

# Representativeness in Business Surveys and Linked Employer Employee Data: Evidence from the SOEP-LEE2 Data for Germany

Wenzel Matiaske\*, Christoph Halbmeier, Martina Maas, Torben Dall Schmidt  
*Helmut Schmidt University/University of the Federal Armed Forces Hamburg*  
 Hamburg, Germany  
 \*matiaske@hsu-hh.de

**Abstract** – A central goal of the dtec.bw project SOEP-LEE2 is to provide high-quality data for research in organisational, business, and economic studies. To this end, the project conducted two multi-year surveys among German establishments. The first survey, SOEP-LEE2-Core, is linkable to employee data from the Socio-Economic Panel (SOEP) and is characterized by a complex sampling procedure. The second survey, SOEP-LEE2-Compare, lacks this linkage possibility and has a simpler sampling design. In this article, we evaluate the sampling designs of the two surveys in term of their statistical representativeness. Our results indicate that the more complex sampling design of SOEP-LEE2-Core is susceptible to selective participation in some of the analysed dimensions, resulting in an overly large proportion of employees from the public sector. In contrast, the simpler design of SOEP-LEE2-Compare is less prone to this type of selection and yields more representative data.

**Keywords** – Survey design, data linkage, employee-first sampling, representativeness, Socio-Economic Panel (SOEP)

## NOMENCLATURE

FEA	Federal Employment Agency
HR	Human resources
IT	Information technology
SOEP	Socio-Economic Panel
SOEP-LEE2	Socio-Economic Panel Linked Employee-Employer Version 2

## I. INTRODUCTION

Socio-economic research increasingly analyses data that are obtained from various sources and linked together. Data linkage enriches the information available in single datasets, enhancing their research potential. If datasets with comparable content are linked, data linkage can also help to assess and improve data quality. The benefits of data linkage can come at the cost of losing statistical representativeness. Not every observation of an initially representative dataset may be linkable. If these observations have different characteristics than the linkable ones, the linked dataset is not representative any longer. As a consequence, analytical findings cannot be easily extrapolated to the initial population, weakening the generalizability of the analysis. The causes, magnitude, and consequences of selective linkage depend on the specific case

at hand and require empirical assessment to determine suitable remedies.

In this article, we assess the statistical representativeness of two establishment surveys that are central to the dtec.bw project SOEP-LEE2 [1], [2]. The first survey, SOEP-LEE2-Core, is a linked employer employee survey with a sampling design that follows the employee-first approach [3], [4]. The survey collects data from establishments that employ individuals taking part in the German Socio-Economic Panel study (SOEP-Core). Every employee in SOEP-Core was asked for contact data of their employer and, if provided, employers were invited to participate in the SOEP-LEE2-Core survey. The sampling procedure is more complex than those of typical establishment surveys and involves different stages of selective participation that can erode statistical representativeness.

We assess the representativeness of the SOEP-LEE2-Core data by comparing them to administrative data from the employment statistics of the Federal Employment Agency (FEA). We find that the SOEP-LEE2-Core data are representative in some, but not all, dimensions. Employees in SOEP-LEE2-Core distribute across East and West Germany in proportions that are comparable to the FEA data. Likewise, they present a good representation of the population in terms of the establishment size, although employees from very large establishments are somewhat underrepresented. In contrast, employees from the public sector are relatively abundant in the SOEP-LEE2-Core data, suggesting that the survey's sampling and linking approach is susceptible to this type of selection.

To understand the selection mechanisms further, we analyse the representativeness of the project's second survey, SOEP-LEE2-Compare. This survey was conducted among establishments drawn as a random sample from the FEA register data. Different to SOEP-LEE2-Core, the sampling procedure did not include the selective step of collecting employer contact information. Fewer steps of selection should, all else equal, lead to more representative data. Indeed, we find that the establishments surveyed for SOEP-LEE2-Compare resemble the population of the FEA data very well. Although the SOEP-LEE2-Compare data also contain a relatively large proportion of public sector establishments, the imbalance is substantially smaller than in SOEP-LEE2-Core.

Our findings are corroborated by an additional analysis of the SOEP-LEE2-Core survey in [5]. There we find that

employees working in the public sector are more willing to reveal employer contact information. In addition, public sector employers are more likely to take part in the subsequent survey. The two effects add up, affecting the representativeness of SOEP-LEE2-Core in this dimension.

In the next section, we provide an overview of the related literature on representativeness in business surveys. In section III, we describe the sampling procedure of SOEP-LEE2-Core and show how the collected data compare to the FEA data. In section IV, we examine the sampling design and representativeness of the SOEP-LEE2-Compare survey. In the last section, we provide some concluding remarks.

## II. THE REPRESENTATIVENESS OF BUSINESS SURVEYS IN THE LITERATURE

In order to generalize the results of a survey to the population, the participants in the survey must be representative of the population. The risk of systematic bias in the sample increases with the proportion of non-participants in a sample [6]. This risk has increased in recent years as the willingness to participate in business surveys has been declining for several years [7], [8]. However, the declining willingness to participate does not necessarily lead to sample bias, but only if non-participation depends on the characteristics of the establishment and is therefore not random. There are two forms of unit nonresponse: non-participation can be due to failure to contact the selected participant and non-participation can be a result of refusal to participate [9], [10]. Studies on non-response indicate that the probability of participation in a survey depends on several factors. These factors fall into four categories: 1. environment, 2. participant, 3. survey design, 4. interviewer [11], [12], [10]. The researcher can influence the reduction of non-response, particularly through the choice of survey design and the selection and training of interviewers [13], [14], [15]. At the environmental or organisational level, the researcher has no influence. Factors at the organisational level that have been identified as influencing non-response include the organisational structure, corporate culture, the internal distribution of information, and the sector [16], [17]. In this context, studies have shown that older and smaller establishments are more likely to participate in a survey than younger and larger establishments [8]. The fact that larger establishments are less likely to participate in surveys is due to the coordination effort required, as several departments and levels of hierarchy are involved in responding to the questionnaire, depending on the focus of the topic [18]. Willingness to participate in surveys also increases in good economic times [19]. With regard to the respondent, studies show that the likelihood of participating in a survey depends on the respondent's knowledge, qualifications, attitudes, and commitment [10], [20]. Respondents can include senior management as well as managers and staff from specific departments such as HR or IT. In some cases, especially when the survey covers several topics, it may be necessary to involve several employees with specific expertise, which can increase the risk of non-response [18]. It should be noted that sampling bias is not uncommon in surveys and that the investigation of non-response requires special attention in personnel and organizational research.

## III. SOEP-LEE2-CORE

### A. Sampling design of SOEP-LEE2-Core

The sampling design of the SOEP-LEE2-Core survey is closely linked to the SOEP-Core study. SOEP-Core is a panel study of German individuals that every year collects survey data about topics such as income, educational attainments, health, and well-being. SOEP-Core regularly receives new samples to compensate for panel attrition and, in some cases, to collect data from specific populations of interest. The various samples are weighted with survey weights to obtain data that are representative for the adult population living in Germany [21].

In its 2021 and 2023 waves, SOEP-Core provided the sampling frame for SOEP-LEE2-Core, following the employee-first sampling procedure [3], [4]. This consisted in asking dependently employed individuals in SOEP-Core about the name and address of their employer with the aim to subsequently survey them. Employees in establishments with fewer than five employees were not considered for sampling due to privacy reasons. In the 2021 wave, the procedure yielded a sample of 1,063 individuals with a successful linkable employer interview, starting from a total of 8,343 eligible individuals. The reduction in sample size was due to two factors. First, some employees had to be excluded because they refused to provide employer contact information. Second, the response rate in the establishment survey was 17.8 percent, resulting in a further reduction of the sample size.

### B. Representativeness of SOEP-LEE2-Core

The diminution of the initial sample begs the question of whether the remaining individuals still represent the total population of employees in Germany well. We assess this by comparing the SOEP-Core and SOEP-LEE2-Core 2021 data to aggregate statistics from the establishment file of the Federal Employment Agency (FEA data for short) [22]. The FEA data are a reliable benchmark because they are assembled from the obligatory employment notifications that employers send to the social security system. Therefore, they comprise all German establishments with at least one employee subject to social security contributions. For the comparison, we use the data as of 30 June 2021, and we exclude establishments with fewer than five employees beforehand. In this section, we use the person-level version of the data, which is the level SOEP-LEE2-Core was designed to be representative for.

The comparison is affected, to a minor degree, by small differences between the FEA data and the sampling frame of SOEP-LEE2-Core. SOEP-LEE2-Core includes civil servants, persons in marginal employment, and family members working in family businesses, whereas the FEA data are limited to employed persons subject to social security contributions [22], [23]. Moreover, SOEP-LEE2-Core includes individuals who lost their job within the six months prior to the SOEP-Core interview, while the FEA data do not contain these recent unemployed. Lastly, there are minor differences in the reference period. SOEP-Core interviews took place between May 2021 and January 2022, whereas the FEA data refer to 30 June 2021. The minor incongruencies are reflected in the total counts of employees in both datasets. The FEA data cover 31,339,827 employees in establishments with at least five employees, whereas SOEP-Core comprises a weighted number of 33,356,544 employees. Despite the differences, we consider the FEA data to be a suitable

benchmark due to its administrative nature and coverage of the grand majority of employees in Germany.

In Table I, we show how employees are distributed in the different datasets. The first column contains the distribution in the FEA data, the second column shows the distribution among those 8,343 SOEP-Core individuals eligible for sampling, and the third column refers to SOEP-LEE2-Core, that is, individuals with a successful and linkable establishment interview. The data shown in columns (2) and (3) are weighted with SOEP survey weights to compensate for the disproportional sampling of SOEP-Core.

TABLE I: DISTRIBUTION OF EMPLOYEES IN THE FEA, SOEP-CORE, AND SOEP-LEE2-CORE DATA IN PERCENT.

	(1)	(2)	(3)	(4)
	<i>FEA data</i>	<i>SOEP-Core</i>	<i>SOEP-LEE2-Core</i>	<i>Difference (1) - (3)</i>
<b>Establishment size</b>				
5-9	8.1	7.4	9.6	-1.5
10-49	25.6	25.2	25.0	0.6
50-249	30.0	27.0	31.8	-1.8
250-499	11.6	10.4	11.0	0.6
500 and more	24.7	26.1	21.2	3.5
Information missing	0.0	4.0	1.5	-1.5
Total	100.0	100.1	100.1	
<b>Region</b>				
East Germany incl. Berlin	18.3	19.1	20.2	-1.9
West Germany	81.7	80.9	79.8	1.9
Total	100.0	100.0	100.0	
<b>Industry</b>				
Agriculture, forestry, fishing	0.6	0.5	0.4	0.2
Mining, energy and water supply, sewage	1.8	1.7	2.0	-0.2
Food, beverages, tobacco, consumer goods	3.6	3.1	3.7	-0.1
Production goods, capital goods, and consumer durables	17.5	14.6	11.6	5.9
Construction	5.4	3.3	3.0	2.4
Wholesale trade	6.1	1.7	3.9	2.2
Retail trade	7.0	6.1	5.6	1.4
Transportation and storage	5.8	4.4	3.3	2.5
Accommodation and food service activities	2.5	1.5	2.1	0.4
Information and communication, financial and insurance activities, business, scientific and professional services	20.8	12.6	12.8	8.0
Public administration, defence, social security, education, human health, and social work	26.0	28.7	43.2	-17.2
Other services	2.9	2.0	4.1	-1.2
Information missing	0.1	19.6	4.3	-4.2

	(1)	(2)	(3)	(4)
	<i>FEA data</i>	<i>SOEP-Core</i>	<i>SOEP-LEE2-Core</i>	<i>Difference (1) - (3)</i>
Total	100.0	99.8	100.0	

Note: In case of SOEP-Core, the region of residence is shown, whereas for the other datasets, the region of the workplace is shown.

The upper pane of the table shows the distribution across the establishment size. In most size categories, the share of employees is very similar in the three datasets. For instance, in the category of establishments with 10 to 49 employees, the share is 25.6 percent according to the FEA data, and 25.0 percent in the SOEP-LEE2-Core data. In the category of establishments with 500 and more employees, the SOEP-LEE2-Core data have a share that is by 3.5 percentage points lower than that of the FEA data, indicating that employees from very large establishments are slightly underrepresented in SOEP-LEE2-Core. This is probably acceptable for most analyses and overall, the SOEP-LEE2-Core data are a good representation of the employee population in terms of the establishment size.

The distribution of employees across regions is again similar in all three datasets. 18.3 percent of employees work in East Germany according to the FEA data, 19.1 percent according to SOEP-Core, and 20.2 percent according to SOEP-LEE2-Core. The differences in percentage shares may relate to the fact that establishments in East Germany have on average fewer employees than West German establishments. Overall, the difference is relatively small, indicating a good regional balance in the SOEP-LEE2-Core data.

The imbalances are more substantial in the industry categories. Most prominently, it is in the category of “public administration, defence, social security, education, human health and social work” industries, where the largest differences between the datasets exist. According to the FEA data, 26.0 percent of employees work in these industries, whereas the SOEP-LEE2-Core data contains 43.2 percent of them. In SOEP-Core, the industry variable has a large share of missing values and provides evidence that is not fully conclusive. However, SOEP-Core shows a share that is much closer to the FEA data (28.7 percent), suggesting that the imbalances are introduced by the sampling design of SOEP-LEE2-Core. Indeed, in [5] we show that employees from the public sector are significantly more likely to report their employer contact information. In addition, public sector employers are also more likely than their private sector counterparts to participate in the subsequent establishment survey, resulting in significant differences in the industry composition. This is something that has to be taken into account when analysing the SOEP-LEE2-Core data.

#### IV. SOEP-LEE2-COMPARE

The SOEP-LEE2 project comprised a second establishment survey, SOEP-LEE2-Compare. It aimed at, among other things, assessing the design and representativeness of traditional establishment surveys with a simple sampling design compared to linkable surveys and employer-first sampling.

### A. Sampling design of SOEP-LEE2-Compare

Different to SOEP-LEE2-Core, SOEP-LEE2-Compare used traditional random sampling and forwent the additional stage of selection inherent in the employer-first sampling. Establishments were drawn at random from the establishment file of the FEA described in the previous section. Before sampling, establishments with less than five employees were excluded for comparability with SOEP-LEE2-Core that did not contain these establishments either. The remaining establishments were divided into 120 strata along three dimensions: 1) five categories of firm size 2) two regions, and 3) 12 categories of industries as depicted in Table I and II. Larger establishments were drawn with disproportional higher probability as we expected them to respond less often than small ones.

The drawn sample consisted of 25,973 establishments that subsequently were contacted and asked to participate in the survey. As in SOEP-LEE2-Core, establishments could choose between answering via telephone or web interview. The survey questions were identical to those of the SOEP-LEE2-Core survey and a total of 2,026 establishments responded, corresponding to a response rate of 7.8 percent.

### B. Representativeness of SOEP-LEE2-Compare

Does a simpler survey design lead to more representative data? To answer this question, we assess the representativeness of SOEP-LEE2-Compare by comparing the surveyed establishments to the total of establishments in the FEA data. The comparison is straightforward because establishments were sampled from these data, and we can compare them in the same dimensions that were used for stratification and are available for all establishments. To take into account the disproportional sampling, we weight the surveyed establishments with the inverse of the sampling probability. Then, establishments in SOEP-LEE2-Compare should show the same distribution of characteristics as all FEA establishments if their response to the survey is not selective.

The results of the comparison are presented in Table II. The upper part of the table shows that establishments surveyed for SOEP-LEE2-Compare have a very similarly distribution of the establishment size than those in the FEA data. For instance, 42.3 percent of the establishments in SOEP-LEE2-Compare belong to the category of establishments with 5 to 9 employees. The percentage is 43.4 in the FEA data. Similarly, there are only very minor deviations between the two datasets among larger establishments, indicating that the surveyed establishments are very representative in terms of size. The regional distribution of establishments resembles each other in the two datasets as well, underscoring the representativeness of the data.

In terms of industries, the results indicate that the SOEP-LEE2-Compare data are somewhat skewed to industries in the category “public administration, defence, social security, education, human health and social work”. They constitute 25.9 percent of the surveyed establishments relative to 19.6 percent in the FEA data. Similar to SOEP-LEE2-Core, the difference suggests that public sector establishments were more likely to consent to the survey. However, the difference is not as large as in SOEP-LEE2-Core, where the extra step of collecting employer contact information additionally distorted the data towards public sector employers. Yet, analyses based

on the SOEP-LEE2-Compare may take into account this distortion towards establishments from the public sector.

## V. CONCLUSION

In this article, we evaluate how the sampling design of different establishment surveys affects their respective statistical representativeness. The first survey, SOEP-LEE2-Core, is a linked employer employee survey whose sampling design involves two stages of selective participation. We show that the collected linkable data is representative in some, but not all dimensions, by comparing them to data from an administrative register. The second survey, SOEP-LEE2-Compare, has a simpler sampling design, but otherwise features the same design characteristics as SOEP-LEE2-Core. The collected data match the distributional characteristics of the register data better than SOEP-LEE2-Core, suggesting that a simpler sampling design is beneficial for statistical representativeness.

TABLE II: DISTRIBUTION OF ESTABLISHMENTS IN THE FEA AND SOEP-LEE2-COMPARE DATA IN PERCENT.

	(1)	(2)	(3)
	<i>FEA data</i>	<i>SOEP-LEE2-Compare</i>	<i>Difference (1) - (2)</i>
<b>Establishment size</b>			
5-9	43.4	42.3	1.1
10-49	44.2	44	.2
50-249	10.5	12	-1.5
250-499	1.2	1.1	.1
500 and more	.7	.6	.1
Total	100	100	0
<b>Region</b>			
East Germany incl. Berlin	19.7	20.1	-.4
West Germany	79.2	79.9	-0.7
Total	100	100	0.0
<b>Industry</b>			
Agriculture, forestry, fishing	1.3	1.6	-0.3
Mining, energy and water supply, sewage	1.2	1.1	0.1
Food, beverages, tobacco, consumer goods	2.8	3.9	-1.1
Production goods, capital goods, and consumer durables	8.3	8.2	0.1
Construction	11.3	8.6	2.7
Wholesale trade	7.9	6.1	1.8
Retail trade	11.8	9.5	2.3
Transportation and storage	4.8	2.9	1.9
Accommodation and food service activities	5.6	5.2	0.4
Information and communication, financial and insurance activities, business, scientific and professional services	20.7	20.0	0.7

	(1)	(2)	(3)
	<i>FEA data</i>	<i>SOEP-LEE2-Compare</i>	<i>Difference (1) - (2)</i>
Public administration, defence, social security, education, human health, and social work	19.6	25.9	-6.3
Other services	4.7	6.9	-2.2
Information missing	0.1	0.0	0.1
Total	100	99.9	0.0

Our results have a number of important implications for users of linked datasets as well as researchers who want to design linkable surveys. First, data linkage often requires the consent of the persons whose data should be linked. Giving consent can be selective, underscoring the importance of careful formulation of data linkage consents. Second, in many cases the linked sample will be smaller than the original sample. The selection process should be analysed thoroughly. Statistical weighting can rebalance the linked sample, which should be feasible, as rich data for modelling the selection process are typically available. Third, there are instances where none of the above is possible, which does not invalidate the analyses on the linked dataset. However, generalizations to the overall population should be made with great caution.

#### ACKNOWLEDGEMENT

This research and the SOEP-LEE2 project are funded by dtec.bw – Digitalization and Technology Research Centre of the Bundeswehr which we gratefully acknowledge. dtec.bw is funded by the European Union – NextGenerationEU.

#### REFERENCES

- [1] W. Matiaske, T. D. Schmidt, C. Halbmeier, M. Maas, C. Schröder, T. Böhm, A. S. Kritikos and S. Liebig, „The Socio-Economic Panel Linked Employer- Employee Survey Version 2 (SOEP-LEE2): Overview and Results from the First Wave,“ in *dtec.bw-Beiträge der Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg: Forschungsaktivitäten im Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr dtec.bw*, Bd. 1, D. Schulz, A. Fay, W. Matiaske and M. Schulz, Hrsg., Hamburg, 2022, pp. 378-382.
- [2] W. Matiaske, T. D. Schmidt, C. Halbmeier, M. Maas, D. Holtmann, C. Schröder, T. Böhm, S. Liebig and A. S. Kritikos, „SOEP-LEE2: Linking Surveys on Employees to Employers in Germany,“ *Jahrbücher für Nationalökonomie und Statistik Data Observer*, 2023.
- [3] M. Weinhardt, P. Jacobebbinghaus and S. Liebig, „Linked-Employer-Employee (LEE) Daten in der Organisationsforschung,“ in *Handbuch empirische Organisationsforschung*, S. Liebig, W. Matiaske and S. Rosenbohm, Hrsg., Wiesbaden, Springer Gabler, 2016, p. 1–28.
- [4] M. Weinhardt, A. Meyermann, S. Liebig and J. Schupp, „The Linked Employer–Employee Study of the Socio-Economic Panel (SOEP-LEE): Content, Design and Research Potential,“ *Jahrbücher für Nationalökonomie und Statistik*, Bd. 237, Nr. 5, pp. 457-467, 2017.
- [5] W. Matiaske, T. D. Schmidt, C. Halbmeier and M. Maas, „Nonresponse and Non-contact in Linked Employee-Employer Surveys: Evidence from the German Socio-Economic Panel Linked Employer-Employee Survey (SOEP-LEE2),“ *Working Paper*, 2024.
- [6] P. L. Alreck and R. B. Settle, *The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey*, Irwin Professional Publishing, 1995.
- [7] D. I. Pielsticker and M. R. Hiebl, „Survey Response Rates in Family Business Research,“ *European Management Review*, Bd. 17, Nr. 1, p. 327–346, 2020.
- [8] C. König and J. W. Sakshaug, „Nonresponse trends in establishment panel surveys: findings from the 2001–2017 IAB establishment panel,“ *Journal for Labour Market Research*, Bd. 57, Nr. 23, 2023.
- [9] R. M. Groves and E. Peytcheva, „The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis,“ *Public Opinion Quarterly*, Bd. 72, Nr. 2, p. 167–189, 2008.
- [10] F. Janik and S. Kohaut, „Why don't they answer? Unit non-response in the IAB establishment panel,“ *Quality & Quantity*, Bd. 46, p. 917–934, 2012.
- [11] R. M. Groves and M. P. Couper, *Nonresponse in Household Interview Surveys*, New York: John Wiley & Sons (Wiley series in survey methodology), 1998.
- [12] D. K. Willimack, E. Nichols and S. Sudman, „Understanding unit and item nonresponse in business surveys,“ in *Survey nonresponse*, R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little, Hrsg., John Wiley & Sons, 2002, p. 213–227.
- [13] National Research Council, *Nonresponse in Social Science Surveys: A Research Agenda*, R. Tourangeau and T. J. Plewes, Hrsg., Washington, D.C.: The National Academies Press, 2013.
- [14] D. K. Willimack and J. L. Dalzell, „An Examination of Non-Contact as a Source of Nonresponse in a Business Survey,“ *AAPOR - ASA Section on Survey Research Methods*, 2006.
- [15] P. Gleiser, J. W. Sakshaug, M. Volkert, P. Ellguth, S. Kohaut and I. Möller, „Introducing Web in a Mixed-Mode Establishment Survey: Effects on Nonresponse,“ *Journal of the Royal Statistical Society Series A (Statistics in Society)*, Bd. 185, Nr. 3, p. 891–915, 2022.
- [16] D. Tomaskovic-Devey, J. Leiter und S. Thompson, „Organizational Survey Nonresponse,“ *Administrative Science Quarterly*, Bd. 39, Nr. 3, pp. 439-457, 1994.
- [17] D. Tomaskovic-Devey, J. Leiter und S. Thompson, „Item Nonresponse in Organizational Surveys,“ *Sociological Methodology*, Bd. 25, pp. 77-110, 1995.
- [18] D. K. Willimack und G. Snijkers, „The Business Context and its Implications for the Survey Response Process,“ in *Designing and Conducting Business Surveys*, G. Snijkers, G. Haraldsen, D. Willimack und J. Jones, Hrsg., Hoboken, NJ, John Wiley & Sons, 2013, p. 39–82.
- [19] C. Seiler, „The determinants of unit non-response in the Ifo Business Survey,“ *ASta Wirtschafts- und Sozialstatistisches Archiv*, Bd. 8, p. 161–177, 2014.
- [20] B. R. Fulton, „Organizations and Survey Research: Implementing Response Enhancing Strategies and Conducting Nonresponse Analyses,“ *Sociological Methods & Research*, Bd. 47, Nr. 2, p. 240–276, 2018.
- [21] J. Goebel, M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder und J. Schupp, „The German Socio-Economic Panel (SOEP),“ *Jahrbücher für Nationalökonomie und Statistik*, Bd. 239, Nr. 2, pp. 345-360, 2019.
- [22] U. Brix und M. Fritsch, „Die Betriebsdatei der Beschäftigtenstatistik der Bundesanstalt für Arbeit,“ in *Das Gründungsgeschehen in Deutschland. Darstellung und Vergleich der Datenquellen*, M. Fritsch und R. Grotz, Hrsg., Heidelberg, Physica-Verlag, 2002, pp. 55-78.
- [23] Bundesagentur für Arbeit, *Grundlagen: Qualitätsbericht – Statistik der sozialversicherungspflichtigen und geringfügigen Beschäftigung. Version 7.12*, Nürnberg: Bundesagentur für Arbeit, 2022.