6 Sampling

6.1 Introduction

The sampling design for the second wave of the HFCS in Austria was specifically developed by the OeNB in collaboration with the survey company IFES (Institut für empirische Sozialforschung GmbH). Sampling is understood as the selection of a set of units (i.e. a sample) from the whole population on the basis of which conclusions can be derived about the behavior of the whole population. The units of the sample should be representative of the whole population; in other words, in expectation an analysis of the sample (using appropriate weights) should lead to the same estimates as an analysis of the whole population. Another criterion of major importance for the HFCS – coverage of households in all regions – is achieved through stratified sampling, i.e. by dividing the country into smaller geographical units from which the sampling units are drawn. Although some degree of statistical uncertainty cannot be ruled out, sampling – together with imputation and weighting – serves to produce the best unbiased estimates (and confidence intervals) and keep uncertainty as low as possible by taking into account restrictions like costs, time and practicability. Therefore, every survey is highly dependent on the quality of its sampling design.

This chapter describes the sampling procedure for the HFCS in Austria and is structured as follows: First, we define the target population (section 6.2) and provide a short overview of the sampling design in a box. This part is followed by a description of the required external data on geography and population (section 6.3). Next, we detail the stratification process (section 6.4) and the two stages of drawing the survey's sample population (section 6.5), which form the main part of the sampling procedure. Section 6.6 completes the chapter with some concluding remarks.

6.2 Target population and sampling frame

The first step in determining the sampling procedure is defining the target population of the survey. The HFCS is intended to cover all households living permanently in Austria, independent of citizenship and/or residence status. According to the ECB, a household in the HFCS is defined as

"a person living alone or a group of people who live together in the same private dwelling and share expenditures, including the joint provision of the essentials of living. Employees of other residents (i.e. live-in domestic servants, au-pairs, etc.) and roommates without other family or partnership attachments to household members (e.g. resident boarders, lodgers, tenants, visitors, etc.) are considered separate households." [1]

More specifically, the following persons are to be regarded as household members, according to the ECB's definition[1]:
"1. persons usually resident, related to other members
 2. persons usually resident, not related to other members
 3. persons usually resident, but temporarily absent from dwelling (for reasons of holiday travel, work, education or similar)
 4. children of household being educated away from home

---

[1] See ECB (2013a), p. 80f.

5. persons absent for long periods, but having household ties: persons working away from home

6. persons temporarily absent but having household ties: persons in hospital, nursing home, boarding school or other institution"

In the case of the HFCS, the target population does not include households that live in institutions such as:

• homes for elderly people,
• military compounds,
• monasteries,
• prisons, and
• boarding schools.

Additionally, the Austrian HFCS does not cover homeless people. People without a residence could not be reached with the survey, as sampling was based on dwellings (see section 6.3). At the same time, the HFCS in Austria is not limited to households officially registered at their main residences.

In order to draw a sample from this target population, we would need a complete list of households in Austria. As such a list does not exist, we use a complete list of postal addresses in Austria as our sampling frame. These external data, explained in more detail below, provide the best possible sampling frame in the sense that (almost) all households in Austria appear in the data (and appear only once) and that the data are highly up to date.
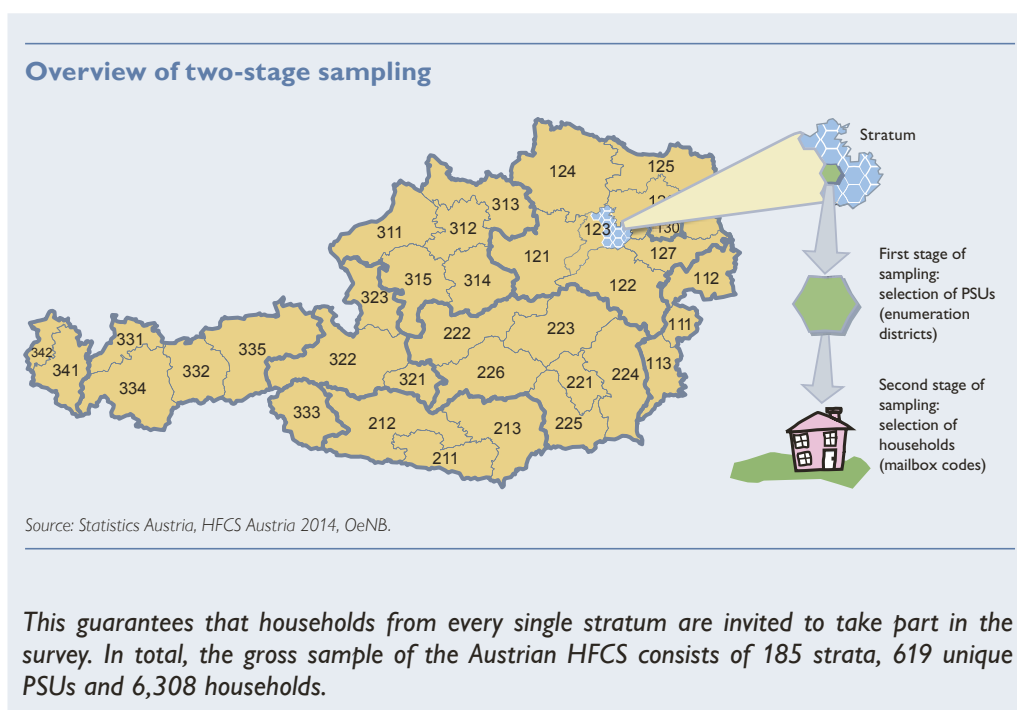
Box 1

Sampling in the HFCS in Austria

The HFCS in Austria is based on a stratified, two-stage cluster sampling design:

"Stratified sampling" ensures that the units of collection – i.e. households, for our purposes – are drawn from all subgroups of the target population. Stratification in the Austrian HFCS was carried out geographically (based on NUTS-3 regions[1]) and by municipality size categories.

"Two-stage cluster sampling" means that, first, primary sampling units (PSUs) are selected from each geographical unit (i.e. stratum) and that, second, secondary sampling units (SSUs) are drawn from within each selected PSU. The two-stage sampling design of the Austrian HFCS (see the infographic below) entails, first, selecting a random sample of enumeration districts (the smallest geographical unit for which statistical data are available) from each stratum and, second, selecting a random sample of households (postal addresses) from each sampled enumeration district. Unlike during the first wave of the HFCS in Austria, the probability of being drawn during the first stage is proportional to the number of households in the respective PSUs. The households constitute the secondary sampling units (SSUs) and are selected at random from within a drawn PSU. The two-stage cluster design reduces costs due to relatively small distances between the 12 households (8 households in strata with over 50,000 inhabitants) selected within each PSU, while it ensures a sufficient number of PSUs within the individual stratum.

[1] *See www.statistik.gv.at/web_en/classifications/regional_breakdown/nuts_units/index.html (accessed on December 9, 2016). Austria is divided into 35 NUTS-3 regions. These regions typically consist of several neighboring political districts or correspond to urban areas including the capital cities of the provinces.*

Overview of two-stage sampling

Stratum

First stage of sampling: selection of PSUs (enumeration districts)

Second stage of sampling: selection of households (mailbox codes)

Source: Statistics Austria, HFCS Austria 2014, OeNB.

This guarantees that households from every single stratum are invited to take part in the survey. In total, the gross sample of the Austrian HFCS consists of 185 strata, 619 unique PSUs and 6,308 households.

6.3 Background – the (external) datasets used

Given the definition of the target population, geographical data, as well as data on households in Austria are needed. A representative draw from the sample requires the target population to be correctly represented by the sampling frame. The frame data are perfect "if every element appears on the list separately, once, only once and nothing else appears on the list" (Kish, 1995, p. 53). In practice it is not possible to achieve this theoretical optimum. The HFCS in Austria has been designed to meet this goal subject to the constraints of the data sources available. The following sections give more details on the data which provided the basis for the sampling used in the HFCS in Austria.

For the HFCS in Austria, we relied on two different sources: We used data from Statistics Austria for the purpose of stratification and for selecting a random sample of PSUs (primary sampling units; in Austria those are the enumeration districts) and we used post office data to draw the households, the actual SSUs (secondary sampling units), at random. The advantage of the post office data is that they are up to date and that the data fit the HFCS definition of households.

6.3.1 Statistics Austria

We used information about the geographical structure of Austria, i.e. data on the NUTS-3 regions, and the enumeration districts (PSUs) from the 2011 register-based census.[2] These enumeration districts are the smallest territorial units in Austria for which basic data characteristics are collected by Statistics Austria by default (each enumeration district contains around 440 dwellings on average).[3]

---

[2] See also www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/volkszaehlungen_ registerzaehlungen_abgestimmte_erwerbsstatistik/index.html (German only) (accessed on December 9, 2016).

[3] The estimated number of households in Austria according to the HFCS definition (3.9 million) divided by the number of enumeration districts (8,821) yields 442.1.

In addition, we relied on the municipality directory of 2013 (to categorize by municipality size) and on Statistics Austria's register-based census from 2011 for the population data.[4] Both datasets were collected and provided by Statistics Austria. The register-based census of 2011 indicates in particular the population (households) of each stratum and thus serves to determine the number of sampling units to be drawn within each stratum. To sum it up, all the information necessary for the stratification and the first stage of the sampling design is based on geographical and population information provided by Statistics Austria.

6.3.2 Austrian Post Office

Once the appropriate primary sampling units have been randomly selected, information on the households is needed to complete the sample selection. The dataset of choice for the purpose of the HFCS was a dataset of postal addresses for sale from the Austrian Post Office, based on the assumption that the number of households living in each building corresponds to the number of postal addresses. Specifically, we used a commercial product called "Adress.Certified" developed by the Austrian Post Office. This address register contains information about individual buildings (including street name, house number and whether the building is used privately or commercially). It can be purchased in combination with a product called "DATA.DOOR," which is a directory of post office-certified address codes (shortened to PAC), i.e. a directory of all addresses in Austria which mail can be delivered to. This information is available in disaggregated form. In Austria, there are about 4,051,000 private mail delivery points. Addresses identified by the Austrian Post Office as vacation homes have already been excluded.

Thus, our starting point was some 4.1 million private mail addresses. Very few remaining commercial addresses and ineligible addresses had to be removed after the first contact by the interviewer (e.g. if the interviewer arrived at the address and noted that it was incorrect or was a commercial building) and were given weights of zero, since they do not belong to the target population (see chapters 4 and 7). Moreover, households at secondary residences whose main residence address was identifiable as such were also excluded from the dataset of the sampling frame or given a weight of zero to ensure that every household was included only once in the list of post office-certified address codes. After removing these commercial addresses and ineligible addresses, the total of all weights comes to roughly 3.86 million; thus, Austria has an estimated 3.9 million households.

The post office data we used do not reflect whether a given address is a household's main residence or not or whether the residence at that address is registered in the Austrian residence registry (Zentrales Melderegister). Yet they provide a realistic picture of households and thus meet the HFCS requirement of reflecting actual living situations. Unlike other data sources (e.g. EU SILC), the post office data cover households at addresses that are registered as a secondary home or that are not registered at all, but fulfill the HFCS definition of a household. They have thus been included in the sampling frame because they have a post-certified address code.[5]

---

[4] *See also Statistics Austria (2013).*

[5] *The post-certified addresses for some 4.1 million households compare with about 3.8 million household addresses documented by other sources (such as the microcensus based on the residence registry; see also section 7.2.4).*

6.3.3 Profile.Address and IFES

To identify the names of the households that correspond to the selected postal addresses – information that is not evident from the datasets described so far – the survey company, IFES, relied on its databases or obtained the corresponding addresses from a commercial provider called "Profile.Address."

This information was needed in the contact phase when households received individualized letters of invitation to participate in the survey.[6]

6.4 Stratification and sample size

6.4.1 Stratification

The Austrian HFCS essentially used two indicators for stratification, the first one being the 35 NUTS-3 regions (see chart 4).

With the exception of the capital city Vienna, each NUTS-3 region was divided further into the following eight categories according to municipality size.
- Up to 2,000 inhabitants
- 2,001 – 3,000 inhabitants
- 3,001 – 5,000 inhabitants
- 5,001 – 10,000 inhabitants
- 10,001 – 20,000 inhabitants
- 20,001 – 50,000 inhabitants
- 50,001 – 1 million inhabitants
- Over 1 million inhabitants

The large "50,001 – 1 million inhabitants" category essentially contains just the provincial capitals. Vienna is a special case as it is the national capital and the only city with more than 1 million inhabitants; it was subdivided into its 23 districts.

Chart 4

NUTS-3 regions
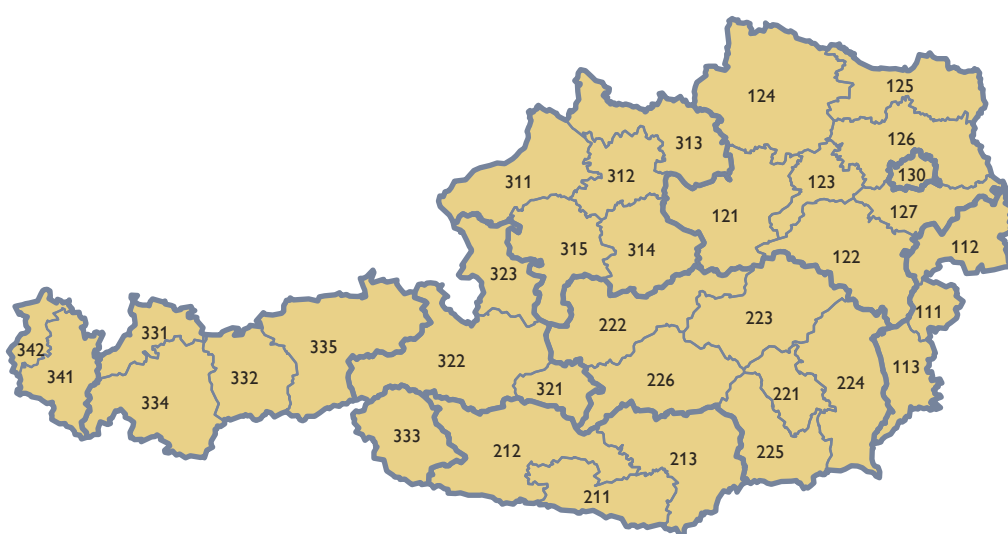
Source: Statistics Austria.

---

[6] For more details on the contact strategy, see section 3.4.

This very fine stratification yielded 193 strata. Where the number or proportional share of households per stratum was too small to allow the selection of enumeration districts, individual strata were merged with neighboring strata to increase the share of households and thus insure the selection of at least one PSU from each stratum. This exercise left the HFCS with 185 strata for sampling covering all households in Austria. The distribution of strata across provinces and municipality size categories can be seen in table 9.

Table 9

Allocation of strata within the sample

	Municipality size[1] (number of inhabitants)								
	up to 2,000	2,001–3,000	3,001–5,000	5,001–10,000	10,001–20,000	20,001–50,000	50,001–1 million	over 1 million	Total
Vienna	0	0	0	0	0	0	0	23	23
Lower Austria	7	7	7	7	6	5	1	0	40
Burgenland	3	2	2	2	1	0	0	0	10
Styria	6	6	5	6	2	1	1	0	27
Carinthia	3	3	3	3	2	1	1	0	16
Upper Austria	5	5	5	5	3	2	1	0	26
Salzburg	3	2	2	2	2	1	1	0	13
Tyrol	5	5	3	3	3	0	1	0	20
Vorarlberg	2	2	2	1	2	1	0	0	10
Total	34	32	29	29	21	11	6	23	185

Source: Statistics Austria (municipality directory 2013).

[1] Municipality size accounts for municipality mergers up to 2013.

Each stratum contained about 50 PSUs on average, which in turn contained around 460 households[7] on average.

6.4.2 Sample size

The variance of estimates based on the underlying data will be smaller the larger the sample is. At the same time, the cost of data collection increases with sample size. Therefore a balance has to be found in order to yield reasonably precise estimates whilst taking into account the given budget constraints. Furthermore, given the focus of the survey and the analyses that the HFCS is to carry out, the HFCS should produce enough observations to allow for an analysis of subpopulations (e.g. indebted households, which are only a (small) fraction of the sampling frame) and provide some insight into the regional differences within Austria. We know from previous OeNB surveys (e.g. the first wave of the HFCS in Austria in 2010 and the Household Survey on Housing Wealth in 2008) that at least 2,000 households need to be successfully interviewed and that the unit nonresponse rate can be expected to reach some 40% to 50% (with expected differences between Vienna and the rest of Austria).[8] With some leeway for extraordinary circum-

---

[7] The difference between this figure and the 440 households per PSU cited earlier is the result of aggregation (see also section 6.5.1). Similar to the data above we used the roughly 3.9 million households according to the HFCS definition in the 8,471 enumeration districts resulting from the aggregation to calculate this average.

[8] The regional differences between participation rates in the first wave of the HFCS in Austria were taken into account in the sample design for the second wave.

stances, the HFCS survey was designed to yield a sample of 3,000 successfully interviewed households and a participation rate of about 35% in Vienna and around 55% in the rest of Austria. These participation rates are estimates based on the experience of past surveys. The participation rates recorded in the first wave in a stratum were used to determine the exact number of households to be drawn in the respective stratum in the second wave.

The targeted net sample of *n = 3,000* was divided between the nine provinces, based on their population share as surveyed by Statistics Austria in the register-based census of 2011[9] (table 10, column 1). These figures, which corresponded to the targeted number of secondary sampling units (SSUs, column 2), were subsequently translated into gross samples of SSUs based on the estimated participation rates (column 3). Due to the shorter distances between buildings, 8 households were selected in Vienna and in strata with more than 50,000 inhabitants, whilst this number was 12 in the rest of Austria (column 4). The number of PSUs to be drawn in each province was calculated on this basis (column 5).

Table 10

Identification of the number of primary sampling units (PSUs) to be drawn

	% of households	Target sample	Gross sample	Number of households per PSU (enumeration district)	Number of PSUs to be drawn
	(1)	(2)	(3)	(4)	(5)
Vienna	23	688	1,960	8	245
Lower Austria	19	599	956	8/12	82
Burgenland	3	96	168	12	14
Styria	14	421	832	8/12	82
Carinthia	7	198	404	8/12	41
Upper Austria	16	485	932	8/12	90
Salzburg	6	184	384	8/12	39
Tyrol	8	244	468	8/12	45
Vorarlberg	4	125	204	12	17
Total	100	3,000	6,308	29	655

Source: Statistics Austria (municipality directory 2013), HFCS Austria 2014, OeNB.

In total the Austrian HFCS sample design produced 655 (619 unique) PSUs across all strata and a gross sample size of 6,308 households that were invited to participate in the HFCS (see box 2 in chapter 7 for information on the number of households interviewed successfully). Drawing the PSUs was done with replacement, which entails some PSUs being drawn multiple times (see section 6.5.1). Drawing possible substitute addresses was explicitly excluded from the HFCS to begin with to ensure that all households from the gross sample would be interviewed with the same commitment so as to prevent data distortions (see also section 4.4.1).

---

[9] Although Statistics Austria's register-based census of 2011 is based on a different definition of households, we used the respective data as an approximation.

6.5 The two stages of the random draw

The Austrian HFCS is based on a stratified, two-stage cluster sampling design:
- stage one: random draw of PSUs (enumeration districts) from each stratum
- stage two: random draw of a predefined number of households (postal addresses) from each PSU

6.5.1 First stage

We chose the smallest territorial units in Austria, the so-called enumeration districts (of which there are 8,821) as the PSUs for the Austrian HFCS. On average, an enumeration district contains 440 households, but there are also PSUs with only a few households. Such units were aggregated with neighboring units to ensure that each PSU contains at least 50 households and that at least one PSU can be chosen per stratum. This aggregation process narrowed the number of PSUs down to 8,471, which then had 460 households on average. The description above shows that the number of PSUs to be drawn in each province is determined a priori by the chosen sample size and stratification. To translate the numbers allocated to each region (table 10) into the desired number of PSUs within a given stratum, the total number of PSUs in the respective province was distributed proportionally according to the number of households in the respective stratum. For example, the 90 PSUs to be drawn in Upper Austria (table 10) were divided up among the 26 strata in this province according to the population share.

After determining how many PSUs were to be drawn per stratum, the PSUs – unlike during the first wave of the HFCS in Austria[10] – were drawn proportionally to their size (measured in terms of the number of households in a PSU).[11] The purpose of this change is to reduce the standard errors of estimates by reducing the variance of the sample design weights (see also section 7.2.2). Likewise, it ensures that every household within a stratum has the same probability of being drawn in the gross sample of the HFCS. PSUs were drawn with replacement, meaning that a PSU can be drawn multiple times. This meant that a total of 655 PSUs were drawn in the second wave of the HFCS in Austria, only 619 of which were unique.

6.5.2 Second stage

With 655 (619 unique) PSUs having been randomly drawn, we turn to the second stage in which households are selected.

Mail delivery points were randomly selected from each PSU drawn, with 8 being chosen in Vienna or in strata with more than 50,000 inhabitants and 12 being chosen in all other strata. In this process, every household in a given PSU has an equal probability of being selected in the sample, which is measured as a ratio of 1 to the number of households in that PSU. This procedure resulted in a gross sample of 6,308 households in Austria.

---

[10] *During the first wave, the probability of drawing a PSU within a stratum was identical for every PSU.*

[11] *Mathematically, the probability of an enumeration district being drawn in a given stratum can be expressed as the number of households in a given PSU divided by the total number of households in this stratum times the number of enumeration districts drawn.*

6.5.3 Practical implementation

Table 11 illustrates how the data from the second stage of the sampling were used, after the PSUs had been chosen in the first stage (column 2): Austrian Post Office data (column 6) were used to determine the appropriate mail delivery point, which gives the address, but not the holder of this address. To identify the name of the household corresponding to the selected postal address, the survey company, IFES, used its own databases or, where necessary, bought the corresponding name from the company "Profile.Address" (column 7).

Table 11

Matching of Statistics Austria data with post office and commercial data (fictitious example)

First stage		Second stage				
Statistics Austria		Austrian Post Office				Profile.Address/IFES
Municipality code (1)	Enumeration district (2)	Postal code (3)	Street (4)	House number (5)	Mail delivery point (PAC) (6)	Name of household (7)
90101	90101001	XXXX	Sample street	6	101255765	John Doe
90101	90101001	XXXX	Sample street	6	101255766	Jane Doe
90101	90101002	XXXX	Sample street	9	101255767	John Doe
90101	90101001	XXXX	Sample street	10	101255768	Jane Doe

Source: Statistics Austria, Austrian Post Office, Profile.Address/IFES.

Since the first contact with a household is very important for a successful interview, every household selected for the HFCS survey received an individualized letter signed by the governor of the OeNB. This letter contained information on the survey and an invitation to take part (see section 3.5.1).[12]

6.6 Concluding remarks

This chapter provides information on the sampling design specifically developed for the second wave of the HFCS in Austria on the basis of the first-wave design. The improvements specified above were implemented as a result of the experience and information gained from the first wave. As described, the survey is based on stratified, two-stage cluster random sampling, consisting of a random draw (proportional to the number of households) of primary sampling units (PSUs; here, enumeration districts) from each stratum plus a random draw (with an identical probability of being drawn within a given PSU) of households (postal addresses as available from an Austrian Post Office database) from the selected PSUs.

The sampling method used for the HFCS has a number of advantages, with the following aspects being particularly important:
- Compared to the first wave, the fact that the probability of being drawn was proportional to the size of the PSU in terms of the number of households improved the efficiency of sampling design by reducing the variance of the design weights.

---

[12] See the online appendix for the invitation letter.

- As sampling does not differentiate between main residences and second homes (as recorded in the residence registry), all households that correspond to the HFCS household definition have a positive probability of being selected.
- The very fine stratification structure ensures that all segments of the Austrian population are represented in the survey.

At the same time, given the topics covered by the HFCS it would be desirable to oversample certain groups of the population, such as wealthy households, to improve the efficiency of estimates for these subgroups. However, the underlying information needed for such oversampling is not available yet.