

## The use of microgeographic data to study nonresponse in the German General Social Survey (ALLBUS) 2012 - A replication and extension of a previous analysis with 2010 data.

Michael Blohm & Achim Koch

GESIS – Leibniz Institute for Social Sciences

The study of response and nonresponse requires data both about respondents and nonrespondents of a survey. There are various sources which might provide such information: individual data from (population) registers, aggregate level data from administrative records (e.g. for small areas like city blocks), or observations by interviewers (e.g. on the type of housing and the immediate vicinity of a household).

Another source of information are microgeographic data from private vendors which are mainly offered for the purpose of direct marketing. These data include a wide array of variables, often on different levels of aggregation (e.g. on the level of street segments, or communities).

Basically, these data can be used both for postsurvey adjustments for nonresponse and for the purpose of control and management of data collection, e.g. by implementing tailored or targeted survey protocols. The latter requires, that the data are available before fieldwork for a survey starts. To be an effective means for postsurvey adjustments, microgeographic data have to be predictive of both the response propensity of sample units **AND** the key survey variables.

The present paper deals with five questions:

1. What predictive power do microgeographic data have in explaining survey participation?
2. What predictive power do microgeographic data have in explaining key survey variables of interest?
3. How valid and reliable are the microgeographic data used?
4. How strong is the predictive power of microgeographic data compared to other auxiliary information like population register - or contact form data?
5. How stable are the relationships we found in an replication study (ALLBUS 2012)

To answer these questions we merged data from the German General Social Survey (ALLBUS) 2010 and the ALLBUS 2012 with microgeographic data. The ALLBUS is a multitopic survey fielded every second year since 1980 (i.e. repeated cross-sections). The survey uses a two stage sampling design with communities as the first stage units, and individuals from the community lists of residents as the second stage units. As a side product of this design, a few variables are available in the sampling frame for all target persons (gender, age, nationality, urbanicity, region).

The microgeographic data we used come from a private company called **microm**. **microm** offers microgeographic information for direct marketing purposes on different levels (Address/building, comprising 2.3HHs on average; street segments, comprising 27.3 HHs on average; voting districts, comprising 450 HHs on average; communities). The data cover several topical areas, like e.g. neighborhood information, mobility, purchasing power, socio economics, socio demographics, attitudes, consumer preferences, media use etc.

The data originate from various sources. Most data are estimations und approximations based on microgeographic segmentation procedures for the addresses. The data are linked with the ALLBUS data at the address level – (buildings). Due to data protection issues the data are pooled (at least 5 households) for small buildings. A potential drawback of the data is that its sources and the way individual variables were generated is not published in detail.

A) In the following we first analyse the relationship between these microgeographic information and survey participation. We distinguish between the two main phases of survey participation, contact and cooperation, and check both bi- and multivariate relationships between contact, cooperation and response and the various microgeographic variables. We compare these results with similar analyses, using the data which are available from the sampling frame of ALLBUS. The question we try to answer is whether the microgeographic variables provide any explanatory power in predicting survey participation, in addition to the frame variables which are available as a standard feature.

B) Subsequently we investigate the relationship between the microgeographic data and a selection of key survey variables from ALLBUS. The key survey variables selected cover a wide array of topics, so that the results might be useful also for other surveys. Again, we compare the strength of the relationships between the microgeographic variables an the key survey variables with the strength of the correlations between the available frame variables and the same set of key survey variables.

C) In a third step we analyse how well some of the microgeographic variables compare with similar data, either collected in the ALLBUS interview or collected by the interviewers. These analyses shed some light on the validity of the microgeographic data.

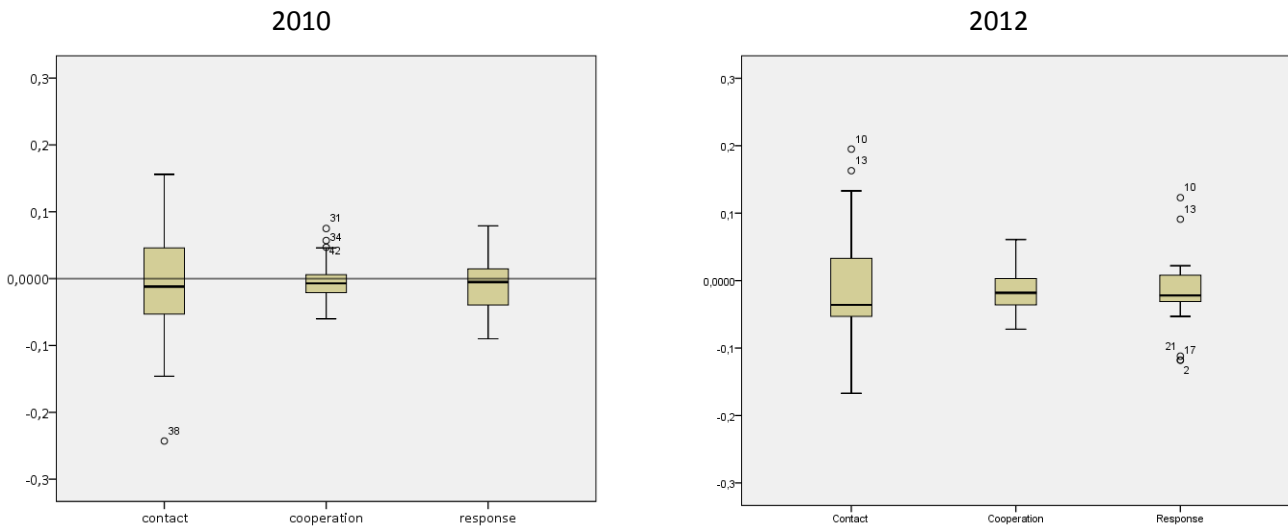
D) Finally, we compare unadjusted estimates for key survey variables with estimates applying different auxiliary variables for nonresponse adjustment (register, contact form, microm) and poststratification procedures (according to the external criterion microcensus).

Data:

Frame and auxiliary information (Z)		Key survey variables (Y)
(Z3) Microgeographic data: Level of house/building (A_...): neighborhood typology socio economic targets advert(ising) refuser Level of street sectors (ST_...): purchasing power Level of vote district (VD_...): unemployment milieu typology Level of community (C_...): proportions denominations	(Z1) Frame, Population registers (R_...): R_sex, R_age, R_citizenship, R_urbanicity, R_Region (West-, East Germany)  (Z2) Contact form data (CF_...): # of contact attempts, type of building, intercom, condition of building, estimation of social class	(Y) Data set of completed interviews (ALLBUS 2010, ALLBUS 2012):  income, education, unemployed, HH-size, church attendance social class, economic situation, health status political interest, membership in union/party, trust in people, vote intention

Main Results:

**A) Figure 1a,b:** Correlations between all frame/auxiliary information and survey participation (Pearson's *r*, grouped by contact, cooperation and response)



**Table 1a,b :** Response propensity models (Nagelkerke's Pseudo R<sup>2</sup>), separately for different sources of auxiliary data.

2010

	Dependent Variable		
	Contact	Coop.	Response
<b>Register</b> R_age, R_sex, R_region, R_citizenship, R_urbanicity	0.103	0.013	0.020
<b>Contact Form</b> CF_contacts, CF_building, CF_condition, CF_intercom, CF_socialclass	0.193	0,061	0.068
<b>microm</b> A_nationality, A_family composition, A_parmacy products, A_garden owner, A_advert refuser, A_pet owner, A_Lifecycle, A_women mag., ST_prop. renters, ST_purchasing power, VD_Urbanicity, C_prop. catholics	0.087	0.011	0.021

2012

	Dependent Variable		
	Contact	Coop.	Response
<b>Register</b> R_age, R_sex, R_region, R_citizenship, R_urbanicity	0.112	0.017	0.029
<b>Contact Form</b> CF_contacts, CF_building, CF_condition, CF_intercom, CF_socialclass	0.145	0,036	0.069
<b>microm</b> A_family, A_exclusive area, A_prop. academics, A_gardenowner, A_fax A_mobilephone, A_pc A_age, A_answermachine A_housetype, A_advertrefuser, C_prop. catholics	0.100	0.010	0.030

**Table 2a,b :** Response propensity models (Nagelkerke's Pseudo R<sup>2</sup>), stepwise inclusion of auxiliary data of different sources.

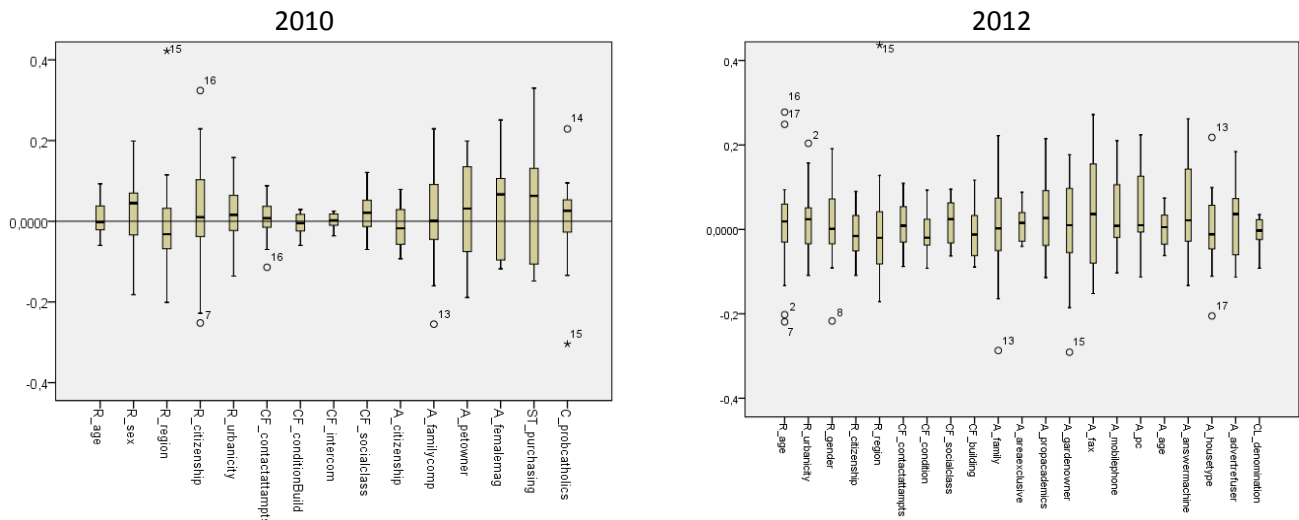
2010

	Dependent Variable		
	Contact	Coop.	Response
<b>(1) Register</b>	0.103	0.013	0.020
<b>(2) Register +Contact Form</b>	0.233	0.073	0.084
<b>(3) Register +Contact Form +microm</b>	0.237	0.080	0.092

2012

	Dependent Variable		
	Contact	Coop.	Response
<b>(1) Register</b>	0.112	0.017	0.029
<b>(2) Register +Contact Form</b>	0.214	0.052	0.086
<b>(3) Register +Contact Form +microm</b>	0.232	0.055	0.090

**B) Figure Ba,b: Correlations between microgeographic + Contact form data (CF...) + population register data (R...), which have significant effects in response propensity models (A), and key survey variables. (Pearson's r, grouped by microgeogr. CF and register variables)**



**C1) Table Ca: Validation: Relationship between microgeographic data and similar ALLBUS 2010 variables**

ALLBUS 2010 Microgeo. Var.	Age	Family compo.	Type of Building	Immi-grant	Un- employed	HH- Income	Social class.	Education
B_Age (head of HH)	.313	-.120	-.133	-.109	-.066	.096	.080	
B_Family composition		.255	-.627	-.158	-.119	.155	.091	-.060
B_Type of building	-.065	-.192	.787	.156	.083	-.161	-.111	
B_Prob. Immigrant	-.117		.298	.217		-.076	-.056	
VD_Prob. Unemployed	-.037	-.078	.385	.076	.153	-.252	-.190	
ST_Purchasing power	.084	.098	-.439	-.126	-.085	.249	.171	.092
B_Status	.079		-.308	-.077	-.133	.330	.279	.179

Weighted by design weight (different selection probabilities in West-, East-Germany); Only significant correlations  $p < .01$

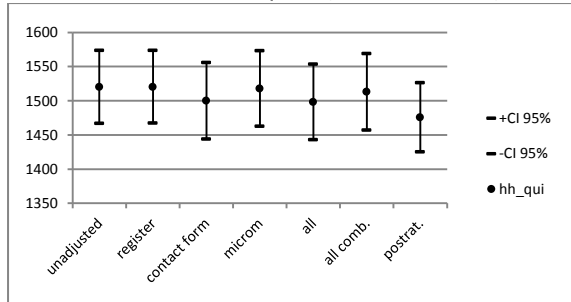
**C2) Table Cb: Validation: Relationship between microgeographic data and similar ALLBUS 2012 variables**

ALLBUS 2012 Microgeo. Var.	Age	Family compo.	Type of Building	Immi-grant	Un- employed	HH- Income	Social class.	Education
B_Age (head of HH)	.269	-.104	-.133	-.100	-.044	.049	.047	-.044
B_Family composition		.291	-.645	-.097	-.111	.109		-.113
B_Type of building	-.085	-.241	.808	.129	.114	-.137	-.062	.065
B_Prob. Immigrant	-.068	-.075	.322	.166				
VD_Prob. Unemployed		-.106	.358	.059	.174	-.219	-.148	.008
ST_Purchasing power	.084	.079	-.368	-.082	-.109	.234	.162	.089
B_Status	.064	.041	-.271	-.036	-.189	.304	.274	.172

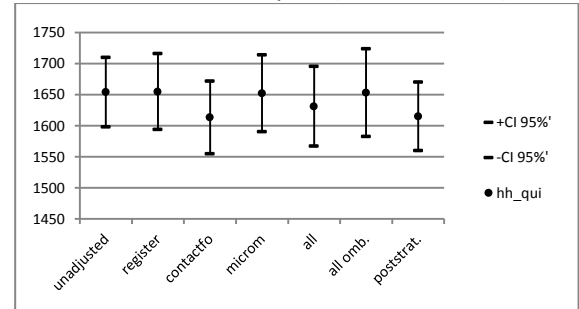
Weighted by design weight (different selection probabilities in West-, East-Germany); Only significant correlations  $p < .01$

**D) Figure Da-h: Effects of weighting:** Comparison of unadjusted estimates for key survey variables with estimates applying different auxiliary variables for nonresponse adjustment (register, contact form, microm) and poststratification procedures (according to the external criterion microcensus).

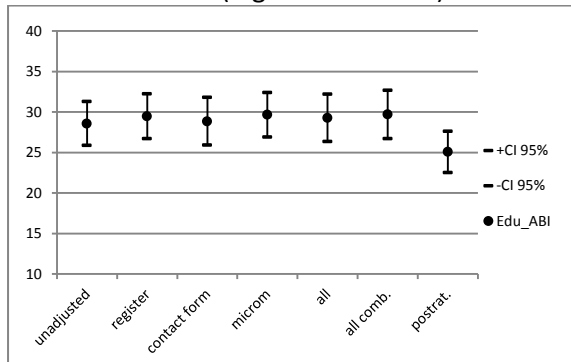
**Household income, equiv. (Mean in Euro) 2010**



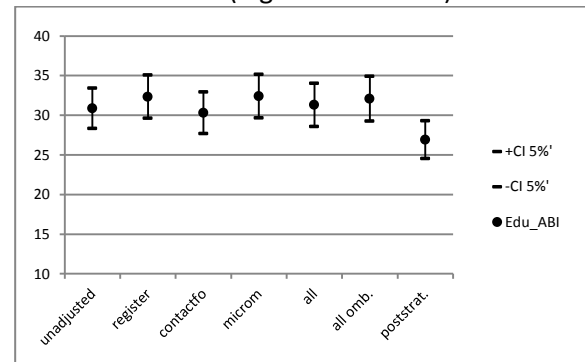
**Household income, equiv. (Mean in Euro) 2012**



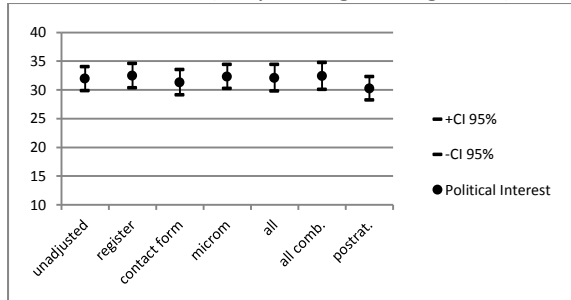
**School certificate (highest level in %) 2010**



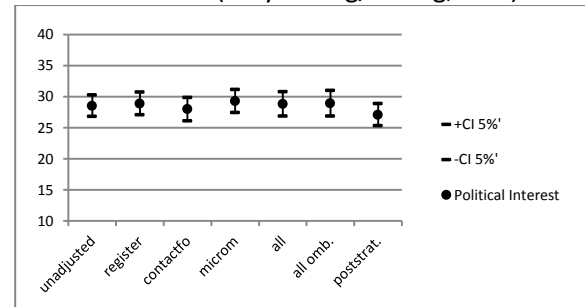
**School certificate (highest level in %) 2012**



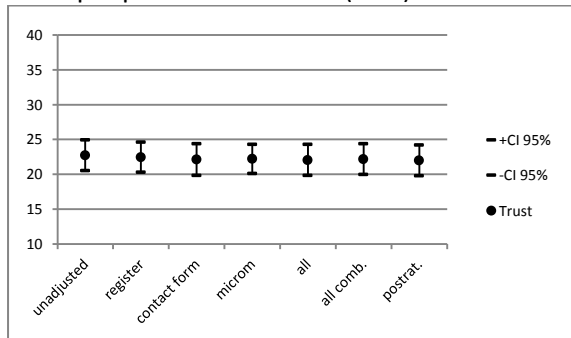
**Political Interest (very strong/strong, in %) 2010**



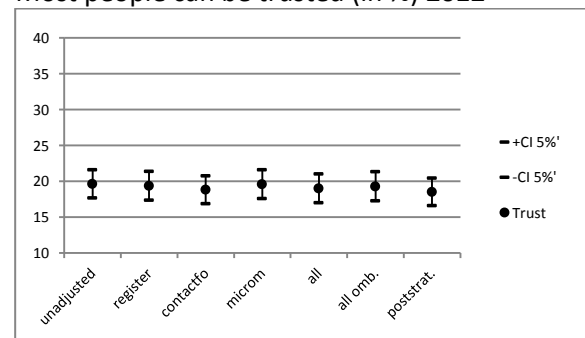
**Political Interest (very strong/strong, in %) 2012**



**Most people can be trusted (in %) 2010**



**Most people can be trusted (in %) 2012**



Summing up our analyses, we arrive at the following conclusions and questions:

- Microgeographic variables help to explain survey participation in ALLBUS to a small extent, in addition to the frame variables (population register) and contact form data available as a standard.
- Microgeographic variables are primarily helpful to explain contactability, the explained variance is much smaller for cooperation and response.
- The correlation between microgeographic variables and ALLBUS key survey estimates is rather low for the large majority of variables.
- Using population register and microgeographic variables for nonresponse adjustment leads to almost no effect on estimates of key survey variables.
- Models are by and large stable over time explaining survey participation
- As predicted power of auxiliary information is rather low, the usefulness of these data for targeting fieldwork efforts is an open question.
- Quality of microgeographic data is an issue to be discussed.

## References

- Groves, R. M. (2006): Nonresponse rates and nonresponse bias in household surveys. *POQ*, 70, 646-675.
- Kreuter, F. et al. (2010): Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *J. R. Statistic. Soc. A*, 173, 389-407.
- Olson, Kristen (2013): Paradata for Nonresponse Adjustment *The ANNALS of the American Academy of Political and Social Science*, 645, p. 142-170.
- Schneekloth, U. & Leven, I. (2003): Woran bemisst sich eine "gute" Allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. *ZUMA-Nachrichten* 53, S. 16-57.
- Schräpler, Jörg-Peter, et al. (2010): Individual and Neighborhood Determinants of Survey Nonresponse: An Analysis Based on a new subsample of the German Socio-Economic Panel (SOEP), Microgeographic Characteristics and Survey-Based Interviewer Characteristics. *SOEP-Papers* 288/2010.
- Smith, T. W. & Kim, J. (2009): An Assessment of the Multi-level Integrated Database Approach. *GSS Methodological Report No. 116*, NORC/University of Chicago.