:

$$RelVar(\hat{Y}) = \beta_1 \left(\frac{1}{Y} - \frac{1}{T_d} \right)$$

For the NSCG data, GVF models were specified for the overall population and for separate subgroups such as race/ethnicity, field of education, degree level, and sex.

To fit the model, between 37 to 152 different totals were estimated for each domain. Direct estimates of the variances for these domain totals were generated using the successive difference method and then converted to relative variances. The estimated domain totals and their directly estimated relative variances were then used to fit the model using weighted least squares with the weight equal to the inverse

of the square of the predicted relative variance (after removing outliers). The results are presented as a table of generalized variance model parameters which can be used to estimate standard errors.

2. Direct Variance Estimates – The Successive Difference Method

Because systematic sampling was used in sample selection, the 1993 NSCG chose the *successive difference method* to calculate their direct variance estimates. The successive difference method estimator is based on squared differences between neighboring sample cases. For illustration, let y_i denote the value of the *I*-th systematic sample unit from a population of size *N*, where i=1,2,...,*n*.. Then, an estimator of the variance of the estimated population total

$$\hat{Y} = \sum_{i=1}^{n} y_i$$

using the successive difference method is:

$$\hat{V}_{SD} = (1-f)\frac{n}{2(n-1)}\sum_{i=2}^{n}(y_i - y_{i-1})^2$$

where *f* denotes the sampling fraction *n/N* and *n* is a total sample size.

3. Examples: How Data Users Can Calculate Standard Errors for the 1993 NSCG

The parameter tables provide a method for approximating standard errors for estimated counts and percents of scientists and engineers for the total NSCG population and for different segments of the population.

Totals: Suppose data from 1993 NSCG are used to estimate persons having a bachelor's degree in engineering and the corresponding estimate is reported as 1,500,000. As the domain for this population is the total engineering population, we look in this table for the 1993 NSCG to determine the values for

$\hat{\beta}_0$ (= -0.000006) and $\hat{\beta}_1$ (= 191).

Then, using Equation III.1, the approximate standard error on the estimated number of 1,500,000 persons is:

$$SE(\hat{Y}) = \sqrt{(-0.000006 \times 1{,}500{,}000^2) + (191 \times 1{,}500{,}000)} = 16{,}523$$

Thus, a 95% confidence interval for the true value for the total number of persons having a bachelor's degree in engineering would be between 1,467,615 and 1,516,523 (1,500,000 +/- 32,385), where 32,385 represents 1.96 times the standard error.

Percentages: Suppose the percentage of the persons with a bachelors degree in engineering who are women is 12.0% among 1,500,000 persons. Obtaining the appropriate value of the parameter $\hat{\beta}_1 = (191)$ from *Parameter Tables: 1993 NSCG* and using Equation III.2, the approximate standard error on 12.0% is

$$SE(\hat{P}) = \sqrt{\frac{191}{1{,}500{,}000} \times 12\,(100-12)} = 0.37\%$$

The 95% confidence interval for the percentage of the persons with a bachelors degree in engineering who are women is 12% +/- 0.725% (where 0.725 equals 1.96 times the estimated standard error).

C. The Survey of Doctorate Recipients (SDR)

1. GVF Model

The 1993 Survey of Doctorate Recipients (SDR) used the basic GVF model or,

$$RelVar(\hat{Y}) = \beta_0 + \frac{\beta_1}{Y}.$$

For the SDR data, GVF models were specified for the scientists and engineers doctoral population overall, for groups defined by field, and for selected subgroups of analytic interest.

To estimate the parameters of the model, totals for a number of different characteristics were estimated for each domain. Direct estimates of the variances for these domain totals were generated using the *Balanced Repeated Replication* Procedure and then converted to relative variances. The estimated domain totals and their relative variances were then used to fit the model through a series of successive iterations of the relative variance function until values of $\hat{\beta}_0$ and $\hat{\beta}_1$ differed negligibly. The final set of $\hat{\beta}_0$ and $\hat{\beta}_1$ estimates were ultimately used to calculate approximate standard errors of totals. The results are presented as a table of GVF model parameters which can be used to estimate standard errors.

Instructions are provided on how data users can calculate standard errors for an estimated total or proportion using these parameters.

2. Direct Variance Estimates – The Balanced Repeated Replication Method

The 1993 SDR chose the *balanced repeated replication method* to calculate their direct variance estimates.[4] Using this procedure, the overall sample is partitioned into many overlapping subsamples (replicates). Suppose it is desired to estimate a population total *Y* from a stratified design. Let *H* denote the number of strata and *R* be the number of replications. Let $\hat{Y}_{hr}$ be the estimator of the *h*-th stratum total Y_h from the *r*-th replicate where $h=1,2,...,H$ and $r=1,2,..,R$. Then, the variance of the *h*-th stratum total estimator $\hat{Y}_h$ for the given domain is expressed as:

$$S_h^2 = \frac{1}{R} \sum_{r=1}^{R} (\hat{Y}_{hr} - \bar{\hat{Y}}_h)^2$$

where

$$\bar{\hat{Y}}_h = \frac{1}{R} \sum_{r=1}^{R} \hat{Y}_{hr}$$

Thus, the variance of the total estimator $\hat{Y}$ is given by

$$\hat{V}_{BRR} = \sum_h S_h^2$$

where:

$$\hat{Y} = \sum_h^H \hat{Y}_h$$

3. Examples: How Data Users Can Calculate Standard Errors for the 1993 SDR

The parameter tables provide a method for approximating standard errors for estimated counts and percents of scientists and engineers for the total population and for different segments of the population.

Totals: If the number of life science Ph.D.s employed in universities and 4-year colleges is reported as 66,900 and one wishes to determine the approximate standard error, one can use the values of

[4]For more information about this topic, see Chapter 3 of *Introduction to Variance Estimation*, by Kirk Wolter (New York: Springer-Verlag, 1985).

$\hat{\beta}_0$ (= −0.000068) and $\hat{\beta}_1$ = (10.5008) for all life science Ph.D.s from *Parameter Tables: 1993 SDR* and Equation III.1. Then the approximate standard error is:

$$SE(\hat{Y}) = \sqrt{-0.000068(66{,}900)^2 + 10.5008(66{,}900)} = 631$$

To develop a 95% confidence interval around this estimate of 66,900, one would add and subtract from the estimated total the standard error multiplied by 1.96, or

$$66{,}900 \pm (1.96 \times 631) = 65{,}663 \text{ to } 68{,}137.$$

This range of 65,663 to 68,137 represents the 95% confidence interval for the estimated total of 66,900.

Percentages: The unemployment rate for physical science doctorate was estimated at 2.1% of the 100,660 individuals who are in the labor force. Using the appropriate value of $\hat{\beta}_1$ (= 11.9638) from *Parameter Tables: 1993 SDR* and the GVF formula in Equation III.2, the standard error of $\hat{P}$ is determined as follows:

$$SE(\hat{P}) = \sqrt{\frac{11.9638}{100{,}660} \times 2.1(100 - 2.1)} = 0.156$$

To develop a 95% confidence interval around this estimate of 2.1%, one would add and subtract from the estimate the standard error multiplied by 1.96, or

$$2.1 \pm (1.96 \times .156) = 1.8 \text{ to } 2.4$$

The range of 1.8% to 2.4% represents the 95% confidence interval for the estimated percent of 2.1.

D. The National Survey of Recent College Graduates (NSRCG)

1. GVF Model

The 1993 National Survey of Recent College Graduates used the basic GVF model (see Chapter III), or

$$RelVar(\hat{Y}) = \beta_0 + \frac{\beta_1}{Y}.$$

For the NSRCG data, GVF models were specified for 1991 bachelor's recipients, 1991 master's recipients, 1992 bachelor's recipients, and 1992 master's recipients. In addition, the models were specified separately within each of these groups for graduates by gender (male/female), major (science/engineering), occupation (scientist/engineer/other), and race (white/black/Hispanic/other).

To estimate the parameters of the model, totals of 54 different characteristics were estimated for each domain. Direct estimates of the variances for these domain totals were generated using the jackknife replication method and then converted to relative variances. The estimated domain totals and their relative variances were then used to fit the model using ordinary least squares to produce estimates of the parameters of the model. The results are presented as a table of GVF model parameters which can be used to estimate standard errors.

2. Direct Variance Estimates -- The Jackknife Replication Method

The 1993 NSRCG chose the *jackknife replication method* to calculate their direct variance estimates. Jackknife replication involves constructing a number of subsamples (replicates) from the full sample and computing the statistics of interest for each replicate. To construct the replicates, 50 stratified subsamples of the full sample were created. Fifty jackknife replicates were then formed by deleting one subsample at a time from the full sample. The variability between the replicate estimates provides an estimate of the variance of the statistic of interest. WesVarPC, a public use computer program developed at Westat and available on the Internet at www.westat.com, was used to calculate direct estimates of standard errors for a number of statistics from the survey using the jackknife methodology.

3. Examples: How Data Users Can Calculate Standard Errors for the 1993 NSRCG

The parameter tables provide a method for approximating standard errors for estimated counts and percents of scientists and engineers for the total population and for different segments of the population.

Totals: Suppose that the number of 1991 bachelor's degree recipients in engineering currently working in a engineering-related job in 1993 was 40,000 ($\hat{Y}$=40,000). The most appropriate domain from *Parameter Tables: 1993 NSRCG* is engineering majors with bachelor's degrees from 1991 and the parameters are $\hat{\beta}_0$=0.000818 and $\hat{\beta}_1$=80.968 .

Then the approximate standard error using the GVF equation, Equation III.1, is:

$$SE(\hat{Y}) = \sqrt{0.000818(40000)^2 + 80.969(40000)} = 2{,}133$$

To develop a 95% confidence interval around the estimate of 40,000, one would add and subtract from the estimate the standard error multiplied by 1.96. This means that the average estimate from all possible samples would be expected 95 times out of 100 to fall within the range of

$$40{,}000 \pm (1.96 \times 2{,}133) = 35{,}819 \text{ to } 44{,}181$$

This range of 35,819 to 44,181 represents the 95% confidence interval for the estimate of 40,000.

Percentage: Suppose that of the 40,000 1991 engineering bachelor's degree recipients, 60% ($\hat{P}$=60) were working in a engineering-related job in 1993. The most appropriate domain from *Parameter Tables: 1993 NSRCG* is engineering majors with bachelor's degrees in 1991 which shows $\hat{\beta}_1$ = 80.969. Then, the approximate standard error using the GVF equation, Equation III.2, is:

$$SE(\hat{P}) = \sqrt{\frac{80.969}{40{,}000} \times 60\,(100-60)} = 2.2\%$$

To develop a 95% confidence interval around this estimate of 60%, one would add and subtract from the estimated percentage the standard error multiplied by 1.96. That is, the estimate from all possible samples would be expected 95 times out of 100 to fall within the range

$$60 \pm (1.96 \times 2.2) = 55.7 \text{ to } 64.3$$

The range of 55.7% to 64.3% represents the 95% confidence interval for the estimated percent of 60.

APPENDIX A

LOOK-UP TABLES:
1993 SESTAT

Scientists and Engineers in 1993 (Total Population):
Approximate Standard Errors for Specified Demographic Groups

Estimated Number	Demographic Group				
	Total Population	Male	Female	White	Nonwhite
100	130	130	160	140	90
200	190	180	230	200	130
500	300	290	370	320	200
750	360	350	450	390	250
1,000	420	410	520	450	290
2,000	590	580	740	630	410
3,000	730	710	900	780	500
4,000	840	820	1,040	900	580
5,000	940	910	1,160	1,000	650
10,000	1,330	1,290	1,640	1,420	920
25,000	2,110	2,040	2,600	2,250	1,480
50,000	2,980	2,900	3,670	3,180	2,150
100,000	4,240	4,120	5,190	4,520	3,190
250,000	6,780	6,600	8,190	7,220	5,690
500,000	9,780	9,530	11,540	10,380	9,370
750,000	12,210	11,920	14,080	12,930	12,900
1,000,000	14,360	14,040	16,190	15,170	-
2,000,000	21,710	21,360	22,540	22,760	-
3,000,000	28,200	27,880	27,170	29,390	-
4,000,000	34,330	34,060	-	35,610	-
5,000,000	40,260	40,060	-	41,590	-
6,000,000	46,060	45,950	-	47,430	-
7,000,000	51,780	51,770	-	53,170	-
8,000,000	57,450	-	-	58,850	-
9,000,000	63,070	-	-	64,480	-
10,000,000	68,670	-	-	70,060	-

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

Bachelor's Scientists and Engineers in 1993:
Approximate Standard Errors for Specified Demographic Groups

Estimated Number	Total Population	Male	Female	White	Nonwhite
100	110	120	160	120	90
200	150	170	230	170	130
500	240	270	360	260	200
750	300	330	450	320	250
1,000	340	380	520	370	290
2,000	490	540	730	530	410
3,000	600	660	890	650	500
4,000	690	760	1,030	750	580
5,000	770	850	1,150	840	650
10,000	1,090	1,200	1,630	1,180	930
25,000	1,740	1,910	2,580	1,880	1,520
50,000	2,480	2,720	3,650	2,680	2,250
100,000	3,570	3,900	5,160	3,860	3,470
250,000	5,930	6,430	8,170	6,400	6,630
500,000	9,020	9,690	11,580	9,700	11,590
750,000	11,770	12,540	14,220	12,620	-
1,000,000	14,380	15,230	16,450	15,390	-
2,000,000	24,270	25,310	23,460	25,860	-
3,000,000	33,870	35,010	-	35,990	-
4,000,000	43,370	44,590	-	46,010	-
5,000,000	52,820	-	-	55,980	-
6,000,000	62,250	-	-	65,920	-

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

Master's Scientists and Engineers in 1993:
Approximate Standard Errors for Specified Demographic Groups

Estimated Number	Total Population	Male	Female	White	Nonwhite
100	140	130	160	150	100
200	200	180	220	210	140
500	320	290	360	340	230
750	400	360	440	410	280
1,000	460	410	500	480	320
2,000	650	580	710	680	450
3,000	790	710	870	830	560
4,000	920	820	1,000	960	640
5,000	1,020	920	1,120	1,070	720
10,000	1,450	1,300	1,590	1,510	1,010
25,000	2,290	2,060	2,500	2,390	1,590
50,000	3,220	2,900	3,520	3,360	2,230
100,000	4,540	4,070	4,940	4,730	3,070
250,000	7,060	6,320	7,600	7,360	-
500,000	9,720	8,630	10,250	10,120	-
750,000	11,570	10,190	11,900	12,020	-
1,000,000	12,960	11,310	12,950	13,440	-
2,000,000	15,860	-	-	16,290	-
3,000,000	15,840	-	-	-	-

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

Doctorate Scientists and Engineers in 1993:
Approximate Standard Errors for Specified Demographic Groups

Estimated Number	Total Population	Male	Female	White	Nonwhite
100	80	80	80	80	90
200	110	110	110	110	130
500	180	170	170	180	210
750	220	210	210	220	250
1,000	260	250	240	250	290
2,000	360	350	350	360	410
3,000	440	420	420	440	500
4,000	510	490	490	510	580
5,000	570	550	540	570	650
10,000	810	770	770	800	900
25,000	1,280	1,220	1,200	1,260	1,350
50,000	1,800	1,710	1,670	1,780	-
100,000	2,520	2,390	2,290	2,480	-
250,000	3,840	3,640	-	3,750	-
500,000	5,110	4,800	-	4,890	-
750,000	5,830	-	-	-	-

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

APPENDIX B

PARAMETER TABLES:
1993 SESTAT

Scientists and Engineers in 1993 (Total Population):
$ß_0$ and $ß_1$ Parameters for Specified Demographic Groups

Field	Parameter	Total	Male	Female	White	Nonwhite
Total Scientists and Engineers						
Total, All Fields	$ß_0$	0.000029	0.000031	-0.000008	0.000029	0.000185
	$ß_1$	176.694901	166.310995	270.328363	200.999510	83.186636
Highest Degree						
Computer/Math sciences	$ß_0$	0.000290	0.000056	0.000759	0.000212	0.001474
	$ß_1$	89.803999	130.807477	145.064552	134.710118	108.843207
Life sciences	$ß_0$	0.000075	0.000113	-0.000019	0.000064	0.000326
	$ß_1$	251.348834	294.314327	196.031161	274.209782	77.239592
Physical sciences	$ß_0$	0.000031	0.000035	0.000319	0.000047	0.000040
	$ß_1$	170.492431	157.997872	176.999935	177.932091	101.409530
Social sciences	$ß_0$	0.000037	0.000134	-0.000144	0.000042	0.000187
	$ß_1$	303.607445	265.723071	394.628593	347.429493	82.273485
Engineering	$ß_0$	0.000067	0.000067	-0.000176	0.000068	0.000545
	$ß_1$	128.775350	114.981848	147.681396	133.658866	84.711138
Non-S&E;	$ß_0$	-0.000035	0.000000	-0.000034	-0.000046	-0.000025
	$ß_1$	256.851348	251.531029	242.245989	287.084207	147.341024
Occupation						
Computer/Math sciences	$ß_0$	-0.000062	-0.000056	-0.000034	-0.000101	0.001076
	$ß_1$	221.850649	204.842453	178.918654	238.246130	83.862026
Life sciences	$ß_0$	-0.000012	0.000009	-0.000163	0.000064	-0.000030
	$ß_1$	138.753466	149.307476	154.145336	141.492622	94.859093
Physical sciences	$ß_0$	-0.000004	0.000050	-0.000356	-0.000006	0.002676
	$ß_1$	130.530544	112.340964	168.710246	130.327850	60.756323
Social sciences	$ß_0$	0.000341	0.000374	0.000348	0.000489	0.002770
	$ß_1$	109.844071	108.543722	153.284390	112.450179	54.155536
Engineering	$ß_0$	0.000027	0.000031	-0.000354	0.000005	0.000069
	$ß_1$	96.102065	101.138205	113.822217	110.789200	77.599636
Non-S&E	$ß_0$	0.000005	-0.000013	-0.000074	-0.000005	0.000193
	$ß_1$	287.167300	291.879264	331.938943	340.295571	86.684411

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

Bachelor's Scientists and Engineers in 1993:
$ß_0$ and $ß_1$ Parameters for Specified Demographic Groups

Field	Parameter	Total	Male	Female	White	Nonwhite
Bachelor's Scientists and Engineers						
Total, All Fields	$ß_0$ $ß_1$	0.000088 118.863944	0.000088 143.422537	0.000005 266.097384	0.000097 139.302457	0.000371 83.053031
Highest Degree						
Computer/ Math sciences	$ß_0$ $ß_1$	0.000422 70.132892	0.000106 127.745725	0.000738 152.674555	0.000291 122.242507	0.002338 101.439493
Life sciences	$ß_0$ $ß_1$	0.000038 256.311044	0.000169 308.566271	-0.000169 194.078346	0.000039 276.025364	0.000530 71.415188
Physical sciences	$ß_0$ $ß_1$	0.000173 194.580077	0.000218 208.152873	0.000477 194.841922	0.000179 222.709686	-0.000012 95.934589
Social sciences	$ß_0$ $ß_1$	0.000087 331.370202	0.000116 339.146125	-0.000103 389.321456	0.000094 379.885266	0.000185 96.437934
Engineering	$ß_0$ $ß_1$	0.000060 181.287222	0.000052 171.156982	-0.000368 166.408076	0.000060 182.283383	0.000963 84.009555
Non-S&E	$ß_0$ $ß_1$	0.000356 179.644002	0.000296 190.185326	-0.000076 247.512112	0.000362 190.880062	-0.000247 138.615251
Occupation						
Computer/ Math sciences	$ß_0$ $ß_1$	0.000039 176.920901	0.000086 176.359400	-0.000366 168.356534	-0.000018 196.706268	0.001745 82.978399
Life sciences	$ß_0$ $ß_1$	0.000143 190.862376	-0.000804 222.784891	-0.000202 192.275687	0.000201 188.614257	0.000903 122.430378
Physical sciences	$ß_0$ $ß_1$	0.000450 121.520966	0.000477 124.818113	-0.001155 154.561669	0.000430 131.479018	0.004938 76.419725
Social sciences	$ß_0$ $ß_1$	0.001093 197.354385	-0.001687 247.241934	0.003517 193.661795	0.001197 234.772686	0.000764 105.807220
Engineering	$ß_0$ $ß_1$	0.000013 115.175340	-0.000004 119.742575	-0.000468 128.914381	-0.000025 126.978454	0.000667 79.376096
Non-S&E	$ß_0$ $ß_1$	0.000084 260.556567	0.000034 290.085479	-0.000018 315.021724	0.000085 316.803300	0.000408 81.397975

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.
SOURCE: 1993 SESTAT Database

Master's Scientists and Engineers in 1993:
β_0 and β_1 Parameters for Specified Demographic Groups

Field	Parameter	Total	Male	Female	White	Nonwhite
Master's Scientists and Engineers						
Total, All Fields	β_0 β_1	-0.000042 210.065366	-0.000042 170.115170	-0.000085 252.392454	-0.000048 228.742977	-0.000090 103.535428
Highest Degree						
Computer/ Math sciences	β_0 β_1	0.000189 188.109917	0.000251 196.743588	-0.000556 209.280519	0.000184 192.467607	-0.000969 90.393643
Life sciences	β_0 β_1	-0.000471 211.007321	-0.001215 238.293608	0.000292 207.138575	-0.000588 219.725866	-0.001509 95.502471
Physical sciences	β_0 β_1	-0.000739 192.222804	-0.000866 201.636944	-0.001955 175.236118	-0.000671 193.434602	-0.001802 110.590972
Social sciences	β_0 β_1	-0.000045 299.153598	0.000165 257.608076	0.000042 261.308182	0.000066 310.119865	-0.000116 80.373159
Engineering	β_0 β_1	0.000046 123.149358	0.000045 134.006323	0.001652 117.056671	-0.000003 129.906814	0.001355 68.378293
Non-S&E	β_0 β_1	-0.000046 268.597292	0.000147 197.098078	-0.000142 272.386679	-0.000062 288.327605	-0.000097 97.974929
Occupation						
Computer/ Math sciences	β_0 β_1	0.000003 183.507581	0.000209 172.440062	0.001152 170.762337	-0.000053 190.912531	0.001313 72.022245
Life sciences	β_0 β_1	-0.000098 140.384105	0.000287 157.766873	-0.000053 172.980525	-0.000199 152.413526	-0.001694 80.585504
Physical sciences	β_0 β_1	0.000270 169.607948	0.000325 142.656068	-0.002395 196.480275	0.000374 159.905450	-0.012569 113.477784
Social sciences	β_0 β_1	0.000113 212.352864	0.000362 199.572394	0.000256 202.907210	0.000346 220.334536	-0.000914 97.532332
Engineering	β_0 β_1	-0.000111 127.044886	-0.000130 141.784548	-0.001691 148.969559	-0.000135 133.622329	-0.001218 71.585296
Non-S&E	β_0 β_1	-0.000056 261.662917	-0.000176 269.768092	-0.000165 285.510025	-0.000050 279.121849	0.000049 102.423660

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

Doctorate Scientists and Engineers in 1993:
$ß_0$ and $ß_1$ Parameters for Specified Demographic Groups

Field	Parameter	Total	Male	Female	White	Nonwhite
Doctorate Scientists and Engineers						
Total, All Fields	$ß_0$ $ß_1$	-0.000028 66.044221	-0.000028 60.131650	-0.000072 59.687153	-0.000034 64.814662	-0.000529 86.208470
Highest Degree						
Computer/ Math sciences	$ß_0$ $ß_1$	-0.000123 23.030148	-0.000212 28.862306	0.000465 24.606270	-0.000170 24.495462	0.005226 10.408216
Life sciences	$ß_0$ $ß_1$	0.000081 51.106271	0.000108 45.234084	-0.000410 53.645852	0.000065 48.051474	-0.002341 67.488260
Physical sciences	$ß_0$ $ß_1$	-0.000087 34.171400	-0.000138 37.689623	-0.000494 19.184464	-0.000079 33.340388	-0.002639 27.069883
Social sciences	$ß_0$ $ß_1$	0.000002 35.663555	-0.000041 37.240698	-0.000283 39.817315	-0.000072 39.619747	0.002247 19.727061
Engineering	$ß_0$ $ß_1$	-0.000222 40.324285	-0.000210 39.622957	-0.001815 31.031940	-0.000288 42.331954	0.005926 34.301077
Non-S&E	$ß_0$ $ß_1$	-0.000207 147.319548	-0.000102 142.760889	-0.000021 111.781146	-0.000303 146.752522	-0.001812 127.309538
Occupation						
Computer/ Math sciences	$ß_0$ $ß_1$	0.000329 62.088195	0.000263 52.351634	-0.000114 87.728705	0.000188 60.430333	-0.002467 69.392321
Life sciences	$ß_0$ $ß_1$	-0.000189 51.473330	-0.000242 45.578983	-0.000772 49.055488	-0.000179 52.508480	0.000739 32.188706
Physical sciences	$ß_0$ $ß_1$	-0.000057 39.013694	-0.000084 38.399035	-0.001622 54.765092	-0.000012 36.745355	-0.003311 34.691400
Social sciences	$ß_0$ $ß_1$	-0.000090 44.939758	0.000009 37.320441	-0.000278 56.385559	-0.000085 49.488283	0.000350 38.588539
Engineering	$ß_0$ $ß_1$	0.000109 33.462653	0.000201 34.340634	-0.000768 18.864315	0.000117 33.644936	-0.003695 43.223131
Non-S&E	$ß_0$ $ß_1$	0.000079 101.369441	-0.000039 96.602795	0.001099 79.175775	0.000113 96.754722	-0.002584 113.496119

NOTE: White category excludes persons of Hispanic origin. Hispanics are included in the Nonwhite category.

SOURCE: 1993 SESTAT Database

APPENDIX C

PARAMETER TABLES:
1993 NSCG

1993 National Survey of College Graduates:
$ß_0$ and $ß_1$ Parameters for Selected Groups

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Total S&E	-0.000007	222
Female	-0.000015	218
Hispanic	-0.000017	93
Black	-0.000016	112
Asian	-0.000003	96
Ph.D. Recipients	-0.000004	127
Biological Scientists	-0.000008	263
Female	-0.000015	229
Hispanic	-0.000018	98
Black	-0.000017	123
Asian	-0.000003	105
Ph.D. Recipients	-0.000005	169
Physical Scientists	-0.000006	208
Female	-0.000016	207
Hispanic	-0.000017	97
Black	-0.000015	110
Asian	-0.000003	106
Ph.D. Recipients	-0.000004	123
Math/Computer	-0.000007	228
Female	-0.000014	208
Hispanic	-0.000016	91
Black	-0.000016	113
Asian	-0.000003	99
Ph.D. Recipients	-0.000004	129

1993 National Survey of College Graduates:
$ß_0$ and $ß_1$ Parameters for Selected Groups
(continued)

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Engineers	-0.000006	191
Female	-0.000010	153
Hispanic	-0.000016	92
Black	-0.000014	101
Asian	-0.000003	92
Ph.D. Recipients	-0.000004	134
Social Scientists	-0.000009	296
Female	-0.000017	260
Hispanic	-0.000019	107
Black	-0.000019	133
Asian	-0.000004	119
Ph.D. Recipients	-0.000004	142

APPENDIX D

PARAMETER TABLES:
1993 SDR

1993 Survey of Doctorate Recipients:
$ß_0$ and $ß_1$ Parameters for Total (All Doctorates)

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Total	-0.000026	13.48
Scientists	-0.000030	13.12
Computer and Math Subtotal	-0.000030	14.77
Computer and Information Science	-0.001287	7.37
Mathematical Science	-0.000592	16.557
Life and Related Science	-0.000068	10.50
Agricultural and Food Science	-0.000965	18.91
Biological and Health Science	-0.000082	9.73
Environmental Science	-0.000825	13.27
Physical and Related Subtotal	-0.000105	11.96
Chemistry (except biochem)	-0.000233	14.16
Geology and Oceanography	-0.000117	5.20
Physics and Astronomy	-0.000260	9.74
Other Physical Sciences	-0.010099	15.67
Social and Related Subtotal	-0.000106	17.13
Economics	0.000155	5.56
Political Science	0.000203	21.69
Psychology	-0.000248	19.43
Sociology and Anthropology	-0.000434	10.81
Other Social Sciences	-0.001757	34.23
Engineering Subtotal	-0.000168	15.45
Aeronautical/Astronautical Engineering	0.003337	21.50
Chemical Engineering	-0.001168	16.10
Civil Engineering	0.023625	6.66
Electrical, Computer Engineering	-0.000705	15.91
Industrial Engineering	-0.005265	18.26
Mechanical Engineering	-0.000772	12.99
Other Engineering	0.001212	12.42

1993 Survey of Doctorate Recipients:
β_0 and β_1 Parameters for Females

Domain	GVF Coefficients	
	Intercept (β_0)	Slope (β_1)
Total	-0.000087	9.45
Scientists	-0.000092	9.62
Computer and Math Subtotal	-0.000092	8.48
Computer and Information Science	-0.010720	9.04
Mathematical Science	-0.002761	8.88
Life and Related Science	-0.000210	7.77
Agricultural and Food Science	-0.008273	18.565
Biological and Health Science	-0.000224	7.68
Environmental Science	0.066940	8.57
Physical and Related Subtotal	-0.000733	10.05
Chemistry (except biochemistry)	-0.001095	10.33
Geology and Oceanography	-0.003137	8.79
Physics and Astronomy	-0.002154	7.29
Other Physical Sciences	-0.024718	11.82
Social and Related Subtotal	-0.000211	10.79
Economics	-0.002650	9.17
Political Science	-0.001252	7.89
Psychology	-0.000398	12.66
Sociology and Anthropology	-0.000914	7.76
Other Social Sciences	-0.003642	19.78
Engineering Subtotal	-0.001140	5.90
Aeronautical/Astronautical Engineering	-0.001140	5.90
Chemical Engineering	0.009568	6.54
Civil Engineering	0.100001	-1.50
Electrical, Computer Engineering	-0.004224	6.70
Industrial Engineering	0.006136	9.38
Mechanical Engineering	-0.006733	6.43
Other Engineering	-0.006641	10.60

1993 Survey of Doctorate Recipients:
β_0 and β_1 Parameters for Whites

Domain	GVF Coefficients	
	Intercept (β_0)	Slope (β_1)
Total	-0.000031	14.76
Scientists	-0.000035	14.28
Computer and Math Subtotal	-0.000035	15.54
Computer and Information Science	-0.000928	55.535
Mathematical Science	-0.000711	17.29
Life and Related Science	-0.000070	10.25
Agricultural and Food Science	-0.000881	16.94
Biological and Health Science	-0.000096	10.21
Environmental Science	0.000183	13.15
Physical and Related Subtotal	-0.000116	11.97
Chemistry (except biochem)	-0.000274	14.45
Geology and Oceanography	0.000055	5.20
Physics and Astronomy	-0.000224	7.73
Other Physical Sciences	-0.013421	16.79
Social and Related Subtotal	-0.000126	19.84
Economics	0.000256	6.10
Political Science	-0.000219	24.35
Psychology	-0.000305	23.16
Sociology and Anthropology	-0.000337	10.90
Other Social Sciences	-0.002411	36.68
Engineering Subtotal	-0.000247	17.95
Aeronautical/Astronautical Engineering	0.006919	15.85
Chemical Engineering	-0.001098	14.49
Civil Engineering	0.035403	9.57
Electrical, Computer Engineering	-0.001047	16.03
Industrial Engineering	-0.010131	22.54
Mechanical Engineering	0.001164	14.26
Other Engineering	0.001562	14.81

1993 Survey of Doctorate Recipients:
ß$_0$ and ß$_1$ Parameters for Asians

Domain	GVF Coefficients	
	Intercept (ß$_0$)	Slope (ß$_1$)
Total	0.000023	11.48
Scientists	0.000109	10.58
Computer and Math Subtotal	0.000109	15.81
Computer and Information Science	-0.002154	11.79
Mathematical Science	-0.000249	17.88
Life and Related Science	-0.000333	8.36
Agricultural and Food Science	-0.006030	14.15
Biological and Health Science	-0.000139	6.64
Environmental Science	-0.039187	8.32
Physical and Related Subtotal	0.000509	12.36
Chemistry (except biochemisty)	0.000399	18.15
Geology and Oceanography	0.021301	10.02
Physics and Astronomy	0.002000	7.58
Other Physical Sciences	0.000509	12.36
Social and Related Subtotal	0.002951	8.76
Economics	-0.003223	13.01
Political Science	0.021517	7.16
Psychology	-0.006483	12.05
Sociology and Anthropology	-0.000588	9.86
Other Social Sciences	0.033866	5.54
Engineering Subtotal	-0.000131	12.97
Engineering	0.014224	21.21
Chemical Engineering	0.007662	10.42
Civil Engineering	0.011251	7.88
Electrical, Computer Engineering	-0.001236	11.27
Industrial Engineering	0.029049	11.79
Mechanical Engineering	-0.005297	18.19
Other Engineering	0.005771	11.63

1993 Survey of Doctorate Recipients:
$ß_0$ and $ß_1$ Parameters for Blacks

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Total	-0.000584	11.32
Scientists	-0.000611	10.76
Computer and Math Subtotal	-0.000611	3.32
Computer and Information Science	0.026504	3.32
Mathematical Science	0.031535	2.90
Life and Related Science	-0.001365	6.74
Agricultural and Food Science	0.013928	11.10
Biological and Health Science	-0.000888	5.13
Environmental Science	-0.001365	6.73
Physical and Related Subtotal	-0.002337	11.33
Chemistry (except biochem)	0.002105	9.59
Geology and Oceanography	-0.002337	11.33
Physics and Astronomy	-0.036344	13.51
Other Physical Sciences	-0.002337	11.33
Social and Related Subtotal	-0.001747	13.99
Economics	0.007075	10.51
Political Science	-0.013924	12.91
Psychology	-0.005128	16.06
Sociology and Anthropology	-0.004615	7.32
Other Social Sciences	0.011698	8.50
Engineering Subtotal	-0.000369	12.93
Aeronautical/Astronautical Engineering	-0.000369	12.63
Chemical Engineering	-0.036776	7.87
Civil Engineering	0.415796	1.22
Electrical, Computer Engineering	-0.033401	17.58
Industrial Engineering	-0.000369	12.63
Mechanical Engineering	0.151036	1.87
Other Engineering	0.023358	8.65

1993 Survey of Doctorate Recipients:
$ß_0$ and $ß_1$ Parameters for Native Americans

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Total	-0.001682	19.59
Scientists	-0.002595	20.84
Computer and Math Subtotal	-0.002595	20.84
Computer and Information Science	-0.0025895	20.84
Mathematical Science	-0.002595	20.84
Life and Related Science	0.023396	6.64
Agricultural and Food Science	0.023396	6.64
Biological and Health Science	0.033779	5.29
Environmental Science	0.023396	6.64
Physical and Related Subtotal	0.075976	1.93
Chemistry (except biochemistry)	0.117198	1.42
Geology and Oceanography	0.075976	1.93
Physics and Astronomy	0.075976	1.93
Other Physical Sciences	0.075976	1.93
Social and Related Subtotal	-0.010340	28.45
Economics	-0.010340	28.45
Political Science	-0.010340	28.45
Psychology	-0.043994	21.89
Sociology and Anthropology	0.237751	2.25
Other Social Sciences	-0.010340	28.45
Engineering Subtotal	0.053005	5.86
Aeronautical/Astronautical Engineering	0.053005	5.86
Chemical Engineering	0.053005	5.86
Civil Engineering	0.053005	5.86
Electrical, Computer Engineering	0.053005	5.86
Industrial Engineering	0.053005	5.86
Mechanical Engineering	0.053005	5.86
Other Engineering	0.053005	5.86

1993 Survey of Doctorate Recipients:
$ß_0$ and $ß_1$ Parameters for Hispanics

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Total	-0.000321	12.82
Scientists	-0.000392	12.70
Computer and Math Subtotal	-0.000392	17.52
Computer and Information Science	0.066164	-0.10
Mathematical Science	-0.019405	19.26
Life and Related Science	-0.003245	10.92
Agricultural and Food Science	0.058986	3.23
Biological and Health Science	-0.003999	10.28
Environmental Science	-0.003245	10.92
Physical and Related Subtotal	0.044566	-0.30
Chemistry (except biochemistry)	0.060822	-0.42
Geology and Oceanography	0.001711	7.80
Physics and Astronomy	0.021624	12.75
Other Physical Sciences	0.044566	-0.30
Social and Related Subtotal	0.002030	8.03
Economics	0.018067	17.31
Political Science	0.018067	9.45
Psychology	0.006116	9.09
Sociology and Anthropology	0.008481	5.13
Other Social Sciences	-0.021878	15.65
Engineering Subtotal	-0.003625	15.56
Aeronautical/Astronautical Engineering	-0.003625	15.56
Chemical Engineering	-0.054127	21.86
Civil Engineering	-0.054127	5.69
Electrical, Computer Engineering	0.027241	12.78
Industrial Engineering	-0.003625	15.56
Mechanical Engineering	0.064669	5.44
Other Engineering	-0.032770	14.30

1993 Survey of Doctorate Recipients:
$ß_0$ and $ß_1$ Parameters for 1991-1992 Cohort

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
Total	-0.000243	11.24
Scientists	-0.000310	11.57
Computer and Math Subtotal	-0.000310	12.79
Computer and Information Science	-0.006365	9.98
Mathematical Science	-0.008632	15.67
Life and Related Science	-0.000733	9.23
Agricultural and Food Science	-0.008613	15.50
Biological and Health Science	-0.000792	8.62
Environmental Science	0.039506	4.13
Physical and Related Subtotal	-0.001722	15.42
Chemistry (except biochemistry)	-0.004239	19.87
Geology and Oceanography	-0.001840	7.66
Physics and Astronomy	-0.004709	12.68
Other Physical Sciences	-0.001722	15.42
Social and Related Subtotal	-0.000963	12.01
Economics	0.017476	13.89
Political Science	0.017476	18.17
Psychology	-0.001529	10.30
Sociology and Anthropology	-0.010177	16.81
Other Social Sciences	0.007272	10.75
Engineering Subtotal	-0.001113	10.12
Aeronautical/Astronautical Engineering	-0.045483	18.76
Chemical Engineering	0.009251	12.65
Civil Engineering	0.009251	11..67
Electrical, Computer Engineering	-0.003072	10.50
Industrial Engineering	0.040079	13.39
Mechanical Engineering	-0.001254	10.70
Other Engineering	-0.000695	7.45

1993 Survey of Doctorate Recipients:
$ß_0$ and $ß_1$ Parameters for Foreign-Born

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_0$)
Total	-0.000268	13.66
Scientists	-0.000317	12.77
Computer and Math Subtotal	-0.003802	16.37
Computer and Information Science	-0.009285	15.55
Mathematical Science	-0.005853	16.24
Life and Related Science	-0.001060	10.84
Agricultural and Food Science	-0.007744	11.39
Biological and Health Science	-0.001167	9.76
Environmental Science	-0.001060	10.84
Physical and Related Subtotal	-0.001565	15.82
Chemistry (except biochemistry)	-0.003472	18.11
Geology and Oceanography	-0.011668	9.59
Physics and Astronomy	-0.004833	18.45
Other Physical Sciences	-0.001565	15.83
Social and Related Subtotal	-0.000310	11.38
Economics	-0.019238	14.30
Political Science	-0.019238	17.48
Psychology	-0.007936	12.62
Sociology and Anthropology	-0.009831	10.09
Other Social Sciences	-0.002515	12.66
Engineering Subtotal	-0.001088	15.16
Aeronautical/Astronautical Engineering	0.019269	28.87
Chemical Engineering	-0.002974	5.85
Civil Engineering	-0.002974	18.95
Electrical, Computer Engineering	-0.002218	10.69
Industrial Engineering	0.001834	4.45
Mechanical Engineering	-0.003204	6.13
Other Engineering	-0.002171	21.82

APPENDIX E

PARAMETER TABLES:
1993 NSRCG

1993 National Survey of Recent College Graduates:
$ß_0$ and $ß_1$ Parameters for 1991 Bachelor's Graduates

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
All Graduates	0.000116	132.515
Male	0.001079	94.871
Female	-0.000010	178.568
Science Majors	0.000411	178.903
Engineering Majors	0.000818	80.969
Scientists	-0.000872	131.591
Engineers	-0.000045	82.807
Other Occupations	0.000451	195.981
White	0.000718	120.830
Black	0.032007	82.327
Hispanic	0.006942	141.348
Asian	0.001170	124.246

1993 National Survey of Recent College Graduates:
$ß_0$ and $ß_1$ Parameters for 1991 Master's Graduates

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
All Graduates	0.013208	11.064
Male	0.000354	38.915
Female	0.001214	37.876
Science Majors	0.002822	26.086
Engineering Majors	-0.001952	41.629
Scientists	0.003328	27.255
Engineers	-0.000006	27.478
Other Occupations	-0.000500	41.688
White	0.001287	37.517
Black	0.019705	43.892
Hispanic	0.000230	26.526
Asian	0.002644	33.742

1993 National Survey of Recent College Graduates:
β_0 and β_1 Parameters for 1992 Bachelor's Graduates

Domain	GVF Coefficients	
	Intercept (β_0)	Slope (β_1)
All Graduates	0.006530	68.747
Male	0.000731	83.675
Female	0.000494	166.358
Science Majors	0.000566	150.333
Engineering Majors	-0.000971	72.909
Scientists	-0.000565	171.386
Engineers	-0.001254	73.203
Other Occupations	0.000117	188.939
White	0.000746	88.749
Black	0.026867	128.155
Hispanic	0.006141	121.118
Asian	0.004119	106.541

1993 National Survey of Recent College Graduates:
$ß_0$ and $ß_1$ Parameters for 1992 Master's Graduates

Domain	GVF Coefficients	
	Intercept ($ß_0$)	Slope ($ß_1$)
All Graduates	0.014906	9.977
Male	-0.000229	42.863
Female	-0.000665	42.195
Science Majors	-0.001252	68.013
Engineering Majors	-0.000408	28.939
Scientists	-0.000062	39.428
Engineers	0.000029	27.230
Other Occupations	0.000463	41.454
White	-0.000806	55.473
Black	0.015080	32.573
Hispanic	-0.001975	25.648
Asian	0.001693	29.712