Bias-variance trade off on the use of non-response weights in inequality estimates

Josep Espasa Reig, LIS - Cross-National Data Center in Luxembourg, espasareig@lisdatacenter.org

Abstract

Income inequality estimates are typically computed using sampling weights produced at household-level. In recent years, some National Statistical Offices have started producing additional weights at individual-level. These contain an extra adjustment by the non-response propensity of individuals. We aim to understand how using person-level weights instead of household-level ones might: 1) Change wage inequality estimates; 2) Reduce their bias; 3) Impact their precision (i.e. Standard Errors).

The analysis uses 35 LIS datasets from Germany (originating from GSOEP) and 32 from the US (originating from CPS-ASEC). In both countries, wage inequality estimates increase slightly if person-level weights are used. In the case of Germany, this increase is more pronounced in the last decade. By comparing the survey estimates with the National Accounts we find evidence that person-level weights might be reducing survey bias. This seems to come at the cost of variance increase. The effective sample size decreases slightly when using person-level weights. In the case of Germany, this effect is again more noticeable in the last decade. Applying person-level weights also tends to increase the Standard Errors (SE) of inequality estimates. This is more pronounced for 'Poverty Gap' and 'Poverty Headcount' indicators.

Keywords: inequality, sampling weights, bias, variance, wages

1. Introduction

1.1. Individual non-response and person-level weights

Falling survey response rates have been a widespread trend that has prevailed across developed countries during the last decades (Beullens, et. al., 2022). Lower response rates are associated with higher non-response biases (Groves and Peytcheva, 2008), and tend to require larger weighting adjustments (Little, 1986).

Survey error has two main components: sampling error and non-sampling error (Assael and Keon, 1982; Biemer and Lyberg, 2003). Adjusting a survey sample with non-response weights aims to decrease the later by reducing the non-response bias. This tends to come at the cost of an increase in sampling error due to a loss in precision of estimates.

In this context, there is a risk of over-adjusting samples. In such case, benefits from bias reduction could be partially and even totally neutralized by the increasing standard errors (Gary, 2007).

In certain situations, using non-response adjustments could even be counter-productive for the quality of estimates. This could happen if the adjustment is modelled using auxiliary variables not related to the dependent variables in the estimate. Thus, if the variables used to model non-response were not related to wages, using non-response weights would be detrimental for estimating wage inequality indicators.

The multi-stage design of most household surveys means that unit non-response can happen both at household or person-level. Here we aim to explore how using person-level weighting instead of household-level adjustments can impact inequality estimates both in terms of bias and variance of these.[1]

1.2. Research questions

The Luxembourg Income Study Cross-National Data Center (LIS) is a research centre devoted to cross-national analysis of income and wealth data. After the data is acquired, harmonized and documented, LIS produces a series of estimates showing inequality measures for many different wealth and income outcomes, spanning a period of 50 years for 52 countries.

The data acquired by LIS typically contains sampling weights computed at household-level. In recent years, some data providers have started producing additional weights at person-level. These contain an extra adjustment by non-response propensity of individuals.

The impact of using person-level in weights instead of household-level ones on income, wages, poverty and inequality indicators is currently unexplored. This paper aims to answer the following research questions:

1. Do wage inequality estimates change when using person-level weights?
2. Do person-level weights decrease the bias of estimates?
3. Does using person-level weights increase the Standard Error of estimates?

We believe that answering these questions is relevant for LIS and other organizations producing similar estimates. This should help them make informed methodological decisions on how to compute better estimates (Biemer and Lyberg, 2003).

1.3. Methodological notes

For all the analyses in this paper we use individual wages. This is a person-level variable. Inequality estimates are frequently computed for wages and can be benchmarked against National Account figures. We prepare the data as we usually do at LIS for the comparable estimates in the LIS Key

[1] Person-level weights typically include the household-level non-response adjustment component as well. The difference with household-level weights tends to be that they also account for non-response at individual level.

Figures and DART indicators (Neugschwender and Espasa-Reig, 2022). We restrict the sample to individuals aged 16 to 64 that have a larger than 0 value in the wage variable. We also cap the bottom and top values to limit the influence of extreme outliers.

2. Weight comparison descriptive statistics

Using person-level weights instead of household-level ones can substantially impact the composition of the sample. Person-level weights give more importance to immigrants, men, people with low levels of education and younger individuals. Most of these trends are in line with the literature on survey non-response.[2]

In Germany, immigrants tend to make a larger proportion of the sample when using person-level weights. This is especially noticeable until 2010. After this year, the refreshment and immigrant boost samples were added to the GSOEP (SOEP 2022). In the German surveys from 1990s, using household-level weights means having around 20-30% less immigrants in the sample (e.g. from 14.9% to 11.4% in 1996). The effect is much smaller in the US, where the underrepresentation of immigrants is always below 0.5 percentage points (p.p.s.) (see Figure 1 in the Annex).

Both in Germany and in the United States, person-level weights increase the proportion of males in the sample. This trend has grown in recent years in both countries and is more noticeable in Germany. In the last available years, the proportion of males increases between 2 to 2.5 p.p.s. when the person-level weights are used instead of household-level ones in the German sample. For the US, the difference is only 0.5 p.p.s. (see Figure 2 in the Annex).

Using person-level weights increases the sample of younger individuals (ages 16 to 33).[3] The effect of weighting on the age composition can be relatively large (Figure 3 in the Annex). For example, in Germany 2015 the proportion of individuals aged 24 to 33 with person-level weights is 21.3%, while with household-level weights it is ~17%. Less substantial are the differences on formal education. The proportion of individuals with high formal education levels is larger when using household-level weights, although differences are relatively small (Figure 4 in the Annex).

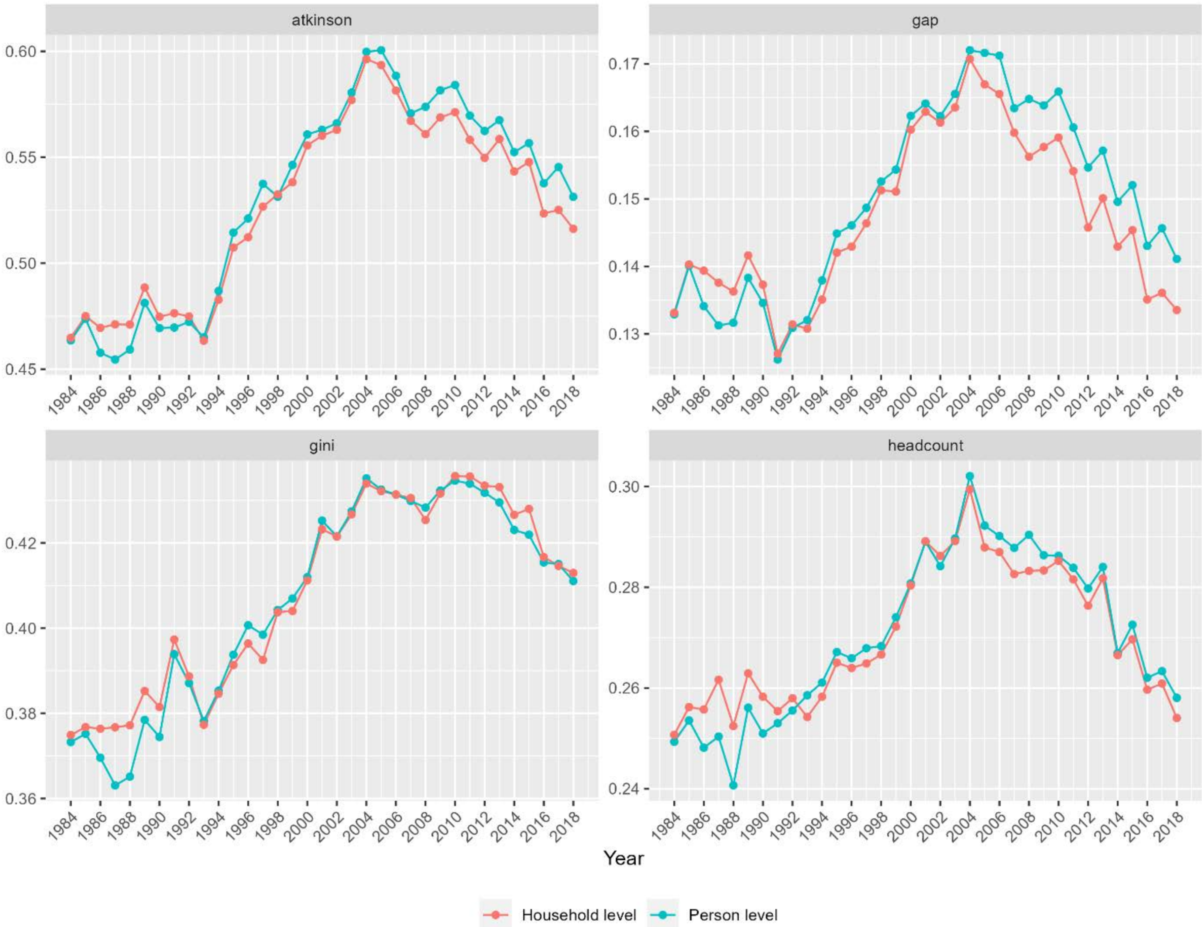
3. Weights and survey bias

3.1. Changes on indicators

[2] For example, the lower response propensity of men and individuals with lower levels of formal education seems to be well-documented (Curtin et al., 2000).
[3] With the exception of Germany until 1992.

When comparing different inequality indicators computed using both person-level and household-level weights, we see that inequality tends to be larger with the first. The figures below show the Gini Coefficient, the Atkinson Index, the Poverty Gap and the Poverty Headcount[4] for Germany (Figure 5) and United States (Figure 6).

Figure 5: Inequality indiators by weight - Germany



For some indicators, the increase in inequality observed in most estimates is consistent for the whole time series. The changes are relatively mild. For example, the Atkinson Index for Germany 2017 increases from 0.53 when applying household-level weights to 0.55 when using person-level ones.

[4] The proportion of people with labour income below 40% of the population median and equivalent to the 'Share of Low Income Workers (<50% of Median)' from DART (DART 2022).

Figure 6: Inequality indiators by weight - US

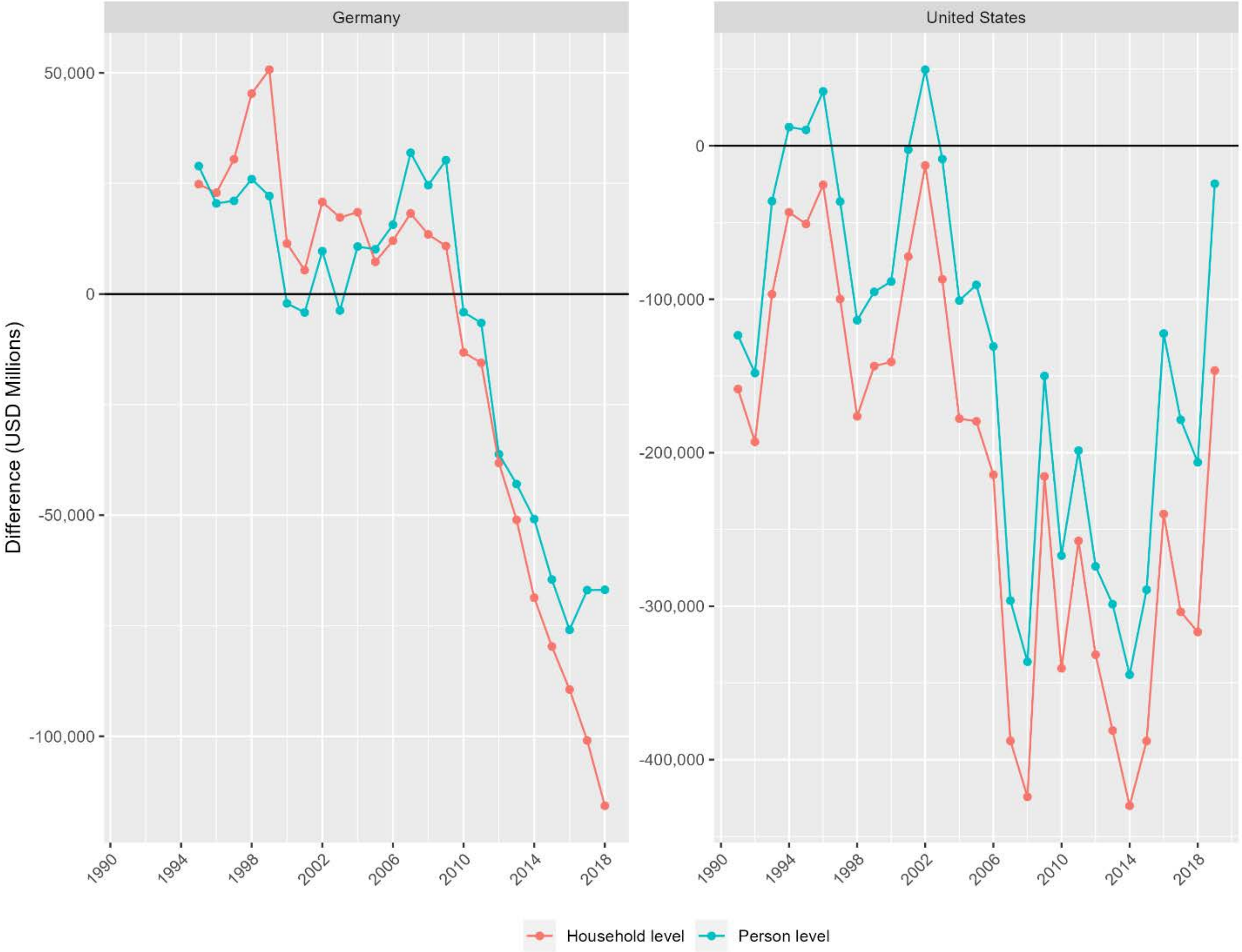
3.2. Bias reduction

In order to assess the possible increase or decrease of the survey bias with the changes presented above, we can compare the produced estimates with those from other sources. The inequality estimates shown can not be directly benchmarked to any source not using survey data. We suggest, therefore, to compare the National Accounts from economic data to the same concepts computed from our survey data.

Figure 7 below shows the differences between the OECD National Accounts 'wages and salaries' measures and our computations of these from survey data, with points above the horizontal line at 0 reflecting over-coverage and the points below under-coverage. We observe smaller differences between the survey data and the 'wages and salaries' figures from the OECD National Accounts[5] when using person-level weights.

[5] With the exception of Germany for the period between 2005-2009.

Figure 7: Differences with OECD National Accounts wages and salaries (NFD11P)



We believe that the reduction in the differences with OECD National Accounts when using person-level weights could provide evidence of a similar decrease in survey bias for the inequality estimates. This would make the indicators computed with person-level weights preferable to those computed with household-level ones.

4. Weights and sampling errors

As mentioned, the quality of survey estimates depends both on sampling and non-sampling errors (Assael and Keon, 1982; Biemer and Lyberg, 2003). To have a full picture of the possible improvements in estimates because of a reduction in survey bias using person-level weights, we also need to consider the variance inflation in estimates. Larger weighting adjustments might come at the expense of increased Standard Errors that can hinder the accuracy of estimates.

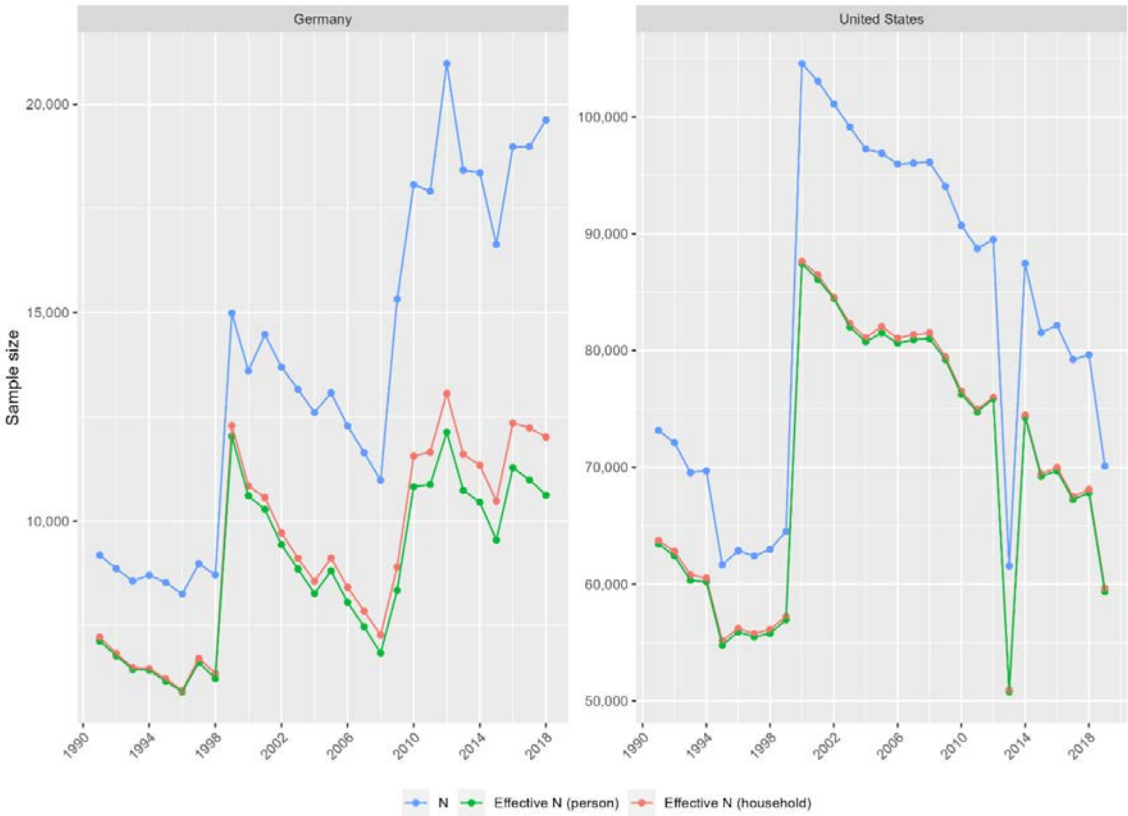
4.1. Sample size and Design Effects

Figure 8 and Figure 9 below compare the Effective Sample Size (n_{eff}) and the Design Effect (D_{eff})[6] of samples applying person and household-level weights. These two measures are related as:

$$n_{eff} = \frac{n}{D_{eff}}$$

We see that the Design Effect is larger when using person-level weights instead of household-level ones. This implies that the Effective Sample Size is smaller in the first case. The variance of the person-level weights is larger and the adjustments made by these are also bigger.

Figure 8: Sample size by weights



The likely reduction in bias comes therefore at the expense of a decrease in precision of estimates. This might be especially true for the last years of German GSOEP surveys, as the ratios between the Design Effects of the weights appear to be the largest.

[6] The Effective Sample Size is the sample size needed to achieve the current precision of the estimate (i.e. Standard Error) if the sample design was that of a simple random sample (i.e. not weighted). The Design Effect is computed as the sample size divided by the Effective Sample Size.

Figure 9: Design effect by weight

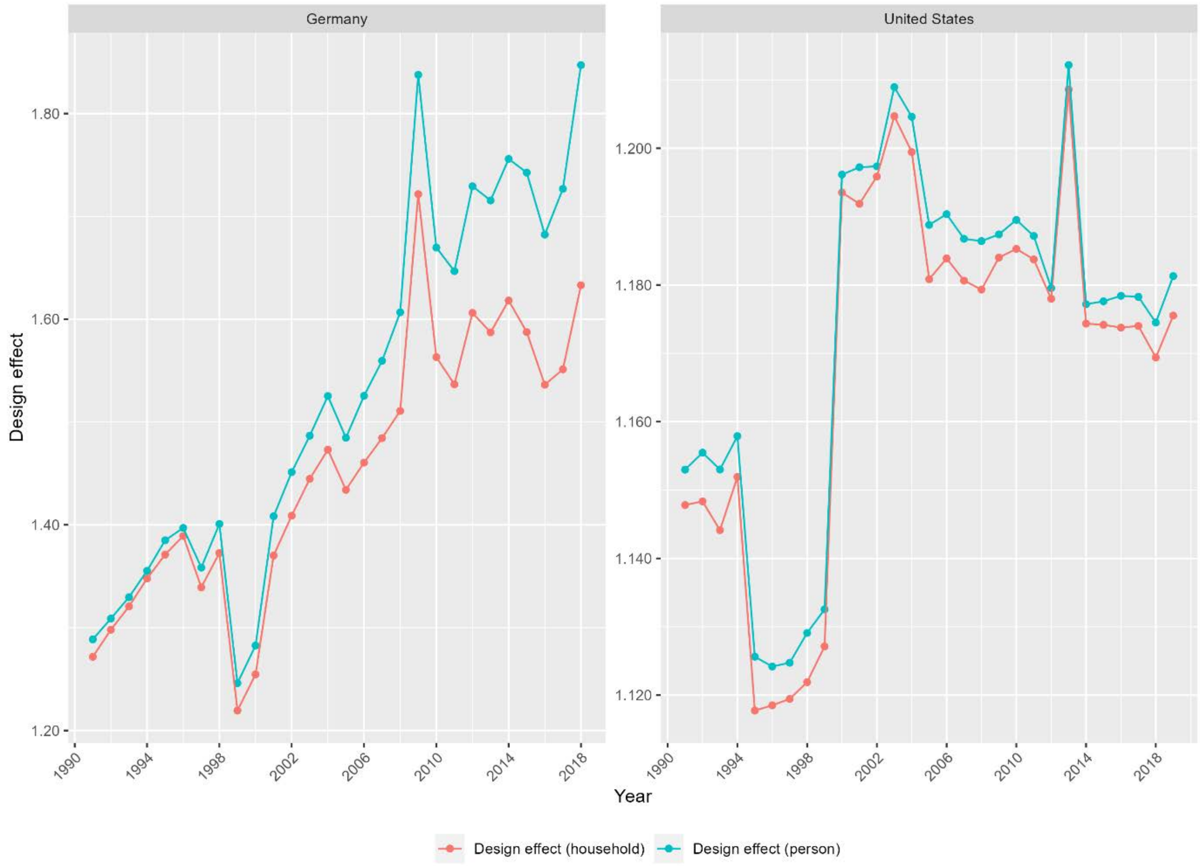


Figure 10 below also shows this decrease in precision. This figure shows the ratio between the indicators computed with person-level weights over household-level ones. For example, the Poverty Gap in Germany 2017 had a Standard Error 1.18 times larger when using person-level weights.[7] The blue line superposed is a smoothing spline showing the overall trend across years.

The data shows that, when using person-level weights, there is a clear increase in the Standard Errors of the inequality indicators for the last years of German data. This is in line with the substantial inflation in the Design Effects for the same period. The most affected indicators seem to be the Poverty Gap and the Atkinson Coefficient. The Gini Index, on the other hand, might be less affected.

[7] With a SE of 0.003476 when using person-level weights and 0.002946 when using the household-level ones.

Figure 10: Ratio of increase of Standard Errors of inequality indicators

5. Conclusion

Using person-level weights with individual non-response components instead of only household-level ones produces changes in inequality estimates. These changes are likely reducing survey bias at the expense of increasing sampling error. The later effect is especially large for some indicators of the most recent German survey years.

To understand if the reduction in bias is worth the inflation in variance we would suggest performing Monte Carlo simulations, which are beyond the scope of this paper. These could mimic different levels of bias reduction combined with the increase in Standard Errors reported above.

6. References

Assael, H. and Keon, J., 1982. Nonsampling vs. Sampling Errors in Survey Research. Journal of Marketing, 46(2), p.114.

Beullens, K., Loosveldt, G., Vandenplas, C. and Stoop, I., 2022. Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts? Available at: https://surveyinsights.org/?p=9673. [Accessed 16 April 2022].

Biemer, P. and Lyberg, L., 2003. Introduction to Survey Quality. 1st ed. John Wiley & Sons.

Curtin, R., Presser, S. Singer, E., 2000. The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly, 64(4), pp.413–428.

DART.lisdatacenter.org, 2022. DartFrontend. [online] Available at: https://dart.lisdatacenter.org/dart [Accessed 26 April 2022].

Gary, P., 2007. Adjusting for Nonresponse in Surveys. In: J. Smart, ed., Higher Education: Handbook of Theory and Research, vol 22, 1st ed. Dordrecht: Springer.

Groves, R. and Peytcheva, E., 2008. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. Public Opinion Quarterly, 72(2), pp.167-189.

Little, R., 1986. Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review / Revue Internationale de Statistique, 54(2), p.139.

Neugschwender, J. and Espasa-Reig, J., 2022. Data Access Research Tool (DART). LIS Technical Working Paper. No.12. [online] Luxembourg. Available at: https://www.lisdatacenter.org/wps/techwps/12.pdf [Accessed 24 April 2022].

SOEP Companion documentation. 2022. The SOEP Samples in Detail [online] Available at: http://companion.soep.de/Target%20Population%20and%20Samples/The%20SOEP%20Samples%20in%20Detail.html [Accessed 25 April 2022].

7. Annex

Figure 1: Proportion of immigrants by weight

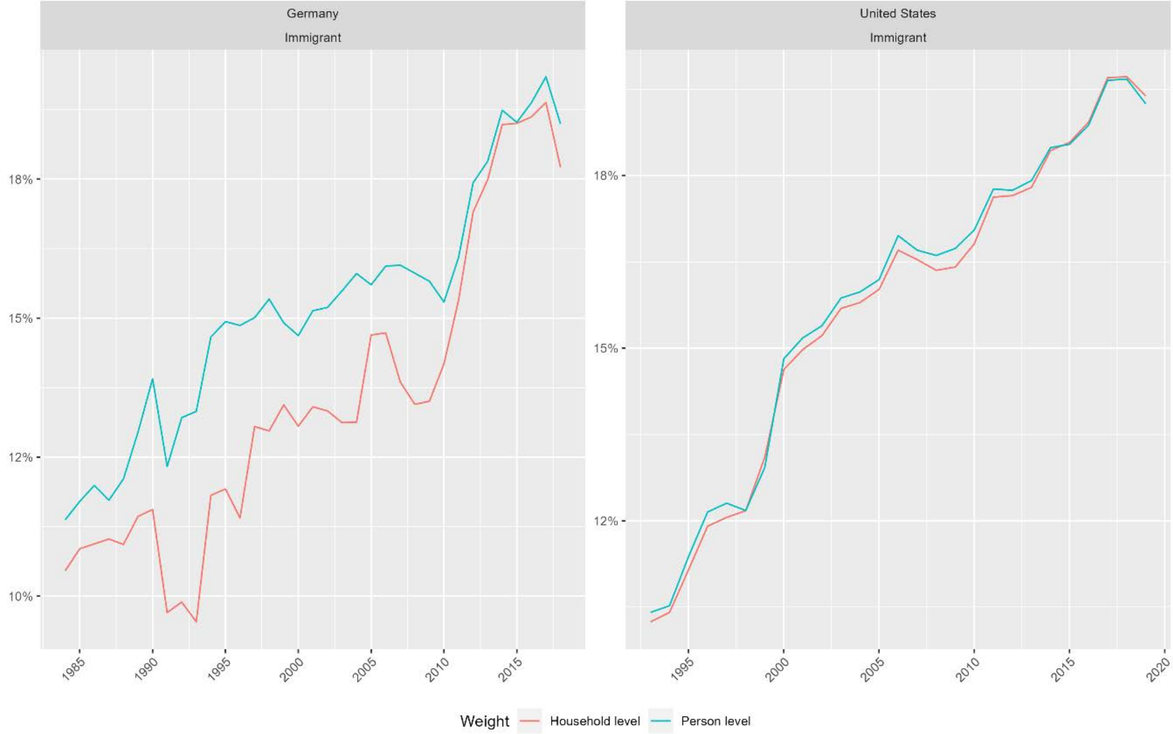


Figure 2: Proportion of males by weight

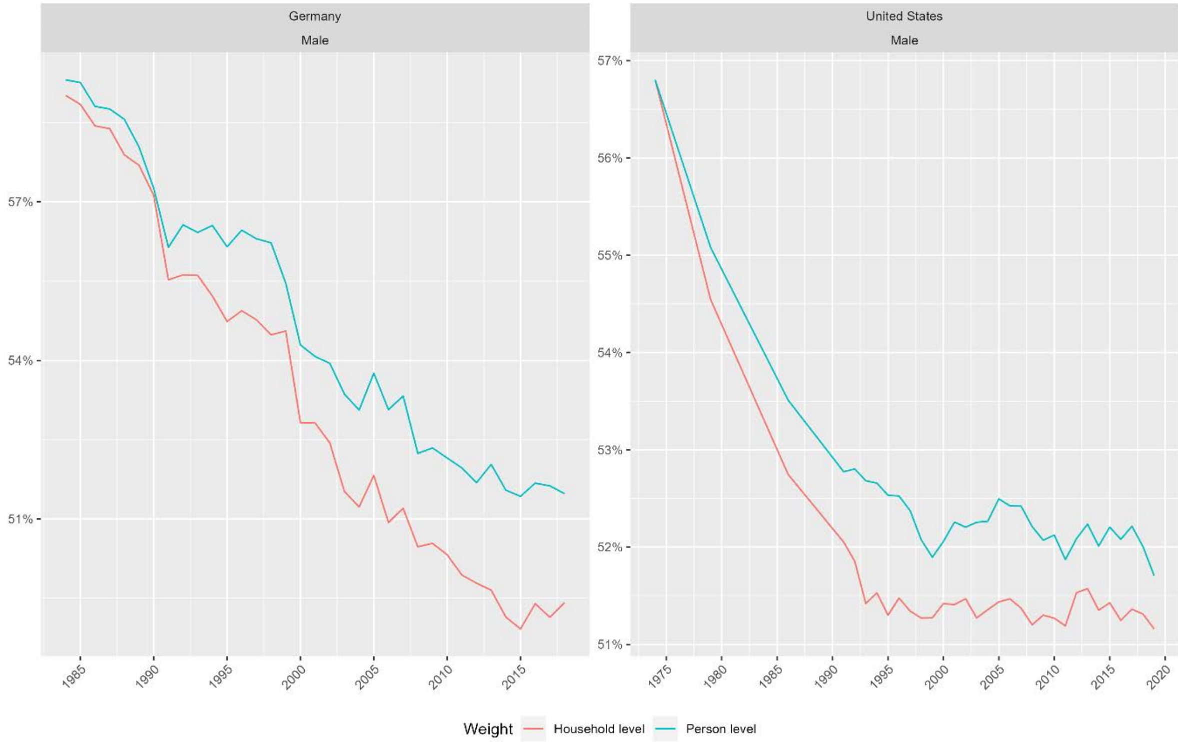


Figure 3: Proportion of age groups by weight

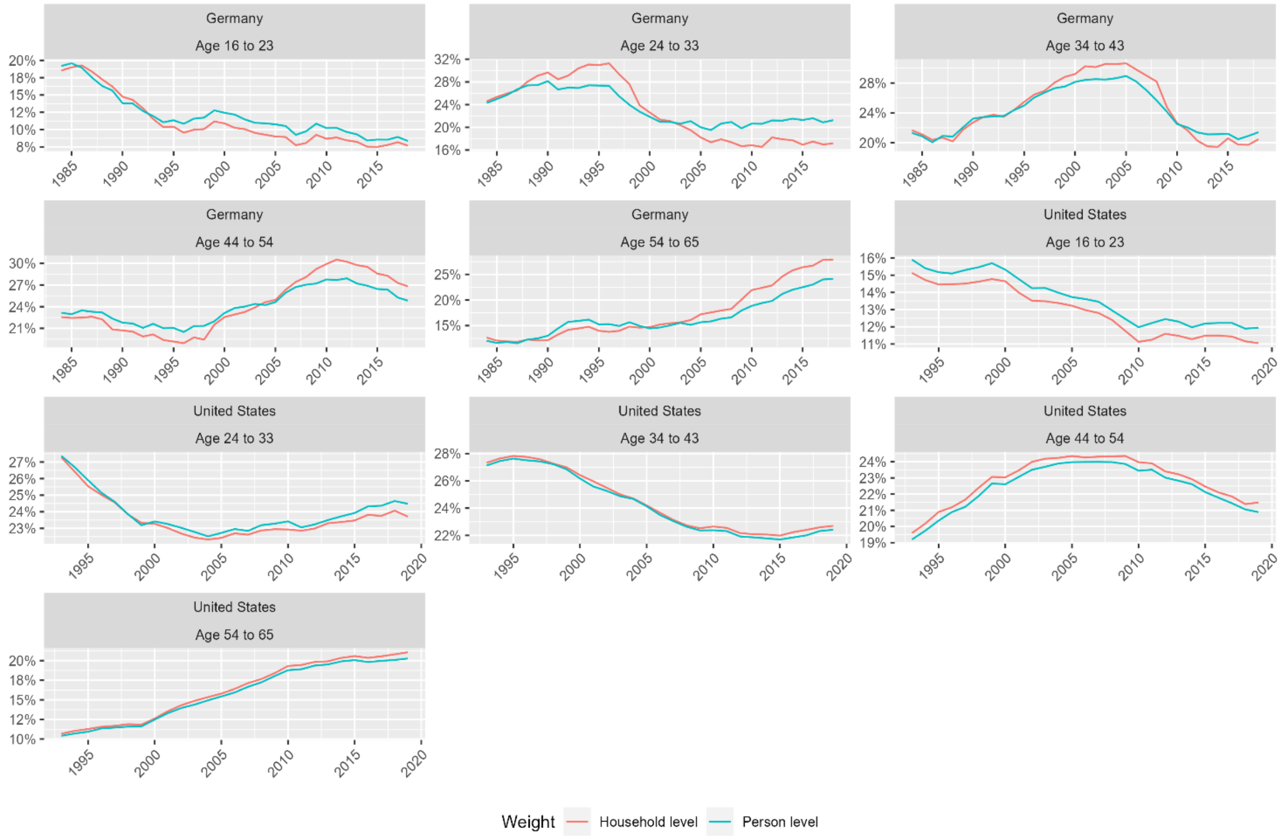


Figure 4: Proportion of education levels by weight