MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0379208 2

ornl

ORNL/TM-12379

OAK RIDGE
NATIONAL
LABORATORY

MARTIN MARIETTA

Prediction and Standard Error Estimation for a Finite Universe Total When a Stratum Is Not Sampled

Tommy Wright

OAK RIDGE NATIONAL LABORATORY
CENTRAL RESEARCH LIBRARY
CIRCULATION SECTION
4500N ROOM 175
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this report, send in name with report and the library will arrange a loan.

MANAGED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/TM-12379

405

Prediction and Standard Error Estimation for a Finite Universe Total When a Stratum Is Not Sampled

January 1994

Tommy Wright
Mathematical Sciences Section
Engineering Physics and Mathematics Division
Oak Ridge National Laboratory
P. O. Box 2008 Building 6012
Oak Ridge, Tennessee 37831-6367

Research sponsored by the Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-84OR21400

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0379208 2

Table of Contents

Abstract vii

1. Introduction 1

    1.1 General Background for the 1990 NTACS 1

    1.2 General Statement of Problem 3

2. Setting Using Age Post-Stratification 4

    2.1 Introduction 4

    2.2 Key Assumption 11

    2.3 Prediction and Sampling from Finite Universes 11

    2.4 Prediction Estimation Model for $\bar{V}_{t'_1}$ Using Post-stratification 12

    2.5 Application to the Post-stratified 1990 NTACS Sample Data 14

3. Setting Using Age Stratification Before Sampling At Times t and t' 17

4. Concluding Remarks 29

Acknowledgements 31

References 32

List of Figures

Figure 1. 1987 TIUS and 1990 NTACS Universes 3

Figure 2. Post-Stratified Universe Setting at Time t 7

Figure 3. Post-Stratified Universe Settings at Times t and $t'(=t+1)$ 8

Figure 4. Stratified Universe Setting at Time t 19

Figure 5. Stratified Universe Setting at Times t and t' 20

List of Exhibits

Exhibit 1. Age Post-stratum Sample Observations and Means at Time t 5

Exhibit 2. Age Post-stratum Sample Observations and Means at Time t' 9

Exhibit 3. Post-stratum Sample Means at Times t and $t'(=t+1)$ 11

Exhibit 4. Age Stratum Sample Means at Time t 18

Exhibit 5. Age Stratum Sample Means at Time t' 20

Exhibit 6. Stratum Sample Means at Times t and t' 22

PREDICTION AND STANDARD ERROR ESTIMATION FOR A FINITE UNIVERSE TOTAL WHEN A STRATUM IS NOT SAMPLED

Tommy Wright
Mathematical Sciences Section
Oak Ridge National Laboratory
P. O. Box 2008 Bldg. 6012
Oak Ridge, Tennessee 37831-6367

ABSTRACT

In the context of a universe of trucks operating in the United States in 1990, this paper presents statistical methodology for estimating a finite universe total on a second occasion when a part of the universe is sampled and the remainder of the universe is not sampled. Prediction is used to compensate for the lack of data from the unsampled portion of the universe. The sample is assumed to be a subsample of an earlier sample where stratification is used on both occasions before sample selection. Accounting for births and deaths in the universe between the two points in time, the detailed sampling plan, estimator, standard error, and optimal sample allocation, are presented with a focus on the second occasion. If prior auxiliary information is available, the methodology is also applicable to a first occasion.

KEY WORDS: Domain Estimation; Post-stratification; Prediction; Stratification; Standard error estimation; Subsampling.

1. INTRODUCTION

1.1 General Background for the 1990 NTACS

The *Truck Inventory and Use Survey (TIUS)* is the oldest national, vehicle-based survey of freight transportation in the United States. The TIUS is conducted by the U.S. Bureau of the Census every five years and collects extensive information on a sample of trucks (including vans and minivans), their typical characteristics, and their use over a year's period.

> "TIUS information is of considerable value to State and Federal transportation agencies in the planning of highway cost allocations, improvements in road conditions, energy consumption, and emergency preparedness. The private sector also needs these data to plan for future vehicle designs and improvements, market studies, and more efficient vehicle usage."
>
> *1987 TIUS Report*
> U.S. Bureau of the Census

The TIUS sample is drawn from vehicle registration files and represents all vehicles except buses, automobiles, mobile homes, motorcycles, and vehicles owned by governments (Federal, State, and local). The 1987 TIUS stratified random sample includes information on approximately 105,000 vehicles which were selected from a sampling frame of over 45,000,000 vehicles. A key parameter for estimation is the total annual *vehicle miles traveled (VMT)* for these 45,000,000+ vehicles.

While the TIUS provides critical information on the composition and typical use of the nation's trucking fleet, it does not provide a detailed picture of the temporal and geographic variation in truck use. This variation has major implications for multimodal transportation policies involving highway investment requirements, equitable tax policy, effective safety programs, and similar issues. It was decided to sponsor a follow-on to the Census Bureau's 1987 TIUS in order to capture the temporal and geographic variation in truck use and to measure other detailed attributes of trucking that were beyond the scope of the TIUS. The follow-on became known as the *1990 Nationwide Truck Activity and Commodity Survey (NTACS)* and was conducted by the U.S. Bureau of the Census over a 12-month period that ended in 1990. The 1990 NTACS sample is a stratified subsample of 44,002 vehicles selected from the approximately 105,000 vehicles which were 1987 TIUS sample respondents. These 44,002 vehicles make it possible to provide estimates for what has been viewed as the *1990 NTACS universe*. For a complete discussion on data limitations with the 1990 NTACS sample, see the *1990 NTACS Public Use File Technical Documentation* (1992).

The *1990 NTACS Selected Tabulations (1993)* describes the 1990 NTACS universe as "...'All trucks operating during the NTACS period (October 29, 1989 - October 27, 1990) and registered in one of the 50 states or the District of Columbia on July 1, 1987, and operating in 1987 as estimated by the 1987 TIUS.' It is important to note that the 1990 NTACS universe does not include trucks operating in the United States during the NTACS period which are less than four years old, i.e., trucks registered after July 1, 1987. The estimated 1990 NTACS universe size (of four year and older trucks) is 41,768,000 trucks.

A clearer view of the 1990 NTACS universe can be obtained by referring to Figure 1 (Hu, Wright, and Miaou, 1990) where:

— the 1987 and 1990 U.S. truck populations are represented by two overlapping figures:

— the intersection, represented by "A", represents the group of trucks that were in operation in 1987 and that were still in operation in 1990;

— the region that is indicated by "B" represents the group of trucks that were operating in 1987 but which were no longer in operation in 1990, i.e. this group of trucks was scrapped between 1987 and 1990; and

— the region indicated by "C" represents the group of trucks that began operating after 1987 (more precisely after July 1, 1987).

The 1987 total U.S. truck population consists of trucks in groups A and B; and the 1990 total U.S. truck population consists of trucks in groups A and C. The trucks in group A represent the 1990 NTACS universe."

Use of the sampling weights and reported annual VMT for the 1990 NTACS sample of 44,002 trucks provides estimates for group A, the 1990 NTACS universe, but not for the entire 1990 U.S. truck universe. Lacking any sample data from those trucks less than four years old in 1990 (trucks in group C), this paper discusses a simple application of ordinary least squares to predict annual VMT for those trucks in group C and hence leading to an estimate of annual VMT for the entire 1990 U.S. truck universe. Because this problem occurs in other settings, a general statement of the problem is given next.

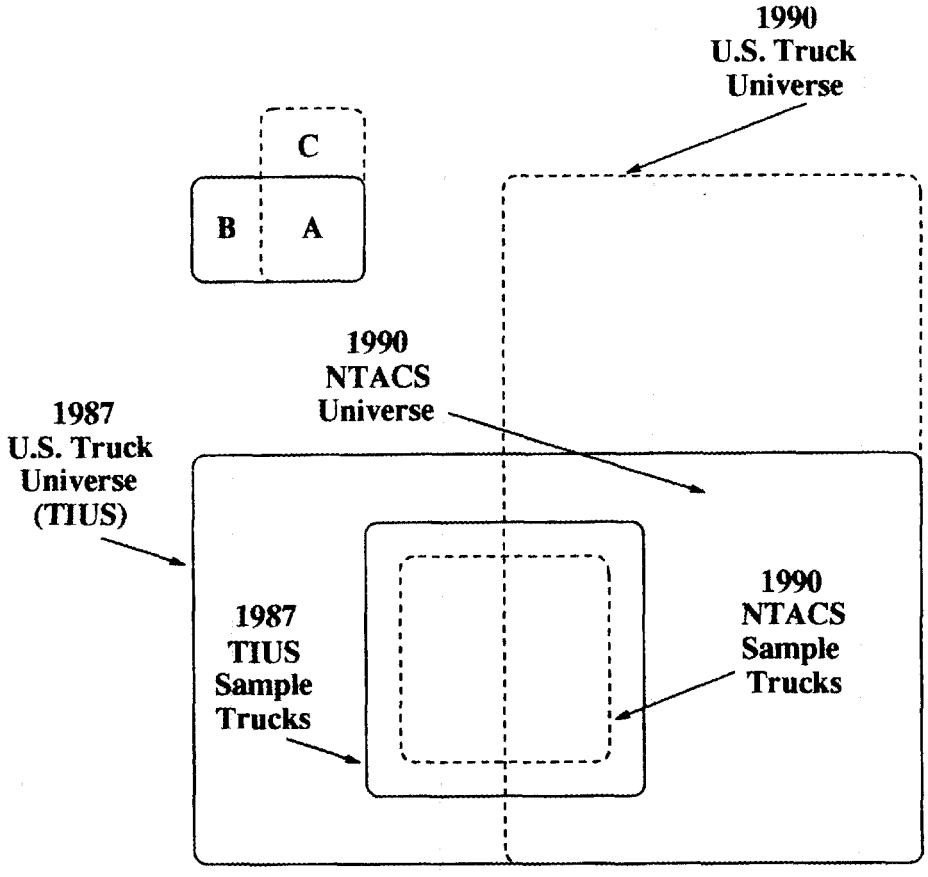

Figure 1. 1987 TIUS and 1990 NTACS Universes

1.2 General Statement of Problem

In periodic (e.g. monthly, quarterly, annual) sample surveys where the same sample units are asked to report data from one reporting period to another, the sample's relation with the target universe becomes less clear with the passing of time because the sample has remained constant while the universe may have changed. This is especially true when the universe experiences births and deaths between reporting periods. The extent of unit deaths in the universe can be approximated by the reported deaths of sample units for a given reporting period, and estimates can be adjusted accordingly. However, for unit births which are not added to the sampling frame nor sampled, the problem is more severe and almost certainly results in negatively biased estimates of key parameters such as totals if ignored. On some occasions, this bias can be removed by supplementing the original sample with a sample of units selected from those units born since the previous reporting period (Wright and Tsao, 1983).

Because resources are limited, supplemental samples for new units for future reporting periods are not always feasible and knowledge of the subject matter investigator is often used to

upwardly adjust estimates believed to be negatively biased. It is often difficult to present measures of goodness, such as sampling error estimates, in support of such adjusted estimates. When supplemental samples are not feasible, this paper illustrates for a common sampling situation a simple prediction estimation approach for making use of the subject matter investigator's knowledge based on ordinary least squares. The common sampling situation can be described as a two phase sampling plan where the original sample which collects relatively inexpensive data on one occasion is followed on another occasion by a subsample which collects more detailed (and often more expensive) data. Between the first and second occasion, it is assumed that the universe has experienced deaths and births, but resources are not available for taking a supplemental sample of the births. This was precisely the case with the 1987 TIUS sample, which was the first phase sample, and the second phase sample was the 1990 NTACS (sub) sample. The remainder of this paper considers two alternatives of an approach in the context of sampling trucks on two different occasions when the average mileage for certain aged trucks on one occasion is believed to be highly correlated with the average mileage for trucks of the same age on a second occasion. Applications to other settings are clear.

The first alternative (Hu, Wright, Miaou, 1990), which is essentially the one actually used to obtain 1990 estimates of VMT for the U.S. truck population, discussed in Section 2 makes use of age post-stratification and prediction; but it has some disadvantages—one is the lack of a known straightforward expression for providing a single estimate of variability for each estimate. The second alternative discussed in Section 3 makes use of age stratification (before sample selection) and prediction; and the estimator of total truck miles traveled has a straightforward method for computing a single estimate of its variability. Concluding remarks are given in Section 4.

2. SETTING USING AGE POST-STRATIFICATION

2.1 Introduction

Assume a finite universe of N_t similar trucks at time t. Let N_{ty} represent the number of similar y year old trucks at time t for $y = 1, 2, ..., Y$. Then $N_t = \sum_{y=1}^{Y} N_{ty}$. The ith y year old truck has associated vehicle miles traveled (VMT) for the year prior to time t given by

$$V_{tyi} \text{ for } i = 1, 2, ..., N_{ty}. \tag{1}$$

If V_t = the total VMT of the N_t trucks for the year prior to time t and

V_{ty} = the total VMT of the y year old trucks for the year prior to time t, then

$$V_t = \sum_{y=1}^{Y} V_{ty} = \sum_{y=1}^{Y} \sum_{i=1}^{N_{ty}} V_{tyi} \quad . \tag{2}$$

The quantities N_t and N_{ty} are assumed known through annual truck registration records, while V_t, V_{ty}, and V_{tyi} are assumed unknown.

At time t, a simple random sample of size n_t is selected from the N_t trucks to estimate V_t yielding the age post-stratum sample observations:

	Truck Age (yrs)						
	1	**2**	**3**	...	***y***	...	***Y***
	V_{t11}	V_{t21}	V_{t31}		V_{ty1}		V_{tY1}
	V_{t12}	V_{t22}	V_{t32}		V_{ty2}		V_{tY2}
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	...
	$V_{t1n_{t1}}$	$V_{t2n_{t2}}$	$V_{t3n_{t3}}$		$V_{tyn_{ty}}$		$V_{tYn_{tY}}$
Age Post-stratum Sample Means at Time *t*	$\hat{\bar{V}}_{t1}$	$\hat{\bar{V}}_{t2}$	$\hat{\bar{V}}_{t3}$	...	$\hat{\bar{V}}_{ty}$	...	$\hat{\bar{V}}_{tY}$

(3)

Exhibit 1. Age Post-stratum Sample Observations and Means at Time *t*.

where $0 \le n_{ty} \le n_t$ and $n_t = \sum_{y=1}^{Y} n_{ty}$. Note that n_t is fixed, while the quantities n_{ty} are random variables.

If w_{tyi} is the sampling weight for the ith y year old sample truck at time t, then an unbiased estimator for V_t is

$$\hat{V}_t = \sum_{y=1}^{Y} \sum_{i=1}^{n_{ty}} w_{tyi} V_{tyi} = \frac{N_t}{n_t} \sum_{y=1}^{Y} \sum_{i=1}^{n_{ty}} V_{tyi} = N_t \hat{\bar{V}}_t \tag{4}$$

because $w_{tyi} = N_t / n_t$ under simple random sampling. It is known (Cochran, 1977) that the sampling variance of $\hat{V}_t$ is given by

$$Var(\hat{V}_t) = \frac{N_t(N_t - n_t)}{(N_t - 1)\, n_t} \sum_{y=1}^{Y} \sum_{i=1}^{N_{ty}} (V_{tyi} - \bar{V}_t)^2 \tag{5}$$

where $\bar{V}_t = V_t / N_t$, with an unbiased estimator for $Var(\hat{V}_t)$ given by

$$\hat{Var}(\hat{V}_t) = \frac{N_t(N_t - n_t)}{(n_t - 1)\, n_t} \sum_{y=1}^{Y} \sum_{i=1}^{n_{ty}} (V_{tyi} - \hat{\bar{V}}_t)^2 \tag{6}$$

where $\hat{\bar{V}}_t = \sum_{y=1}^{Y} \sum_{i=1}^{n_{ty}} V_{tyi} / n_t$. A visual summary of the setting at time t is given in Figure 2. The simple random sample of n_t trucks is randomly distributed among the truck age post-strata.

Now assume that at time t' in the future a subsample of size $n_{t'}$ is to be selected from the n_t units to estimate $V_{t'}$, which is the total VMT for the universe at time t' of size $N_{t'}$. That is, the $n_{t'}$ units constitute a simple random sample with respect to the n_t units, and hence the $n_{t'}$ units constitute a simple random sample with respect to the N_t units (Wright and Tsao, 1985; Hedayat and Sinha, 1991). For simplicity, throughout the remainder of this paper, except for the example in Subsection 2.5, we will assume that t' occurs one year after t, i.e. $t' = t+1$. Because of births (new model trucks) and deaths (scrapped trucks), special care is needed — especially because a sample from the new model trucks is unavailable and the $n_{t'}$ trucks will contain some trucks that have been scrapped between times t and t'. Refer to Figure 3 for the settings of the universes at times t and t' and note that $n_{t'} = n_{t'2} + n_{t'3} + \ldots + n_{t',Y+1}$.

Let $N_{(tt')}$ be the number of trucks in operation at time t which are still in operation at time t'. In other words, $N_{(tt')}$ is the number of trucks in the intersection of the universe at

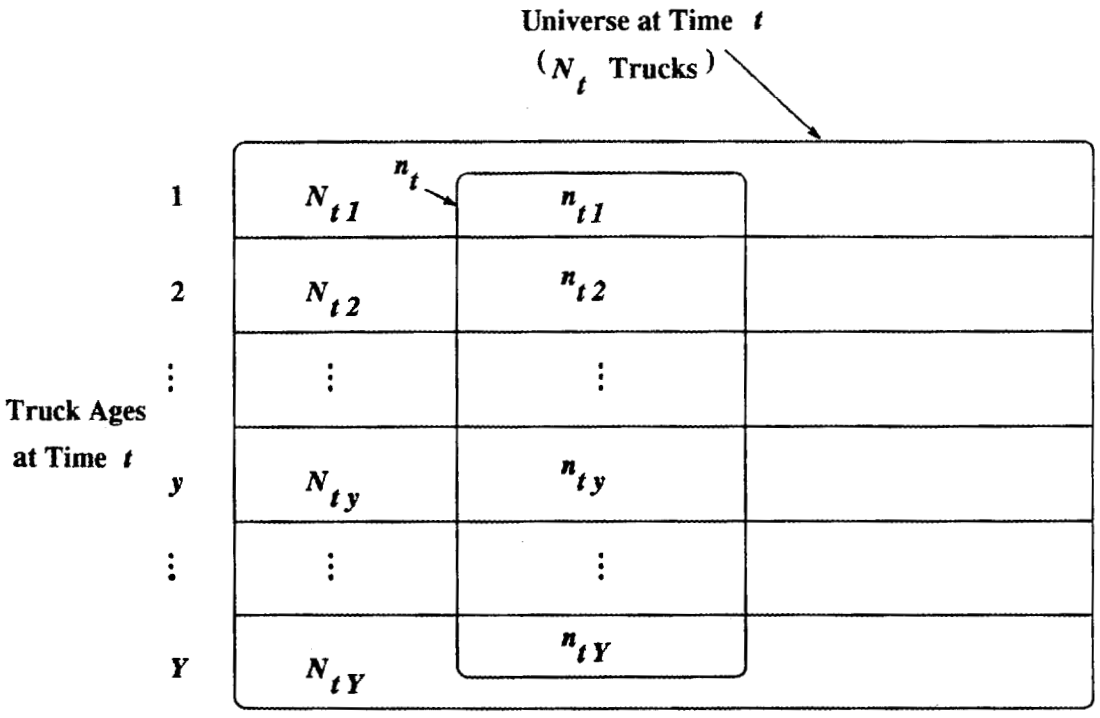

Figure 2. Post-Stratified Universe Setting at Time *t*.

time t with the universe at time t'. From Figure 3, there are $N_{(tt')}$ trucks in the region composed of {4, 5, 6}.

For estimation of $V_{t'}$, the VMT at time t', there are two concerns:

(i) deaths - the trucks in operation at time t which are no longer in operation at time t', and

(ii) births - the trucks in operation at time t' which were not in operation at time t.

Concerning deaths, it is assumed that the number of trucks scrapped during the 1 year period from t to t' is

$$N_t - N_{(tt')} \quad . \tag{7}$$

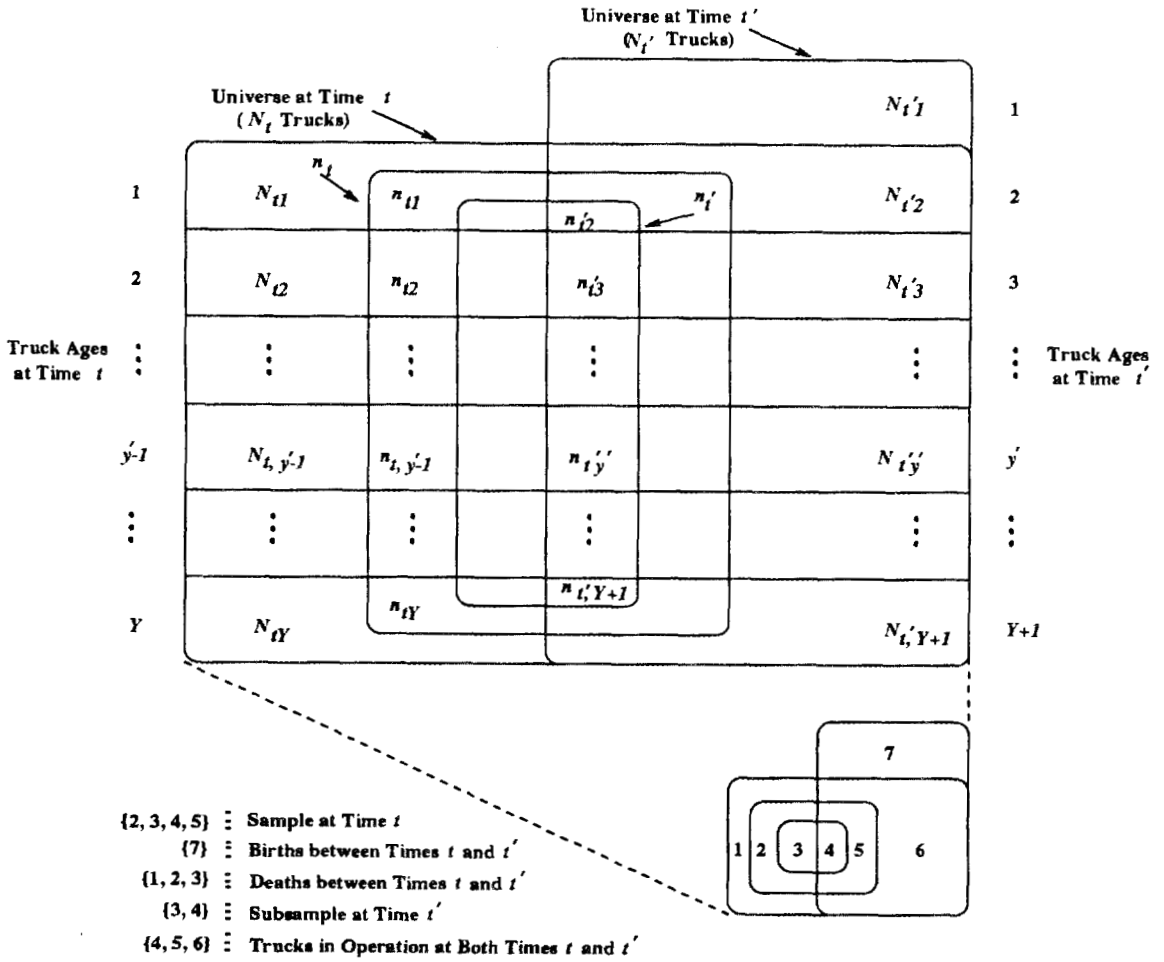

Figure 3. Post-Stratified Universe Settings at Times t and $t'(=t+1)$

Thus these $N_t - N_{(tt')}$ trucks are no longer in the target universe at time t'. Concerning births, the number of new trucks one year (and younger) at time t' is

$$N_{t'1} = N_{t'} - N_{(tt')} \quad . \tag{8}$$

We assume that $N_t - N_{(tt')}$ and $N_{t'1} = N_{t'} - N_{(tt')}$ are both known from truck registration and sales records.

Let $N_{t'y'}$ be the number of y' year old trucks in the universe at time t' which were $y'-1$ years old at time t for $y' = 2, 3, 4, ..., Y+1$. Because of deaths, note that $N_{t'y'} \leq N_{t,y'-1}$ for $y' = 2, 3, 4, ..., Y+1$. (If y' is a positive integer, reference to a y' year old truck means that the truck is greater than $y'-1$ years old and less than or equal to y' years old.) Thus at

time t',

$$N_{t'} = N_{t'1} + \sum_{y'=2}^{Y+1} N_{t'y'} = N_{t'1} + N_{(t')} = N_{t'1} + N_{t'2^+} \tag{9}$$

In an analogous way, we have

$$V_{t'} = V_{t'1} + \sum_{y'=2}^{Y+1} V_{t'y'} = V_{t'1} + V_{t'2^+} \tag{10}$$

At time t', the age post-stratified subsample data from the simple random sample of size $n_{t'}$ can be presented as:

	Truck Age (yrs)					
	2	**3**	...	y'	...	$Y+1$
	$V_{t'21}$	$V_{t'31}$		$V_{t'y'1}$		$V_{t',Y+1,1}$
	$V_{t'22}$	$V_{t'32}$	...	$V_{t'y'2}$	...	$V_{t',Y+1,2}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$V_{t'2n_{t'2}}$	$V_{t'3n_{t'3}}$		$V_{t'y'n_{t'y'}}$		$V_{t',Y+1,n_{t',Y+1}}$
Age Post-stratum Sample Means at Time t'	$\hat{\bar{V}}_{t'2}$	$\hat{\bar{V}}_{t'3}$	...	$\hat{\bar{V}}_{t'y'}$	...	$\hat{\bar{V}}_{t',Y+1}$

Exhibit 2. Age Post-stratum Sample Observations and Means at Time t'.

where $0 \le n_{t'y'} \le n_{t'}$ and $n_{t'} = \sum_{y'=2}^{Y+1} n_{t'y'}$. As with the sample at time t, $n_{t'}$ is fixed while the quantities $n_{t'y'}$ are random variables. The subsample at time t' includes two types of trucks—(1) those trucks in the sample which are still in operation at time t' (region {4} in Figure 3) and (2) those trucks in the sample which are no longer in operation at time t' (region {3} in Figure 3). For those trucks in the subsample of $n_{t'}$ trucks which are no longer in operation at time t', the sample value of VMT is zero; the same is true for those trucks in the universe (at time t) that were scrapped between times t and t'.

Because the subsample of $n_{t'}$ trucks is a simple random sample with respect to the universe of N_t trucks at time t and because domain estimation can be employed for the $N_{t'2^+}$ trucks as discussed on pages 36-37 of Cochran (1977), it is appropriate to use the sample weights $w_{t'y'i} = N_t / n_{t'}$ to obtain as an unbiased estimator of $V_{t'2^+}$ the estimator,

$$\hat{V}_{t'2^+} = \frac{N_t}{n_{t'}} \sum_{y'=2}^{Y+1} \sum_{i=1}^{n_{t'y'}} V_{t'y'i} \quad . \tag{11}$$

The sampling variance of $\hat{V}_{t'2^+}$ is given by

$$Var(\hat{V}_{t'2^+}) = \frac{N_t (N_t - n_{t'})}{(N_t - 1)\, n_{t'}} \left\{ \sum_{y'=2}^{Y+1} \sum_{i=1}^{N_{t'y'}} V_{t'y'i}^2 - \frac{(V_{t'2^+})^2}{N_t} \right\} \quad , \tag{12}$$

and an unbiased estimator of $Var(\hat{V}_{t'2^+})$ is given by

$$\hat{Var}(\hat{V}_{t'2^+}) = \frac{N_t (N_t - n_{t'})}{(n_{t'} - 1)\, n_{t'}} \sum_{y'=2}^{Y+1} \sum_{i=1}^{n_{t'y'}} \left(V_{t'y'i} - \frac{\sum_{y'=2}^{Y+1} \sum_{i=1}^{n_{t'y'}} V_{t'y'i}}{n_{t'}} \right)^2 \quad . \tag{13}$$

From (10), it remains to estimate $V_{t'1}$ to complete our estimation of $V_{t'}$. However, as noted from Figure 3, we have no sample data from the $N_{t'1}$ units whose total VMT at time t' is $V_{t'1}$. Note that

$$V_{t'1} = N_{t'1} \bar{V}_{t'1} \quad , \tag{14}$$

where $\bar{V}_{t'1}$ is the average VMT for 1 year old trucks at time t'. Because $N_{t'1}$ is known, it is enough to estimate $\bar{V}_{t'1}$. Exhibit 3 shows known estimates of means based on sample data from time t as well as time t'. Note that an estimate of $\bar{V}_{t'1}$ would occupy the space indicated by "?" in Exhibit 3.

	Truck Age (yrs)								
Time	1	2	3	4	...	y	...	Y	Y+1
t	$\hat{\bar{V}}_{t1}$	$\hat{\bar{V}}_{t2}$	$\hat{\bar{V}}_{t3}$	$\hat{\bar{V}}_{t4}$	...	$\hat{\bar{V}}_{ty}$	...	$\hat{\bar{V}}_{tY}$	—
t'	?	$\hat{\bar{V}}_{t'2}$	$\hat{\bar{V}}_{t'3}$	$\hat{\bar{V}}_{t'4}$	...	$\hat{\bar{V}}_{t'y}$	...	$\hat{\bar{V}}_{t'Y}$	$\hat{\bar{V}}_{t',Y+1}$

Exhibit 3. Post-stratum Sample Means at Times t and t' (=t+1).

2.2 Key Assumption

It is assumed that the average mileage of similar y year old trucks at time t is correlated with average mileage of similar y year old trucks at time t'.

2.3 Prediction and Sampling from Finite Universes

Prediction has proved to be very useful in problems relating to sampling from finite universes when certain data are unavailable. Prediction models are frequently used to justify the choice of a particular estimator, sampling plan, or both. Examples of methodology addressed with assistance from a prediction approach include: (i) model-based sampling (Royall 1970, 1988; Wright 1983; Cassel, Särndal, and Wretman 1977; Särndal, Swenson, and Wretman 1992; Bolfarine and Zacks 1992); (ii) analysis with weighted survey data (Zieschang 1990; Little 1991); (iii) incomplete data (Little and Rubin 1987; Madow, Olkin, and Rubin 1983; Schaible 1983); and (iv) estimation for small areas (Platek, Rao, Särndal, and Singh 1987; and, Purcell and Kish 1979) — especially improved small area census counts (Ericksen and Kadane 1985; Freedman and Navidi 1986; Wright 1993).

One characteristic common to these and other methods is that there must be a link established between what is available and what is not available, and regression methods help in this regard.

The model of this paper developed in Section 3 in the context of a universe of trucks uses age as the link between two points in time to predict miles for new trucks by using current and previous data from older trucks with stratification before sample selection. Before considering that model, Subsections 2.4 and 2.5 describe a prediction approach in the context of post-stratification which was used with the 1990 NTACS sample data.

2.4 Prediction Estimation Model for $\bar{V}_{t'1}$ Using Post-stratification

The method presented in this subsection is a variation of the method proposed by Hu, Wright, and Miaou (1990). With reference to Exhibit 3, the following given pairs of estimates (observations) — $(\hat{\bar{V}}_{t2}, \hat{\bar{V}}_{t'2})$; $(\hat{\bar{V}}_{t3}, \hat{\bar{V}}_{t'3})$; ...; $(\hat{\bar{V}}_{tY}, \hat{\bar{V}}_{t'Y})$ — will be used to fit the following first - order linear model employing the method of least squares:

$$\hat{\bar{V}}_{t'y} = \beta_0 + \beta_1 \hat{\bar{V}}_{ty} + \varepsilon_y \quad \text{for } y = 2, 3, ..., Y \tag{15}$$

where ε_y is a random variable (error) with $E(\varepsilon_y) = 0$; $Var(\varepsilon_y) = \sigma^2$, and $Cov(\varepsilon_l, \varepsilon_k) = 0$ for $l \neq k$. In this formulation, we view $\hat{\bar{V}}_{ty}$ $(y = 2, ..., Y)$ as being fixed constants while the $\hat{\bar{V}}_{t'y}$ $(y = 2, 3, ..., Y)$ are random variables with $E(\hat{\bar{V}}_{t'y}) = \beta_0 + \beta_1 \hat{\bar{V}}_{ty}$ and $Var(\hat{\bar{V}}_{t'y}) = \sigma^2$.

As in the usual linear regression approach, β_0, β_1, and σ^2 are unknown parameters with respective least square estimates (Draper and Smith, 1981):

$$\hat{\beta}_1 = \frac{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{\bar{V}}}_t)(\hat{\bar{V}}_{t'y} - \hat{\bar{\bar{V}}}_{t'})}{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{\bar{V}}}_t)^2},$$

$$\hat{\beta}_0 = \hat{\bar{\bar{V}}}_{t'} - \hat{\beta}_1 \hat{\bar{\bar{V}}}_t \text{, and}$$

$$\hat{\sigma}^2 = \frac{\sum_{y=2}^{Y} (\hat{\bar{V}}_{t'y} - \hat{\bar{V}}^*_{t'y})^2}{Y-3} \tag{16}$$

where $\hat{\bar{\bar{V}}}_t = \sum_{y=2}^{Y} \hat{\bar{V}}_{ty} / (Y-1)$, $\hat{\bar{\bar{V}}}_{t'} = \sum_{y=2}^{Y} \hat{\bar{V}}_{t'y} / (Y-1)$, and the predicted averages at time t' are

$$\hat{\bar{V}}^*_{t'y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\bar{V}}_{ty} \quad \text{for } y = 2, 3, ..., Y. \tag{17}$$

To estimate $\bar{V}_{t'1}$, take the prediction of $\bar{V}_{t'1}$, which is the predicted value

$$\bar{V}^*_{t'1} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\bar{V}}_{t1} \tag{18}$$

at $\hat{\bar{V}}_{t1}$. If the model in (15) is correct, the variance of $\bar{V}^*_{t'1}$ is

$$Var(\bar{V}^*_{t'1}) = \frac{\sigma^2}{Y-1} + \frac{(\hat{\bar{V}}_{t1} - \hat{\bar{V}}_t)^2 \sigma^2}{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{V}}_t)^2} \quad , \tag{19}$$

and an estimator of $Var(\bar{V}^*_{t'1})$ is

$$V\hat{a}r(\bar{V}^*_{t'1}) = \frac{\hat{\sigma}^2}{Y-1} + \frac{(\hat{\bar{V}}_{t1} - \hat{\bar{V}}_t)^2 \hat{\sigma}^2}{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{V}}_t)^2} \quad . \tag{20}$$

Thus our estimator of $V_{t'}$ from (18) and (11) is

$$\hat{V}_{t'} = \hat{V}_{t'1} + \hat{V}_{t'2^+} = N_{t'1} \bar{V}^*_{t'1} + \frac{N_t}{n_{t'}} \sum_{y'=2}^{Y+1} \sum_{i=1}^{n_{t'y'}} V_{t'y'i} \quad . \tag{21}$$

<u>*Comment*</u>: From (21), it is clear that the estimator $\hat{V}_{t'}$ has two components — $\hat{V}_{t'1}$ and $\hat{V}_{t'2^+}$. In this post-stratified setting, the variability of $\hat{V}_{t'}$ is given in terms of the variability of each component.

— If the model in (15) is correct, the estimator $\hat{V}_{t'1}$ is an unbiased least squares estimator of $V_{t'1}$, and an estimator of its variability is obtained by multiplying (20) by $N^2_{t'1}$, i.e. $N^2_{t'1} V\hat{a}r(\bar{V}^*_{t'1})$. This estimator makes use of all of the observed post-stratified sample means at times t and t' except $\hat{\bar{V}}_{t',Y+1}$. To assess the extent to which the model in (15) is appropriate as a predictor of $\bar{V}_{t'1}$, we compute the Pearson correlation coefficient between the observed post-stratified sample means at time t and at time t', i.e.

$$r_{tt'} = \frac{\sum_{y=2}^{Y} (\hat{\hat{V}}_{ty} - \hat{\hat{V}}) (\hat{\hat{V}}_{t'y} - \hat{\hat{V}}_{t'})}{\sqrt{\sum_{y=2}^{Y} (\hat{\hat{V}}_{ty} - \hat{\hat{V}}_{t})^2 \sum_{y=2}^{Y} (\hat{\hat{V}}_{t'y} - \hat{\hat{V}}_{t'})^2}} \quad . \tag{22}$$

It is desired that $r_{tt'}$ be near one. Of course, one can also examine the residuals $\hat{\hat{V}}_{t'y} - \hat{\hat{V}}^{*}_{t'y}$ for $y = 2, 3, ..., Y$ (Draper and Smith, 1981).

— The estimator $\hat{V}_{t'2^{\cdot}}$ is a sample design unbiased estimator of $V_{t'2^{\cdot}}$ with an estimator of its sampling variability given in (13). This (domain) estimator makes use of all of the sample data at time t' and the universe size at time t.

2.5 Application to the Post-stratified 1990 NTACS Sample Data

The $t = 1987$ and $t' = 1990$ U.S. truck universes can each be partitioned into nine subuniverses each of which is composed of a homogeneous collection of trucks. The 1990 total vehicle miles traveled (VMT) for the entire U.S. universe, a primary parameter, was estimated by summing estimates of the 1990 total vehicle miles traveled for these nine subuniverses. In this subsection, we describe the estimation and prediction steps in estimating the 1990 VMT for the trucks in one of those subuniverses which is,

The collection of all trucks operating in the U.S. during 1990 which are described as being long haul commodity-carrying truck-tractors.

A similar approach was used for the other eight subuniverses *(1990 NTACS Selected Tabulations, 1993). To avoid awkward notation for this subuniverse, throughout this subsection and from this point on, we will refer to the subuniverse as "the universe" and use t for 1987 and t' for 1990.* Note here that $t' = t+3$ and that we will be predicting for the times $t+1$, $t+2$, and $t+3$ (i.e. 1988, 1989, and 1990). The objective is to give the details for the estimation of $V_{t'}$, the total 1990 VMT for all long haul commodity-carrying truck-tractors operating in the U.S.

We assume the selection of a simple random sample of $n_t = 12{,}563$ trucks from the approximately $N_t = 407{,}757$ long haul commodity-carrying truck-tractors in $t = 1987$ followed by the selection of a simple random (sub)sample of $n_{t'} = 2{,}372$ trucks for $t' = 1990$. Thus no sample was selected from new trucks for 1988, 1989, nor 1990. From truck registration records, assume

that the number of new trucks operating in 1990 for these years are $N_{t'3} = 71{,}948$; $N_{t'2} = 51{,}555$; and $N_{t'1} = 37{,}024$ respectively. A display of post-stratified sample means at times t and t' corresponding to Exhibit 3 is given below:

Post-Stratum Sample Means

	Time	
Truck Age (yrs)	t = 1987	t' = 1990
1	93316	?
2	104420	?
3	102462	?
4	97236	109651
5	88028	101130
6	79123	96714
7	81778	79393
8	73336	76882
9	71645	75119
10	68305	71490
11	—	64574
12	—	58771
13	—	56888
14⁺	—	53771

(14⁺ includes trucks that are at least 14 years old.)

Note that $\hat{\bar{V}}_{t'1}$, $\hat{\bar{V}}_{t'2}$, and $\hat{\bar{V}}_{t'3}$ are not given in the table. Using only the seven pairs of means for $y' = 4, 5, 6, 7, 8, 9, 10$, the prediction equation using (17) is obtained as

$$\hat{\bar{V}}^{*}_{t'y} = -20{,}419 + 1.3465\ \hat{\bar{V}}_{ty}$$

with $\hat{\sigma}^2 = 46{,}235{,}640$. It is important to note that the correlation coefficient between the seven pairs of means at times t and t' is $r_{tt'} = .9103$. The predicted values of $\bar{V}_{t'1}$, $\bar{V}_{t'2}$, and $\bar{V}_{t'3}$ are obtained as

$$\bar{V}^*_{t'1} = 105{,}231$$
$$\bar{V}^*_{t'2} = 120{,}183$$
$$\text{and} \quad \bar{V}^*_{t'3} = 117{,}546.$$

Assuming that the model is correct and given the values $\hat{\bar{V}}_{ty}$ $(y = 1,\ 2,\ 3)$, the estimated standard errors for these predictions are

$$\sqrt{V\hat{a}r(\bar{V}^*_{t'1})} = 4{,}600$$
$$\sqrt{V\hat{a}r(\bar{V}^*_{t'2})} = 7{,}261$$
$$\text{and} \quad \sqrt{V\hat{a}r(\bar{V}^*_{t'3})} = 6{,}769 \quad .$$

The estimate for $V_{t'4^+}$ is given using (11) by:

$$\hat{V}_{t'4^+} = \frac{N_t}{n_{t'}} \sum_{y'=4}^{14^+} \sum_{i=1}^{n_{t'y'}} V_{t'y'i}$$
$$= 31{,}661{,}686{,}995 \text{ miles}$$

with estimated standard error given by (13):

$$\sqrt{V\hat{a}r\ (\hat{V}_{t'4^+})} = \left\{ \frac{N_t(N_t - n_{t'})}{(n_{t'}-1)\, n_{t'}} \sum_{y'=4}^{14^+} \sum_{i=1}^{n_{t'y'}} \left(V_{t'y'i} - \frac{\sum_{y'=4}^{14^+} \sum_{i=1}^{n_{t'y'}} V_{t'y'i}}{n_{t'}} \right)^2 \right\}^{1/2}$$
$$= \left\{ \frac{407{,}757\,(407{,}757 - 2372)}{2372}\ (3{,}462{,}307{,}302) \right\}^{1/2}$$
$$\approx 491{,}201{,}872.5 \quad \text{miles} \quad .$$

Thus by (21), an estimate of $V_{t'}$ is

$$\begin{aligned}\hat{V}_{t'} &= \hat{V}_{t'1} + \hat{V}_{t'2} + \hat{V}_{t'3} + \hat{V}_{t'4^+} \\ &= N_{t'1}\,\bar{V}^*_{t'1} + N_{t'2}\,\bar{V}^*_{t'2} + N_{t'3}\,\bar{V}^*_{t'3} + \frac{N_t}{n_t}\sum_{y'=4}^{14^+}\sum_{i=1}^{n_{t'y'}} V_{t'y'i} \\ &= 37{,}024\,(105{,}231) + 51{,}555\,(120{,}183) + 71{,}948\,(117{,}546) + 31{,}661{,}686{,}995 \\ &= 50{,}210{,}993{,}712 \text{ miles.}\end{aligned}$$

where its goodness is reflected in the goodness of its components as given by

$$N_{t'y}\sqrt{V\hat{a}r\,(\bar{V}^*_{t'y})} \text{ for } y = 1,\ 2,\ 3 \text{ and } \sqrt{V\hat{a}r\,(\hat{V}_{t'4^+})} \quad .$$

3. SETTING USING AGE STRATIFICATION BEFORE SAMPLING AT TIMES t AND t'

One problem with the post-stratification approach discussed in Section 2 and just demonstrated for the example is that we lack one expression for the variance of $\hat{V}_{t'}$, $Var\,(\hat{V}_{t'})$; and one component of $\hat{V}_{t'}$ is viewed as being predicted where the expression of variance is true for this component only if the prediction model is correct. Also recall that the n_{ty} and $n_{t'y'}$ quantities are random variables. Hence it is possible that no trucks for a particular age may be realized in the sample either at time t and/or time t'. In this section, we illustrate how these difficulties can be removed with a stratified approach before selection of sample which is conditioned on the age sample means from time t.

As in Section 2, we have a finite universe of N_t similar trucks at time t. Unlike the approach of Section 2, here the universe is stratified by truck age before sampling. All similar trucks y years old are assigned to stratum y where $y = 1,\ 2,\ \ldots,\ Y$. Let N_{ty} be the known number of similar y year old trucks in stratum y at time t. As noted earlier,

$$V_t = \sum_{y=1}^{Y} V_{ty} = \sum_{y=1}^{Y}\sum_{i=1}^{N_{ty}} V_{tyi} \tag{23}$$

where V_t, V_{ty}, and V_{tyi} are all as defined in (1) and (2). (While some of the notation of Section 2 will be used throughout this section, the meanings of estimators and sample sizes will be different largely due to the difference in sampling plans between the two sections and due to the

fact that the quantities n_{ty} and $n_{t'y'}$ in Section 3 will be fixed, unlike the n_{ty} and $n_{t'y'}$ in Section 2 which are random variables.)

At time t, a stratified random sample of size n_t is selected to estimate V_t yielding the strata sample means:

Truck Age Stratum	1	2	3	...	y	...	Y
Stratum Sample Mean at Time t	$\hat{\bar{V}}_{t1}$	$\hat{\bar{V}}_{t2}$	$\hat{\bar{V}}_{t3}$	...	$\hat{\bar{V}}_{ty}$	...	$\hat{\bar{V}}_{tY}$

Exhibit 4. Age Stratum Sample Means at Time t.

where the size of the sample from stratum y at time t is n_{ty}. Here, $n_t \left(= \sum_{y=1}^{Y} n_{ty} \right)$ and the n_{ty} are fixed. An unbiased estimator of V_t is

$$\hat{V}_{t(s)} = \sum_{y=1}^{Y} \hat{V}_{ty(s)} = \sum_{y=1}^{Y} N_{ty} \hat{\bar{V}}_{ty} \quad . \tag{24}$$

(The "s" subscript indicates stratification before sampling.) The sampling variance of $\hat{V}_{t(s)}$ is

$$Var(\hat{V}_{t(s)}) = \sum_{y=1}^{Y} N_{ty} (N_{ty} - n_{ty}) S_{ty}^2 / n_{ty} \tag{25}$$

where $S_{ty}^2 = \sum_{i=1}^{N_{ty}} (V_{tyi} - \bar{V}_{ty})^2 / (N_{ty} - 1)$. An unbiased estimator of $Var(\hat{V}_{t(s)})$ is

$$\hat{Var}(\hat{V}_{t(s)}) = \sum_{y=1}^{Y} N_{ty} (N_{ty} - n_{ty}) s_{ty}^2 / n_{ty} \tag{26}$$

where $s_{ty}^2 = \sum_{i=1}^{n_{ty}} (V_{tyi} - \hat{\bar{V}}_{ty})^2 / (n_{ty} - 1)$.

The visual stratified setting at time t is given in Figure 4 and should be contrasted with Figure 2.

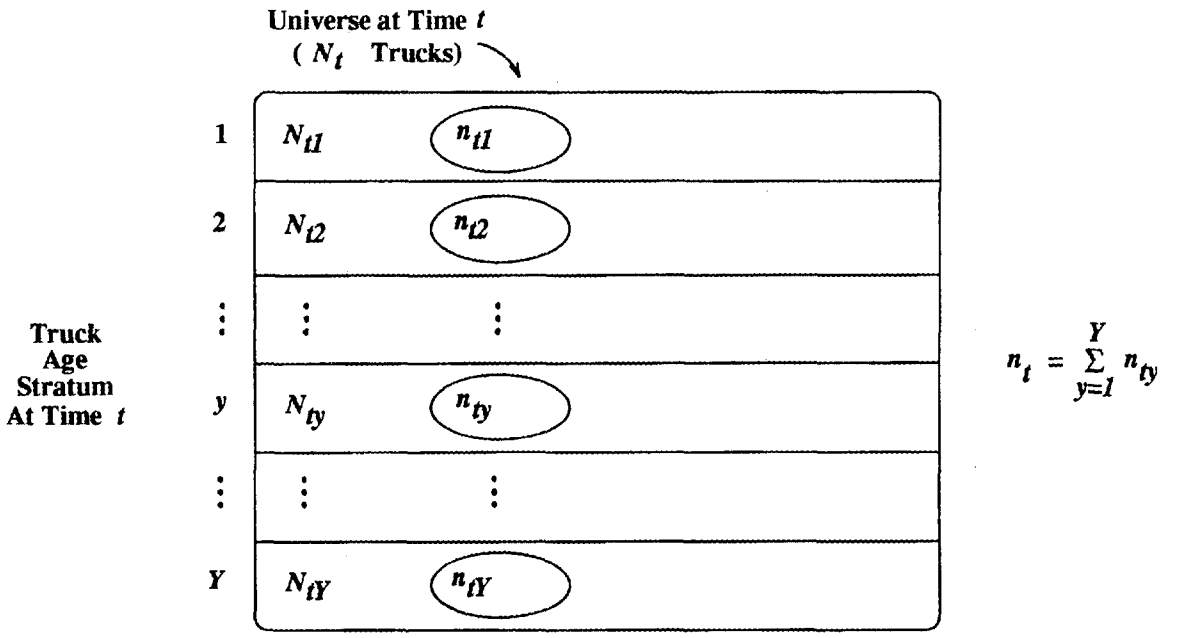

Figure 4. Stratified Universe Setting at Time *t*.

Now at time t' (assume $t' = t+1$) in the future, a simple random sample of size $n_{t'y'}$ is selected from the $n_{t,y'-1}$ sample units ($y' = 2, 3, ..., Y+1$) to estimate $V_{t'}$, which is the total VMT for the universe of size $N_{t'}$ at time t'. (As before, we take $t' = t+1$ for simplicity.) The $n_{t'} = \sum_{y'=2}^{Y+1} n_{t'y'}$ sample units constitute a stratified random sample from the N_t units because the $n_{t'y'}$ units constitute a simple random sample from the $N_{t,y'-1}$ units at time t. With the deaths, births, and stratification by age of the universe at time t', see Figure 5 and contrast it with Figure 3.

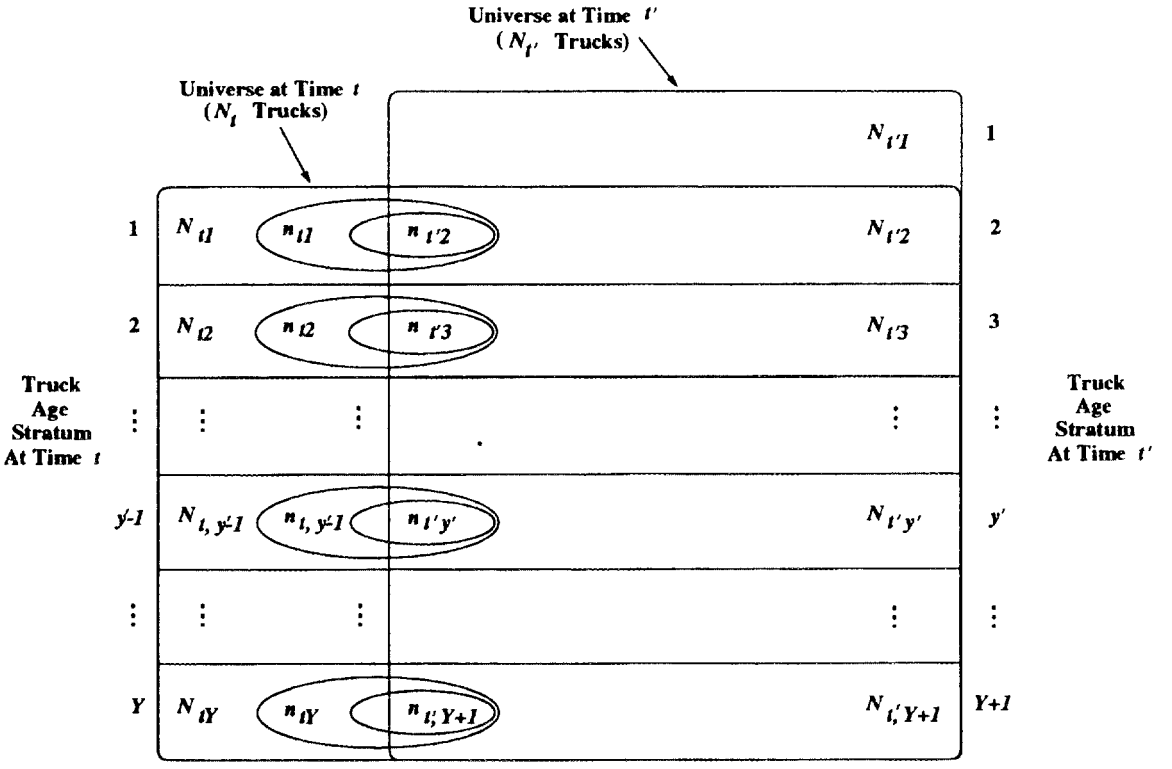

Figure 5. Stratified Universe Setting at Times t and t'.

As noted earlier in Section 2, it is assumed that the number of trucks scrapped during the 1 year period from t to t' is $N_t - N_{(u')}$. Concerning births, the number of new trucks one year (and younger) at time t' is $N_{t'1} = N_{t'} - N_{(u')}$. Both $N_t - N_{(u')}$ and $N_{t'1}$ are known as is $N_{t'y'}$ for $y' = 2, 3, 4, \ldots, Y+1$. Also $N_{t'}$ and $V_{t'}$ can be written as in (9) and (10), where in particular,

$$V_{t'} = V_{t'1} + V_{t'2^+} \quad .$$

At time t', the observed stratum sample means are computed as given in (27) below and are:

Truck Age Stratum	2	3	4	...	y'	...	$Y+1$
Stratum Sample Mean at Time t'	$\hat{V}_{t'2}$	$\hat{V}_{t'3}$	$\hat{V}_{t'4}$	...	$\hat{V}_{t'y'}$	...	$\hat{V}_{t',Y+1}$

Exhibit 5. Age Stratum Sample Means at Time t'.

Here, $n_{t'}$ and $n_{t'y'}$ are fixed where $y' = 2, \ldots, Y+1$. For those trucks in the sample (and universe) which are no longer in operation at time t', the sample (and universe) values are zero. Note the

absence of sample data from stratum 1 at time t'.

An unbiased estimator of $V_{t'2^+}$ is

$$\begin{aligned}
\hat{V}_{t'2^+(s)} &= \sum_{y'=2}^{Y+1} \hat{V}_{t'y'} \\
&= \sum_{y'=2}^{Y+1} \frac{N_{t,\,y'-1}}{n_{t'y'}} \left(\sum_{i=1}^{n_{t'y'}} V_{t'y'i} \right) \\
&= \sum_{y'=2}^{Y+1} N_{t'y'} \left(\frac{\frac{N_{t,\,y'-1}}{n_{t'y'}} \sum_{i=1}^{n_{t'y'}} V_{t'y'i}}{N_{t'y'}} \right) \\
&= \sum_{y'=2}^{Y+1} N_{t'y'} \, \hat{\bar{V}}_{t'y'} \quad .
\end{aligned} \tag{27}$$

The sampling variance of $\hat{V}_{t'2^+(s)}$ is given by

$$\begin{aligned}
Var(\hat{V}_{t'2^+(s)}) &= \sum_{y'=2}^{Y+1} Var(\hat{V}_{t'y'}) = \sum_{y'=2}^{Y+1} N_{t,\,y'-1} \, (N_{t,y'-1} - n_{t'y'}) \, \frac{S^2_{t'y'}}{n_{t'y'}} \\
\text{where } S^2_{t'y'} &= \frac{\sum_{i=1}^{N_{t,y'-1}} (V_{t'y'i} - \bar{V}_{t'y'})^2}{N_{t,\,y'-1} - 1} \quad \text{and} \quad \bar{V}_{t'y'} = \frac{\sum_{i=1}^{N_{t,y'-1}} V_{t'y'i}}{N_{t,\,y'-1}} \quad .
\end{aligned} \tag{28}$$

Because $V_{t'y'i} = 0$ for those trucks scrapped between times t and $t'(=t+1)$, the expression for $Var(\hat{V}_{t'y'})$ follows from domain estimation as discussed by Cochran (1977, pp 36-37).

An unbiased estimator of $Var(\hat{V}_{t'2^+(s)})$ is

$$\begin{aligned}
\hat{Var}(\hat{V}_{t'2^+(s)}) &= \sum_{y'=2}^{Y+1} N_{t,\,y'-1} \, (N_{t,\,y'-1} - n_{t'y'}) \, \frac{s^2_{t'y'}}{n_{t'y'}} \\
\text{where } s^2_{t'y'} &= \frac{\sum_{i=1}^{n_{t'y'}} (V_{t'y'i} - \tilde{\bar{V}}_{t'y'})^2}{n_{t'y'} - 1} \quad \text{and} \quad \tilde{\bar{V}}_{t'y'} = \frac{\sum_{i=1}^{n_{t'y'}} V_{t'y'i}}{n_{t'y'}} \quad .
\end{aligned} \tag{29}$$

For $V_{t'1}=N_{t'1}\ \bar{V}_{t'1}$, we again use prediction for $\bar{V}_{t'1}$. Refer to Exhibit 6 for known estimates of stratum sample means at times t and t' and contrast with Exhibit 3.

	Truck Age Stratum								
Time	1	2	3	4	...	y	...	Y	$Y+1$
t	$\hat{\bar{V}}_{t1}$	$\hat{\bar{V}}_{t2}$	$\hat{\bar{V}}_{t3}$	$\hat{\bar{V}}_{t4}$	...	$\hat{\bar{V}}_{ty}$	...	$\hat{\bar{V}}_{tY}$	—
t'	?	$\hat{\bar{V}}_{t'2}$	$\hat{\bar{V}}_{t'3}$	$\hat{\bar{V}}_{t'4}$	...	$\hat{\bar{V}}_{t'y}$	...	$\hat{\bar{V}}_{t'Y}$	$\hat{\bar{V}}_{t',Y+1}$

Exhibit 6. Stratum Sample Means at Times t and t'

Proceeding as before in Section 2 with the $Y-1$ pairs $(\hat{\bar{V}}_{t2}, \hat{\bar{V}}_{t'2})$; ...; $(\hat{\bar{V}}_{tY}, \hat{\bar{V}}_{t'Y})$ and assuming a linear model analogous to (15),

$$\hat{\bar{V}}_{t'y} = \beta_{0(s)} + \beta_{1(s)}\,\hat{\bar{V}}_{ty} + \varepsilon_y \tag{30}$$

where ε_y is a random variable (error) with $E(\varepsilon_y)=0$; $Var(\varepsilon_y)=\sigma^2_{(s)}$, and $Cov(\varepsilon_l, \varepsilon_k)=0$ for $l \neq k$, we predict $\bar{V}_{t'1}$ by

$$\bar{V}^*_{t'1(s)} = \hat{\beta}_{0(s)} + \hat{\beta}_{1(s)}\,\hat{\bar{V}}_{t1} \tag{31}$$

at $\hat{\bar{V}}_{t1}$ with $\hat{\beta}_{0(s)}$, $\hat{\beta}_{1(s)}$, and $\hat{\sigma}^2_{(s)}$ given analogously as $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}^2$ in (16). Also $Var(\bar{V}^*_{t'1(s)})$ and $\hat{Var}(\bar{V}^*_{t'1(s)})$ are given analogously as $Var(\bar{V}^*_{t'1})$ and $\hat{Var}(\bar{V}^*_{t'1})$ are by (19) and (20) respectively.

Thus when we stratify before sampling on each occasion as noted in Figure 5, we estimate $V_{t'}$ by

$$
\begin{aligned}
\hat{V}_{t'(s)} &= \hat{V}_{t'1(s)} + \hat{V}_{t'2^*(s)} \\
&= N_{t'1}\ \bar{V}^*_{t'1(s)} + \hat{V}_{t'2^*(s)} \\
&= N_{t'1}\ (\hat{\beta}_{0(s)} + \hat{\beta}_{1(s)}\ \hat{\bar{V}}_{t1}) + \sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'} \\
&= N_{t'1}\ \left\{(\hat{\bar{V}}_{t'} - \hat{\beta}_{1(s)}\ \hat{\bar{V}}_{t}) + \hat{\beta}_{1(s)}\ \hat{\bar{V}}_{t1}\right\} + \sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'} \\
&= N_{t'1}\ \left[\hat{\bar{V}}_{t'} + (\hat{\bar{V}}_{t1} - \hat{\bar{V}}_{t})\ \hat{\beta}_{1(s)}\right] + \sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'} \\
&= N_{t'1}\left\{\frac{\sum_{y=2}^{Y} \hat{\bar{V}}_{t'y}}{Y-1} + (\hat{\bar{V}}_{t1} - \hat{\bar{V}}_{t})\ \frac{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t})\ \hat{\bar{V}}_{t'y}}{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t})^2}\right\} + \sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'} \\
&= N_{t'1} \sum_{y=2}^{Y} \left\{\frac{1}{Y-1} + (\hat{\bar{V}}_{t1} - \hat{\bar{V}}_{t})\ \frac{(\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t})}{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t})^2}\right\} \hat{\bar{V}}_{t'y} + \sum_{y=2}^{Y+1} N_{t'y}\ \hat{\bar{V}}_{t'y} \\
&= \sum_{y=2}^{Y} \left[N_{t'1}\left\{\frac{1}{Y-1} + (\hat{\bar{V}}_{t1} - \hat{\bar{V}}_{t})\ \frac{(\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t})}{\sum_{y=2}^{Y} (\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t})^2}\right\} + N_{t'y}\right] \hat{\bar{V}}_{t'y} + N_{t',Y+1}\ \hat{\bar{V}}_{t',Y+1} \\
&= \sum_{y=2}^{Y} f_y\ \hat{\bar{V}}_{t'y} + N_{t',Y+1}\ \hat{\bar{V}}_{t',Y+1}
\end{aligned}
$$

where $f_y = N_{t'1}\left\{\frac{1}{Y-1} + \left(\hat{\bar{V}}_{t1} - \hat{\bar{V}}_{t}\right) \frac{\left(\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t}\right)}{\sum_{y=2}^{Y} \left(\hat{\bar{V}}_{ty} - \hat{\bar{V}}_{t}\right)^2}\right\} + N_{t'y}$.

Thus our stratified estimator of $V_{t'}$ is

$$\hat{V}_{t'(s)} = \sum_{y=2}^{Y} f_y\ \hat{\bar{V}}_{t'y} + N_{t',Y+1}\ \hat{\bar{V}}_{t',Y+1} \tag{32}$$

where f_y is a function of sample data at time t and $N_{t'y}$ for $y=1, 2, ..., Y$.

If the model in (30) is correct, then clearly given the results from time t ,

$$
\begin{aligned}
\mathrm{E}_D\ \mathrm{E}_M\left[\hat{V}_{t'(s)}\right] &= \mathrm{E}_D\ \mathrm{E}_M\left[\hat{V}_{t'1(s)}+\hat{V}_{t'2\tau(s)}\right] \\
&= \mathrm{E}_D\ \mathrm{E}_M\left[N_{t'1}\ \bar{V}^{*}_{t'1(s)}+\sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'}\right] \\
&= \mathrm{E}_D\left[N_{t'1}\ \mathrm{E}_M\left(\bar{V}^{*}_{t'1(s)}\right)+\sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'}\right] \\
&= \mathrm{E}_D\left[N_{t'1}\ \bar{V}_{t'1}+\sum_{y'=2}^{Y+1} N_{t'y'}\ \hat{\bar{V}}_{t'y'}\right] \\
&= N_{t'1}\ \bar{V}_{t'1}+\sum_{y'=2}^{Y+1} N_{t'y'}\ \mathrm{E}_D\left(\hat{\bar{V}}_{t'y'}\right) \\
&= N_{t'1}\ \bar{V}_{t'1}+\sum_{y'=2}^{Y+1} N_{t'y'}\ \bar{V}_{t'y'} \qquad (33) \\
&= \sum_{y'=1}^{Y+1} N_{t'y'}\ \bar{V}_{t'y'} \\
&= \sum_{y'=1}^{Y+1} V_{t'y'} \\
&= V_{t'}
\end{aligned}
$$

where E_D means expectation relative to the sampling design and E_M means expectation relative to the model given the results from time t. Because $\mathrm{E}_D\ \mathrm{E}_M\left(\hat{V}_{t'(s)}\right) = V_{t'}$ given the results from time t and given the model in (30) is correct, we say that $\hat{V}_{t'(s)}$ is a <u>conditionally design-model unbiased</u> estimator of the total $V_{t'}$.

As noted earlier, the motivation for the method of Section 3 leading to the estimator $\hat{V}_{t'(s)}$ is to be able to provide a single measure of standard error for the estimated VMT at time t'. To assess the reliability of $\hat{V}_{t'(s)}$ given only the results from time t in terms of sampling design variability, we have

$$
\begin{aligned}
Var_D\left(\hat{V}_{t'(s)}\right) &= Var_D\left[\sum_{y=2}^{Y} f_y \hat{\bar{V}}_{t'y} + N_{t',Y+1} \hat{\bar{V}}_{t',Y+1}\right] \\
&= \sum_{y=2}^{Y} f_y^2 \, Var_D\left(\hat{\bar{V}}_{t'y}\right) + N_{t',Y+1}^2 \, Var_D\left(\hat{\bar{V}}_{t',Y+1}\right) \\
&= \sum_{y=2}^{Y} \frac{f_y^2}{N_{t'y}^2} \, Var_D\left(\hat{V}_{t'y}\right) + \frac{N_{t',Y+1}^2}{N_{t',Y+1}^2} \, Var_D\left(\hat{V}_{t',Y+1}\right) \\
&= \sum_{y=2}^{Y} \left(\frac{f_y}{N_{t'y}}\right)^2 N_{t,y-1}\left(N_{t,y-1} - n_{t'y}\right) \frac{S_{t'y}^2}{n_{t'y}} + N_{tY}\left(N_{tY} - n_{t',Y+1}\right) \frac{S_{t',Y+1}^2}{n_{t',Y+1}}
\end{aligned}
\tag{34}
$$

where the $S_{t'y}^2 = S_{t'y'}^2$ $(y = y' = 2, 3, \ldots, Y+1)$ are as defined in (28). Note that the expression for $Var_D\left(\hat{V}_{t'(s)}\right)$ in (34) holds whether or not the model in (30) is correct.

Similarly, and given the results from times t and t', an (sampling design) unbiased estimator of $Var_D\left(\hat{V}_{t'(s)}\right)$ is given by

$$
V\hat{a}r_D\left(\hat{V}_{t'(s)}\right) = \sum_{y=2}^{Y} \left(\frac{f_y}{N_{t'y}}\right)^2 N_{t,y-1}\left(N_{t,y-1} - n_{t'y}\right) \frac{s_{t'y}^2}{n_{t'y}} + N_{ty}\left(N_{tY} - n_{t',Y+1}\right) \frac{s_{t',Y+1}^2}{n_{t',Y+1}} \tag{35}
$$

where the $s_{t'y}^2 = s_{t'y'}^2$ $(y = y' = 2, 3, \ldots, Y+1)$ are as defined in (29). Hence the reported estimated standard error for $\hat{V}_{t'(s)}$ is $\sqrt{V\hat{a}r_D\left(\hat{V}_{t'(s)}\right)}$.

Lemma 1 gives conditions under which the sampling error of $\hat{V}_{t'(s)}$ is less than the sampling error of $\hat{V}_{t'}$ (or more precisely, $\hat{V}_{t'2^+}$). For convenience and from (12), rewrite $Var\ (\hat{V}_{t'2^+})$ as

$$
Var\ (\hat{V}_{t'2^+}) = \frac{N_t(N_t - n_{t'})}{n_{t'}} S_{t'}^2
$$

where $S_{t'}^2$ is an overall variance for the 2 year and older trucks at time t'.

Lemma 1. Let $S^2_{t'y(\max)} = \max_y \left\{S^2_{t'y}\right\}$ and assume that $S^2_{t'y(\max)} \le S^2_{t'}$. Under the stratification approach, assume proportional allocation of the sample at time t', i.e.,

$$n_{t'y} = n_{t'} \frac{N_{t,y-1}}{N_t} \qquad \text{for } y = 2, 3, \ldots, Y+1.$$

If $f_y/N_{t'y} \approx 1$ for all y, then

$$Var_D\,(\hat{V}_{t'(s)}) \le Var\,(\hat{V}_{t'2^+}) \quad .$$

Proof: From (34) and assuming proportional allocation of the sample at time t', $Var_D\,(\hat{V}_{t'(s)})$ simplifies to

$$Var_D\,(\hat{V}_{t'(s)}) = \left\{\frac{N_t - n_{t'}}{n_{t'}}\right\}\left\{\sum_{y=2}^{Y}\left(\frac{f_y}{N_{t'y}}\right)^2 N_{t,y-1}\,S^2_{t'y} + N_{tY}\,S^2_{t',Y+1}\right\} \quad .$$

By definition of $S^2_{t'y(\max)}$, an upper bound on $Var_D\,(\hat{V}_{t'(s)})$ is

$$\begin{aligned} Var_D\,(\hat{V}_{t'(s)}) &\le \left\{\frac{N_t - n_{t'}}{n_{t'}}\right\}\left\{\sum_{y=2}^{Y}\left(\frac{f_y}{N_{t'y}}\right)^2 N_{t,y-1}\,S^2_{t'y(\max)} + N_{tY}\,S^2_{t'y(\max)}\right\} \\ &\approx \left\{\frac{N_t - n_{t'}}{n_{t'}}\right\}\left\{\sum_{y=2}^{Y} N_{t,y-1}\,S^2_{t'y(\max)} + N_{tY}\,S^2_{t'y(\max)}\right\} \qquad \text{if } \frac{f_y}{N_{t'y}} \approx 1 \\ &= \frac{N_t(N_t - n_{t'})}{n_{t'}}\,S^2_{t'y(\max)} \end{aligned}$$

Because $\qquad S^2_{t'y(\max)} \le S^2_{t'}$,

$$\frac{N_t(N_t - n_{t'})}{n_{t'}}\,S^2_{t'y(\max)} \le \frac{N_t(N_t - n_{t'})}{n_{t'}}\,S^2_{t'}$$

which implies $Var_D\ (\hat{V}_{t'(s)}) \leq Var\ (\hat{V}_{t'2^*})$ and the result follows.

Comment: Because $\hat{V}_{t'} = \hat{V}_{t'1} + \hat{V}_{t'2^*}$ by (21), Lemma 1 implies conditions when $\hat{V}_{t'(s)}$ is very likely to be a more precise estimator than $\hat{V}_{t'}$. If instead of using the proportional allocation assumed in Lemma 1 optimal allocation as obtained in Lemma 2 is used, the precision of $\hat{V}_{t'(s)}$ is even greater.

Sample Size and Allocation at Time t'.

At time t, the determination of optimal n_t and its optimal allocation to n_{t1}, n_{t2}, ..., n_{tY} can be obtained from the usual Neyman-Tschuprow Approach (Neyman 1934; Tschuprow 1923) by minimizing $Var_D\ (\hat{V}_{t(s)})$ subject to a cost constraint.

At time t' and conditional on the results from time t, Lemma 2 provides the details for a similar determination of optimal $n_{t'}$ and its optimal allocation to $n_{t'2}$, $n_{t'3}$, ..., $n_{t',Y+1}$.

Lemma 2. Let a cost function be given by $C = \sum_{y=2}^{Y+1} c_y\ n_{t'y}$ where C is fixed and c_h is the cost of a single observation in stratum y at time t'. Then $Var_D\ (\hat{V}_{t'(s)})$ is minimized subject to $C = \sum_{y=2}^{Y+1} c_y\ n_{t'y}$ when

$$(i) \quad n_{t'} = \frac{C \sum_{y=2}^{Y+1} \sqrt{\frac{d_y\ N_{t,\,y-1}}{c_y}}\ S_{t'y}}{\sum_{y=2}^{Y+1} \sqrt{c_y\ d_y\ N_{t,\,y-1}}\ S_{t'y}} \quad \text{, and} \tag{36}$$

$$(ii) \quad n_{t'y} = n_{t'} \left(\frac{\sqrt{\frac{d_y\ N_{t,\,y-1}}{c_y}}\ S_{t'y}}{\sum_{y=2}^{Y+1} \sqrt{\frac{d_y\ N_{t,\,y-1}}{c_y}}\ S_{t'y}} \right) \quad \text{for } y = 2, 3, ..., Y+1$$

$$\text{where} \quad d_y = \begin{cases} \left(\dfrac{f_y}{N_{t'y}}\right)^2 N_{t,\,y-1} & \text{for } y = 2, 3, \ldots, Y \\ N_{tY} & \text{for } y = Y+1 \end{cases} .$$

Proof: We want to find $n_{t'2}, n_{t'3}, \ldots, n_{t'Y}, n_{t',\,Y+1}$ (also $n_{t'}$) to minimize $Var_D\left(\hat{V}_{t'(s)}\right)$ subject to

$C = \sum_{y=2}^{Y+1} c_y\, n_{t'y}$. By the method of Lagrange multipliers, let

$$\begin{aligned} W\left(n_{t'2}, n_{t'3}, \ldots, n_{t',\,Y+1}\right) &= Var_D\left(\hat{V}_{t'(s)}\right) + \lambda\left(\sum_{y=2}^{Y+1} c_y\, n_{t'y} - C\right) \\ &= \sum_{y=2}^{Y}\left(\frac{f_y}{N_{t'y}}\right)^2 N_{t,\,y-1}\left(N_{t,y-1} - n_{t'y}\right)\frac{S^2_{t'y}}{n_{t'y}} + N_{tY}\left(N_{tY} - n_{t',\,Y+1}\right)\frac{S^2_{t',\,Y+1}}{n_{t',\,Y+1}} + \lambda\left(\sum_{y=2}^{Y+1} c_y\, n_{t'y} - C\right) \\ &= \sum_{y=2}^{Y+1}\frac{d_y\, N_{t,\,y-1}\, S^2_{t'y}}{n_{t'y}} - \sum_{y=2}^{Y+1} d_y\, S^2_{t'y} + \lambda\left(\sum_{y=2}^{Y+1} c_y\, n_{t'y} - C\right) . \end{aligned}$$

where $d_y = \left(f_y/N_{t'y}\right)^2 N_{t,\,y-1}$ for $y=2, 3, \ldots, Y$ and $d_{Y+1} = N_{tY}$.

Thus $$\frac{\partial W}{\partial n_{t'y}} = -\frac{d_y\, N_{t,\,y-1}\, S^2_{t'y}}{n^2_{t'y}} + \lambda\, c_y \quad \textit{for } y=2, 3, \ldots, Y+1 \tag{37}$$

and $$\frac{\partial W}{\partial \lambda} = \sum_{y=2}^{Y+1} c_y\, n_{t'y} - C .$$

Setting the partial derivatives in (37) equal to zero yields

$$\begin{aligned} n_{t'y} &= \sqrt{\frac{d_y\, N_{t,\,y-1}}{\lambda\, c_y}}\; S_{t'y} \quad \text{for } y=2, 3, \ldots, Y+1 \\ \text{and} \quad C &= \sum_{y=2}^{Y+1} c_y\, n_{t'y} , \end{aligned} \tag{38}$$

which becomes

$$c_y\, n_{t'y} = \sqrt{\frac{c_y\, d_y\, N_{t,\,y-1}}{\lambda}}\; S_{t'y} \quad \text{for } y=2, 3, \ldots, Y+1. \tag{39}$$

Summing the expressions in (39) gives

$$\sum_{y=2}^{Y+1} c_y \, n_{t'y} = \frac{1}{\sqrt{\lambda}} \sum_{y=2}^{Y+1} \sqrt{c_y \, d_y \, N_{t,\,y-1}} \; S_{t'y}$$

which is equivalent to

$$\sqrt{\lambda} = \frac{\sum_{y=2}^{Y+1} \sqrt{c_y \, d_y \, N_{t,\,y-1}} \; S_{t'y}}{C} \quad . \tag{40}$$

Substituting (40) into (38) leads to

$$n_{t'y} = \frac{C}{\sum_{y=2}^{Y+1} \sqrt{c_y \, d_y \, N_{t,\,y-1}} \; S_{t'y}} \sqrt{\frac{d_y \, N_{t,\,y-1}}{c_y}} \; S_{t'y} \quad \text{for } y=2,\ 3,\ \ldots,\ Y+1 \quad . \tag{41}$$

Summing the expressions in (41) gives optimal $n_{t'}$

$$n_{t'} = \sum_{y=2}^{Y+1} n_{t'y} = C \, \frac{\sum_{y=2}^{Y+1} \sqrt{\frac{d_y \, N_{t,\,y-1}}{c_y}} \; S_{t'y}}{\sum_{y=2}^{Y+1} \sqrt{c_y \, d_y \, N_{t,\,y-1}} \; S_{t'y}} \tag{42}$$

Hence from (41) and (42), the optimal $n_{t'y}$ are given by

$$n_{t'y} = n_{t'} \left(\frac{\sqrt{\frac{d_y \, N_{t,\,y-1}}{c_y}} \; S_{t'y}}{\sum_{y=2}^{Y+1} \sqrt{\frac{d_y \, N_{t,\,y-1}}{c_y}} \; S_{t'y}} \right) \quad \text{for } y=2,\ 3,\ \ldots,\ Y+1 \quad . \tag{43}$$

as was to be demonstrated.

4. CONCLUDING REMARKS

The objective of this paper has been to provide at time t' the estimate of a total (32) and to derive a single expression for its standard error, which follows from (34), when a stratum of a finite universe is not sampled at time t'. Prediction was used to compensate for not sampling this stratum. In this paper, the strata were determined by truck ages, and the unsampled stratum consisted of new universe trucks since time t. Other variables for stratification are possible as

long as the correlation between strata means at times t and t' is high. The method of Section 3 shows explicitly via (32), (34), and (35) how to handle births and deaths of universe units between times t and time t'. Subsampling at time t' is desirable because contact has been previously made with potential respondents at time t. Hence there can be a savings in terms of not having to establish new relations. (In a recent and related paper pointing to realized benefits if the current estimate(s) at occasion t' can be based on estimates and data from previous occasion(s) t where $t < t'$, Nandram and Sedransk (1993) consider estimators of means at occasion t' where the population unit values of the variable of interest are likely to differ relatively little from the previous occasions $t'-1$, $t'-2$, ...) A few specific remarks follow.

Remark 1. Stratifying by age at times t and t' and using the method of Section 3 ensures that n_{ty} and $n_{t'y}$ are nonzero. This is not guaranteed under the method of Section 2.

Remark 2. Because stratification is by age at times t and t' and given the linear model in (30) is used, the principal control that the experimenter has in decreasing $Var_D\left(\hat{V}_{t'(s)}\right)$ is to increase the sample sizes $n_{t'y}$. At time t', optimal $n_{t'y}$ for minimizing $Var_D\left(\hat{V}_{t'(s)}\right)$ subject to cost constraints are given in (36) where $S_{t'y}$ are estimated using prior information, very likely from time t. It is also seen from $f_y\,(y=2, \ldots, Y)$ where

$$f_y = N_{t'1}\left\{\frac{1}{Y-1} + \left(\hat{\bar{V}}_{t1} - \hat{\bar{\Delta}}_t\right)\frac{\left(\hat{\bar{V}}_{ty} - \hat{\bar{\Delta}}_t\right)}{\sum_{y=2}^{Y}\left(\hat{\bar{V}}_{ty} - \hat{\bar{\Delta}}_t\right)^2}\right\} + N_{t'y} \quad , \tag{44}$$

that $Var_D\left(\hat{V}_{t'(s)}\right)$ decreases as $\hat{\bar{V}}_{t1}$ approaches $\hat{\bar{\Delta}}_t$.

Remark 3. Age throughout this paper is in terms of years. There is nothing in the methodology development to prevent use of other time increments if available for age such as months, quarters of a year, etc.

Remark 4. Similar results to those of Section 3 can be obtained for cases $t' = t + z$ where z is an integer such that $1 < z \leq Y - 2$ where Y is the largest truck age at time t. (See the example of Subsection 2.5.)

Remark 5. Finally, throughout this entire paper, the sample at time t' is viewed as a subsample of an earlier sample. However, given some prior values $\hat{\bar{V}}_{t2}, \hat{\bar{V}}_{t3}, \ldots, \hat{\bar{V}}_{tY}$, there is nothing in the theory and development in Section 3 to prevent one from using the results for a first occasion sample at time t' ! With this view, one way to obtain the values $\hat{\bar{V}}_{t2}, \hat{\bar{V}}_{t3}, \ldots, \hat{\bar{V}}_{tY}$ is to sample prior to time t' as was the approach of this paper. Other methods for obtaining prior auxiliary values $\hat{\bar{V}}_{t2}, \hat{\bar{V}}_{t3}, \ldots, \hat{\bar{V}}_{tY}$ exist.

ACKNOWLEDGMENTS

The author is grateful to his colleagues Pat Hu and Shaw-Pin Miaou of the ORNL Center for Transportation Analysis for insightful discussions on the post-stratification model which served as partial motivation for the stratification model before sampling which is the chief focus of this paper. An earlier draft of this manuscript was greatly improved by many review comments from John Beauchamp, Mathematical Sciences Section and Shaw -Pin Miaou. Some of the computations of the example in subsection 2.5 are the results of earlier work by Stacy Davis and Jennifer Young, both also of the ORNL Center for Transportation Analysis. Special thanks are extended to William Bostic Jr. of the U.S. Bureau of the Census and Rolf Schmitt of the U.S. Department of Transportation who helped with the interpretation of the 1990 NTACS and its data limitations. Finally, special appreciation goes to Marjorie Harris of the Mathematical Sciences Section for her expertise, creativity in producing Figures 1-5, and patience which have all made possible the production of this paper in its present form. This work was supported by the Federal Highway Administration of the U.S. Department of Transportation and was completed during the summer of 1993 while the author was a research fellow at the U. S. Bureau of the Census..

REFERENCES

Bolfarine, H. and Zacks, S. (1992), *Prediction Theory for Finite Populations*, New York: Springer-Verlag.

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1977), *Foundations of Inference in Survey Sampling*, New York: John Wiley and Sons.

Cochran, W. G. (1977), *Sampling Techniques*, New York: John Wiley and Sons.

Draper, N. R. and Smith, H. (1981), *Applied Regression Analysis*, New York: John Wiley and Sons.

Ericksen, E. P. and Kadane, J. B. (1985), "Estimating the Population in a Census Year: 1980 and Beyond", *Journal of the American Statistical Association*, **80**, 98-131.

Freedman, D. A. and Navidi, W. C. (1986), "Regression Models for Adjusting the 1980 Census", *Statistical Science*, **1**, 3-39.

Hedayat, A. S. and Sinha, B. K. (1991), *Design and Inferences in Finite Population Sampling*, New York: John Wiley and Sons.

Hu, P., Wright, T., and Miaou, S. -P. (1990), "Nationwide Truck Activity and Commodity Survey: Methodologies for Controlling Statistical Biases Caused by (Three) - Year Lag Between 1987 and (1990)," Unpublished Oak Ridge National Laboratory Technical Memorandum, Oak Ridge, Tennessee 37831.

Little, R. J. A. (1991), "Inference with Survey Weights," *Journal of Official Statistics*, **7**, 405-424.

Little, R. J. A. and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*, New York: John Wiley and Sons.

Madow, W. G., Olkin, I., and Rubin, D. B. (1983), (Editors). *Incomplete Data in Sample Surveys, Vol 2, Theory and Bibliographies*, New York: Academic Press.

Nandram, B. and Sedransk, J. (1993), "Empirical Bayes Estimation for the Finite Population Mean on the Current Occasion," *Journal of the American Statistical Association*, **88**, 994-1000.

Neyman, J. (1934), "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, **97**, 558-606.

1987 Census of Transportation: Truck Inventory and Use Survey-United States, TC87 - T - 52, U. S. Dept. of Commerce, Bureau of the Census, Washington, D. C.

1990 Nationwide Truck Activity and Commodity Survey Selected Tabulations (1993), Oak Ridge National Laboratory Technical Report #12361, Oak Ridge, Tennessee 37831.

Platek, R., Rao, J. N. K., Särndal, C. E., and Singh, M. P. (1987), *Small Area Statistics*, New York: John Wiley and Sons.

Purcell, N. P. and Kish, L. (1979), "Estimation for Small Domains," *Biometrics*, 35, 365-384.

Royall, R. M. (1970), "On Finite Population Sampling Theory Under Certain Linear Regression Models," *Biometrika*, 57, 377-387.

Royall, R. M. (1988), "The Prediction Approach to Sampling Theory" in *Handbook to Statistics: Sampling, Vol. 6*, (P. R. Krishnaiah and C. R. Rao, Editors), Elsevier Science Publishers B. V., 399-413.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Schaible, W. L. (1983), "Estimation of Finite Population Totals from Incomplete Sample Data: Prediction Approach" in *Incomplete Data in Sample Surveys, Vol. 3, Proceedings of the Symposium*, New York: Academic Press.

Technical Documentation for the 1990 Nationwide Truck Activity and Commodity Survey Public Use File (1992), Oak Ridge National Laboratory Technical Report #12188, Oak Ridge, Tennessee 37831.

Tschuprow, A. A. (1923), "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations," *Metron, 2*, 461-493; 646-683.

Wright, R. L. (1983), "Finite Population Sampling with Multivariate Auxiliary Information," *Journal of the American Statistical Association*, 78, 879-884.

Wright, T. (1993), "CENSUSPLUS: A Sampling and Prediction Approach for the 2000 Census of the United States" (Draft), unpublished paper written while the author was a visitor at the U. S. Bureau of the Census during the summer of 1993.

Wright, T. and Tsao, H. (1983), "A Frame on Frames: An Annotated Bibliography" in *Statistical Methods and the Improvement of Data Quality*, (T. Wright, Editor), New York: Academic Press, 25-72.

Wright, T. and Tsao, H. (1985), "Some Useful Notes on Simple Random Sampling," *The Journal of Quality Technology*, 17, 67-73.

Zieschang, K. D. (1990), "Sampling Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey," *Journal of the American Statistical Association*, 85, 986-1001.

INTERNAL DISTRIBUTION

1. B. R. Appleton
2. C. Bayne
3. J. J. Beauchamp
4. K. O. Bowman
5. C. C. Brandt
6. M. S. Bronzini
7. T. Darland
8. S. C. Davis
9. D. J. Downing
10. D. M. Flanagan
11. E. L. Frome
12. D. L. Greene
13. T. L. Hebble
14. P. Hu
15. E. Leach
16. W. E. Lever
17. S. -P. Miaou
18. M. D. Morris
19. C. E. Oliver
20. G. Ostrouchov

21-25. S. Raby

26. R. L. Schmoyer

27-31. R. F. Sincovec

32-36. R. C. Ward

37-41. T. Wright

42. J. Young
43. Central Research Library
44. K-25 Applied Tech. Library
45. ORNL Patent Office
46. Y-12 Technical Library

47-48. Laboratory Records Department

49. Laboratory Records Dept. - RC

EXTERNAL DISTRIBUTION

50. Ms. Maxine Anderson-Brown, Program Manager, DIR, Room 2270-3, U.S. Bureau of the Census, Washington, D. C. 20233.

51. Mr. William Bostic, Jr., Business Division, U.S. Bureau of the Census, Washington, D. C. 20233.

52. Professor Roger W. Brockett (EPMD Advisory Committee), Wang Professor of Electrical Engineering and Computer Science, Division of Applied Sciences, Harvard University, Cambridge, MA 02138.

53. Professor Donald J. Dudziak, Department of Nuclear Engineering Physics, North Carolina State University, Raleigh, North Carolina 27695-7909.

54. Dr. Robert E. Fay, Senior Mathematical Statistician, U.S. Bureau of the Census, Washington, D. C. 20233-4001.

55. Mr. Charlie Goodman, Federal Highway Administration, Office of Policy Development, HPP-12, 400 7th Street, S.W., Washington, D. C. 20590.

56. Dr. James E. Leiss (EPMD Advisory Committee), Route 2, Box 142C, Broadway, VA 22815.

57. Professor Neville Moray (EPMD Advisory Committee), Department of Mechanical and Industrial Engineering, University of Illinois, 1206 West Green Street, Urbana, IL 61801.

58. Dr. David Nelson, Scientific Computing Staff, Applied Mathematical Sciences, Office of Energy Research, U.S. Department of Energy, Washington, D. C. 20585.

59. Dr. Richard M. Royall, Biostatistics Department, Johns Hopkins University, 615 N. Wolfe, Baltimore, MD 21205.

60. Dr. Jagdiah S. Rustagi, IBM Research Center, 605/503, 5600 Cattle Road, San Jose, CA 95193.

61. Dr. Jerome Sacks, NISS, P.O. Box 14162, Research Triangle Park, North Carolina 27709-4162.

62. Mr. Wesley L. Schaible, Research and Evaluation Department, Bureau of Labor Statistics, 441 G. Street, N.W., Room 2126, Washington, D. C. 20210.

63. Dr. Rolf Schmitt, Bureau of Transportation Statistics, U.S. Department of Transportation, 400 7th Street, S.W., Washington, D. C. 20590.

64. Dr. L. R. Shenton, Office of Computing and Information Services, Boyd Graduate Studies Building, University of Georgia, Athens, Georgia 30602.

65. Mr. John Thompson, Decennial Statistical Studies Division, U.S. Bureau of the Census, Washington, D. C. 20233.

66. Dr. Robert D. Tortora, Office of the Director, U.S. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. 20233-0001.

67. Dr. How J. Tsao, Mgmt. Services Division, Eastman Kodak, Rochester, NY 14652-3302.

68. Mr. Preston J. Waite, Demographic Statistical Methods Division, U.S. Bureau of the Census, Washington, D. C. 20233.

69. Dr. Ray A. Waller, S-1, Statistics, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, New Mexico 87545.

70. Professor Mary F. Wheeler (EPMD Advisory Committee), Rice University, Department of Mathematical Sciences, P.O. Box 1892, Houston, TX 77251.

71. Professor Don Ylvisaker, Department of Mathematics, University of California, Los Angeles, California 90024.

72. Dr. Shelemyahu Zacks, Mathematical Sciences Department, State University of New York, Binghamton Center, Binghamton, NY 13902-6000.

73. Professor A. Zaslavsky, Department of Statistics, Harvard University's Science Center, 1 Oxford Street, Cambridge, Massachusetts 02138.

73. Office of Assistant Manager for Energy Research and Development, U.S. Department of Energy, Oak Ridge Operations Office, P.O. Box 2001, Oak Ridge, Tennessee 37831-8600.

74-75. Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, Tennessee 37831-0062.