Estimating total alcohol consumption in the Monitor survey

– a technical description of the estimation method

Mats Nyfjäll & Björn Trolldal

Centralförbundet för alkohol- och narkotikaupplysning
Rapport 182
Stockholm 2019

PREFACE

The Swedish Council for Information on Alcohol and Other Drugs (CAN) is an independent national competence center. Our foremost task is to increase and disseminate knowledge about trends in the consumption of alcohol and other drugs, and related harm. We do this mainly by conducting research, publishing articles and reports as well as by arranging courses and conferences. Our main national surveys are: Alcohol and Drug Use Among Students; the Monitor Study; and Habits and Consequences.

CAN is part of Swedish civil society and includes 50 member organizations. The members of our board are appointed by the Swedish Research Council for Health, Working Life and Welfare (Forte), the Swedish Research Council, the Public Health Agency of Sweden, the National Board of Health and Welfare and at the CAN's annual meeting. The Swedish Government appoints the board's chairman and deputy chairman.

The main purpose of the Monitor Study is to calculate total alcohol consumption in Sweden. This is done by adding the amount of alcohol procured from unregistered sources to the amount shown in published data on registered sales. The amount of alcohol that comes from unregistered sources is captured through a continuous survey of Sweden's inhabitants. Alcohol from these sources is measured at the acquisition level and not at the consumption level. Measuring acquisition from unregistered sources requires us to make certain assumptions and perform some calculations. These procedures are described in detail, and in plain language, in our published reports in Swedish. However, in the present report the methodology is described in technical terms and with statistical/mathematical notation. The parameters estimated in the study are formalized, as is the sampling design. In addition, the method used to compensate for non-response is described.

The report was written by Mats Nyfjäll at Statisticon in Uppsala, with support from Björn Trolldal at CAN. Gösta Forsman at Statisticon reviewed the report. The part of the Monitor study that covers alcohol is funded by Systembolaget.

Stockholm May, 2019

Charlotta Rehnman Wigstad

Director

INDEX

1 BACKGROUND

The Swedish Council for Information on Alcohol and Other Drugs (CAN) is a non-governmental organization. Their main tasks are to follow the drug trends in Sweden and to inform the public and educate professionals on alcohol and other drugs. This is e.g. done by publishing national reports and performing surveys. One such important survey is the Monitor survey.

The main purpose of the Monitor survey is to calculate the total quantity of alcohol and tobacco consumed, or more precise acquired, in Sweden. The acquisition of alcohol can be divided into (i) registered acquisition and (ii) unregistered acquisition. The registered acquisition is from Systembolaget, restaurants and grocery stores[1]. Quantities for registered acquisition is available from registers whereas unregistered acquisition is not.

The unregistered alcohol acquisition consists primarily of travelers' imports, but also of purchases of alcohol that have been smuggled into the country, home production and purchases via the internet. For the acquisition of tobacco, the travelers' import also constitutes the largest unregistered source.

In order to calculate the quantities of the unregistered parts of acquisition, telephone interviews with a random sample of people aged 17-84 are carried out on an ongoing basis.

Over a year, more than 18 000 interviews are conducted in the Monitor survey. In addition to questions about the travelers' import of alcohol and tobacco, issues of consumption habits are also included.

The Monitor survey has been going on continuously since 2000. CAN has been responsible for the Monitor measurement since 2013. Systembolaget (alcohol) and the Ministry of Health and Social Affairs (tobacco) are financing the survey.

Several reports describe results from the survey, see e.g. Trolldal and Leifman (2017). Appendix 2 in that report describes the methodology in detail. However, the description is verbal without a statistical/mathematical notation. The present report is a step towards filling that gap. By describing the methodology in technical terms, procedures and methods can become more transparent, at least for those who are familiar with reading mathematical text.

The present report only deals with the acquisition of alcohol, not tobacco. Another delimitation is that only the technical aspects of the Monitor survey is treated, not conceptual parts. For example, in this report we do not repeat the four starting points in appendix 2 in Trolldal and Leifman (2017). The *purpose of this report* is thus to *describe* the estimation procedure in technical terms. We will formalize the parameter the survey is

[1] Only low alcohol beer at grocery stores (2.8 % to 3.5 % alcohol by volume)

estimating as well as the sampling design and estimator(s). Moreover, the report will also describe the method used to compensate for non-response. However, we will not discuss uncertainty due to frame imperfections and measurement errors.

The disposition of the report is that we start by formalizing the parameters in chapter 2. In chapter 3, we describe the sampling design and in chapter 4 the estimators are given. Chapter 5 is a chapter to help reader connect estimates presented in Trolldal and Leifman (2017) with the estimators in chapter 4.

2 PARAMETERS

The acquisition of alcohol comes from two sources: (i) registered and (ii) unregistered acquisition. In Sweden, the registered acquisition is mainly done at Systembolaget. In addition, the alcohol sold at restaurants and in grocery stores contributes to the registered acquisition. The unregistered acquisition has four main sources

- Travelers' import (resandeinförsel)
- Purchases of smuggled alcohol
- Internet acquisition
- Home production (hemtillverkning)

Based on the Monitor survey, the total quantity of unregistered acquisition of alcohol in the country is estimated. The registered acquisition is obtained from register data, mainly from Systembolaget.

In order to describe the parameters the Monitor survey is estimating we start by introducing some notation. Additional notation will be introduced later on. The notation is somewhat extensive. To facilitate a quick reference, we summarized the notation in appendix 1.

2.1 Basic notation

The target population for the Monitor survey is all registered Swedish citizens in ages 17 to 84 years old. The reference period for the survey is calendar year, e.g. 2017. Let $U = (1,2,\dots,k,\dots,N)$ denote the target population where N is the population size of individuals and k is a running index for individuals in the population[2]. Since each individual k can acquire alcohol several times during the reference period (a year), an individual can be seen as a cluster where the cluster size is the number of times an acquisition is made. Let N_k denote the number of acquisitions individual k does during the reference period. The population of acquisitions is thus of size $\sum_{k\in U} N_k$.

We introduce variables y and z that are associated with unregistered acquisition of alcohol. Let z be associated with each separate acquisition of unregistered acquisition that individual k does. That means that summing each acquisition for individual k over the reference period gives the total acquisition done by individual k, i.e[3].

$$y_k = \sum_{\iota=1}^{N_k} z_\iota \qquad (1)$$

[2] During a whole year, the population size varies every day due to deaths, births and migration. We do not take that complicating aspect into account in the notation.

[3] The running index is Greek letter iota ι.

Correspondingly, we introduce variable x associated with registered acquisition of alcohol. Let z' be associated with each separate acquisition of registered acquisition that individual k does. Hence,

$$x_k = \sum_{\iota=1}^{N_k} z'_\iota \qquad (2)$$

is the total acquisition of registered alcohol for individual k over the reference period[4].

The construction of y_k and x_k might seem unconventional but the purpose is to avoid (the somewhat burdensome) cluster notation so that the parameters (and estimators) can be expressed in terms of $y{:}\,s$ and $x{:}\,s$.

The notation y and x follows the survey sample notation tradition that y is a study variable and x is an auxiliary variable used (in this case) in the estimation.

Both y and x are being defined in a conceptual manner of unregistered and registered acquisition of alcohol, in appendix 2 a more operational definition is presented.

Since there are different types of alcohol and different acquisition modes, we need subindexes to keep track. We use subindex i for type of alcohol and j for mode of acquisition, se appendix 1 for an explanation of the indexes. Hence, y_{ijk} denotes the (total) unregistered acquisition of type of alcohol i from acquisition mode j for individual k during the reference period. This variable is expressed in volume liters, not in pure alcohol. Correspondingly, x_{ijk} is interpreted as the (total) registered acquisition of type of alcohol i from acquisition mode j for individual k during a year.

Please note that unregistered alcohol is associated with subindex $j = 1,2,3,4$ and registered alcohol is associated with subindex $j = 5,6,7$. See appendix 1 for an overview of the notation.

2.2 **Parameters**

Now, the parameters can be defined. Summing all y_{ijk} in the population U, i.e.

$$t_{y_{ij}U} = \sum\nolimits_{k\in U} y_{ijk} \qquad (3)$$

gives the parameter *total* quantity of unregistered acquired alcohol for type of alcohol i and acquisition mode j. Correspondingly,

[4] Please note that N_k in (1) and (2) usually differs.

$$t_{x_{ij}U} = \sum_{k \in U} x_{ijk} \tag{4}$$

is the parameter *total* quantity of registered acquired alcohol for alcohol sort i and acquisition mode j. Both $t_{y_{ij}U}$ and $t_{x_{ij}U}$ is expressed in volume liters of alcohol, i.e. not in pure alcohol.

Another important parameter is the quantity of acquired pure alcohol expressed as per capita. Let α_{ij} denote the (average) alcohol strength, i.e. the percentage of pure alcohol, for type of alcohol i and acquisition mode j. For the majority of acquisition modes the strength according to Systembolaget ($j = 5$) is used. For example, for spirits ($i = 4$) according to Systembolaget the strength is $\alpha_{i=4,j=5} = 0{,}373$. See table 8 in appendix 1 for the (average) alcohol strength. Then

$$Q_{ij}^{UNREG} = t_{y_{ij}U} \times \alpha_{ij} \tag{5}$$

is the *total* quantity of unregistered acquired pure alcohol for type of alcohol i and acquisition mode j. Please note that $\alpha_{i,j=5}$, i.e. the strength according to Systembolaget, is used for all acquisition modes except restaurants ($j = 6$) and grocery stores ($j = 7$).

Dividing (5) by the population size gives the per capita measure. However, the division is not done by N, the population size of individuals 17-84 years old, but rather by the number of individuals at age 15 and older, which is a national and international standard procedure. Denote this number by N_{15+}. The per capita parameter of unregistered acquisition of pure alcohol (for type of alcohol i and acquisition mode j) is given by

$$\frac{Q_{ij}^{UNREG}}{N_{15+}} = \frac{t_{y_{ij}U} \times \alpha_{ij}}{N_{15+}} \tag{6}$$

Correspondingly, the parameter of registered acquisition of pure alcohol is given by

$$\frac{Q_{ij}^{REG}}{N_{15+}} = \frac{t_{x_{ij}U} \times \alpha_{ij}}{N_{15+}} \tag{7}$$

3 SAMPLING DESIGN

The sampling frame in the survey is PAR Konsument which is based on the (larger) frame named SPAR, which in turn is a complete (highly accurate) register of all Swedish citizens. The survey is done by telephone so it is convenient with a sampling frame that contains telephone numbers. Now, SPAR does not contain telephone numbers but PAR Konsument does. Trolldal and Leifman (2017) indicate that PAR Konsument contain some 70 percent of all individuals in the population, so there is undercoverage in the frame.

We first describe the sampling design in words, then in a more formal way. Each month, a large sample is drawn from the frame. The sample is stratified by gender and age groups and the sample size is one million individuals. A simple random sample[5] is drawn within each stratum. The allocation is proportional which means that the sample is self-weighted, i.e. all individuals have the same inclusion probability. From this large base sample, a simple random sample is drawn every week[6]. This means that the monthly base sample is used for four or five week samples depending on the number of weeks in the month. During a year, 52 different week samples are drawn from 12 different base samples. The 52-week samples are not coordinated, so a selected individual can be selected twice. If this occurs within the same base sample, the duplicate is removed from the sample. If this occurs in different base samples the duplicate is not removed.

An aspect regarding the data collection might be mentioned. For a given week-sample, the sampled individuals are contacted by telephone at most during a four weeks' period. For example, an individual in the week-3-sample is contacted during week 3 to 6 (at most). If no reply is obtained during that period, no further contact is done and the individual is classified as non-respondent. If contact is obtained and an interview is done, the respondent is asked about acquisition during the last 30 days. The date the interview is performed determines to which period (see below) in the estimation process an individual is allocated.

The sampling design with a base sample from the frame and then weekly samples from the base sample is in fact a *two-phase* sampling procedure (with stratification by gender and age in the first phase). In the second phase, independent weekly samples (four or five) of individuals are drawn by a simple random sampling procedure from the first phase base sample. In the estimation, a post-stratification procedure is used on a monthly basis. This means that from a randomization based perspective the design, as well as the estimator, should formally encompass both the first and second phase samples together with the post stratification. However, since the first phase is proportionally allocated this

---

[5] Without replacement

[6] If an individual is selected week 1 and the selected again week 2 the individual is removed from the sample. Hence, we can say that the sampling procedure is without replacement.

sample can be regarded (from a practical viewpoint) as a simple random sample from the frame[7]. Moreover, the four (or five) week samples from the base sample can together (for practical purposes) also be regarded as a simple random sample from the base sample[8]. Hence, in the technical description below we regard the four (our five) samples in a month from one base sample as one simple random sample from the frame.

In general terms, let $s = (1,2, \dots, k, \dots, n)$ denote the (whole) sample over a year and n its size. The base sample is drawn each month. Let $p = 1,2, \dots, P$ be an index for *period* (month). There are $P = 12$ periods (base samples) during a year. Since the base sampling procedure is repeated each month, the monthly P samples can be regarded as stratified samples from the population[9]. With an index for period the sample is s_p and the size n_p and the corresponding population notation is U_p with size N_p.

The sample s_p is considered a simple random sample from the population $U_p = (1, \dots, k, \dots, N_p)$, rather than from the frame (PAR Konsument) which is approximately of size $0.7 \times N_p$.

Note that the population size in period p is approximately equal to the population size over the whole year, i.e. $N_p \approx N$.

Summing the P samples gives $s = \cup_{p=1}^{P} s_p$ and $n = \sum_{p=1}^{P} n_p$.

Non-response occurs. Let $r_p = (1,2, \dots, k, \dots, m_p)$ denote the response set and m_p its size[10] in period p. Correspondingly, summing the P samples gives $r = \cup_{p=1}^{P} r_p$ and $m = \sum_{p=1}^{P} m_p$. During a typical year m is around 18 000 individuals.

---

7 The inclusion probability in a proportionally allocated stratified simple random sample is approximately equal to the inclusion probability in simple random sampling.

8 Drawing four independent simple random samples from a population and joining them together is formally not equivalent to one simple random sample (four times as large as the individual samples) from the population, but the difference is for practical purposes negligible.

9 We use p as index for stratification variable period rather than the more conventional index h. In the estimation a post stratification is done and there the index h is used.

10 It is the date an individual respond to the survey that determines which period he or she is allocated to.

4 ESTIMATION

4.1 Post stratification and weighting

To account for non-response a post stratification procedure is used. Post stratification is a well-known and widely used method to compensate for non-response. Se for example Särndal and Lundström (2005) for a reference.

The post stratification variables are gender (two categories), age (four categories) and region[11] (three categories). The three variables are cross classified. Let $h = 1,2, \dots, H$ be an index for post strata where $H = 2 \times 4 \times 3 = 24$ is the number of post strata.

Within each period the response set r_p is divided into H sets $r_{hp} = (1, \dots, m_{hp})$ where m_{hp} is the number of responses in post stratum h in period p. The population in each post stratum is denoted U_{hp} with size N_{hp}.

Since the sample in each period s_p is regarded as a simple random sample from the population U_p the weighting factor in the estimation is due to the post stratification to compensate for non-response. The traditional design based way[12] of calculating the weight would be N_{hp}/n_{hp} but since non-response occurs the weight[13] is

$$w_{k,trad} = \frac{N_{hp}}{m_{hp}} \tag{8}$$

CAN calculates a weight variable that is not according to (8). But the estimation process is done in such a way that the final estimator in fact uses a weight according to (8), with one small exception. This will be demonstrated below.

The weight CAN use is calculated according to

$$w_k = \frac{N_h}{N} \times \frac{m_p}{m_{hp}} \tag{9}$$

Note that the population size N per year is used[14] instead of the population size at each period N_p (in the denominator) as well as the population size in each post strata N_h per year instead of size at each period N_{hp} (in the numerator). However, this has a minor effect in the estimation, probably negligible. The reason is that the latter is not publicly available. If the factor N_{hp}/N_p would be used instead of N_h/N in (9), the weighting factor

---

[11] The regional dimension is the H-region division in Sweden.

[12] That is using a traditional Horvitz-Thompson estimator and a simple random sampling design with post strata

[13] This is often called straight expansion within (post) strata (in Swedish “rak uppräkning inom (post) strata).

[14] The source for population size is official population counts published by Statistics Sweden and the population measure for a certain year is the population December 31 previous year.

in the estimation would in effect be the same as (8). This is the small exception that was pointed out in the previous paragraph.

It is illustrative to calculate some sums of w_k over different response sets. For example summing w_k in the set r_p, i.e. within each period p, gives

$$\sum_{k \in r_p} w_k = \sum_{h=1}^{H} \sum_{k \in r_{hp}} \frac{N_h}{N} \times \frac{m_p}{m_{hp}} = m_p \tag{10}$$

Summing w_k in the set r, i.e. over each period p, gives

$$\sum_{k \in r} w_k = \sum_{p=1}^{P} \sum_{h=1}^{H} \sum_{k \in r_{hp}} \frac{N_h}{N} \times \frac{m_p}{m_{hp}} = m \tag{11}$$

This means that the weight w_k is scaled so the sum over all respondents in a year equals the response size m.

4.2 Point estimators

In expression (5) $t_{y_{ij}U}$ is the total quantity of unregistered alcohol in volume liters for type of alcohol i and acquisition mode j. Estimating this total is central in the estimation of the parameter Q_{ij}^{UNREG} in (5) and in the per capita parameter in (6). We start by summarizing the form of the estimator for $t_{y_{ij}U}$. Then, in section 4.2.1 below, the estimator $\hat{t}_{y_{ij}U}$ is explicitly stated.

To summarize, the estimator for $t_{y_{ij}U}$ can be written as

$$\hat{t}_{y_{ij}U} \times R_{ij} \tag{12}$$

where $\hat{t}_{y_{ij}U}$ is an estimator using weights w_k in (9) and R_{ij} is (an inflating) ratio[15] that account for several aspects. For example:

- It is well known, se Trolldal and Leifman (2017), that self-reported acquisition of alcohol (during the last month) tends to be under-estimated. This is true both for acquisition of unregistered as well as for registered alcohol. Since the quantity of registered alcohol can be obtained from reliable register data (mainly Systembolaget) and the survey ask questions regarding acquisition of unregistered as well as alcohol bought at Systembolaget the amount of under-reporting can be estimated. This is encapsulated into R_{ij}, as will be shown later, in a sort of ratio type estimation.

[15] We could have used a "hat" over R_{ij}, i.e. $\hat{R}_{ij}$, but since the ratio is not estimating anything we omitted it, to not clutter the notation unnecessarily.

- Earlier studies have shown that consumers of large amounts of alcohol are under-represented in the response set. Hence, an extra effort is done to account for this under-representation in the estimation process. This aspect is also encapsulated into R_{ij}.

When $t_{y_{ij}U}$ is estimated, the parameter of quantity of pure alcohol Q_{ij}^{UNREG} in (5) and in the per capita parameter in (6) can be estimated as

$$\hat{Q}_{ij}^{UNREG} = \hat{t}_{y_{ij}U} \times R_{ij} \times \alpha_{ij} \tag{13}$$

and

$$\frac{\hat{Q}_{ij}^{UNREG}}{N_{15+}} = \frac{\hat{t}_{y_{ij}U} \times R_{ij} \times \alpha_{ij}}{N_{15+}} \tag{14}$$

Below, we explicitly describe the construction of $\hat{t}_{y_{ij}U}$ and R_{ij}.

4.2.1 The estimator $\hat{t}_{y_{ij}U}$

As a first step in the estimation procedure a mean per period (month) p is calculated according to

$$\hat{\bar{y}}_{ijU_p} = \frac{\sum_{k \in r_p} w_k y_{ijk}}{\sum_{k \in r_p} w_k} = \frac{1}{N} \sum_{h=1}^{H} N_h \bar{y}_{ijr_{hp}} \tag{15}$$

where $\bar{y}_{ijr_{hp}} = (1/m_{hp}) \sum_{k \in r_{hp}} y_{ijk}$ is the ordinary mean for variable y_{ijk} in the set r_{hp}. A couple of aspects serve mentioning:

- In this expression y_{ijk} is the total acquisition of type of alcohol i by mode of acquisition j for individual k during <u>period</u> p, i.e. not the whole year as was stated earlier in section 2.1. That means that the definition of y_k in (1), summing all individual acquisitions for individual k during a year now means summing all individual acquisitions for individual k during a <u>month</u>. However, in the survey the interviewer ask the respondent only about the *last* acquisition (for each mode j)[16]. In appendix 2 the operational construction of y_{ijk} is described in more detail, see e.g. expression (58) regarding travelers' import. One important aspect of this procedure is that the total acquisition for individual k during period p is *estimated* rather than registered[17].

[16] To ask for only the last acquisition make the questionnaire less burdensome

[17] It might have been more appropriate to account for this estimation by the notation $\hat{y}_{ijk}$ instead of y_{ijk}. However, this is not done in order keep the notation as simple as possible.

- The “weights” in (15) are N_h/N rather than N_{hp}/N_p in this expression. As mentioned earlier this has marginal effect.
- The estimator $\hat{\bar{y}}_{ijU_p}$ can be seen as an estimator of the parameter $\bar{y}_{ijU_p} = (1/N_p)\sum_{k\in U_p} y_{ijk}$ the true mean of the acquisition of type of alcohol i for acquisition mode j (expressed in volume liters[18]). The estimates obtained by (15) are not of particular interest other than as an intermediate calculation step and are not presented in any of CAN:s publications.

In an additional intermediate calculation the means $\hat{\bar{y}}_{ijU_p}$ in (15) are multiplied by $N \times 0.01$. The multiplication by 0.01 converts the unit from centiliters to liters.

Next, the values are summed over all P periods giving

$$\hat{t}_{y_{ijU}} = \sum_{p=1}^{P} \hat{\bar{y}}_{ijU_p} \times N \times 0.01 = 0.01 \times \sum_{p=1}^{P}\sum_{h=1}^{H} N_h \bar{y}_{ijr_{hp}} \qquad (16)$$

Note that although this is an estimator it is not the estimator of $t_{y_{ijU}}$ in (3) since R_{ij} in (12) need to be multiplied with $\hat{t}_{y_{ijU}}$ to form the (used) estimator.

Note also that (16) is the traditional design based post stratified estimator one would arrive with by using (8) as weights, i.e.

$$\hat{t}_{y_{ijU,trad}} = \sum_{k\in r} w_{k,trad} y_{ijk} \times 0.01 = 0.01 \times \sum_{p=1}^{P}\sum_{h=1}^{H}\sum_{k\in r_{hp}} \frac{N_{hp}}{m_{hp}} y_{ijk} = \\ = 0.01 \times \sum_{p=1}^{P}\sum_{h=1}^{H} N_{hp}\bar{y}_{ijr_{hp}} \approx \hat{t}_{y_{ijU}} \qquad (17)$$

The only difference is the use of N_{hp} instead of N_h but since $N_{hp} \approx N_h$ the estimators are very similar.

4.2.2 The estimator $\hat{t}_{x_{ijU}}$

In the survey, the respondents are asked about their acquisition of registered alcohol at Systembolaget too. Variable x, instead of y, is used to indicate acquisition of all registered alcohol. In section 2.1 we defined x_{ijk} as the total acquisition of registered alcohol for individual k over the reference period (for alcohol type i and mode of

[18] At this point in the calculation process the unit is centiliters rather than liter. That does not affect the principle.

acquisition j). Since the period in the survey is month, x_{ijk} is the total quantity acquired per month. This is equivalent to the construction of variable y described in section 4.2.1, see also appendix 2 and expression (62). Since the respondents are only asked about their acquisition of registered alcohol at Systembolaget ($j = 5$), the appropriate variable is $x_{i,j=5,k}$.

The estimator $\hat{t}_{x_{i,j=5}U}$ is formed identically as $\hat{t}_{y_{ij}U}$ in (16), i.e.

$$\hat{t}_{x_{i,j=5}U} = 0.01 \times \sum_{p=1}^{P} \sum_{h=1}^{H} N_h \bar{x}_{i,j=5,r_{hp}} \tag{18}$$

where $\bar{x}_{i,j=5,r_{hp}} = (1/m_{hp}) \sum_{k \in r_{hp}} x_{i,j=5,k}$ is the ordinary mean for variable $x_{i,j=5,k}$ in the set r_{hp}.

Remark 1: $\hat{t}_{x_{i,j=5}U}$ in (18) is not used to estimate $t_{x_{i,j=5}U}$, the true quantity of registered alcohol at Systembolaget (for type of alcohol i). Rather, $t_{x_{i,j=5}U}$ is obtained from reliable transaction registers at Systembolaget. $\hat{t}_{x_{ij}U}$ is used to form a ratio type of estimator in R_{ij} in (12).

Remark 2: Acquisitions from restaurants ($j = 6$) and grocery stores ($j = 7$) is not estimated based on the Monitor survey. The true parameter values for $t_{x_{ij}U}$ for $j = 6,7$ are obtained from external sources[19], see section 4.2.5.

4.2.3 The construction of the ratio R_{ij}

The construction of R_{ij} in (12) is somewhat complex and are done slightly different depending on type of alcohol i and acquisition mode j. In words the procedure can be described as

1. First heavy consumers[20] are empirically established based on responses to the survey, and estimates for this domain are done
2. Secondly, a preliminary adjustment ratio is calculated. This preliminary adjustment (ratio) is calculated in two versions
3. Thirdly, the preliminary adjustment ratio is "fine-tuned" leading to a final R_{ij}

Below, we describe the procedure chronologically.

---

[19] Note that for grocery stores ($j = 7$) only alcohol type $i = 7$, low alcohol beer, applies.

[20] The identification of a heavy consumer is based on *consumption* patterns rather than *acquisition* patterns

4.2.3.1 Definition of heavy consumers, and estimator

The definition of a heavy consumer is data driven, i.e. depends on collected data. The quantity of pure alcohol consumed during the last month (period p) is estimated based on questions asked in the survey. The questions are not in terms of pure alcohol but the units in the question asked is converted into pure alcohol and summed over all type of alcohol. Let c_{100k} denote the consumed quantity of 100 percent pure alcohol for individual k in period p. Let $Percentile_{99}$ denote the 99:th empirical (unweighted) percentile in the response set r_p. We introduce an indicator for the heavy consumers according to

$$Ind_k = \begin{cases} 1 & \text{if } c_{100k} > Percentile_{99} \\ 0 & \text{if } c_{100k} \leq Percentile_{99} \end{cases} \tag{19}$$

Since the threshold $Percentile_{99}$ is data driven it varies between periods.

Next step is to apply a mean estimator similar to (15), but only for the heavy consumers and summing over all periods, to get a mean value for the whole year. This is equivalent to a domain estimator where the domain[21] is defined by the indicator variable (19). For domain estimation, it is practical to introduce a domain variable y_{dijk} defined as

$$y_{dijk} = \begin{cases} y_{ijk} & \text{if } Ind_k = 1 \\ 0 & \text{if } Ind_k = 0 \end{cases} \tag{20}$$

In words, variable y_{dijk} takes the value of y_{ijk} within the domain (i.e. the heavy consumers) and zero outside the domain. The domain variable is used in the estimator

$$\hat{\bar{y}}_{ijU_d} = \sum_{p=1}^{P} \hat{\bar{y}}_{ijU_{dp}} = \frac{1}{N} \sum_{p=1}^{P} \sum_{h=1}^{H} \frac{N_h}{m_{hp}} \sum\nolimits_{k \in r_{hp}} y_{dijk} = \frac{1}{N} \sum_{p=1}^{P} \sum_{h=1}^{H} N_h \, \bar{y}_{dijr_{hp}} \tag{21}$$

where $\bar{y}_{dijr_{hp}} = (1/m_{hp}) \sum_{k \in r_{hp}} y_{dijk}$ is the ordinary mean for domain variable y_{dijk} in the response set r_{hp}. The estimator $\hat{\bar{y}}_{ijU_d}$ gives an estimate of *average* quantity of acquired unregistered alcohol for heavy consumers during a year (for type of alcohol i and acquisition mode j). If variable y_{dijk} in (21) is replaced by y_{ijk}, i.e. the ordinary $y-$ variable without a domain indicator, we obtain

$$\hat{\bar{y}}_{ijU} = \sum_{p=1}^{P} \hat{\bar{y}}_{ijU_p} = \frac{1}{N} \sum_{p=1}^{P} \sum_{h=1}^{H} N_h \, \bar{y}_{ijr_{hp}} \tag{22}$$

---

[21] A domain (or domain of study) is a subpopulation of interest e.g. “men” and “women” as two subpopulations of “all persons”. Belonging to domain “men” or “women” does not depend on data, but is fixed in advance, as well as the size of the domain. Belonging to domain “heavy consumers” does depend on data. Hence, the domain heavy consumers is not a domain in a traditional meaning. However, it is convenient to borrow notation from domain theory to describe the procedure.

where $\bar{y}_{ijr_{hp}} = (1/m_{hp}) \sum_{k \in r_{hp}} y_{ijk}$ is the ordinary mean for variable y_{ijk} in the response set r_{hp}. Note that $\hat{\bar{y}}_{ijU_p}$ was introduced earlier in (15). It is apparent that the only difference between (21) and (22) is which variable, y_{dijk} or y_{ijk}, that is plugged in to the estimator.

Estimators (21) and (22) are used to calculate the preliminary adjustment ratio. This is described in the next section.

4.2.3.2 Preliminary adjustment ratio

The adjustment ratio is *principally* of the form

$$\frac{\text{True value of acquisition from Systembolaget}}{\text{Estimated value of acquisition from Systembolaget based on Monitor survey}} \tag{23}$$

Multiplying this ratio with the estimator $\hat{t}_{y_{ij}U}$ in (16) gives (in spirit) the well-known ratio estimator, see for example Särndal et. al. (1992). However, the final (fine-tuned) version of the ratio is not calculated according to the school book. Let us look in to the steps involved in the process.

Version 1 – <u>no</u> heavy consumer adjustment

Since the ratio (23) is based on the acquisition from Systembolaget we use estimator (18) with subindex $j = 5$ for the denominator

$$\hat{t}_{x_{i,j=5}U} = 0.01 \times \sum_{p=1}^{P} \sum_{h=1}^{H} N_h \bar{x}_{i,j=5,r_{hp}} \tag{24}$$

This is an estimator based on the survey for the registered acquisition of alcohol (for type of alcohol i and acquisition mode from Systembolaget, $j = 5$). The true value of registered acquisition of alcohol from Systembolaget in the numerator of (23) is $t_{x_{i,j=5}U}$. However, the true value $t_{x_{i,j=5}U}$ is adjusted to account for the fact that many Norwegians acquire alcohol at Systembolaget, especially close to the border in northern part of the west coast. Hence, the true value is adjusted down with 5 percent; $t_{x_{i,j=5}U} \times 0.95$.

Version 1 of the adjustment ratio is then given by

$$v_{i,j=5,l=1} = \frac{t_{x_{i,j=5}U} \times 0.95}{\hat{t}_{x_{i,j=5}U}} \tag{25}$$

The new subindex $l = 1$ indicates that this is the first version of the preliminary adjustment ratio. Also note that subindex $j = 5$ indicates that this concerns acquisition at Systembolaget. Version 1 does not take the heavy consumers into any special consideration, which version 2 does.

Version 2 –heavy consumer adjustment

The adjustment for heavy consumers involves a couple of steps. The idea behind the steps is to account for the fact that heavy consumers are believed to be under-represented in the response set.

First, from the expression (21) and (22) calculate the ratio

$$prop_{di,=5} = \frac{\hat{\bar{y}}_{i,j=5,U_d}}{\hat{\bar{y}}_{i,j=5,U} \times 100} \tag{26}$$

This is an estimator of the proportion[22] of acquisition for heavy consumers (the subindex d indicates domain) to the total population at Systembolaget $j = 5$ (for type of alcohol i). The rationale behind this interpretation is provided in appendix 3.

For example, regarding spirits ($i = 4$) the estimates for 2017 are

$$prop_{d,i=4,j=5} = \frac{\hat{\bar{y}}_{i=4,j=5,U_d}}{\hat{\bar{y}}_{i=4,j=5,U} \times 100} = \frac{1314.6}{149.6 \times 100} \approx 0.088 \tag{27}$$

The numerator indicates that the heavy consumers on average acquire 1314.6 centiliters of spirits from Systembolaget during 2017. The denominator indicates that the whole population on average acquires 149.6 centiliters of spirits from Systembolaget during 2017. The ratio indicates that the heavy consumers stand for approximately 8.8 percent of all acquisition from Systembolaget of spirits.

Remark: Both the numerator and the denominators underestimate their true population counterparts. It is believed that the underestimation is roughly equal in both groups so the ratio is believed to be more accurate.

Secondly, multiply the estimated proportion with the estimator for acquisition from Systembolaget (18) which gives

$$\hat{t}_{x_{i,j=5,U_d}} = \hat{t}_{x_{i,j=5,U}} \times prop_{di,j=5} \tag{28}$$

i.e. an estimator of the total quantity in volume liters of acquired alcohol from Systembolaget by heavy consumers (for alcohol type i).

This estimator does not take into consideration that the heavy consumers are underrepresented in the response set. The estimator (28) is based on the definition that heavy consumers are the consumers with the one percent largest consumption of

[22] The notation $prop$ is used for proportion, since the more common (and short) notation p already is used to denote period.

alcohol. The estimator (28) can be interpreted as the quantity in volume liters of acquired alcohol from Systembolaget for heavy consumers (for alcohol type i) if they constitute *one percent* of the population. Based on results from earlier studies (Kühlhorn et al, 1999), this group of individuals are believed to be underrepresented in the response set. Therefor the estimator is multiplied with a factor 3, i.e. $\hat{t}_{x_{i,j=5,U_d}} \times 3$.

At the same time the respondents that are not considered heavy consumer are "weighted down" with a factor 97/99 according to

$$\hat{t}_{x_{i,j=5,U_{d=not\ HC}}} \times \frac{97}{99} \tag{29}$$

Please note the notation $U_{d=not\ HC}$ which indicated the domain "not" heavy consumers (HC), that is the complement to the domain heavy consumers.

Thirdly, combining these gives the second version ($l = 2$) of the preliminary adjustment factor

$$v_{i,j=5,l=2} = \frac{t_{x_{i,j=5}U} \times 0.95}{\hat{t}_{x_{i,j=5,U_{d=HC}}} \times 3 + \hat{t}_{x_{i,j=5,U_{d=not\ HC}}} \times \frac{97}{99}} \tag{30}$$

Please note the resemblance between (25) and (30). The numerator is the same but the denominators differ. The denominator is a sum of two domain estimators; heavy consumers ($d = HC$) and not heavy consumers ($d = not\ HC$) which combined constitutes the whole population. But multiplying the estimators with factors 3 and $\frac{97}{99}$ gives a sum in the denominator that is larger than the denominator in (25). In (30) the heavy consumers are given a larger weight. Since the denominator in version $l = 2$ in (30) is larger than the denominator in version $l = 1$ in (25), $v_{i,j=5,l=2}$ will be smaller than $v_{i,j=5,l=1}$.

Numerical examples for version 1 and 2

It might be helpful for interpretation with some numerical examples. Continuing with spirits ($i = 4$) and 2017 the version 1 preliminary adjustment ratio given by (25) is given by

$$v_{i=4,j=5,l=1} = \frac{t_{x_{i=4,j=5}U} \times 0.95}{\hat{t}_{x_{i=4,j=5}U}} = \frac{19\ 189\ 382 \times 0.95}{11\ 742\ 122} \approx 1.55 \tag{31}$$

This preliminary adjustment ratio can be interpreted. Regarding spirits ($i = 4$) there is a self-reported underestimation regarding the acquisition at Systembolaget approximately equal to $\frac{1}{1.55} \approx\ 0.64$, i.e. 64 percent, giving rise to a preliminary adjustment ratio that augment the estimate of acquired unregistered alcohol by a factor of 1.55.

Version 2 is given by first calculating (26) which is numerically given in (27). Then secondly calculating (28)

$$\hat{t}_{x_{i=4,j=5,U_d}} = \hat{t}_{x_{i=4,j=5,U}} \times prop_{di,j=5} = 11\,742\,122 \times 0.088 \approx 1\,031\,831 \tag{32}$$

Thirdly, calculate

$$v_{i,j=5,l=2} = \frac{19\,189\,382 \times 0.95}{1\,031\,831 \times 3 + (11\,742\,122 - 1\,031\,831) \times \frac{97}{99}} \approx 1.34 \tag{33}$$

We note that when the heavy consumers are given a larger weight the version 2 of the preliminary adjustment ratio is 1.34, compared to 1.55 for version 1.

The preliminary adjustment ratio is calculated in two versions for all types of alcohol.

4.2.3.3 Fine-tuning of the preliminary adjustment ratio - final adjustment ratio

The fine-tuning of the preliminary adjustment ratio involves calculating moving averages. This is done in order to stabilize the ratio from temporary fluctuations. Regarding traveler's import of spirits, wine and beer as well as purchases of smuggled spirits and beer the fine-tuning involves calculating moving averages as well as some other adjustments. For the other beverages and acquisition sources, the preliminary adjustment ratio is used as the final ratio and the fine-tuning only involves calculating moving averages. Table 1 provides a summary.

Acquisition mode travelers' import $(j = 1)$ **and wine, beer and spirits** $(i = 1,2,4)$

In (26) the proportion of acquisition of alcohol from Systembolaget ($j = 5$) for heavy consumers (domain d) was defined. Replacing the specific index $j = 5$ with a general j gives the proportion for other acquisition modes

$$prop_{dij} = \frac{\hat{\bar{y}}_{ijU_d}}{\hat{\bar{y}}_{ijU} \times 100} \tag{34}$$

where the numerator is given by (21) and the denominator by (22). For example, $prop_{d,i=1,j=1} = \frac{116.2}{179.5 \times 100} \approx 0.006$, indicating that the heavy consumers stands for 0,6 percent of the of acquisition of wine from travelers' import. This can be compared with the acquisition from their share of the acquisition of spirits from Systembolaget that was 8.8 percent, see expression (27).

A moving average of (34) is formed. We introduce subindex t for time[23] (year), hence

$$prop_{dij}^{MOV} = prop_{dijt}^{MOV} = \frac{p_{dijt} + p_{dij,t-1} + p_{dij,t-2}}{3} \quad (35)$$

is a moving average[24] over three years. A remark regarding the notation:

- $prop_{dijt}^{MOV}$ in the middle equality contains subindex t for time which is natural and indicates that it is connected to time period t. In the leftmost expression in (35), i.e. $prop_{dij}^{MOV}$, the subindex t is omitted. Nowhere else in the report there has been a need for an index for time, it is only in connection to the moving averages. Since we do not want to burden the notation with a time-index in every notation we omit the t in $prop_{dij}^{MOV}$. Thus, $prop_{dij}^{MOV}$ always means the latest year t. This applies for all moving averages in the report.

A similar moving average is formed for $prop_{di,j=5}$, the acquisition from Systembolaget. Since (35) encompass j in general we need not explicitly state the moving average for $prop_{di,j=5}$.

A moving average is also formed regarding the preliminary adjustment ratio $v_{i,j=5,l}$ in (25) and (30). Adding a subindex t for time gives

$$v_{i,j=5,l}^{MOV} = \frac{v_{i,j=5,lt} + v_{i,j=5,l,t-1} + v_{i,j=5,l,t-2}}{3} \quad (36)$$

Note that this moving average is only done for $j = 5$, i.e. the acquisition from Systembolaget.

Now, the final (fine-tuned) adjustment ratio can be stated as

$$R_{ij} = \frac{prop_{dij}^{MOV}}{prop_{di,j=5}^{MOV}} \times \left(v_{i,j=5,l=1}^{MOV} - v_{i,j=5,l=2}^{MOV}\right) + v_{i,j=5,l=2}^{MOV} \quad (37)$$

Note that this adjustment ratio applies for travelers' import ($j = 1$) and wine, beer and spirits ($i = 1,2,4$). Below, we give some comments on the interpretation and the rational for the ratio.

A numerical example might be illustrative and help the interpretation. For spirits ($i = 4$) and acquisition mode ($j = 1$) and $t = 2017$ we have

---

[23] Please note the distinction between subindex t (without subindex) for time and t_{yU} the parameter total quantity of unregistered acquired alcohol. The latter has subindex which can vary; t_{yU}, t_{xU}, $t_{y_{ij}U}$ and can also contain a "hat" for estimation $\hat{t}_{y_{ij}U}$.

[24] Note that the moving average is not centered around the middle value.

$$prop^{MOV}_{d,i=4,j=1} = \frac{0.036 + 0.028 + 0.025}{3} \approx 0.030 \quad (38)$$

This shows that in 2017 the heavy consumer acquires 3.6 percent of spirits ($i = 4$) of all travelers' imports ($j = 1$). Corresponding estimates for 2016 and 2015 are 0.028 and 0.025 respectively giving a moving average of 3.0 percent.

The corresponding estimate for Systembolaget ($j = 5$) is

$$prop^{MOV}_{d,i=4,j=5} = \frac{0.088 + 0.097 + 0.084}{3} \approx 0.090 \quad (39)$$

It can be noted that heavy consumers acquires 8.8 percent of spirits ($i = 4$) of all acquisitions at Systembolaget ($j = 5$) the year 2017, a much bigger number than 3.6 percent. The moving average is 9 percent.

The moving average for (36) version $l = 1$ is

$$v^{MOV}_{i,j=5,l=1} = \frac{1.55 + 1.73 + 1.83}{3} \approx 1.71 \quad (40)$$

whereas (36) for version $l = 2$ is

$$v^{MOV}_{i,j=5,l=2} = \frac{1.34 + 1.47 + 1.61}{3} \approx 1.47 \quad (41)$$

Plugging this into (37) gives

$$R_{i=4,j=1} = \frac{0.03}{0.09} \times (1.71 - 1.47) + 1.47 \approx 0.08 + 1.47 \approx 1.55 \quad (42)$$

Taking a closer look at (37) it turns out that it in many cases can be characterized as

$$R_{ij} = small\_number + v^{MOV}_{i,j=5,l=2} \quad (43)$$

where $small\_number$ is the first term in (37). Hence, R_{ij} depends mostly on $v^{MOV}_{i,j=5,l=2}$. But $small\_number$ has the function of adding a smaller or larger value to $v^{MOV}_{i,j=5,l=2}$. Thus R_{ij} is "stretched" by a small or large amount depending on $prop^{MOV}_{dij}$, $prop^{MOV}_{di,j=5}$ and $v^{MOV}_{i,j=5,l=1}$. To understand how this stretching is done we look at some examples.

Example 1: if $prop^{MOV}_{dij} \approx prop^{MOV}_{di,j=5}$ then their ratio is 1. $prop^{MOV}_{dij}$ is the (moving average of the) proportion acquired alcohol for heavy consumers compared to the general public (for alcohol type i and acquisition mode j). $p^{MOV}_{di,j=5}$ is the same proportion but regarding acquisition at Systembolaget. If the proportions are the same, the acquisition pattern for

heavy consumers is the same when comparing acquisition mode j with Systembolaget. This gives

$$R_{ij} = 1 \times \left(v^{MOV}_{i,j=5,l=1} - v^{MOV}_{i,j=5,l=2}\right) + v^{MOV}_{i,j=5,l=2} = v^{MOV}_{i,j=5,l=1} \quad (44)$$

i.e. version $l = 1$ of the preliminary adjustment factor with no special account for heavy consumers. This is reasonable because if $prop^{MOV}_{dij} \approx prop^{MOV}_{di,j=5}$ we do not want to do any special adjustment for heavy consumers, which is what $v^{MOV}_{i,j=5,l=1}$ does (se e.g. expression (25) which is the building block in the moving average).

Example 2: if $prop^{MOV}_{dij} < prop^{MOV}_{di,j=5}$, as in (42), then their ratio is small giving rise to expression (43). If the proportion heavy consumers acquire from e.g. travelers' import are small compared to their proportion acquired from Systembolaget the adjustment factor depends (almost) solely on version $l = 2$ of the adjustment factor, i.e. $v^{MOV}_{i,j=5,l=2}$. This is also reasonable because if the acquisition made by heavy consumers for alcohol type i for travelers' import are relatively modest then we do not need to have the larger expansion factor that version $l = 1$ gives. It suffice with the expansion that version $l = 2$ gives. In other words, the underrepresentation of heavy consumers in the response set does not matter that much since they do not acquire that particular alcohol type from this acquisition mode (travelers' import in the example).

Example 3: if $prop^{MOV}_{dij} > prop^{MOV}_{di,j=5}$, then their ratio is large. If e.g. the ratio is 2.5 then

$$R_{ij} = 2.5 \times \left(v^{MOV}_{i,j=5,l=1} - v^{MOV}_{i,j=5,l=2}\right) + v^{MOV}_{i,j=5,l=2} \quad (45)$$

Studying the expression we see that $v^{MOV}_{i,j=5,l=2}$ is the starting point and then we add 2.5 times the difference between version 1 and 2 of v. In (42) the difference is $1.71 - 1.47 = 0.24$ so this difference is multiplied with a factor 2.5 that stretches R_{ij} to become larger, in this case even larger than $v^{MOV}_{i,j=5,l=1} = 1.71$. This is also reasonable because if the acquisition made by heavy consumers for alcohol type i for travelers' import are relatively large then we want to compensate for them being under represented in the response set by the larger expansion factor that version $l = 1$ gives.

Acquisition mode travelers' import $(j = 1)$, **cider and fortified wine** $(i = 3{,}5)$

Regarding cider and fortified wine the final adjustment ratio takes as simpler form than for wine, beer and spirits, namely

$$R_{ij} = v^{MOV}_{i,j=5,l=1} \quad (46)$$

i.e. the moving average of version 1 of the preliminary adjusting ratio according to (25). This means that the preliminary adjustment ratio in this case also is the final adjustment ratio.

All unregistered acquisition modes $(j = 1,2,3,4)$ **and all types of alcohol** $(i = 1,2,3,4,5)$

Above acquisition mode $j = 1$ and type of alcohol $i = 1,2,3,4,5$ was described in detail. The other acquisition modes are done similarly. Table 1 summarizes the calculation of R_{ij} for all unregistered acquisition modes and types of alcohol.

Table 1. Summary of calculation of final adjustment ratio for all unregistered acquisition modes and types of alcohol

j	**Acquisition mode**	i	**Type of alcohol**	**Final adjustment ratio**	**See**
1	Travelers' import	1,2,4	Wine, beer and spirits	$R_{ij} = \frac{prop_{dij}^{MOV}}{prop_{di,j=5}^{MOV}} \times \left(v_{i,j=5,l=1}^{MOV} - v_{i,j=5,l=2}^{MOV}\right) + v_{i,j=5,l=2}^{MOV}$	(37)
1	Travelers' import	3,5	Cider, fortified wine	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(46)
2	Smuggled	2,4	Beer, spirits	$R_{ij} = \frac{prop_{dij}^{MOV}}{prop_{di,j=5}^{MOV}} \times \left(v_{i,j=5,l=1}^{MOV} - v_{i,j=5,l=2}^{MOV}\right) + v_{i,j=5,l=2}^{MOV}$	(37)
2	Smuggled[25]	1,3	Wine, cider	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(46)
3	Internet	1,2,3,4,5	Wine, beer, spirits, cider, fortified wine	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(46)
4	Home production	1,2	Wine, beer	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(46)
4	Home production[26]	4	Spirits	Special estimator, see below	(51)

4.2.4 The estimator for Q_{ij}^{UNREG} and per capita estimator

In section 4.2.1 to 4.2.3 all building blocks for estimating Q_{ij}^{UNREG} according to (5) has been formalized. The estimator for all types of alcohol and unregistered acquisition modes, except for home production $(j = 4)$ and spirits $(i = 4)$, is done according to

$$\hat{Q}_{ij}^{UNREG} = \hat{t}_{y_{ijU}} \times R_{ij} \times \alpha_{ij} \qquad (47)$$

[25] Smuggled fortified wine is not estimated
[26] Cider and fortified wine is not estimated regarding home production

where $\hat{t}_{y_{ij}U}$, the estimator of total unregistered alcohol (for type of alcohol i and acquisition mode j) is given by (16), R_{ij} is given in table 1 and α_{ij} is given in table 8 in appendix 1. Please not that $\alpha_{i,j=5}$, i.e. the strength according to Systembolaget, is used for all acquisition modes except restaurants ($j = 6$) and grocery stores ($j = 7$).

The estimation for home production of spirits is done in a different way which is described below, but first the per capita estimator is described.

The per capita parameter is given by (6) so replacing Q_{ij}^{UNREG} by its estimator $\hat{Q}_{ij}^{UNREG}$ from (47) gives

$$\frac{\hat{Q}_{ij}^{UNREG}}{N_{15+}} = \frac{\hat{t}_{y_{ij}U} \times R_{ij} \times \alpha_{ij}}{N_{15+}} \tag{48}$$

Special calculation of $\hat{Q}_{i=4,j=4}^{UNREG}$ for home production ($j = 4$) and spirits ($i = 4$)

The procedure is described in words in Trolldal and Leifman (2017), page 42, and is summarized here. Let pH denote the (estimated) proportion of consumed spirits from home production compared to all consumption of spirits. Please note that pH is regarding consumption rather than acquisition. A usual estimate of pH is around 2 to 3 percent. Below, we make some comments on how pH is calculated.

The procedure involves a couple of steps. **First**, calculate an estimate of total acquisition[27] of pure alcohol regarding spirits ($i = 4$). This is done by summing the acquisition of pure alcohol per capita over all acquisition modes, except home production[28]. Denote this quantity ÖS as in Trolldal and Leifman (2017), i.e.

$$\ddot{O}S = \frac{\hat{Q}_{i=4,j=1}^{UNREG}}{N_{15+}} + \frac{\hat{Q}_{i=4,j=2}^{UNREG}}{N_{15+}} + \frac{\hat{Q}_{i=4,j=3}^{UNREG}}{N_{15+}} + \frac{\hat{Q}_{i=4,j=5}^{UNREG}}{N_{15+}} + \frac{\hat{Q}_{i=4,j=6}^{UNREG}}{N_{15+}} \tag{49}$$

Secondly, inflate this quantity by (one minus) the proportion of consumed spirits from home production compared to all consumption of spirits pH according to

$$\frac{\ddot{O}S}{1 - pH} \tag{50}$$

The difference is the estimator for $Q_{i=4,j=4}^{UNREG}/N_{15+}$ the per capita measure, i.e.

$$\frac{\hat{Q}_{i=4,j=4}^{UNREG}}{N_{15+}} = \frac{\ddot{O}S}{1 - pH} - \ddot{O}S = \ddot{O}S \times \frac{pH}{1 - pH} \tag{51}$$

---

[27] Not consumption in this case, which pH is

[28] Grocery stores is also excluded since they are not allowed to sell spirits in Sweden

A numeric example regarding 2017 might be helpful. We have

$$\text{Ö}S = 0.62 + 0.13 + 0.04 + 0.86 + 0.14 = 1.79 \tag{52}$$

These estimates can be found in Trolldal and Leifman (2017) in table 16[29]. The estimate pH for 2017 is 0.034, which gives

$$\frac{\hat{Q}_{i=4,j=4}^{UNREG}}{N_{15+}} = \frac{1.79}{1-0.034} - 1.79 = 1.85 - 1.79 = 0.06 \tag{53}$$

which is the published estimate in Trolldal and Leifman (2017) in table 16. This estimate is regarding consumption of home produced spirits as opposed to all other statistics, which is regarding acquisition of alcohol. Se Trolldal and Leifman (2017) for a discussion on this topic.

Remark: the estimator pH is formed by a moving average over the three last years similar to $v_{i,j=5,l}^{MOV}$ in (36). Before the moving average is calculated both the total consumption of sprits and consumption of homemade spirits are corrected for over representation in the non-response set among the heavy consumers. Since the per capita consumption of spirits from home production is relatively small compared to total acquisition we omit the (lengthy) technical details regarding the construction of pH.

4.2.5 The true parameter for Q_{ij}^{REG} and per capita

The parameter Q_{ij}^{REG}/N_{15+} for registered acquisition modes and types of alcohol per capita was earlier stated in (7) and need no additional explanation. We can underline that this estimator is regarding register acquisition from Systembolaget ($j = 5$), restaurants ($j = 6$) and grocery stores ($j = 7$). In Trolldal and Leifman (2017), page 40, it is stated that the registered acquisition regarding Systembolaget comes directly from Systembolaget. The acquisition from restaurants is based on wholesalers' reported information published by the Public Health Agency of Sweden (Folkhälsomyndigheten). The acquisition from grocery stores is calculated by the company Delfi, on behalf of the Swedish Brewers Association (Sveriges Bryggerier) and this concerns only low alcohol beer with between 2.8% and 3.5 % alcohol by volume.

[29] Swedish wording: ”Tabell 16. Den totala alkoholanskaffningen uppdelad på anskaffningskälla och dryck, i liter ren alkohol per invånare 15 år och äldre, 2001–2017”.

4.3 **Variance estimators**

In Trolldal and Leifman (2017) the uncertainty in the estimates for unregistered acquisition is not calculated. Due to the construction of the estimators, especially the construction of R_{ij} in table 1, deriving analytic expressions for the variance for e.g. $\hat{Q}_{ij}^{UNREG}$ in (47) is a complex task. If variance estimators are sought, perhaps a bootstrap procedure can be used. This report does not go any further into this matter.

5 CONNECTION BETWEEN ESTIMATORS AND ESTIMATES

I "Bilaga 1" (appendix 1) in Trolldal and Leifman (2017) there are several tables with estimates. To facilitate the transition between estimators and their corresponding estimates we provide here some guidance in table 2.

Table 2. Connection between estimators and estimates in Trolldal and Leifman (2017). Table references is regarding Trolldal and Leifman (2017)

Table	Estimates based on estimator (or parameter)	Expression
9,12, 13, 14, 15, 16	$\frac{\hat{Q}_{ij}^{UNREG}}{N_{15+}}$ and $\frac{Q_{ij}^{REG}}{N_{15+}}$	(48) and (7) and summations
10	Q_{ij}^{REG}	Numerator in (7)
11	$\hat{t}_{y_{ij}U} \times R_{ij}$	Part of numerator in (48)
19	$\hat{t}_{y_{ij}U}$	(16) (No adjustment with R_{ij})
21	R_{ij}	See table 1

We give some examples. In table 9, the total quantity of pure alcohol for both unregistered and registered acquisition for all types of alcohol is given. This unregistered acquisition is obtained by summing $\hat{Q}_{ij}^{UNREG}/N_{15+}$ over all i and j, i.e.

$$\frac{\hat{Q}^{UNREG}}{N_{15+}} = \frac{1}{N_{15+}} \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{Q}_{ij}^{UNREG} \tag{54}$$

In a similar manner the registered acquisition is obtained

$$\frac{Q^{REG}}{N_{15+}} = \frac{1}{N_{15+}} \sum_{i=1}^{I} \sum_{j=1}^{J} Q_{ij}^{REG} \tag{55}$$

Remark 1: all indexes j for acquisition does not apply in (54) nor in (55).

Adding them together gives the total acquisition regardless of acquisition mode, i.e.

$$\frac{\hat{Q}}{N_{15+}} = \frac{\hat{Q}^{UNREG}}{N_{15+}} + \frac{Q^{REG}}{N_{15+}} \tag{56}$$

Remark 2: since $\hat{Q}/N_{15+}$ contains one part that is estimated and one part that is registered based (true) parameter value their sum is an estimate and we use a hat in $\hat{Q}/N_{15+}$ to indicate that.

For example, the year 2017 $\hat{Q}^{UNREG}/N_{15+}$ is estimated to 1.96 and Q^{REG}/N_{15+} is 7.07 which gives the sum $\hat{Q}/N_{15+} = 9.03$. In a similar manner all estimates in table 9, and the other tables, can be obtained by summing over appropriate indexes.

One more example: Acquisition of unregistered wine is obtained by summing over the unregistered acquisition sources ($j = 1,2,3,4$) for wine ($i = 1$)

$$\frac{\hat{Q}_{i=1}^{UNREG}}{N_{15+}} = \frac{1}{N_{15+}} \sum_{j=1}^{4} \hat{Q}_{i=1,j}^{UNREG} \tag{57}$$

From table 9 in the report $\hat{Q}_{i=1}^{UNREG}/N_{15+} = 0.40$.

A couple of more remarks: Table 19 contains the unadjusted quantities of acquisition of unregistered alcohol (for type of alcohol i and acquisition mode j) in volume liters (not pure alcohol). These estimates are based on the estimator $\hat{t}_{y_{ij}U}$ in (16) without the adjustment ratio R_{ij}. These estimates are believed to underestimate the true parameter value by large. For example, for spirits and travelers' import $\hat{t}_{y_{i=4,j=1}U} = 9.0$. From table 20 the adjustment factor $R_{i=4,j=1} = 1.55$ and multiplying them gives $9.0 \times 1.55 = 13.9$ which is equal to the adjusted estimate published in table 11.

REFERENCES

Kühlhorn E, Hibell B, Larsson S, Ramstedt M & Zetterberg HL (1999). Alkoholkonsumtionen i Sverige under 1990-talet. Stockholm: OAS, Socialdepartementet.

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

Särndal CE & Lundström S (2005). *Estimation in Surveys with Nonresponse*. Wiley.

Trolldal B & Leifman H (2017). Alkoholkonsumtionen i Sverige 2017. CAN report 175. ISBN 97-7278-285-3.

APPENDIX 1 – SUMMARY OF NOTATION

Table 3 gives a summary of used notation. In the table, we do not attach subindex to variables, since it would mean that the table would be even longer. For example m_h and m_{hp} are not in the table. However, m and index h and p are listed in the table, so by combining the listed variables and indexes all versions of variables used in the report can be obtained.

Table 3. Summary of notation

d	=	Index for domain; in this report the only domain is heavy consumers
h	=	Index for post strata
H	=	Number of post strata
i	=	Index for type of alcohol
I	=	Number of types of alcohol
j	=	Index for mode of acquisition
J	=	Number of acquisition modes
k	=	Index for individuals in target population
l	=	Index for version of handling heavy consumers
p	=	Index for period
P	=	Number of periods
ι	=	Index for acquisitions *within* individual k (Greek letter iota)
y	=	Variable for unregistered acquisition of alcohol (sum for an individual over all acquisitions in a period)
y_d	=	Domain variable, equals y in the domain and zero outside the domain
z	=	Variable for unregistered acquisition of alcohol (each acquisition)
x	=	Variable for registered acquisition of alcohol
z'	=	Variable for registered acquisition of alcohol (each acquisition)
U	=	Target population for the Monitor survey (set of individuals)
N	=	Population size of target population U
N_k	=	Number of acquisitions for individual k during the reference period (year or month[30])
t_{yU}	=	Parameter total quantity of unregistered acquired alcohol (volume liters, not pure alcohol)
t_{xU}	=	Parameter total quantity of registered acquired alcohol (volume liters, not pure alcohol)
Q^{UNREG}	=	Parameter total quantity of unregistered acquired pure alcohol
Q^{REG}	=	Parameter total quantity of registered acquired pure alcohol
α	=	Average alcohol strength, i.e. the percentage of pure alcohol
N_{15+}	=	Population size of age 15 and older in Sweden
s	=	Sample from the population (set of individuals)
n	=	Sample size
r	=	Response set
m	=	Response set size
w	=	Weight variable
$\hat{t}_{yU}$	=	Estimator for t_{yU}

[30] The length of the reference period should be clear from the context

R = Inflating adjustment ratio; account for several aspects, primarily under-reporting of acquisition
$\hat{t}_{xU}$ = Estimator for t_{xU}
$\hat{Q}^{UNREG}$ = Estimator for Q^{UNREG}
Ind = Indicator for the heavy consumers
c_{100} = Consumed quantity of 100 percent pure alcohol
$Percentile_{99}$ = The 99:th percentile
$\hat{\bar{y}}_U$ = Mean estimator
$\hat{\bar{y}}_{U_d}$ = Mean estimator for domain
$prop_d$ = Proportion of acquisition for heavy consumers compared to the total population
$prop_d^{MOV}$ = Three year moving average of $prop_d$
v = Preliminary adjustment factor, used in calculating R
v^{MOV} = Three year moving average of v
$\ddot{O}S$ = Estimate of total acquisition except home production of pure alcohol regarding spirits
pH = Proportion of consumed spirits from home production compared to all consumption of spirits

In appendix 2, the operational definition of variables y and x are made. This requires some additional variables which are listed in table 4.

Table 4. Additional variables in appendix 2

g = number of individuals travelling together on most recent trip (travelers' import)
a = number of trips crossing the Swedish border last 30 days (travelers' import)
q = quantity of peddled alcohol (smuggling acquisition)
b = number of times smuggled alcohol is acquired last 30 days (smuggling acquisition)
f = number of times of internet acquisitions during the last 30 days (internet acquisition)
e = number of times acquiring alcohol from Systembolaget during the last 30 days

Table 5 gives an explanation of index i.

Table 5. Categories for index i

i	Type of alcohol
1	Wine
2	Beer
3	Cider
4	Spirits
5	Fortified wine
6	Low alcohol beer (folköl)

Table 6 gives an explanation of index j.

Table 6. Categories for index j

j	Mode of acquisition
1	Travelers' import (resandeinförsel)
2	Smuggled
3	Internet
4	Home production (hemtillverkning)
5	Systembolaget
6	Restaurants
7	Grocery stores (only $i = 6$, low alcohol beer)

Table 7 gives an explanation of index l.

Table 7. Categories for index l

l	Ways of handling heavy consumers
1	No special nonresponse compensation for heavy consumers
2	Special nonresponse compensation for heavy consumers

Table 8 gives an numeric values for variable α. Each year, the values are revised (often only by a small amount). Not that $\alpha_{i,j=5}$, i.e. the strength according to Systembolaget, is used in the estimation for all acquisition modes except restaurants ($j = 6$) and grocery stores ($j = 7$).

Table 8. Average strength of alcohol 2017, α_{ij}

j	Acquisition mode	i	Type of alcohol	α_{ij}
5	Systembolaget	4	Spirits	0,3729
5	Systembolaget	1	Wine	0,1279
5	Systembolaget	5	Fortified wine	0,1604
5	Systembolaget	2	Beer	0,0556
5	Systembolaget	3	Cider	0,0509
7	Grocery stores	7	Low alcohol beer	0,0331
6	Restaurants	4	Spirits	0,3095
6	Restaurants	1	Wine	0,1140
6	Restaurants	2	Beer	0,0525

APPENDIX 2 – OPERATIONAL DEFINITION OF STUDY VARIABLES

The operational definition of y_{ijk} och z_{ijk} (unregistrated alcohol) and x_{ijk} och z'_{ijk} (registrated alcohol) is done slightly different for different acquisition modes. In section 2.1 it was stated that since each individual k can acquire alcohol several times during the reference period (year) an individual can be seen as a cluster where the cluster size is the number of times an acquisition is made. Let N_k denote the number of acquisitions individual k does during a year and N_{kp} the number of acquisition during a period (month).

In section 2.1 we used ι (Greek iota) as a running index for different acquisitions. In the survey only questions about the most recent acquisition is made. This means we do not have to use the running index ι. Instead we use the ordinary index k.

We divide the description by acquisition mode.

Travelers' import $(j = 1)$

In the survey, the respondent is asked about the most recent trip from abroad to Sweden. This constitutes travelers' import. For the most recent trip, define the following variables asked in the survey:

- $z_{i,j=1,k}$ is the quantity (in centiliters[31]) acquired of unregistered alcohol at the most recent trip k from abroad to Sweden[32], for type of alcohol i and travelers' import $(j = 1)$
- g_k is the number of individuals travelling together (as a group) on the most recent trip. For example, the respondent and the spouse gives $g_k = 2$.
- a_k is the number of trips crossing the Swedish border last 30 days.

The total acquisition for unregistered type of alcohol i and travelers' import $(j = 1)$ individual k has done during the period, i.e. the y –value, is calculated as

$$y_{i,j=1,k} = \frac{z_{i,j=1,k} \times a_k}{g_k} \tag{58}$$

Note that if $a_k > 1$ this is an *estimate* of all travelers' import acquisitions made by individual k during period p. However, we omit the "hat" over y to facilitate the notation.

Remark: CAN has implemented rules if the y – variable in (58) takes too large values. We do not describe these rules in this report.

---

[31] The respondents can actually answer in any unit, e.g. number of bottles, but this is converted into centiliters

[32] Note that an answer of zero acquired alcohol is a valid answer

Smuggled alcohol $(j = 2)$ **- *Purchases of alcohol that have been smuggled into the country***

In the survey the respondent is asked about the most recent purchase of alcohol, that have been smuggled into the country, and if the respondent in its turn has sold any part of it. For the most recent acquisition, define the following variables asked in the survey:

- $z_{i,j=2,k}$ is the quantity (in centiliters) acquired of unregistered smuggled alcohol at the most recent acquisition k
- $q_{i,j=2,k}$ quantity of peddled alcohol in combination with the most recent acquisition for type of alcohol i and acquisition mode j
- b_k the number of times individual k has acquired smuggled alcohol last 30 days.

The total acquisition for unregistered type of alcohol i regarding smuggled alcohol $(j = 2)$ individual k has done during the period, i.e. the y –value, is calculated as

$$y_{i,j=2,k} = (z_{i,j=2,k} - q_{i,j=2,k}) \times b_k \qquad (59)$$

This is an *estimate* of all smuggled acquisitions made by individual k during period p. Note that if $z_{i,j=2,k} = q_{i,j=2,k}$ all acquired smuggled alcohol is peddled.

Remark: expression (59) is not calculated for fortified wine since it is smuggled to negligible extent.

Remark: CAN has implemented rules if variable y in (59) takes too large values. We do not describe these rules in this report.

Internet $(j = 3)$

In the survey the respondent is asked about the most recent internet acquisition, with the exception of Systembolaget. For the most recent acquisition, define the following variables asked in the survey:

- $z_{i,j=3,k}$ is the quantity (in centiliters) acquired of unregistered alcohol from internet (most recent acquisition)
- $f_{i,j=3,k}$ is the number of times of internet acquisitions during the last 30 days

The total acquisition for unregistered type of alcohol i and internet $(j = 3)$ individual k has done during the period, i.e. the y –value, is calculated as

$$y_{i,j=3,k} = z_{i,j=3,k} \times f_{i,j=3,k} \quad (60)$$

Remark: CAN has implemented rules if variable y in (60) takes too large values. We do not describe these rules in this report.

Home production ($j = 4$)

In the survey the respondent is asked about the home production during the last 30 days. A criterion is that the home produced alcohol should have been completed during the last 30 days. The total quantity of completed alcohol is asked. This means we do not have to distinguish between individual k and all possible acquisitions. Define the following variables asked in the survey:

- $z_{i,j=4,k}$ is the quantity (in liters) of home produced (completed) alcohol during the last 30 days.

The total acquisition for unregistered type of alcohol i and home production ($j = 4$) for individual k during the period, i.e. the y −value, is calculated as

$$y_{i,j=4,k} = z_{i,j=4,k} \times 100 \quad (61)$$

The factor 100 converts the unit liters in the question into centiliters to harmonize with other types of alcohol.

Remark 1: CAN has implemented rules if variable y in (61) takes too large values. We do not describe these rules in this report.

Remark 2: expression (61) only applies to wine ($i = 1$) and beer ($i = 2$). Regarding spirits ($i = 4$) a (completely) different procedure is used. See expression (51). No questions regarding home production of cider nor fortified wine is asked.

Registered alcohol ($j = 5$)

In the survey the respondent is asked about the most recent acquisition from Systembolaget ($j = 5$). Regarding acquisition from restaurants and grocery stores, no questions are asked in the survey. For the most recent acquisition, define the following variables asked in the survey:

- $z'_{i,j=5,k}$ is the quantity (in centiliters) acquired registered alcohol regarding last acquisition from Systembolaget.

- e_k is the number of times acquiring alcohol from Systembolaget during the last 30 days

The total acquisition of registered alcohol i at Systembolaget ($j = 5$) individual k has done during the period, i.e. the y −value, is calculated as

$$x_{i,j=5,k} = z'_{i,j=5,k} \times e_k \quad (62)$$

This is an *estimate* of all acquisitions at Systembolaget made by individual k during period p.

APPENDIX 3 – RATIONALE FOR INTERPRETATION OF (26)

Expression (26) is interpreted as estimator of the percentage of acquisition for heavy consumers to the total population at Systembolaget $j = 5$ (for type of alcohol i). The rationale for this interpretation is motivated here.

The heavy consumers constitute a domain in the population. The domain is defined as the individuals in the response set with the one percent largest consumption of pure alcohol (se section 4.2.3.1 and the use of the 99:th percentile $Percentile_{99}$). The *total* acquisition of alcohol (not pure alcohol) for this domain (heavy consumers) at Systembolaget divided by total acquisition of alcohol for the whole population ought to give the sought percentage. Both these parameters are estimated from the survey, so we write

$$\frac{\hat{t}_{i,j=5,U_d}}{\hat{t}_{i,j=5,U}} = \frac{\hat{\bar{y}}_{i,j=5,U_d} \times N_d}{\hat{\bar{y}}_{i,j=5,U} \times N} \approx \frac{\hat{\bar{y}}_{i,j=5,U_d} \times \widehat{N}_d}{\hat{\bar{y}}_{i,j=5,U} \times \widehat{N}} \tag{63}$$

Since the domain is defined by the 99:th percentile $Percentile_{99}$ in the response set the following equality holds approximately

$$\frac{\widehat{N}_d}{\widehat{N}} \approx \frac{1}{100} \tag{64}$$

Which motivates the multiplication of $1/100$ in expression (26).