Contents

1 Introduction

Every summer since 1996, the Swedish National Road Administration (SNRA) conducts a traffic survey on urban roads. The roads are thought of as partitioned into one-meter road sites, that are the population elements. Data are collected for a random sample of sites (during selected twentyfour-hour periods) by use of a measurement equipment installed on the road. The principal aim of the survey is to estimate the average speed, R, on the roads.

A three-stage sampling design, with stratification in each stage, is employed to select sites for observation. We are interested in evaluating the current allocation of the total sample over sampling stages. Our method of doing this is to estimate the components of the total variance of the estimator of R arising from each sampling stage, and analyze their relative sizes. As means for a possible re-allocation of the sample, we also present formulas for optimal sampling sizes.

At present, in all but the first sampling stage, only one sampling unit per stratum is selected. Since units are drawn with replacement in stage one, the total variances of the estimators can still be estimated. The variance contributions from each sampling stage are however inseparable. We circumvent this problem by making use of a fictitious sampling design and some experimental data. In this way, the demanded variance component estimates are calculated for a domain of study.

Throughout this report, possible nonsampling errors, which may bias and increase the variance of the survey estimators, are ignored. Impact of sampling frame errors is investigated in [4]; impact of missing data in [5].

2 Variables and parameters

The main study variables are the traffic flow, y, and the travel time, z. For a given road site and time period, the traffic flow is the number of passing vehicles, and the travel time is the total time the vehicles take to pass the site. Let U denote the target population "in space" – the set of one-meter road sites that make up the urban roads – and U_Υ the target population "in time" – the set of twentyfour-hour periods that make up the time period of study. The population total of y is given by $\sum_{U_\Upsilon} \sum_U y_k^\upsilon$, where y_k^υ equals the traffic flow in site $k \in U$ during twentyfour-hour period $\upsilon \in U_\Upsilon$. Correspondingly, the z total is given by $\sum_{U_\Upsilon} \sum_U z_k^\upsilon$. Since the total vehicle mileage is a measure of distance, and the total travel time a measure of time, their ratio is a measure of speed.

In this report, we ignore possible time variability in y and z. That is, we consider only the special case when $y_k^\upsilon = y_k$ and $z_k^\upsilon = z_k$ for all $\upsilon \in U_\Upsilon, k \in U$. Hence, we will hereafter drop the time index and refer to $t_y = \sum_U y_k$ as total vehicle mileage and $t_z = \sum_U z_k$ as total travel time. The main survey goal is to estimate the ratio $R = t_y/t_z$; the average speed on the roads.

3 Sampling design

Road sites are selected for observation by means of a three-stage sampling design. A brief description, based on [4, Ch. 2], of the different stages, will now be given.

The primary sampling units (PSUs) are the N_I population centers in Sweden, labeled $i = 1, ..., N_I$. The ith PSU is represented by its label i. Thus, we denote the set of PSUs as $U_I = \{1, ..., i, ..., N_I\}$. Population center $i \in U_I$ is partitioned into N_{IIi} small areas, labeled $q = 1, ..., N_{IIi}$, that represent the secondary sampling units (SSUs). The set of SSUs formed by the subdivision of i is denoted $U_{IIi} = \{1, ..., q, ..., N_{IIi}\}$. Finally, the roads in small area q in population center i are viewed as partitioned into N_{iq} one-meter road sites (representing the tertiary sampling units – the TSUs). This set of sites is denoted U_{iq}.

The sample s of road sites is selected from the population U of urban roads in the following way.

Stage I A probability-proportional-to-size (pps) sample of PSUs is drawn with probability proportional to the number of inhabitants. At every draw, p_i is the probability of selecting the ith PSU. Let i_ν denote the PSU selected in the νth draw, $\nu = 1, ..., m_I$, where m_I is the number of draws. The probability of selecting i_ν is denoted p_{i_ν}. If the ith PSU is selected in the νth draw, then $p_{i_\nu} = p_i$. The vector of selected PSUs, $(i_1, ..., i_\nu, ..., i_{m_I})$, is the resulting ordered sample os_I.

Stage II For every i_ν that is a component of os_I, a simple random (SI*)* sample s_{IIi_ν} of SSUs of size n_{IIi_ν} is selected.

Stage III An SI sample $s_{i_\nu q}$ of sites of size $n_{i_\nu q}$ is drawn for every small area $q \in s_{IIi_\nu}$.

To simplify, both in the above description and further, we basically ignore some features of the design. In particular, we ignore that *stratified* sampling is used in each stage (a few details on the stratification are still provided in Section 5.3). Moreover, we ignore that the three largest PSUs define a take-all stratum in stage I; and finally, that SSUs are selected with pps within one stratum (Residential Areas) in stage II.

In practice, within stratum, the sample sizes in each stage are $m_I = 10$, $n_{IIi_\nu} = 1$ and $n_{i_\nu q} = 1$.

4 The survey estimators

In this section, we present an estimator $\hat{t}_a$ of the population total

$$t_a = \sum\nolimits_U a_k = \sum_{i=1}^{N_I} t_{ai} = \sum_{i=1}^{N_I} \sum\nolimits_{U_{IIi}} t_{aiq}$$

where a_k is the value of an arbitrary study variable a for road site $k \in U$, $t_{ai} = \sum_{U_{IIi}} t_{aiq}$ and $t_{aiq} = \sum_{U_{iq}} a_k$. We further introduce an estimator $\hat{R}$ of the ratio $R = t_y/t_z$. The variances of $\hat{t}_a$ and $\hat{R}$, and the components of those variances due to each sampling stage, are also given.

In the speed survey, PSUs are selected *with* replacement and SSUs and TSUs *without* replacement. In order to construct estimators which are unbiased with respect to all three sampling stages, the "p-expanded with replacement" estimation principle (first used by Hansen and Hurwitz [2]; treated for instance in [7, Section 2.9]), and the Horvitz-Thompson estimation principle (usually ascribed to Horvitz and Thompson [3]; treated extensively in [7]) are combined. Throughout this report, estimators of population entities are denoted by a hat, and the subscripts 'pwr' and 'π' used to indicate "p-expanded with replacement" estimators and Horvitz-Thompson estimators, respectively.

4.1 The estimator $\hat{t}_a$ of t_a

From [7, Result 4.5.1], an unbiased estimator of t_a is given by

$$\hat{t}_a = \frac{1}{m_I} \sum_{\nu=1}^{m_I} \frac{\hat{t}_{\pi a i_\nu}}{p_{i_\nu}} \tag{1}$$

where $\hat{t}_{\pi a i_\nu} = (N_{IIi_\nu}/n_{IIi_\nu}) \sum_{s_{IIi_\nu}} \hat{t}_{\pi a i_\nu q}$ and $\hat{t}_{\pi a i_\nu q} = (N_{i_\nu q}/n_{i_\nu q}) \sum_{s_{i_\nu q}} a_k$. (If $i \in U_I$ was selected in the νth draw, then $\hat{t}_{\pi a i_\nu} = \hat{t}_{\pi a i}$ and $\hat{t}_{\pi a i_\nu q} = \hat{t}_{\pi a i q}$.)

The variance of $\hat{t}_a$ is given by

$$V\left(\hat{t}_a\right) = \frac{1}{m_I} \sum_{i=1}^{N_I} p_i \left(\frac{t_{ai}}{p_i} - t_a \right)^2 + \frac{1}{m_I} \sum_{i=1}^{N_I} \frac{V_{ai}}{p_i} \tag{2}$$

where V_{ai} is the variance of $\hat{t}_{\pi a i}$ with respect to the last two sampling stages:

$$V_{ai} = V_{aIIi} + \frac{N_{IIi}}{n_{IIi}} \sum_{U_{IIi}} V_{aiq} \tag{3}$$

where

$$\begin{aligned} V_{aIIi} &= N_{IIi}^2 \frac{1 - f_{IIi}}{n_{IIi}} S_{t_a U_i}^2; \quad f_{IIi} = n_{IIi}/N_{IIi}; \\ S_{t_a U_i}^2 &= \sum_{U_{IIi}} \left(t_{aiq} - t_{ai}/N_{IIi} \right)^2 / \left(N_{IIi} - 1 \right) \end{aligned}$$

for $i \in U_I$, and

$$\begin{aligned} V_{aiq} &= N_{iq}^2 \frac{1 - f_{iq}}{n_{iq}} S_{aU_{iq}}^2; \quad f_{iq} = n_{iq}/N_{iq}; \\ S_{aU_{iq}}^2 &= \sum_{U_{iq}} \left(a_k - t_{aiq}/N_{iq} \right)^2 / \left(N_{iq} - 1 \right) \end{aligned}$$

for $q \in U_{IIi}; i \in U_I$.

The variance $V\left(\hat{t}_a\right)$ can equivalently be written as the sum of three components, mirroring the variation arising from each sampling stage:

$$V\left(\hat{t}_a\right) = V_{\mathrm{PSU}}\left(\hat{t}_a\right) + V_{\mathrm{SSU}}\left(\hat{t}_a\right) + V_{\mathrm{TSU}}\left(\hat{t}_a\right) \tag{4}$$

where

$$V_{\mathrm{TSU}}\left(\hat{t}_a\right) = \frac{1}{m_I} \sum_{i=1}^{N_I} \frac{1}{p_i} \frac{N_{IIi}}{n_{IIi}} \sum_{U_{IIi}} V_{aiq}, \tag{5}$$

$$V_{\mathrm{SSU}}\left(\hat{t}_a\right) = \frac{1}{m_I} \sum_{i=1}^{N_I} \frac{V_{aIIi}}{p_i} \tag{6}$$

and

$$V_{\mathrm{PSU}}\left(\hat{t}_a\right) = \frac{1}{m_I} \sum_{i=1}^{N_I} p_i \left(\frac{t_{ai}}{p_i} - t_a \right)^2. \tag{7}$$

4.2 The estimator $\hat{R}$ of R

From [6, Section 6.8.2], an approximately unbiased estimator of R is given by

$$\hat{R} = \frac{\hat{t}_y}{\hat{t}_z} = \frac{\sum_{\nu=1}^{m_I}\left(\hat{t}_{\pi y i_\nu}/p_{i_\nu}\right)}{\sum_{\nu=1}^{m_I}\left(\hat{t}_{\pi z i_\nu}/p_{i_\nu}\right)}. \tag{8}$$

Define the new study variable $E = y - Rz$. Using Taylor linearization, the approximate variance $AV\left(\hat{R}\right)$ of $\hat{R}$ is given by

$$\begin{aligned} AV\left(\hat{R}\right) &= \frac{1}{t_z^2} V\left(\hat{t}_E\right) \\ &= \frac{1}{t_z^2}\left[\frac{1}{m_I}\sum_{i=1}^{N_I} p_i\left(\frac{t_{Ei}}{p_i} - t_E\right)^2 + \frac{1}{m_I}\sum_{i=1}^{N_I}\frac{V_{Ei}}{p_i}\right] \\ &= \frac{1}{t_z^2}\left(\frac{1}{m_I}\sum_{i=1}^{N_I}\frac{t_{Ei}^2}{p_i} + \frac{1}{m_I}\sum_{i=1}^{N_I}\frac{V_{Ei}}{p_i}\right) \end{aligned} \tag{9}$$

where the last equality holds since $t_E = 0$. The variance V_{Ei} is obtained from Equation (3) by letting the variable a equal E.

The approximate variance of $\hat{R}$ can be decomposed into

$$\begin{aligned} AV\left(\hat{R}\right) &= AV_{\text{PSU}}\left(\hat{R}\right) + AV_{\text{SSU}}\left(\hat{R}\right) + AV_{\text{TSU}}\left(\hat{R}\right) \\ &= \frac{V_{\text{TSU}}\left(\hat{t}_E\right)}{t_z^2} + \frac{V_{\text{SSU}}\left(\hat{t}_E\right)}{t_z^2} + \frac{V_{\text{PSU}}\left(\hat{t}_E\right)}{t_z^2} \end{aligned} \tag{10}$$

where $V_{\text{TSU}}\left(\hat{t}_E\right)$, $V_{\text{SSU}}\left(\hat{t}_E\right)$ and $V_{\text{PSU}}\left(\hat{t}_E\right)$ are obtained from Equations (5)-(7) by (again) letting a equal E.

5 Estimation of variance components

In this section, the problem of estimating sampling stage variance components for the speed survey is treated. By way of introduction, we present the estimators which had been applicable if the sampling sizes had exceeded one in each sampling stage. We continue by treating a situation where the sample sizes exceed one in the first and third sampling stages, but are equal to one in stage two. Under these circumstances, not all components can be estimated (more precisely, the variance contributions from the first and second sampling stages can not be separated). We show how to make use of a fictitious design to enable estimation of all components. These formulas are finally used to calculate variance component estimates from a set of experimental data.

5.1 At least two observations in each sampling stage

Assume that two or more units had been selected from each stratum in each sampling stage in the speed survey. For this situation, estimators of the components of $V\left(\hat{t}_a\right)$ and $AV\left(\hat{R}\right)$ are available. Although in reality we do not face this favorable situation, an investigation of the estimators that ideally could have been used still serves as natural starting-point for our work.

5.1.1 Estimation of the components of $V\left(\hat{t}_a\right)$

In order to estimate the variance $V\left(\hat{t}_a\right)$ of $\hat{t}_a$, it is not necessary to estimate each of its components separately. From [7, Result 4.5.1], $V\left(\hat{t}_a\right)$ is unbiasedly estimated by

$$\hat{V}_{3\text{st}}\left(\hat{t}_a\right) = \frac{1}{m_I\left(m_I - 1\right)} \sum_{\nu=1}^{m_I} \left(\frac{\hat{t}_{\pi a i_\nu}}{p_{i_\nu}} - \hat{t}_a\right)^2 . \tag{11}$$

The computationally simple formula is thanks to the fact that sampling with replacement is used at the first sampling stage. We are however interested in estimating each variance component separately. By slight modification of [7, Result 4.4.3], unbiased estimators of the variance components $V_{\text{TSU}}\left(\hat{t}_a\right)$, $V_{\text{SSU}}\left(\hat{t}_a\right)$ and $V_{\text{PSU}}\left(\hat{t}_a\right)$ are given, respectively, by

$$\hat{V}_{\text{TSU}}\left(\hat{t}_a\right) = \frac{1}{m_I^2} \sum_{\nu=1}^{m_I} \frac{1}{p_{i_\nu}^2} \left(\frac{N_{IIi_\nu}}{n_{IIi_\nu}}\right)^2 \sum\nolimits_{s_{IIi_\nu}} \hat{V}_{ai_\nu q} \tag{12}$$

where

$$\begin{aligned} \hat{V}_{ai_\nu q} &= N_{i_\nu q}^2 \frac{1 - f_{i_\nu q}}{n_{i_\nu q}} S_{as_{i_\nu q}}^2; \quad f_{i_\nu q} = n_{i_\nu q}/N_{i_\nu q}; \\ S_{as_{i_\nu q}}^2 &= \frac{1}{n_{i_\nu q} - 1} \sum\nolimits_{s_{i_\nu q}} \left(a_k - \frac{t_{ai_\nu q}}{n_{i_\nu q}}\right)^2 \end{aligned}$$

for $q \in i_\nu$ and every i_ν that is a component of os_I,

$$\hat{V}_{\text{SSU}}\left(\hat{t}_a\right) = \frac{1}{m_I^2}\sum_{\nu=1}^{m_I}\frac{\hat{V}_{ai_\nu}}{p_{i_\nu}^2} - \hat{V}_{\text{TSU}}\left(\hat{t}_a\right) \tag{13}$$

where

$$\begin{aligned}
\hat{V}_{ai_\nu} &= N_{IIi_\nu}^2\frac{1-f_{IIi_\nu}}{n_{IIi_\nu}}S_{\hat{t}_a s_{IIi_\nu}}^2 + \frac{N_{IIi_\nu}}{n_{IIi_\nu}}\sum\nolimits_{s_{IIi_\nu}}\hat{V}_{ai_\nu q};\\
f_{IIi_\nu} &= n_{IIi_\nu}/N_{IIi_\nu};\\
S_{\hat{t}_a s_{IIi_\nu}}^2 &= \frac{1}{n_{IIi_\nu}-1}\sum\nolimits_{s_{IIi_\nu}}\left(\hat{t}_{\pi a i_\nu q} - \frac{\hat{t}_{\pi a i_\nu}}{n_{IIi_\nu}}\right)^2
\end{aligned}$$

for every i_ν that is a component of os_I, and

$$\hat{V}_{\text{PSU}}\left(\hat{t}_a\right) = \hat{V}_{\text{3st}}\left(\hat{t}_a\right) - \hat{V}_{\text{SSU}}\left(\hat{t}_a\right) - \hat{V}_{\text{TSU}}\left(\hat{t}_a\right). \tag{14}$$

5.1.2 Estimation of the components of $AV\left(\hat{R}\right)$

Define a new variable $e = y - \hat{R}z$. From [6, Section 6.8.2], an estimator of the variance $AV\left(\hat{R}\right)$ of $\hat{R}$ is given by

$$\begin{aligned}
\hat{V}_{\text{3st}}\left(\hat{R}\right) &= \frac{1}{\hat{t}_z^2}\hat{V}_{\text{3st}}\left(\hat{t}_e\right)\\
&= \frac{1}{\hat{t}_z^2}\frac{1}{m_I\left(m_I-1\right)}\sum_{\nu=1}^{m_I}\left(\frac{\hat{t}_{\pi e i_\nu}}{p_{i_\nu}} - \hat{t}_e\right)^2\\
&= \frac{1}{\hat{t}_z^2}\frac{1}{m_I\left(m_I-1\right)}\sum_{\nu=1}^{m_I}\left(\frac{\hat{t}_{\pi y i_\nu}}{p_{i_\nu}} - \hat{R}\frac{\hat{t}_{\pi z i_\nu}}{p_{i_\nu}}\right)^2 \qquad (15)
\end{aligned}$$

where the last equality holds since $\hat{t}_e = 0$.

Estimators of the variance components $AV_{\text{TSU}}\left(\hat{R}\right)$, $AV_{\text{SSU}}\left(\hat{R}\right)$ and $AV_{\text{PSU}}\left(\hat{R}\right)$ are given, respectively, by

$$\hat{V}_{\text{TSU}}\left(\hat{R}\right) = \frac{\hat{V}_{\text{TSU}}\left(\hat{t}_e\right)}{\hat{t}_z^2};\quad \hat{V}_{\text{SSU}}\left(\hat{R}\right) = \frac{\hat{V}_{\text{SSU}}\left(\hat{t}_e\right)}{\hat{t}_z^2};\quad \hat{V}_{\text{PSU}}\left(\hat{R}\right) = \frac{\hat{V}_{\text{PSU}}\left(\hat{t}_e\right)}{\hat{t}_z^2} \tag{16}$$

where $\hat{V}_{\text{TSU}}\left(\hat{t}_e\right)$, $\hat{V}_{\text{SSU}}\left(\hat{t}_e\right)$ and $\hat{V}_{\text{PSU}}\left(\hat{t}_e\right)$ are obtained from Equations (12)-(14) by letting $a = e$.

5.2 One observation in the second stage

We now turn to a design with greater likeness to the real speed survey design than the one dealt with in Section 5.1. It is still assumed that two or more units are selected from each stratum in the first and third sampling stages; the sample size is however now equal to one within stratum in the second stage.

A sample of size one in stage two does not prevent us from estimating the total variance of $\hat{t}_a$, or the last-stage component $V_{\text{TSU}}\left(\hat{t}_a\right)$, as in Section 5.1. Hence, we are also still able to estimate $V_{\text{PSU}}\left(\hat{t}_a\right) + V_{\text{SSU}}\left(\hat{t}_a\right)$. The small sample size does however preclude us from estimating $V_{\text{PSU}}\left(\hat{t}_a\right)$ and $V_{\text{SSU}}\left(\hat{t}_a\right)$ separately. The corresponding estimation problem holds for $\hat{R}$. We choose to tackle the problem as follows. First we note that for $V_{\text{SSU}}\left(\hat{t}_a\right) \neq 0$,

$$V_{\text{PSU}}\left(\hat{t}_a\right) + V_{\text{SSU}}\left(\hat{t}_a\right) = V_{\text{SSU}}\left(\hat{t}_a\right)\left(C\left(\hat{t}_a\right) + 1\right) \tag{17}$$

where $C\left(\hat{t}_a\right) = V_{\text{PSU}}\left(\hat{t}_a\right)/V_{\text{SSU}}\left(\hat{t}_a\right)$. In the same manner, for $AV_{\text{SSU}}\left(\hat{R}\right) \neq 0$,

$$AV_{\text{PSU}}\left(\hat{R}\right) + AV_{\text{SSU}}\left(\hat{R}\right) = AV_{\text{SSU}}\left(\hat{R}\right)\left(C\left(\hat{R}\right) + 1\right) \tag{18}$$

where $C\left(\hat{R}\right) = AV_{\text{PSU}}\left(\hat{R}\right)/AV_{\text{SSU}}\left(\hat{R}\right)$. Next, we formulate a fictitious sampling design, formulated so as to fulfil the criteria

i. closely related to the one actually in use, and

ii. admitting separate estimation of each sampling stage component.

Finally, we derive estimators of (the closest equivalents to) the ratios $C\left(\hat{t}_a\right)$ and $C\left(\hat{R}\right)$ under the fictitious design, and use those to estimate $V_{\text{PSU}}\left(\hat{t}_a\right)$ and $V_{\text{SSU}}\left(\hat{t}_a\right)$, $AV_{\text{PSU}}\left(\hat{R}\right)$ and $AV_{\text{SSU}}\left(\hat{R}\right)$, separately.

5.2.1 Formulation of a fictitious sampling design

Under our fictitious design, PSUs and TSUs are selected *without* replacement; SSUs *with* replacement, as follows.

Stage I' First, the ordered sample os_I is drawn as in stage I in Section 3. The set of distinct PSUs which occur at least twice in os_I then make up the stage I' set-sample $s_{I'}$ of PSUs of size $n_{I'}$.

Stage II' For every $i \in s_{I'}$, a sample of SSUs is drawn with simple random sampling with replacement (SIR). At every draw, $p_q = 1/N_{IIi}$ is the probability of selecting the qth SSU. Let $q_{\nu'}$ denote the SSU selected in the ν'th draw, $\nu' = 1, ..., m_{II'i}$, where m_{IIi} is the number of draws. The probability of selecting $q_{\nu'}$ is denoted $p_{q_{\nu'}}$. If the qth SSU is selected in the ν'th draw, then $p_{q_{\nu'}} = p_q$. The vector of selected SSUs, $\left(q_{1'}, ..., q_{\nu'}, ..., q_{m_{II'i}}\right)$, is the resulting ordered sample $os_{II'i}$.

Stage III' For every $q_{\nu'}$ that is a component of $os_{II'i}$, an SI sample $s_{iq_{\nu'}}$ of TSUs of size $n_{iq_{\nu'}}$ is selected.

Note the resemblance to the actual design in Section 3. What we have done here is to transform the ordered sample in stage I into a set sample, and 'move' the with-replacement sampling one step down the stage hierarchy from the first to the second sampling stage. The main advantage of this procedure is that we gain access to more than one SSU drawing.

The sampling method specified for the first stage is not of standard type. Hence, for estimation purposes, the relevant first and second order inclusion probabilities $\pi_{I'i}$ and $\pi_{I'ij}$ need to be derived. As shown in Appendix A, the probability $\pi_{I'i}$ that PSU i will be included in $s_{I'}; i \in U_I$, is given by

$$\pi_{I'i} = 1 - (1 - p_i)^{m_I} \left(1 + m_I \frac{p_i}{1 - p_i}\right) \tag{19}$$

and the probability $\pi_{I'ij}$ that both PSU i and j will be included in $s_{I'}; i, j \in U_I$, by

$$\begin{aligned} \pi_{I'ij} &= 1 - (1 - p_i)^{m_I} \left(1 + m_I \frac{p_i}{1 - p_i}\right) - (1 - p_j)^{m_I} \left(1 + m_I \frac{p_j}{1 - p_j}\right) \\ &\quad - (1 - p_i - p_j)^{m_I - 1} m_I \left[p_i + p_j + (m_I - 1) \frac{p_i p_j}{1 - p_i - p_j}\right] \end{aligned} \tag{20}$$

5.2.2 Use of the fictitious design

For the sake of completeness, we start by presenting the estimators of t_a and R and their variances under the fictitious design, and continue by giving the estimators of the variances and their sampling stage components. The impatient reader is encouraged to proceed directly to subsection 'Estimation of the components of $V\left(\hat{t}_a\right)$ and $AV\left(\hat{R}\right)$', where our proposal for estimation of the components of $V\left(\hat{t}_a\right)$ and $AV\left(\hat{R}\right)$ by help of the fictitious design is summarized.

The estimator $\hat{t}_a'$ of t_a Under the fictitious design, from [7, Result 4.4.1], an unbiased estimator of t_a is given by

$$\hat{t}_a' = \sum\nolimits_{s_{I'}} \frac{\hat{t}_{\mathrm{pwr}ai}'}{\pi_{I'i}} \tag{21}$$

where $\hat{t}_{\mathrm{pwr}ai}' = (N_{IIi}/m_{II'i}) \sum_{\nu'=1}^{m_{II'i}} \hat{t}_{\pi aiq_{\nu'}}'$, $\hat{t}_{\pi aiq_{\nu'}}' = \left(N_{iq_{\nu'}}/n_{iq_{\nu'}}\right) \sum_{s_{iq_{\nu'}}} a_k$, and $N_{iq_{\nu'}}$ is the number of one-meter road sites in small area $q_{\nu'}$. (If the qth SSU was selected in the ν'th draw, then $\hat{t}_{\pi aiq_{\nu'}}' = \hat{t}_{\pi aiq} = (N_{iq}/n_{iq}) \sum_{s_{iq}} a_k$.)

The variance of $\hat{t}_a'$ is given by

$$V'\left(\hat{t}_a'\right) = \sum\sum\nolimits_{U_I} \Delta_{I'ij} \frac{t_{ai}}{\pi_{I'i}} \frac{t_{aj}}{\pi_{I'j}} + \sum\nolimits_{U_I} \frac{V_{ai}'}{\pi_{I'i}} \tag{22}$$

where $\Delta_{I'ij} = \pi_{I'ij} - \pi_{I'i}\pi_{I'j}$, V_{ai}' is the variance of $\hat{t}_{\mathrm{pwr}ai}'$ with respect to the last two sampling stages:

$$V_{ai}' = V_{aIIi}' + \frac{N_{IIi}}{m_{II'i}} \sum_{q=1}^{N_{IIi}} V_{aiq} \tag{23}$$

where

$$V_{aIIi}' = \frac{N_{IIi}\left(N_{IIi} - 1\right)}{m_{II'i}} S_{t_a U_i}^2 .$$

Equivalently, the variance $V'\left(\hat{t}_a'\right)$ can be written as

$$V'\left(\hat{t}_a'\right) = V_{\mathrm{PSU}}'\left(\hat{t}_a'\right) + V_{\mathrm{SSU}}'\left(\hat{t}_a'\right) + V_{\mathrm{TSU}}'\left(\hat{t}_a'\right) \tag{24}$$

where

$$V_{\mathrm{TSU}}'\left(\hat{t}_a'\right) = \sum\nolimits_{U_I} \frac{1}{\pi_{I'i}} \frac{N_{IIi}}{m_{II'i}} \sum_{q=1}^{N_{IIi}} V_{aiq}, \tag{25}$$

$$V_{\mathrm{SSU}}'\left(\hat{t}_a'\right) = \sum\nolimits_{U_I} \frac{V_{aIIi}'}{\pi_{I'i}}, \tag{26}$$

and

$$V_{\mathrm{PSU}}'\left(\hat{t}_a'\right) = \sum\sum\nolimits_{U_I} \Delta_{I'ij} \frac{t_{ai}}{\pi_{I'i}} \frac{t_{aj}}{\pi_{I'j}}. \tag{27}$$

The estimator $\hat{R}'$ of R Under the fictitious design, from [7, Result 5.6.2], an approximately unbiased estimator of R is given by

$$\hat{R}' = \frac{\hat{t}'_y}{\hat{t}'_z} = \frac{\sum_{s_{I'}} \frac{\hat{t}'_{\text{pwr}yi}}{\pi_{I'i}}}{\sum_{s_{I'}} \frac{\hat{t}'_{\text{pwr}zi}}{\pi_{I'i}}}. \tag{28}$$

The estimator $\hat{R}'$ has the approximate (Taylor) variance

$$\begin{aligned} AV'\left(\hat{R}'\right) &= \frac{1}{t_z^2} V'\left(\hat{t}'_E\right) \\ &= \frac{1}{t_z^2}\left(\sum\sum\nolimits_{U_I} \Delta_{I'ij} \frac{t_{Ei}}{\pi_{I'i}} \frac{t_{Ej}}{\pi_{I'j}} + \sum\nolimits_{U_I} \frac{V'_{Ei}}{\pi_{I'i}}\right) \\ &= \frac{1}{t_z^2}\left(\sum\sum\nolimits_{U_I} \Delta_{I'ij} \frac{t_{yi} - Rt_{zi}}{\pi_{I'i}} \frac{t_{yj} - Rt_{zj}}{\pi_{I'j}} + \sum\nolimits_{U_I} \frac{V'_{Ei}}{\pi_{I'i}}\right) \end{aligned} \tag{29}$$

where V'_{Ei} is obtained from Equation (23) by letting the variable a equal E.

The approximate variance $AV'\left(\hat{R}'\right)$ can equivalently be written as

$$AV'\left(\hat{R}'\right) = AV'_{\text{PSU}}\left(\hat{R}'\right) + AV'_{\text{SSU}}\left(\hat{R}'\right) + AV'_{\text{TSU}}\left(\hat{R}'\right) \tag{30}$$

$$= \frac{V'_{\text{PSU}}\left(\hat{t}'_E\right)}{t_z^2} + \frac{V'_{\text{SSU}}\left(\hat{t}'_E\right)}{t_z^2} + \frac{V'_{\text{TSU}}\left(\hat{t}'_E\right)}{t_z^2} \tag{31}$$

where $V'_{\text{TSU}}\left(\hat{t}'_E\right)$, $V'_{\text{SSU}}\left(\hat{t}'_E\right)$ and $V'_{\text{PSU}}\left(\hat{t}'_E\right)$ are obtained from Equations (25)-(27) by letting $a = E$.

Estimation of the components of $V'\left(\hat{t}'_a\right)$ From [7, Result 4.4.1], under the fictitious design, an unbiased estimator of $V'\left(\hat{t}'_a\right)$ is given by

$$\hat{V}'_{3\text{st}}\left(\hat{t}'_a\right) = \sum\sum\nolimits_{s_{I'}} \frac{\Delta_{I'ij}}{\pi_{I'ij}} \frac{\hat{t}'_{\text{pwr}ai}}{\pi_{I'i}} \frac{\hat{t}'_{\text{pwr}aj}}{\pi_{I'j}}. \tag{32}$$

By slight modification of [7, Result 4.4.3], unbiased estimators of $V'_{\text{TSU}}\left(\hat{t}'_a\right)$, $V'_{\text{SSU}}\left(\hat{t}'_a\right)$ and $V'_{\text{PSU}}\left(\hat{t}'_a\right)$ are given, respectively, by

$$\hat{V}'_{\text{TSU}}\left(\hat{t}'_a\right) = \sum\nolimits_{s'_I} \frac{1}{\pi_{I'i}^2} \left(\frac{N_{IIi}}{m_{II'i}}\right)^2 \sum\nolimits_{\nu'=1}^{m_{II'i}} \hat{V}'_{aiq_{\nu'}} \tag{33}$$

where

$$\begin{aligned}\hat{V}'_{aiq_{\nu'}} &= N^2_{iq_{\nu'}}\frac{1-f_{iq_{\nu'}}}{n_{iq_{\nu'}}}S^2_{as_{iq_{\nu'}}}; \quad f_{iq_{\nu'}} = n_{iq_{\nu'}}/N_{iq_{\nu'}};\\ S^2_{as_{iq_{\nu'}}} &= \frac{1}{n_{iq_{\nu'}}-1}\sum\nolimits_{s_{iq_{\nu'}}}\left(a_k - \frac{t_{a_{iq_{\nu'}}}}{n_{iq_{\nu'}}}\right)^2\end{aligned}$$

for every $q_{\nu'}$ that is a component of $os_{II'i}$ and $i \in s'_I$,

$$\hat{V}'_{\text{SSU}}\left(\hat{t}'_a\right) = \sum\nolimits_{s'_I}\frac{\hat{V}'_{ai}}{\pi^2_{I'i}} - \hat{V}'_{\text{TSU}}\left(\hat{t}'_a\right) \tag{34}$$

where

$$\begin{aligned}\hat{V}'_{ai} &= \frac{N^2_{IIi}}{m_{II'i}}S^2_{\hat{t}'_a os_{II'i}};\\ S^2_{\hat{t}'_a os_{II'i}} &= \frac{1}{m_{II'i}-1}\sum\nolimits_{\nu'=1}^{m_{II'i}}\left(\hat{t}_{\pi aiq_{\nu'}} - \frac{\sum_{\nu'=1}^{m_{II'i}}\hat{t}_{\pi aiq_{\nu'}}}{m_{II'i}}\right)^2\end{aligned}$$

for $i \in s'_I$, and

$$\hat{V}'_{\text{PSU}}\left(\hat{t}'_a\right) = \hat{V}'_{\text{3st}}\left(\hat{t}'_a\right) - \hat{V}'_{\text{SSU}}\left(\hat{t}'_a\right) - \hat{V}'_{\text{TSU}}\left(\hat{t}'_a\right). \tag{35}$$

Estimation of the components of $AV'\left(\hat{R}'\right)$ Define the variable $e' = y - \hat{R}'z$. Under the fictitious design, from [7, Result 5.6.2], an estimator of $AV'\left(\hat{R}'\right)$ is given by

$$\begin{aligned}\hat{V}'_{\text{3st}}\left(\hat{R}'\right) &= \frac{1}{\left(\hat{t}'_z\right)^2}\hat{V}'_{\text{3st}}\left(\hat{t}'_{e'}\right)\\ &= \frac{1}{\left(\hat{t}'_z\right)^2}\sum\sum\nolimits_{s_{I'}}\frac{\Delta_{I'ij}}{\pi_{I'ij}}\frac{\hat{t}'_{\text{pwr}e'i}}{\pi_{I'i}}\frac{\hat{t}'_{\text{pwr}e'j}}{\pi_{I'j}}\\ &= \frac{1}{\left(\hat{t}'_z\right)^2}\sum\sum\nolimits_{s_{I'}}\frac{\Delta_{I'ij}}{\pi_{I'ij}}\frac{\hat{t}'_{\text{pwr}yi} - \hat{R}'\hat{t}'_{\text{pwr}zi}}{\pi_{I'i}}\frac{\hat{t}'_{\text{pwr}yj} - \hat{R}'\hat{t}'_{\text{pwr}zj}}{\pi_{I'j}}.\end{aligned} \tag{36}$$

Estimators of the variance components $AV'_{\text{TSU}}\left(\hat{R}'\right)$, $AV'_{\text{SSU}}\left(\hat{R}'\right)$ and $AV'_{\text{PSU}}\left(\hat{R}'\right)$ are given, respectively, by

$$\hat{V}'_{\text{TSU}}\left(\hat{R}'\right) = \frac{\hat{V}'_{\text{TSU}}\left(\hat{t}'_{e'}\right)}{\left(\hat{t}'_z\right)^2}; \quad \hat{V}'_{\text{SSU}}\left(\hat{R}'\right) = \frac{\hat{V}'_{\text{SSU}}\left(\hat{t}'_{e'}\right)}{\left(\hat{t}'_z\right)^2}; \quad \hat{V}'_{\text{PSU}}\left(\hat{R}'\right) = \frac{\hat{V}'_{\text{PSU}}\left(\hat{t}'_{e'}\right)}{\left(\hat{t}'_z\right)^2} \tag{37}$$

where $\hat{V}'_{\mathrm{TSU}}\left(\hat{t}'_{e'}\right)$, $\hat{V}'_{\mathrm{SSU}}\left(\hat{t}'_{e'}\right)$ and $\hat{V}'_{\mathrm{PSU}}\left(\hat{t}'_{e'}\right)$ are obtained from Equations (33)-(35) by letting $a = e'$.

Estimation of the components of $V\left(\hat{t}_a\right)$ and $AV\left(\hat{R}\right)$ As estimators of $C\left(\hat{t}_a\right)$ and $C\left(\hat{R}\right)$, we suggest using the estimators of the corresponding population entities under the fictitious design. That is, estimate $C\left(\hat{t}_a\right)$ by

$$\hat{C}'\left(\hat{t}_a\right) = \frac{\hat{V}'_{\mathrm{PSU}}\left(\hat{t}'_a\right)}{\hat{V}'_{\mathrm{SSU}}\left(\hat{t}'_a\right)} \tag{38}$$

and $C\left(\hat{R}\right)$ by

$$\hat{C}'\left(\hat{R}\right) = \frac{\hat{V}'_{\mathrm{PSU}}\left(\hat{R}\right)}{\hat{V}'_{\mathrm{SSU}}\left(\hat{R}\right)} = \frac{\hat{V}'_{\mathrm{PSU}}\left(\hat{t}'_{e'}\right)}{\hat{V}'_{\mathrm{SSU}}\left(\hat{t}'_{e'}\right)}. \tag{39}$$

The resulting estimators of $V_{\mathrm{SSU}}\left(\hat{t}_a\right)$ and $V_{\mathrm{PSU}}\left(\hat{t}_a\right)$ are

$$\hat{V}_{\mathrm{SSU}}\left(\hat{t}_a\right) = \frac{\hat{V}_{\mathrm{3st}}\left(\hat{t}_a\right) - \hat{V}_{\mathrm{TSU}}\left(\hat{t}_a\right)}{\hat{C}'\left(\hat{t}_a\right) + 1} \tag{40}$$

and

$$\hat{V}_{\mathrm{PSU}}\left(\hat{t}_a\right) = \hat{V}_{\mathrm{3st}}\left(\hat{t}_a\right) - \hat{V}_{\mathrm{SSU}}\left(\hat{t}_a\right) - \hat{V}_{\mathrm{TSU}}\left(\hat{t}_a\right) \tag{41}$$

respectively, where $\hat{V}_{\mathrm{TSU}}\left(\hat{t}_a\right)$ is given by Equation (12) and $\hat{V}_{\mathrm{3st}}\left(\hat{t}_a\right)$ by Equation (11). In the same manner, our suggested estimators of $AV_{\mathrm{SSU}}\left(\hat{R}\right)$ and $AV_{\mathrm{PSU}}\left(\hat{R}\right)$ are given by

$$\hat{V}_{\mathrm{SSU}}\left(\hat{R}\right) = \frac{\hat{V}_{\mathrm{3st}}\left(\hat{R}\right) - \hat{V}_{\mathrm{TSU}}\left(\hat{R}\right)}{\hat{C}'\left(\hat{R}\right) + 1} \tag{42}$$

and

$$\hat{V}_{\mathrm{PSU}}\left(\hat{t}_a\right) = \hat{V}_{\mathrm{3st}}\left(\hat{R}\right) - \hat{V}_{\mathrm{SSU}}\left(\hat{R}\right) - \hat{V}_{\mathrm{TSU}}\left(\hat{R}\right) \tag{43}$$

respectively (with $\hat{V}_{\mathrm{TSU}}\left(\hat{R}\right)$ as in Equation (16) and $\hat{V}_{3\mathrm{st}}\left(\hat{R}\right)$ as in Equation (15)).

There is no guarantee for $\hat{V}'_{\mathrm{PSU}}\left(\hat{t}'_a\right)$ or $\hat{V}'_{\mathrm{SSU}}\left(\hat{t}'_a\right)$ to take on positive values. In case any of them is negative, it does not make sense to calculate $\hat{C}'\left(\hat{t}_a\right)$, and the variance component estimators in Equations (40)-(41) must be abandoned. Correspondingly, the estimators in Equations (42)-(43) should not be used if $\hat{V}'_{\mathrm{PSU}}\left(\hat{R}\right)$ or $\hat{V}'_{\mathrm{SSU}}\left(\hat{R}\right)$ is negative.

5.3 Calculation of variance component estimates from real data

In the speed survey, the sample size within stratum is one in both the second and the third sampling stage. To render variance component estimation in accordance with Section 5.2.2 possible, within the frame of the main survey 2001, some experimental data were however collected. Here, the design and outcome of this experiment are presented.

5.3.1 Data collection and processing

The collection of experimental data was restricted to one PSU stratum: the South-Eastern SNRA region and the size class Large Major Population Centers of Municipality. From this set of population centers, as part of the main survey, a sample of 75 road sites was selected. Our plan was to double this number by selecting an additional road site from each chosen small area (within PSU drawing). For various reasons (such as missing data problems), for six chosen small areas, data were obtained from less than two road sites. We decided to exclude these small areas from the experiment, which left us with 69×2 observations on traffic flow and travel time. These 69 observation pairs are distributed among four SSU strata or *development types* (city, industrial, residential, and other) and three TSU strata or *road types* (M70=major roads with a speed limit of 70 kilometers per hour (km/h), M50=major roads with a speed limit of 50 km/h, and other roads) as shown in Table 1. We see that throughout, we have very few observations on M50 roads. Therefore, this stratum is left out of further consideration.

Development type	Road type	Number of pairs
City	M70	10
City	M50	0
City	Other	6
Industrial	M70	8
Industrial	M50	3
Industrial	Other	10
Residential	M70	5
Residential	M50	2
Residential	Other	8
Other	M70	7
Other	M50	1
Other	Other	9

Table 1: The number of observation pairs, for each combination of SSU and TSU stratum.

5.3.2 Results

For each combination of SSU and TSU stratum, variances and variance components are estimated in accordance with Section 5.2.2. The estimates for $\hat{R}$ are presented in Table 2; the corresponding estimates for $\hat{t}_y$ and $\hat{t}_z$ in Appendix B. To simplify the estimation task, a few shortcuts are taken. As mentioned in Section 3, SSUs are really selected with pps rather than SI sampling within stratum Residential Areas. This exception is disregarded, and the estimates for residential areas calculated as if SI sampling was used. Also, our estimates refer to a single (arbitrary) twentyfour-hour period within the time period of study, rather than the whole period.

In Table 2 and Appendix B, there are lots of hyphens and asterisks replacing numbers. The hyphens are used for cases where the difference $\hat{V}_{3\mathrm{st}} - \hat{V}_{\mathrm{TSU}}$ is negative: then, we do not attempt to estimate $\hat{V}_{\mathrm{SSU}}$ or $\hat{V}_{\mathrm{PSU}}$. The asterisks are used when $\hat{V}_{3\mathrm{st}} - \hat{V}_{\mathrm{TSU}}$ is positive but $\hat{V}'_{\mathrm{PSU}}$ or $\hat{V}'_{\mathrm{SSU}}$ is negative. If $\hat{V}'_{\mathrm{PSU}}$ is negative, $\hat{V}_{\mathrm{SSU}}$ is calculated as $\hat{V}_{3\mathrm{st}} - \hat{V}_{\mathrm{TSU}}$ whereas $\hat{V}_{\mathrm{PSU}}$ is marked with an asterisk. Correspondingly, if $\hat{V}'_{\mathrm{SSU}}$ is negative, $\hat{V}_{\mathrm{PSU}}$ is calculated as $\hat{V}_{3\mathrm{st}} - \hat{V}_{\mathrm{TSU}}$ whereas $\hat{V}_{\mathrm{SSU}}$ is marked with an asterisk. If $\hat{V}'_{\mathrm{PSU}}$ *and* $\hat{V}'_{\mathrm{SSU}}$ are negative (occurs only once), $\hat{V}_{\mathrm{PSU}}$ and $\hat{V}_{\mathrm{SSU}}$ are both marked with asterisks.

Since R is the most important parameter, we focus on Table 2. Typically,

Development type	Road type	$\hat{V}_{\text{3st}}\left(\hat{R}\right)$	$\hat{V}_{\text{TSU}}\left(\hat{R}\right)$	$\hat{V}_{\text{SSU}}\left(\hat{R}\right)$	$\hat{V}_{\text{PSU}}\left(\hat{R}\right)$
City	M70	5.6367	7.5538	–	–
City	Other	10.8389	11.2937	–	–
Industrial	M70	1.2929	1.1497	*	*
Industrial	Other	2.1851	5.9771	–	–
Residential	M70	9.4960	3.0601	6.4359	*
Residential	Other	9.4112	13.1166	–	–
Other	M70	5.9345	1.3650	4.5695	*
Other	Other	6.6365	5.6771	0.9594	*

Table 2: Estimates (in km/h) of the approximate variance of $\hat{R}$ and its components, for various combinations of SSU and TSU strata.

$\hat{V}_{\text{TSU}}\left(\hat{R}\right)$ is nearly as large (or even larger) than $\hat{V}_{\text{3st}}\left(\hat{R}\right)$. Thus, our main conclusion is that $AV_{\text{TSU}}\left(\hat{R}\right)$ seems to predominate among the components of the variance of $\hat{R}$.

6 Optimal allocation over sampling stages

In this section, formulas for determination of optimal sampling sizes in each sampling stage are presented. As in most parts of the report, the stratification in each sampling stage is ignored. Extension of the theory presented here to the stratified case is however straight-forward.

6.1 Conditions and general solution

The conditions for allocation are the following. The variance of $\hat{t}_a$ and the approximate variance of $\hat{R}$ both fit into the general variance expression

$$V = \frac{A_1}{x_1} + \sum_{i=1}^{N_I} \frac{A_{IIi}}{x_{IIi}} + \sum_{i=1}^{N_I} \sum\nolimits_{U_{IIi}} \frac{A_{iq}}{x_{iq}} \tag{44}$$

where the x's are given by

$$
\begin{aligned}
x_1 &= m_I && (45)\\
x_{IIi} &= m_I n_{IIi}; \quad i \in U_I && (46)\\
x_{_{iq}} &= m_I n_{IIi} n_{iq}; \quad q \in U_{IIi}; i \in U_I, && (47)
\end{aligned}
$$

and the A's are constants with respect to the x's.

Let C_I denote the cost of conducting one PSU drawing. Within PSU $i \in U_I$, the listing cost per SSU, and the cost of selecting one SSU, are denoted C^l_{IIi} and C^s_{IIi}, respectively. In the same manner, within selected SSU $q \in U_{IIi}$ and PSU $i \in U_I$, the listing cost per TSU, and the cost of observing a selected TSU, are denoted C^l_{iq} and C^s_{iq}, respectively. The variable costs of the survey can now be described by the linear function

$$
\begin{aligned}
VC &= m_I C_I + \sum_{\nu=1}^{m_I} N_{IIi_\nu} C^l_{IIi_\nu} + \sum_{\nu=1}^{m_I} n_{IIi_\nu} C^s_{IIi_\nu} + \sum_{\nu=1}^{m_I} \sum\nolimits_{s_{IIi_\nu}} N_{i_\nu q} C^l_{i_\nu q} \\
&\quad + \sum_{\nu=1}^{m_I} \sum\nolimits_{s_{IIi_\nu}} n_{i_\nu q} C^s_{i_\nu q} \qquad (48)
\end{aligned}
$$

If PSU $i \in U_I$ was selected in the νth draw, then $C^l_{IIi_\nu} = C^l_{IIi}$, $C^s_{IIi_\nu} = C^s_{IIi}$, $C^l_{i_\nu q} = C^l_{iq}$ and $C^s_{i_\nu q} = C^s_{iq}$.

In practice, the listing and selection costs C^l_{IIi} and C^s_{IIi} may be approximately constant over PSUs. The listing costs C^l_{iq} are however certain to vary substantially between SSUs. We recall that the TSUs are one-meter road sites and the SSUs small geographical areas. For each chosen small area, the list of PSUs is prepared from a city map. Using the road intersections as breakpoints, the map road network is partitioned into links, and the link lengths determined manually by use of a map measurer. This listing procedure can be very time-consuming in areas with complex road networks, whilst quite quick in areas containing only a few roads. (The differences in listing costs are mitigated, but hardly removed, by the stratification of SSUs.)

Since VC is a random variable (it depends on the random samples os_I and s_{IIi_ν}), we do not want to base an optimization problem directly upon it. Instead, we follow established practice and use its expectation. Under the sampling design described in Section 3, the expected value of VC is given by

$$\begin{aligned} EVC &= m_I C_I + m_I \sum_{i=1}^{N_I} p_i N_{IIi} C^l_{IIi} + m_I \sum_{i=1}^{N_I} p_i n_{IIi} C^s_{IIi} \\ &+ m_I \sum_{i=1}^{N_I} p_i \sum\nolimits_{U_{IIi}} \frac{n_{IIi}}{N_{IIi}} N_{iq} C^l_{iq} \\ &+ m_I \sum_{i=1}^{N_I} p_i \sum\nolimits_{U_{IIi}} \frac{n_{IIi}}{N_{IIi}} n_{iq} C^s_{iq} \end{aligned} \tag{49}$$

Equivalently, the expected variable cost can be expressed as

$$EVC = x_1 a_1 + \sum_{i=1}^{N_I} x_{IIi} a_{IIi} + \sum_{i=1}^{N_I} \sum\nolimits_{U_{IIi}} x_{iq} a_{iq} \tag{50}$$

with the x's as in Equations (45)-(50), and the a's given by

$$\begin{aligned} a_1 &= C_I + \sum_{i=1}^{N_I} p_i N_{IIi} C^l_{IIi} && (51) \\ a_{IIi} &= p_i \left(C^s_{IIi} + \sum\nolimits_{U_{IIi}} \frac{N_{iq}}{N_{IIi}} C^l_{iq} \right); \quad i \in U_I && (52) \\ a_{iq} &= p_i \sum\nolimits_{U_{IIi}} \frac{C^s_{iq}}{N_{IIi}}; \quad q \in U_{IIi}; i \in U_I && (53) \end{aligned}$$

The allocation problem has two possible formulations:

- Minimize the variance V in Equation (44) with respect to x_1, x_{IIi} and x_{iq} under the expected cost constraint

  $$EVC = C_0$$

  where EVC is given by Equation (50), or

- minimize the expected cost EVC in Equation (50) with respect to x_1, x_{IIi} and x_{iq} under the variance constraint

  $$V = V_0$$

  where V is given by Equation (44).

We restrict our attention here to the second case. If the A's are all greater than zero, from [1, p. 15], this minimization problem has the analytical solution

$$\begin{aligned} x_1 &= K\sqrt{\frac{A_1}{a_1}} & (54) \\ x_{IIi} &= K\sqrt{\frac{A_{IIi}}{a_{IIi}}}; \quad i \in U_I & (55) \\ x_{iq} &= K\sqrt{\frac{A_{iq}}{a_{iq}}}; \quad q \in U_{IIi}; i \in U_I & (56) \end{aligned}$$

where

$$K = \frac{1}{V_0}\left(\sqrt{a_1 A_1} + \sum_{i=1}^{N_I} \sqrt{a_{IIi} A_{IIi}} + \sum_{i=1}^{N_I} \sum\nolimits_{U_{IIi}} \sqrt{a_{iq} A_{iq}}\right).$$

The resulting optimal sampling sizes are given by

$$\begin{aligned} m_I &= K\sqrt{\frac{A_1}{a_1}} & (57) \\ n_{IIi} &= \sqrt{\frac{A_{IIi} a_1}{a_{IIi} A_1}}; \quad i \in U_I & (58) \\ n_{iq} &= \sqrt{\frac{A_{iq} a_{IIi}}{a_{iq} A_{IIi}}}; \quad q \in U_{IIi}; i \in U_I & (59) \end{aligned}$$

(For a solution of the minimization problem if the A's are *not* all greater than zero, see [1, p. 15]).)

6.2 Solutions for t_a and R

The general variance expression in Equation (44) turns into the variance of $\hat{t}_a$ if the A's are defined as

$$\begin{aligned} A_1 &= \sum_{i=1}^{N_I} p_i \left(\frac{t_{ai}}{p_i} - t_a\right)^2 - \sum_{i=1}^{N_I} \frac{N_{IIi}}{p_i} S^2_{t_a U_i} & (60) \\ A_{IIi} &= \frac{N_{IIi}}{p_i}\left(N_{IIi} S^2_{t_a U_i} - \sum\nolimits_{U_{IIi}} N_{iq} S^2_{aU_{iq}}\right); \quad i \in U_I & (61) \\ A_{iq} &= \frac{N_{IIi}}{p_i} N^2_{iq} S^2_{aU_{iq}}; \quad q \in U_{IIi}; i \in U_I & (62) \end{aligned}$$

Insertion in Equations (57)-(59) gives the solution for t_a (if the A's are all greater than zero). Similarly, in order to transform Equation (44) into the approximate variance of $\hat{R}$, define the A's as

$$
\begin{aligned}
A_1 &= \frac{1}{t_z^2}\left[\sum_{i=1}^{N_I} p_i\left(\frac{t_{Ei}}{p_i}-t_E\right)^2-\sum_{i=1}^{N_I}\frac{N_{IIi}}{p_i}S^2_{t_EU_i}\right] && (63)\\
A_{IIi} &= \frac{1}{t_z^2}\frac{N_{IIi}}{p_i}\left(N_{IIi}S^2_{t_EU_i}-\sum\nolimits_{U_{IIi}}N_{iq}S^2_{EU_{iq}}\right); \quad i\in U_I && (64)\\
A_{iq} &= \frac{1}{t_z^2}\frac{1}{p_i}N_{IIi}N_{iq}^2S^2_{EU_{iq}}; \quad q\in U_{IIi}; i\in U_I && (65)
\end{aligned}
$$

If the A's are all greater than zero, the solution for R is obtained from Equations (57)-(59).

6.3 On use of the solutions

The formulas for optimal sample sizes in Equations (57)-(59) are not very complicated. The real problems start when they are to be used. Since R is the parameter of main interest, it ought to govern the allocation. But let us take a second glance at Equations (63)-(65). Among several unknown population entities, the A's include the population variances of t_E, $S^2_{t_EU_i}$, for all PSUs, as well as the population variances of E, $S^2_{EU_{iq}}$, for all SSUs. Use of Equations (57)-(59) thus require estimates of all these variances. The survey data will not suffice for calculating the demanded estimates, but large amounts of additional data need to be collected.

7 Conclusions and final remarks

Is the current allocation of s over sampling stages the most efficient, or is there room for improvements? In order to answer this question, we underwent the theoretical work of Sections 5.1-5.2, and conducted the experiment reported in Section 5.3. Our efforts resulted in the variance component estimates for $\hat{R}$ presented in Table 2. From this table, it looks as if the final sampling stage contributes the most to the total variance of $\hat{R}$. We conclude that, for unchanged size of s, the precision of $\hat{R}$ would probably improve if the sample sizes in stage three were increased, and the number of drawings in stage one decreased correspondingly (there is no room for decreasing the sample sizes in stage two, since they are already at minimum).

Our advise on re-allocation of the total sample is, by necessity, quite vague. The theoretical tools for choosing the sampling sizes in an optimal manner are provided in Section 6. However, the formulas presented there involve lots of unknown population quantities, and hence may be hard to use in practice.

References

[1] S. Danielsson, *Optimal allokering vid vissa klasser av urvalsförfaranden*, PhD thesis, Stockholm university, Stockholm, 1975. (In Swedish).

[2] M. H. Hansen and W. N. Hurwitz, *On the theory of sampling from finite populations*, Annals of Mathematical Statistics, 14 (1943), pp. 333–362.

[3] D. G. Horvitz and D. J. Thompson, *A generalization of sampling without replacement from a finite universe*, Journal of the American Statistical Association, 47 (1952), pp. 663–685.

[4] A. Isaksson, *Frame coverage errors in a vehicle speed survey: Effects on the bias and variance of the estimators*, Linköping Studies in Arts & Science, Thesis No. 843, Linköpings universitet, 2000.

[5] ——, *Weighting class adjustments for missing data in a vehicle speed survey*, Research report LiU-MAT-R-2000-01, Linköpings universitet, 2002.

[6] D. Raj, *Sampling Theory*, McGraw-Hill, New York, 1968.

[7] C.-E. Särndal, B. Swensson, and J. Wretman, *Model Assisted Survey Sampling*, Springer, New York, 1992.

A Derivations of fictitious first-stage inclusion probabilities

Let r_i denote the number of times PSU i occurs in the ordered sample $os_I; i \in U_I$. Note that r_i is a binomial(m_I, p_i)-distributed variable. Since only PSUs which occur at least twice in os_I are included in $s_{I'}$,

$$\begin{aligned}
\pi_{I'i} &= \Pr\left(i \in s_{I'}\right) = \Pr\left(r_i \geq 2\right) = 1 - \Pr\left(r_i = 0\right) - \Pr\left(r_i = 0\right) \\
&= 1 - \frac{m_I!}{0!m_I!} p_i^0 \left(1 - p_i\right)^{m_I} - \frac{m_I!}{1!\left(m_I - 1\right)!} p_i^1 \left(1 - p_i\right)^{m_I - 1} \\
&= 1 - \left(1 - p_i\right)^{m_I} \left(1 + m_I \frac{p_i}{1 - p_i}\right)
\end{aligned}$$

and

$$\begin{aligned}
\pi_{I'ij} &= \Pr\left(i \& j \in s_{I'}\right) = \Pr\left[\left(i \in s_{I'}\right) \cap \left(j \in s_{I'}\right)\right] \\
&= 1 - \Pr\left[\overline{\left(i \in s_{I'}\right) \cap \left(j \in s_{I'}\right)}\right] \\
&= 1 - \Pr\left[\left(i \notin s_{I'}\right) \cup \left(j \notin s_{I'}\right)\right] \\
&= 1 - \left\{\Pr\left(i \notin s_{I'}\right) + \Pr\left(j \notin s_{I'}\right) - \Pr\left[\left(i \notin s_{I'}\right) \cap \left(j \notin s_{I'}\right)\right]\right\} \\
&= 1 - \left\{\Pr\left(r_i < 2\right) + \Pr\left(r_j < 2\right) - \Pr\left[\left(r_i < 2\right) \cap \left(r_j < 2\right)\right]\right\} \\
&= 1 - \{\Pr\left(r_i = 0\right) + \Pr\left(r_i = 1\right) + \Pr\left(r_j = 0\right) + \Pr\left(r_j = 1\right) \\
&\quad - \Pr\left(r_i = 0, r_j = 0\right) - \Pr\left(r_i = 0, r_j = 1\right) \\
&\quad - \Pr\left(r_i = 1, r_j = 0\right) - \Pr\left(r_i = 1, r_j = 1\right)\} \\
&= 1 - \left\{\frac{m_I!}{0!m_I!}\left[p_i^0 \left(1 - p_i\right)^{m_I} + p_j^0 \left(1 - p_j\right)^{m_I}\right]\right. \\
&\quad + \frac{m_I!}{1!\left(m_I - 1\right)!}\left[p_i^1 \left(1 - p_i\right)^{m_I - 1} + p_j^1 \left(1 - p_j\right)^{m_I - 1}\right] \\
&\quad - \frac{m_I!}{0!0!m_I!} p_i^0 p_j^0 \left(1 - p_i - p_j\right)^{m_I} \\
&\quad - \frac{m_I!}{0!1!\left(m_I - 1\right)!}\left[p_i^0 p_j^1 \left(1 - p_i - p_j\right)^{m_I - 1} + p_i^1 p_j^0 \left(1 - p_i - p_j\right)^{m_I - 1}\right] \\
&\quad \left. - \frac{m_I!}{1!1!\left(m_I - 2\right)!} p_i^1 p_j^1 \left(1 - p_i - p_j\right)^{m_I - 2}\right\} \\
&= 1 - \left(1 - p_i\right)^{m_I}\left(1 + m_I \frac{p_i}{1 - p_i}\right) - \left(1 - p_j\right)^{m_I}\left(1 + m_I \frac{p_j}{1 - p_j}\right) \\
&\quad - \left(1 - p_i - p_j\right)^{m_I - 1} m_I \left[p_i + p_j + \left(m_I - 1\right)\frac{p_i p_j}{1 - p_i - p_j}\right]
\end{aligned}$$

B Variance component estimates for $\hat{t}_y$ and $\hat{t}_z$

In the following tables, estimates of the approximate variances of $\hat{t}_y$ and $\hat{t}_z$ and their components, for various combinations of SSU and TSU strata, are presented. The estimates for $\hat{t}_y$ are given in thousands of kilometers; the estimates for $\hat{t}_z$ in thousands of hours.

Development type	Road type	$\hat{V}_{3\text{st}}(\hat{t}_y)$	$\hat{V}_{\text{TSU}}(\hat{t}_y)$	$\hat{V}_{\text{SSU}}(\hat{t}_y)$	$\hat{V}_{\text{PSU}}(\hat{t}_y)$
City	M70	2005.9518	562.4455	$*$	1443.5063
City	Other	6.8357	5.8788	0.9569	$*$
Industrial	M70	600.3100	384.2024	$*$	216.1076
Industrial	Other	4245.1520	4564.1312	–	–
Residential	M70	17167.6676	606.4022	16561.2654	$*$
Residential	Other	28.8047	5.3361	23.4686	$*$
Other	M70	672.9285	72.2111	600.7174	$*$
Other	Other	0.4747	0.2510	0.2237	$*$

Development type	Road type	$\hat{V}_{3\text{st}}(\hat{t}_z)$	$\hat{V}_{\text{TSU}}(\hat{t}_z)$	$\hat{V}_{\text{SSU}}(\hat{t}_z)$	$\hat{V}_{\text{PSU}}(\hat{t}_z)$
City	M70	0.9143	0.2054	$*$	0.7089
City	Other	0.0039	0.0032	0.0007	$*$
Industrial	M70	0.2697	0.1963	$*$	0.0734
Industrial	Other	1.4209	1.3942	$*$	0.0267
Residential	M70	6.5975	0.1195	6.4780	$*$
Residential	Other	0.0215	0.0019	0.0196	$*$
Other	M70	0.3748	0.0291	0.3457	$*$
Other	Other	0.0003	0.0002	0.0001	$*$