6 Sampling

6.1 Introduction

The sampling design of the HFCS in Austria was specifically developed by the OeNB in collaboration with the Institut für empirische Sozialforschung GmbH – IFES. Sampling means selecting 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) will 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 assured 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 is not avoidable, sampling – together with imputation and weighting – serves to produce the best unbiased estimates (and confidence intervals) and to keep uncertainty as low as possible under restrictions like costs, time and practicability. Therefore, a sound and clear sampling design is essential and the starting point for any survey.

This chapter 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 sample population (section 6.5), which build the main part of the sampling procedure. Section 6.6 contains concluding remarks.

6.2 Target Population and Sample Frame

The first step in determining the sampling procedures 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 as 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
 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"

---

[1] See ECB (2011, p. 7).

In the case of the HFCS, the target population does not include households that are institutionalized, such as households living in
• 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 below). 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 the complete list of private mailboxes in Austria as our frame population. 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 the sample – i.e. households for our purposes – are drawn from all the desired parts of the population. Stratification in the Austrian HFCS was done by geography (based on NUTS-3 regions[1]) and by municipality size categories.*
• *"Two-stage cluster sampling" means that, in a first step, primary sampling units (PSUs) are selected from each geographical unit (i.e. stratum) and that, in a second step, secondary sampling units (SSUs) are drawn within each PSU. The two-stage sampling design of the Austrian HFCS (see the chart in this box) consists, first, of selecting a random sample of enumeration districts (the smallest geographical unit for which statistical data are available) from each stratum and, second, of selecting a random sample of households (mailbox*

Overview of Two-Stage Sampling
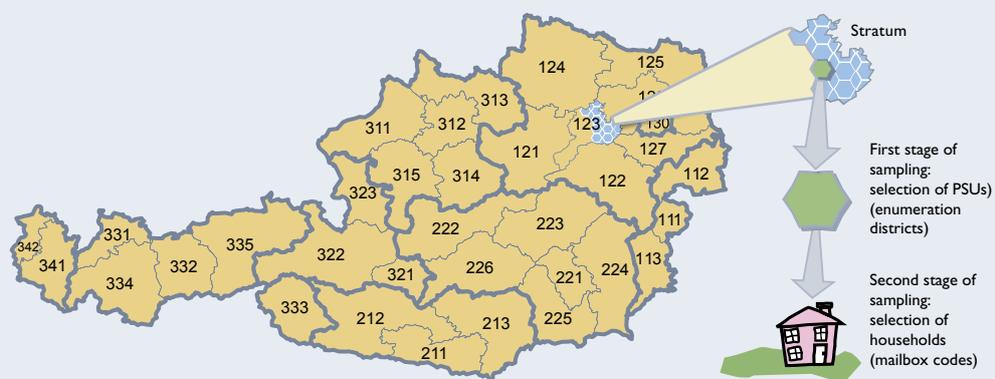
Source: Statistics Austria, HFCS Austria 2010, OeNB.

---

[1] *See www.statistik.gv.at/web_de/klassifikationen/regionale_gliederungen/nuts_einheiten/index.html (accessed on January 22, 2013). 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.*

> *codes) from each sampled enumeration district. The households thus constitute the secondary sampling units (SSUs). The two-stage cluster design reduces costs due to relatively small distances between the 12 households (8 households in Vienna) selected within each PSU while it ensures a sufficiently high number of PSUs within the individual stratum.*
>
> *Thus we ensure 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 170 strata, 422 PSUs and 4,436 households.*

6.3 Background – The (External) Data Sources

Given the definition of the target population, proper information on the geography and on households in Austria is needed. A perfectly representative draw of our sample presupposes that the target population is represented correctly by the frame population. 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). This theoretical optimum cannot be achieved in practice. The HFCS in Austria has been designed to meet this goal subject to the constraints of the available data sources. The following sections provide more details on the data on which sampling was based 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 data from the Post Office for the actual draw of SSUs (Secondary Sampling Units), namely households. 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 patterns of Austria, i.e. data on the NUTS-3 regions, and the enumeration districts (PSUs) from the population census 2001.[2] These enumeration districts are the smallest territorial units in Austria for which basic data characteristics are collected by default by Statistics Austria (each enumeration district contains around 410 dwellings on average).

In addition, we relied on the municipality directory of 2006 (to categorize the municipalities) and on the Micro census of 2009 (for the population data).[3] Both datasets are collected and provided by Statistics Austria. The Micro census 2009 indicates in particular the population size 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.

---

[2] *See www.statistik.at/web_de/statistiken/bevoelkerung/volkszaehlungen_registerzaehlungen/index.html (accessed on January 22, 2013).*

[3] *See Statistics Austria (2011a).*

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 on mailbox codes 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 mailboxes in the respective estate. Specifically, we used a commercial product called Adress.Certified developed by the Austrian Post Office. This address register captures information at the building level (including the information on street, house number and wether the estate is privately or commercially used). It can be bought in combination with a product called DATA.DOOR, which is a directory of mail delivery points (abbreviated PAC, for post office-certified address codes, in table 10), i.e. a directory of all addresses (mailbox codes) in Austria to which mail can be delivered. This information is available at a disaggregated level. In Austria, there are about 3,930,000 private mail delivery points and 233,000 commercial mail delivery points.

Thus, our starting point was some 3.9 million private mail addresses. Very few remaining commercial addresses had to be withdrawn after the first contact by the interviewer (e.g. if the interviewer arrived at the address and noted that this was an invalid address or a commercial building; see chapter 4) and were given weights of zero since they do not belong to the target population.[4]

The post office data we used do not reflect whether a given address is a household's main residence or not. Yet they provide a realistic picture of households and thus meet the HFCS requirement of reflecting actual living situations. In particular, unlike other data sources the post office data cover households at addresses that are registered only as a secondary home but fulfill the HFCS definition of a household. They have thus been included in the frame population because they have a post-certified address code.[5] At the same time, secondary homes inhabited by households whose main residence address was identifiable as such were removed in order to ensure that every household in Austria appeared only once in the remaining list of the post-certified address codes (see also chapter 7).

6.3.3 Profile.Address and IFES

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

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

---

[4] *Having deducted these commercial addresses, the total sum of weights is around 3.77 million; hence the total estimated number of households in Austria is 3.8 million.*

[5] *The post-certified addresses for some 3.9 million households compare with about 3.6 million household addresses documented by other sources (such as the Micro census based on the residence registry).*

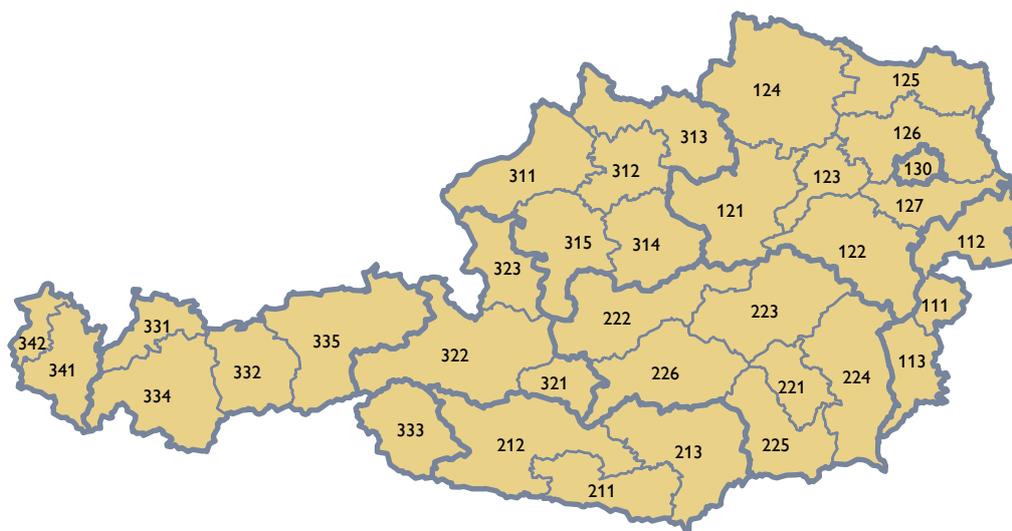[6] *For more details on the contact strategy, see section 3.4.*

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 (chart 5).

Chart 5

NUTS-3 Regions

Source: Statistics Austria.

With the exception of the capital city Vienna, each NUTS-3 region was divided further by the following 8 municipality population size categories.
- 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 the urban areas with the provincial capitals. Vienna stands out as it is the national capital and the only city with more than 1 million inhabitants; it was simply subdivided into its 23 districts.

This very fine stratification yielded 193 strata. Where the number or rather proportional share of households per stratum was too small to have allowed the selection of enumeration districts, individual strata were merged with neighboring strata to increase the share of households and hence insure the selection of PSUs within each stratum. This exercise left the HFCS with 170 strata that covered all households in Austria. See table 8 for the distribution of strata across provinces and municipality size categories.

Table 8

Allocation of Strata within the Sample

| Province | Municipality size (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 Mio	over 1 Mio	Total
Vienna	0	0	0	0	0	0	0	23	23
Lower Austria	7	7	7	7	5	5	1	0	39
Burgenland	2	2	2	1	1	0	0	0	8
Styria	5	5	5	5	3	1	1	0	25
Carinthia	3	3	1	3	2	1	1	0	14
Upper Austria	5	5	4	5	3	2	1	0	25
Salzburg	2	2	2	2	2	0	1	0	11
Tyrol	3	3	3	4	3	0	1	0	17
Vorarlberg	1	1	2	1	2	1	0	0	8
Total	28	28	26	28	21	10	6	23	170

Source: Statistics Austria (municipality directory 2006).

Each stratum contained about 50 PSUs on average, which in turn contained around 470 households on average (see also section 6.5.1).

6.4.2 Sample Size

As a rule, the variance of estimates based on the underlying data will be the smaller, the larger the sample is. At the same time, the cost of data collection increases with size. This trade-off has to be balanced to yield reasonably precise estimates subject to the given budget constraints. Furthermore, given the focus of the survey and the analyses HFCS was supposed to facilitate, 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 frame population) and to provide some insights into the regional differences within Austria. We know from previous OeNB surveys (Household Survey on Housing Wealth, 2008, and Survey of Household Financial Wealth, 2004) that at least 2,000 households need to be successfully interviewed and that the unit nonresponse rate can be expected to reach some 40% (with expected differences between Vienna and the rest of Austria).[7] With some room for extraordinary events, the HFCS survey was therefore designed to yield a sample of successful interviews of 2,700 households and a participation rate of 50% for Vienna and 65% for the rest of Austria. These participation rates are rough estimates based on the experience of past surveys.

The net sample of n = 2,700 was divided between the nine provinces, based on their population shares observed by Statistics Austria in the Micro census of 2009[8] (table 9). These figures, which corresponded to the targeted number of secondary sampling units (SSUs), were translated in a next step into gross samples of SSUs based on the estimated participation rates. Within each enumeration district, 8 households were subsequently drawn in Vienna and 12 households in the rest of

---

[7] The participation rate per stratum in the first HFCS wave will have to be taken into account in future surveys.

[8] Although the Micro census collected by Statistics Austria in 2009 is based on a different definition of households than the HFCS, we used this information as an approximation.

Table 9

Identification of the Number of PSUs to Be Drawn

Province	% of house-holds	Target sample	Target parti-cipation rate in %	Gross sample[1]	Number of households per PSU (enumeration district)	Number of PSUs to be drawn
Vienna	23	629	50	1,258	8	157
Lower Austria	18	499	65	767	12	64
Burgenland	3	84	65	129	12	11
Styria	14	376	65	578	12	48
Carinthia	7	179	65	275	12	23
Upper Austria	16	437	65	672	12	56
Salzburg	6	168	65	258	12	21
Tyrol	8	216	65	333	12	28
Vorarlberg	4	113	65	174	12	14
Total	100	2,700				422

Source: Statistics Austria (municipality directory 2006), HFCS Austria 2010, OeNB.

[1] *In sum, the targeted gross sample does not match 4,436 exactly due to rounding.*

Austria, with the respective difference reflecting the shorter distances between buildings in Vienna. Thus, we arrived at the PSUs to be drawn in each province.

In sum, the Austrian HFCS sample design produced 422 PSUs across all strata and a gross sample size of 4,436 households that were invited to participate in the HFCS (see box 2 in chapter 7 for information on the number of successfully interviewed households). Drawing possible substitute addresses was explicitly excluded in the HFCS to begin with, in order to ensure that all households of the gross sample would be interviewed with the same commitment so as to prevent data distortions (see also section 4.4.1).

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 (mailbox codes) from each PSU

6.5.1 First Stage

As the PSUs of the Austrian HFCS we chose the smallest territorial units in Austria, the so-called enumeration districts, of which there are 8,745 in Austria. On average, an enumeration district contains 410 households, but there are also PSUs with only very few households. These 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 8,407, which then had 470 households on average. Each PSU has the same probability to be drawn (irrespective of size) from a given stratum, i.e. there is no "probability proportional to size" selection – a PSU comprising numerous households has the same probability to be chosen within a stratum as a PSU with fewer households.

From the description above it is clear that the number of PSUs to be drawn in each province is determined a priori by the chosen sample size. To translate the numbers allocated to each province (table 9) into the desired number of PSUs within a given stratum, the former were divided up in proportion to the number of households in that stratum. For example, the 56 PSUs to be drawn in Upper Austria (table 9) were divided up among the 25 strata in this province according to population shares.

With the number of PSUs to be drawn per stratum having been established, they were selected in next step by simple random sampling without replacement.

6.5.2 Second Stage

With 422 PSUs having been randomly drawn, we turn to the second stage in which households are selected.

As outlined above, eight or twelve mail delivery points were randomly selected from each PSU drawn, depending on whether the PSU was domiciled in Vienna or elsewhere in Austria. In this process, every household in a given PSU had an equal probability of being selected into the sample. This procedure resulted in a gross sample of 4,436 households in Austria.

6.5.3 Practical Implementation

Table 10 illustrates how the second stage (following the draw of PSUs) worked: Austrian Post Office data were used to arrive at the appropriate mail delivery point, which essentially contains the address but not the holder of this address. To identify the household's name corresponding to the selected mailbox code the survey firm, IFES, relied on its databases or, where necessary, bought the corresponding name from a commercial provider called "ProfileAddress."

Table 10

Matching of Statistics Austria Data with Post Office and Commercial Data (Fictitious Example)

First stage		Second stage				Profile.Address/IFES
Statistics Austria		Austrian Post Office				
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 a personal letter signed by the governor of the OeNB. This letter contained information on the survey and an invitation to take part (chapter 3).[9]

---

[9] *See the online appendix for the invitation letter.*

6.6 Concluding Remarks

This chapter provides information on the sampling design specifically developed for the HFCS in Austria that has now been applied for the first time. As described, the survey is based on stratified clustered two-stage cluster random sampling, consisting of a random draw of primary sampling units (here, enumeration districts) from each stratum plus a random draw of households (mailbox codes 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 two aspects being particularly important:

- Sampling is based on a broad-based concept of the target population (without differentiating between main residences and second homes), as a result of which all households that correspond to the HFCS household definition have a positive probability of being selected.
- The stratification structure is very fine, as a result of which all segments of the Austrian population are represented in the survey.

At the same time, since one of the topics covered by the HFCS is the wealth structure of households one would want to oversample certain groups of the population such as indebted or wealthy households in order to improve the efficiency of estimates of these subgroups. However, the underlying information needed for such oversampling is unfortunately not available as yet.