A Guide on How Survey Weights of the GBP are Used (and Constructed)

This guide gives instructions on how to use the survey weights of the German Business Panel (GBP). We explain why weighting the data is important. We then give a short overview on how the survey weights are generated. Finally, we give advice on how to use the survey weights for the data analysis process.

1 Introduction: What are We Weighting for?

When conducting a survey, usually not all statistical units of a population are questioned about a topic or issue, but only a part of the population, the so-called sample. This is also the case for the GBP. In order to be able to make statements about the whole German firm population, we construct weights to make our sample as representative as possible on selected dimensions for the universe of active firms in Germany. The weighting of survey data refers to the procedure by which individual survey units (i.e. legally-independent firms in the GBP) are assigned a factor of their (relative) importance when estimating relevant statistics of the population, e.g., the mean revenue of the German firm population (Sand and Kunz, 2020). The aim of the weighting process is to get firm-level weights ω_i, which are used as a multiplicative factor for each observation i to make the estimated parameters of the sample resemble the true population parameters as closely as possible. According to Gabler et al. (2011), the main reasons for using weights are the following: (i) to reduce potential bias in estimated values of the population due to unequal inclusion probabilities of units, (ii) to account for non-response in the survey, and (iii) to adjust certain characteristics of units in the sample to the corresponding population distribution by subsequent stratification. Following Sand and Kunz (2020), we can therefore differentiate between weighting procedures that take into account the potentially different selection probabilities of individual units ((i); i.e. *design weighting*) and those that reduce possible biases due to non-response at the target unit level or that are applied for poststratification ((ii) and (iii); *calibration*).

(i) **Unequal Inclusion Probability** One potential reason for bias in the parameter estimates are unequal inclusion probabilities of individual units in the survey. To tackle this issue, *design weights* have to be used. They are defined as the inverse of the selection or inclusion probability π_i (i.e. $\frac{1}{\pi_i}$; see, e.g., the Horvitz-Thompson-Estimator for the use of *design weights* (Horvitz and Thompson, 1952)). Design

weighting is always necessary when there are different inclusion probabilities of individual survey units in the sample (Gabler et al., 2011).

(ii) Unit Non-response & (iii) Subsequent Stratification Respondent behavior can lead to bias in the estimates due to *unit non-response* (i.e. a unit does not participate) and *item non-response* (i.e. a unit does not answer specific questions). A bias in the estimates due to unit non-response as well as item non-response might occur when respondents and non-respondents differ systematically in regard to certain firm characteristics, e.g., if firms with higher revenue or employment size are more likely to take part in the survey (or have a lower number of item non-response) compared to firms with lower revenues and a lower number of employees. Additionally, as the sample is drawn from *Bureau van Dijk Orbis*[1]-database and not from the original German firm population, estimated parameters can be biased due to an incomplete sample frame (Sand and Kunz, 2020). To account for unit and item non-response and to reduce the bias and variance of the estimated parameters, *calibration* weighting (and *data imputation*) is recommended (Sand and Kunz, 2020) and used in the construction process of the weights described in Section 2.

---

1 https://www.bvdinfo.com/de-de/unsere-losungen/daten/international/orbis

2   Construction of the GBP Weights

To make valid statements about the whole firm population, we construct weights to make our survey representative compared to the universe of active firms in Germany on selected firm characteristics.[2] The following 4 dimensions are taken into account in the weight generating process: **industry Sector** (1-digit WZ08[3]), **employees**[4] (subject to social insurance contributions), **revenues**[5] and **firm location**[6].

    We use the *raking* method of iterative proportional fitting to calculate the survey weights (Kolenikov, 2014). As unit non-response can lead to an instability of the raking algorithm due to empty cells and, therefore, requires a data set with non-missing data items for our 4 calibration variables (industry Sector (1-digit WZ08), employees, revenues, firm location), we use the MICE algorithm (Multivariate Imputation by Chained Equations (van Buuren and Groothuis-Oudshoorn, 2011)) for multiple imputations (replacement values) of the calibaration variables before constructing the survey weights with the *raking* method.

---

2   In the weight generating process, we focus on completed interviews as this requires a lower number of imputations, as more survey items are answered by participants. Moreover, due to item non-response, observations with a lower progress are often not included in the data analysis. Assigning weights to these observations is therefore typically less relevant.

3   Klassifikation der Wirtschaftszweige (WZ 2008) of the German Federal Statistical Office. This classification is compatible with the Statistical Classification of Economic Activities in the European Community (Nomenclature statistique des activités économiques dans la Communauté européenne (NACE) Revision 2).

4   Employee Categories: 0 to 9 employess, 10 to 49 employees, 50 to 249 employees, 250 and more employees.

5   Revenue Categories: 0 to 2 million, more than 2 to 10 million, more than 10 to 50 million, more than 50 million.

6   We classify a firm as belonging to West Germany if the headquarter of the firm is located in the federal states Baden-Württemberg, Bayern, Berlin, Bremen, Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland and Schleswig-Holstein. If a firm is located in the federal states Sachsen, Sachsen-Anhalt, Thüringen, Brandenburg, Mecklenburg-Vorpommern, we categorise it as belonging to East Germany.

The preliminary data imputation is based on two imputation methods: multinomial logit regression (*polyreg*) and predictive mean matching (*pmm*). Following van Buuren and Groothuis-Oudshoorn (2011), we use the *polyreg* method for factor variables with fewer categories (≤ 15 categories; i.e. employees (subject to social insurance contributions), revenues and firm location) and a *pmm* for factor variables with many categories (> 15 categories; i.e. industry Sector (1-digit WZ08)). The primary motivation for this differentiation in imputation methods is the following (see van Buuren and Groothuis-Oudshoorn (2011), p.41 f.): The default imputation method for categorical variables in MICE is the *polyreg* method. When using the *polyreg* method, the MICE function will create a dummy variable for each category, leading to large matrices especially for variables with many categories, which can result in time-consuming imputations, instability of the algorithm and problems with empty cells. Therefore, for variables with many categories (i.e. more than 15 categories), the *pmm* method is recommended where the variable is coded as numeric. Moreover, the *pmm* method only imputes values which are already observed in the data, which preserves the original categories of the respective variable.

The calibration weights are based on the *raking* algorithm (Deming and Stephan, 1940; Kolenikov, 2014). The algorithm relies on the marginal distribution of industry sector (1-digit WZ08), employees (subject to social insurance contributions), revenues and firm location from the 2019 business register of the Federal Statistical Office (AFiD-Panel Unternehmensregister, URS) and iteratively fits survey weights, such that the marginal distributions of the four calibration variables industry sector (1-digit WZ08), employees (subject to social insurance contributions), revenues and firm location in the weighted sample approximate the respective marginal distributions of the full population as closely as possible (Kolenikov, 2014).

Finally, we account for observations with high survey weights due to cells which are highly under-represented relative to the full population. We trim the calibrated weights at the 5 and 95 percentile of the weight distribution, keeping the sum of the weights constant.[7] With this procedure, we can lower the variance of the survey weights and reduce the influence of extreme weights in the data analysis. However, we accept a slightly higher bias in our estimates compared to the population parameters as the lower variance comes at the cost of a higher bias (Sand and Kunz, 2020).[8] We assess the validity of our weights by calculating the design effect factor as proposed by Kish (1995). Our final survey weights for *Wave 1* have a design effect factor of 1.4.[9] Calibration weights are constructed specifically for each wave.

---

7   The algorithm for trimming the survey weights was thankfully provided by Dr. Matthias Sand from *GESIS* (Department Survey Design and Methodology).

8   This reasoning can be illustrated by a trade-off argument using the mean squared error (*MSE*), which consists of the squared bias of an estimator from the true population parameter and the variance of the estimate. By trimming, one reduces the variance of an estimator but at the same time induces an greater bias in the estimate compared to the true population parameter one wants to estimate.

9   For more information, please see https://backend.gbpanel.org/app/uploads/2021/10/Fragebogen.pdf

3   Application in Data Analysis

Like mentioned above, researchers should use the GBP survey weights when they want to make statements about the overall German firm population. By using the GBP survey weights, the results account for the population distribution of the dimensions industry sector (1-digit WZ08), employees, revenues and firm location.[10] The resulting weights are included in the data sets under the variable name *weight*. Data users can specify survey weights in their data analysis with commonly-used statistical software like *Stata* and *R*. In *Stata*, one can use the *svyset* command and prefix *svy* to specify the survey design for the data analysis. For the specification of the survey weights, one can use the following command: *svyset [pweight=weight]* .[11] Moreover, researchers can also specify the weights in *Stata* in single commands, e.g., *reg y v1 v2 v3 [pweight= weight], options*. In *R*, the package *survey* with the *svydesign* function can be used for data analysis. For more information, see Kaminska (2020).

In the following, we show how omitting weights can influence point estimates using data from *Wave 1* of the GBP. In the example below, we present estimates for the prediction of the employee distribution and the revenue change in Germany using the weighted and unweighted sample. The following code in *Stata* can be applied:

---

10 In general, it is not recommended to use the GBP weights when using research-specific sub-samples. With sample splits, the target population distribution may differ from the overall German firm population. In this case, we recommend that researchers construct their own survey weights, which fit the respective analysis.

11 We use *pweight* as we want to draw inference for the general population. *pweight* is the right command to use when the analysis is at the level of the individual observation (i.e. the legally-independent firm), whereas *aweight* assumes that the sample consists of group means. For certain commands such as *summarize*, *Stata* only allows *aweight*. Estimated mean and standard deviation using *pweight* and *aweight* are the same, but not the standard error (& confidence interval/ test statistics). In regression analysis, the *pweight* option gives the same results as the *aweight* option with robust standard errors (*vce(robust)*). Therefore, if the *pweight* option is not available, one can also use *aweight* in multi-variable analyses with the vce(*robust)* option. See https://www.stata.com/support/faqs and https://www.parisschoolofeconomics.eu/docs for a more detailed discussion of when to use *pweight* and *aweight*.

```
    *** Specifying sample design in Stata
    svyset [pweight=weight]


    *** Declare the respective employee categories
    gen ccgic6_new = .
    replace ccgic6_new = 1 if ccgic6 <=3
    replace ccgic6_new = 2 if ccgic6 == 4 | ccgic6 == 5
    replace ccgic6_new = 3 if ccgic6 == 6
    replace ccgic6_new = 4 if ccgic6 == 7 | ccgic6 == 8 | ccgic6 == 9


    *** Compare weighted and unweighted results

    ** Employees
    drop if ccgic6 < 0 \\ (drop observations with missing codes)
    svy: proportion ccgic6_new
    proportion ccgic6_new

    ** Revenue Change
    drop if ccd1==-9998 \\ (drop observations with missing codes)
    svy: mean ccd1
    mean ccd1
```

Table 1 presents the results of the unweighted and weighted estimated proportion and standard errors of employees in our sample. First, we observe that the unweighted distribution underrepresents firms with a lower number of employees and overrepresents larger firms compared to the German firm universe. Second, using weights approximates the firm population with regard to the employee distribution very well. As the weight generation is based on the marginal distributions of industry sector, employees, revenues and firm location, the analysis of the GBP data with weights will make estimates of survey variables (almost) equally distributed to these four dimensions for the German population. For example, when applied to the revenue change of firms during the Corona crisis in our sample (variable *ccd1*), we see that weighting the data increases the estimated revenue decrease from 18% to roughly 20%, i.e. accounting for the population dimensions of revenues, employees, economic sector and firm location increases the estimated population revenue change during the Corona crisis. This difference between the unweighted and weighted results might stem from differential non-response related to revenue change or from an incomplete sample frame as the sample is drawn among other sources from *Bureau van Dijk Orbis*-database and not from the original German firm population. Thus, it is recommended to take weighting into account to minimize the risk that wrong conclusions for the population are drawn due to potential estimation biases.

Table 1: Unweighted vs. Weighted Sample

	Unweighted Sample		Weighted Sample		Destatis (2019)[‡]
	Proportion	Std. Err.	Proportion	Std. Err.	Proportion
Employees (in %)					
0-9 employees	65.67	0.46	87.07	0.28	87.40
10-50 employees	25.43	0.42	10.28	0.25	10.06
50-249 employees	6.44	0.24	2.14	0.10	2.07
250 and more employees	2.46	0.15	0.51	0.04	0.47
Revenue Change (in %)					
Scale [-100, 100 (and more)]	-18.13	0.44	-19.64	0.58	

Note: [‡] Data comes from 2019 business register of the Federal Statistical Office (AFiD-Panel Unternehmensregister, URS). See also https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Unternehmen/Unternehmensregister/ Tabellen/unternehmen-rechtsformen-wz08.html.

References

Deming, W. E. and Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. The Annals of Mathematical Statistics, 11(4):427–444.

Gabler, S., Häder, S., Lehnhoff, I., and Mardian, E. (2011). Weighting for Unequal Inclusion Probabilities and Nonresponse in Dual Frame Telephone Surveys. Telephone Surveys in Europe. Springer, Berlin.

Horvitz, D. G. and Thompson, D. J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. Journal of the American Statistical Association, 47(260):663–685.

Kaminska, O. (2020). Guide to Using Weights and Sample Design Indicators with ESS Data. Technical report, Institute for Social and Economic Research, University of Essex.

Kish, L. (1995). Methods for Design Effects. Journal of Official Statistics, 11(1):55–77. Kolenikov, S. (2014). Calibrating Survey Data using Iterative Proportional Fitting (Raking). The Stata Journal, 14(1):22–59.

Sand, M. and Kunz, T. (2020). Gewichtung in der Praxis (Version 1.0). GESIS Survey Guidelines. GESIS - Leibniz-Institut für Sozialwissenschaften, Mannheim.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3):1–67.