Estimation Techniques in the German Labor Force Survey (LFS)

Dr. Kai Lorentz
Federal Statistical Office of Germany
Group C1 - Mathematical and Statistical Methods
Email: kai.lorentz@destatis.de

The German branch of the European Union‘s Labor Force Survey (LFS) is integrated in the German Microcensus

So we are in fact dealing with the German Microcensus and its methodology.

The German Microcensus – a brief introduction

- The Microcensus is an annual sample survey of the population of Germany that includes 1% of the populace, or about 800,000 people, every year.
- It is the largest annual household survey in Europe and thus a major source of information for policy makers and researchers.
- Its main objective is to provide data for official representative statistics about the population living in Germany. It also provides a reference frame for many smaller surveys.
- Questions on employment constitute an integral part of the survey. The German branch of the European Union‘s Labor Force Survey is integrated in the Microcensus.

The sampling design of the Microcensus (and LFS)

- The Microcensus is a stratified one-stage cluster sample with constant selection probability of 1% in each stratum.
- Within the strata, a variant of simple random sampling is performed.
- In order to improve the precision of estimates for annual trends, the samples of two consecutive years have a planned overlap of 75%. A selected cluster remains in the sample for four years.
- The clusters, so-called sampling districts, consist of small (on average about nine) collections of households. All persons that belong to a household in a selected cluster are interviewed.
- Depending on the size of the buildings involved, a sampling district may be made up of several, exactly one, or parts of a building.

The strategy for estimation in the Microcensus (and LFS)

- The Horvitz-Thompson estimator, i.e. direct estimation using the design weights, is too imprecise in many cases of interest.
- Bias due to nonresponse is to be expected and should be treated, even though unit nonresponse is moderate at about 5%.
- Consequently, our main objectives are the reduction of the sampling variance and the treatment of nonresponse bias.
- Our strategy to address both problems is to use available auxiliary information in conjunction with the technique of general regression estimation.
- We now describe in more detail the estimation process in the most important cases, the quarterly and annual estimates.

The strategy for estimation in the Microcensus (and LFS)

- Our goal is to produce an estimation weight for each household in the sample. These weights can then be applied to compute estimates for all variables.
- In order to obtain the same weight for all people in one household, we replace the auxiliary variables used in the computation by their household averages.
- We proceed in two steps. In each step, we use the available auxiliary information to determine new household weights that are better suited for estimation than the previous ones.
- In the following table we briefly characterize the two steps.

The strategy for estimation in the Microcensus (and LFS)

	Step 1 (Nonresponse Adjustment)	Step 2 (Adjustment to population totals)
Goal	Reduction of the bias due to known nonresponse	Reduction of sampling variance and bias due to unknown nonresponse
Initial weight	Sampling weight, i.e. inverse selection probability	Intermediate weight
Resulting weight	Intermediate weight	Final weight for estimation
Effect	Enlarge sampling weights to compensate for households that could not be reached.	Adjust the weights to better match the distribution of the auxiliary characteristics in the population.
Auxiliary Information	Data on nonresponse households collected by the interviewers.	Data taken from Germany‘s continually updated population projection.

The strategy for estimation in the Microcensus (and LFS)

- **In both steps, the new weights are determined by the method of general regression estimation.**
- **Details about the nature of the auxiliary information will be given later.**
- **The two step process just outlined is applied to the quarterly data. The weights for the annual estimates are obtained from the quarterly weights by simply multiplying by 4.**

The Horvitz-Thompson estimator

Consider a probability sample S from a population U labeled by 1, 2, … , N. Denote by π_k the probability that k is in S and by π_{kl} the probability that both k and l belong to S. We assume $\pi_k > 0$ and $\pi_{kl} > 0$ for all k and l.

- The Horvitz-Thompson estimator (a.k.a. π-estimator) $HT(t_y)$ of the population total $t_y = y_1 + y_2 + \ldots + y_N$ of a variable y is defined by

$$HT(t_y) = \sum_{k \in S} \frac{y_k}{\pi_k}$$

- $HT(t_y)$ is unbiased with respect to the sampling design. Its variance is given by

$$V(HT(t_y)) = \sum_{k,\, l \in U} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) y_k \, y_l$$

- The variance of $HT(t_y)$ is often unnecessarily large. In the presence of auxiliary information more precise estimators can be constructed.

The general regression estimator (GREG)

- **We consider a linear regression model**

$$y_k = A^t x_k + \varepsilon_k ,$$

where $E(\varepsilon_k) = 0$, $Var(\varepsilon_k) = \sigma_k^2$, and $Cov(\varepsilon_k,\varepsilon_l) = 0$.

- **We estimate the vector of regression coefficients A from the sample S by**

$$\hat{A} = \left(\sum_{k \in S} \frac{x_k \, x_k^t}{\sigma_k^2 \, \pi_k} \right)^{-1} \sum_{k \in S} \frac{x_k \, y_k}{\sigma_k^2 \, \pi_k}$$

- **Using $\hat{A}$, we can now create proxy values $\hat{A}^t x_k$ for y_k and estimate t_y.**

The general regression estimator (GREG)

- **We define the general regression estimator GREG(t_y) associated with the linear model defined on the previous slide by**

$$GREG(t_y) = \sum_{k \in U} \hat{A}^t x_k + \sum_{k \in S} \frac{y_k - \hat{A}^t x_k}{\pi_k}$$

with $\hat{A}$ as above.

- **With $\hat{y}_k = \hat{A}^t x_k$, $e_k = y_k - \hat{y}_k$, the GREG can be expressed more elegantly as**

$$GREG(t_y) = \sum_{k \in U} \hat{y}_k + \sum_{k \in S} \frac{e_k}{\pi_k}$$

The general regression estimator (GREG)

- **The following alternative representations of GREG(t_y) are often used:**

$$GREG(t_y) = \sum_{k \in U} \hat{y}_k + \sum_{k \in S} \frac{e_k}{\pi_k}$$

$$= HT(t_y) + \hat{\underline{A}}^t \left(t_{\underline{x}} - HT(t_{\underline{x}}) \right)$$

$$= \sum_{k \in S} \frac{g_k}{\pi_k} y_k$$

with the so-called g-weights

$$g_k = 1 + \left(t_{\underline{x}} - HT(t_{\underline{x}}) \right)^t \left(\sum_{l \in S} \frac{\underline{x}_l \underline{x}_l^t}{\sigma_l^2 \pi_l} \right)^{-1} \frac{\underline{x}_k}{\sigma_k^2}$$

The general regression estimator (GREG)

The GREG can be used with aggregate auxiliary information

- **In order to compute GREG(t_y), it is sufficient to know x_k for k in the sample S and the vector of totals $t_x = x_1 + \ldots + x_N$.**
- **This follows from the representation**

$$GREG(t_y) = HT(t_y) + \hat{A}^t (t_x - HT(t_x))$$

since the Horvitz-Thompson estimators as well as Â are computed from the sample alone.

- **This makes it possible to use the GREG in cases where x_k is not known for elements outside the sample but where the totals t_x are available, often from an external source.**
- **Example: If x_{ik} are binary indicators for membership in subgroups of the population, then t_x contains the group sizes, which might be known from a different source.**

The general regression estimator (GREG)

- The g-weights associated with a GREG can vary widely or even be negative. The description of a GREG as a calibration estimator can be used avoid such undesirable effects.
- We define the g-weights associated with a GREG and contained in an interval [a,b] by minimizing the distance measure under calibration equations and the additional constraint that $a \leq g_k \leq b$ for all k in S.
- In the Microcensus, such bounds on the g-weights are enforced to avoid extremely large or negative weights.

Application of GREGs in the Microcensus (and LFS)

Quarterly and annual estimates

- **As outlined before, we proceed in two steps, first addressing the (known) nonresponse and then adjusting the sample weights to better reflect the distribution of the auxiliary variables in the population.**
- **In both steps, the weights of elements in the sample are adjusted using a GREG. Thus we begin with the sampling weights $1/\pi_k$ (= 1/400 in the case of the quarterly estimates), then obtain intermediate weights in Step 1 and from these the final weights in Step 2.**

Application of GREGs in the Microcensus (and LFS)

Step 1: adjusting for known household nonrepsonse

- **The interviewers collect basic information on households in the sample for which full survey data could not be obtained.**
- **We use the information provided by the interviewers to model the (known) nonresponse in the Microcensus.**
- **We do so by setting up a multi-way ANOVA model and applying the associated GREG described above.**
- **On the following slide, we list the characteristics used in the model and the geographical level at which they are applied.**

Application of GREGs in the LFS

Model for compensation of known household nonresponse (Step 1)

Geographical level	Variables included in the model defining the GREG
State (NUTS-1)	•First year in the sample (yes/no) •Contained in the stratum of new buildings (yes/no)
Regional adjustment group	•Size of household (1, 2, 3 or larger) •Nationality (German/not German) •Status of residence (main residence/not main residence) *Additionally for households consisting of 1 person only:* •Gender (male/female) •Age (younger than 60 years/60 years or older)
Regional stratum subgroup	Total number of (private) households

Application of GREGs in the Microcensus (and LFS)

- A separate (simple) model is used to adjust for nonrepsonse of persons living in collective accomodations.
- To ensure significant bias reduction, variables that are highly correlated with nonrepsonse have been included in the model.
- In order to avoid small sample bias, variables in the model are automatically reduced to a higher geographical level when small cell counts (less than 10) are encountered.
- The quotient of the initial weights and the resulting intermediate weights are sometimes interpreted as response probabilities.

Application of GREGs in the Microcensus (and LFS)

Step 2: adjusting the intermediate weights to known population totals

- **In this second step, we employ auxiliary information about the population as a whole to decrease the sampling variance.**
- **We begin with the intermediate weights computed in Step 1 and apply a GREG to obtain the final weights.**
- **Again, the GREG we use is defined by a multiple-way ANOVA model. The characteristics considered in the model and the geographical level at which they are applied are given in the following table.**

Application of GREGs in the LFS

Model for computing the final quarterly estimation weights (Step 2)

Geographical level	Variables included in the model defining the GREG
State (NUTS-1)	•Age (unter 15 years/15-44 years/45 and older), additionally differentiated by gender •Nationality (German, Turkish, EU25, not EU25) × gender •Soldier (yes/no) •Total population „by month“
NUTS-2	•Nationality (German/not German) × gender
Regional adjustment group	Total population

The group sizes required to compute the GREG are obtained from Germany‘s continually updated population projection, the ministry of defense, and the central register of foreigners in Germany.

Application of GREGs in the Microcensus (and LFS)

- After Step 2 is performed, the obtained final quarterly weights can be used to compute estimates for all desired target variables.
- As mentioned before, the weights for annual estimates are obtained from the quarterly weights by multiplication by 4.
- Both steps in the estimation process are performed using the Bascula software by Statistics Netherlands.
- We also use SAS and the macro CLAN by Statistics Sweden for several estimation tasks.
- Reference: Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005 by *Afentakis and Bihler* (2005).

DESTATIS
wissen.nutzen.

YOU ARE WELCOME!

Dr. Kai Lorentz

Telephone: +49/(0) 611 / 75 25 89

kai.lorentz@destatis.de

www.destatis.de