Schmollers Jahrbuch 128 (2008), 65 – 73
Duncker & Humblot, Berlin

Which Progress for Poverty Studies can we expect from new large Data Sources?

By Jürgen Friedrichs

Abstract

Large data sources would allow us to test the impact of neighborhood characteristics, such as poverty rates, on the attitudes and behavior ot residents. The article explores the feasibility of existing large data sets for such a purpose. Unfortunately, none of the three sets reviewed, the Microcensus, the ALLBUS and the SOEP, allows for such multi-level analyses, because data cannot be regionalized due to data protection or insufficient sample size. To overcome these problems in a limited sense, it is suggested to pursue a "puzzle strategy" to combine data from different existing data sets.

JEL Classifications: C31, C33, C42, I32

Introduction

Over the last twenty years there has been a rapid development in the research on context effects, e.g, on school composition and educational achievement (e. g., Kristen, 2002; van Tubergen / Maas / Flap, 2004; Vartanian, 1999), on voting patterns (e.g., Brown, 1982; Falter, 1991) or networks (as contexts) on juvenile delinquency (e. g., Haynie, 2001). A significant part of the research has been devoted to neighborhoods effects (Dietz, 2000; Leventhal / Brooks-Gunn, 2000; Sampson / Morenoff / Gannon-Rowley, 2002).

In my view, three developments have contributed to this increasing number of studies. First, the methodological paradigm of micro-macro analysis (Blalock, 1984; Blau, 1994; Hernes, 1989; for urban research: Galster / Quercia / Cortes, 2000), namely the influential work of Coleman (1987, 1990, 2000). Coleman posits that any outcome at the macro level Y_j should not be explained by macro-level variables X_i, instead by a context effect from X_i to a micro-level variable x_i, further, a theory, e.g., rational choice theory, linking x_i to a behavior outcome y_i at the micro-level, and the aggregation of outcomes y_i to the macro-level variable Y_i.

Second, the statistical methods of hierarchical linear modeling (HLM), allowing to determine context from individual effects on individual level

outcomes (Hox, 1996; Snijders / Bosker, 1999). Third, an increase in poverty rates in many countries combined with a spatial concentration of poverty in few but growing urban areas or neighborhoods (e.g., Jargowsky, 1996; Wilson, 1987). This, in turn has led to national and regional programs to alleviate the living conditions of residents in deprived of poor urban areas, such as the German program "The Social City" or the U.S. Gautreaux and MTO programs.

Thus, the study of neighborhood effects is not solely a scholarly problem, but has, due to its implications, become an issue of vital interest for national urban policies.

1. The Problem: Data Requirements

Studies of neighborhood contexts effects require data on two levels: neighborhood and residents. Neighborhood data most often come from official statistics or are constructed from individual-level data, such as the poverty quota. Individual data have to come from surveys; among the dependent variables often used in neighborhood studies are deviant behavior, ranging from crime to teenage pregnancy, school achievement, incomes, migration. To make analyses even more complicated, it is advisable to introduce a meso level, for instance schools; we then study the direct impacts of neigborhoods and school characteristics on individuals and the indirect effect on neighborhood on individuals (cf. Friedrichs / Galster / Musterd, 2003).

Typically, such data are not available from official sources, but have to be collected for the purpose of the respective research. Examples are the seminal studies on collective efficacy and violence (Sampson / Raudenbush / Earls, 1997; Sampson / Morenoff / Earls, 1999) or the German study by Oberwittler (2003, 2004). However, there are exceptions, for instance U.S. studies on the impact of neighborhood's share of homeowners on residents have used the General Social Survey, the U.S. Panel Study of Income Dynamics (PSID), or the U.S. Public Use Microsample (PUMS) of the 1980 Census of Population and Housing (Boehm / Schlottmann, 1999, 2002; Harkness / Newman, 2002).

In the following, I will examine research strategies based on three types of sources: large data sets, combined data sets, and new surveys, starting with a basic example of a research design.

2. A Basic Example

In his contribution to "Inner-City Poverty in U.S. Cities" (Lynn / McGeary, 1990), Weicher (1990, 69) supplies some data on the distribution of poor and

non-poor persons over urban areas, classifying them as poor (lowest quartile of share of recipients of social assistance) and non-poor. The data encompass the 100 largest central cities in the U.S. in 1980. Inferring the rest of the data yields the distributions presented in Table 1.

Table 1

Persons living in poor or non-poor areas by poverty status of person, U.S. cities, 1980

1A: Absolute numbers, in millions

Residence area	Status Person		Total
	Poor	Non Poor	
Poverty	5,2 A	10,3 B	15,5
Non-poverty	2,9 C	29,1 D	32,0
Total	8,1	39,4	47,5

1B: In percent

Residence area	Status Person		Total
	Poor	Non Poor	
Poverty	10,9 A	21,7 B	32,6
Non-poverty	6,1 C	61,3 D	67,4
Total	17,0	83,0	100,0

Source of data: Weicher 1990: 69.

As a preliminary test whether the distributions are comparable to German conditions, we calculated the distribution for the urban 85 districts of Cologne. The results are reported in Table 2.

Table 2

Persons living in poor or non-poor areas by poverty status of person, Cologne, 2005

2A: Absolute numbers

Residence area	Status Person		Total
	Poor	Non Poor	
Poverty	60,457 A	86,109 B	146,566
Non-poverty	165,064 C	711,471 D	876,535
Total	225,521	797,580	1,03,1201

2B: In percent

Residence area	Status Person		Total
	Poor	Non Poor	
Poverty	5,9 A	8,4 B	14,3
Non-poverty	16,1 C	69,6 D	85,7
Total	22,0	78,0	100,0

Notes: "Poor = Social assistance recipients.

"Poverty area" = All districts of the lowest quartile of percentage of social assistance recipients.

Source: Author's calculations.

Comparing the percentages of the U.S. cities and the Cologne data, Tables 1B and 2B exhibit very similar distributions. Although this is a preliminary replication of the U.S. results (different periods, number of cities), results obtained for Cologne point into the same direction. Of course, the categorization of the two variables can be refined by grouping the variables into three or four categories.

The data reveal some interesting facts: First, more poor live in non-poor areas than in poor areas (cells C vs. A). Second, more non-poor than poor persons are living in poor areas (cells B vs. A). This raises some important research questions:

1. How do persons fare in cell A vs. cell B?
2. How do persons fare in cell B vs. cell D?
3. How do persons fare in cell A vs. cell C?

The first question pertains to the impact of poverty areas on poor and non-poor residents. This problem is addressed in a research project on the living conditions and attitudes toward crime in a study of a deprived area in Cologne, Vingst-Hoehenberg (Blasius / Friedrichs, 2007; Friedrichs / Blasius, 2006).

The second question, again, refers to a neighborhood effect, but now comparing non-poor in poor vs. non-poor neighborhoods. This may compliment studies devoted to the first question. Such comparisons are crucial for programs such as the "Social City" program in Germany, since so far only deprived neighborhoods, to which the policy measures are directed, are analyzed – without comparisons with non-poor areas or correcting for sample bias. Hence, we do not know whether the findings (e.g., correlations, trends) obtained for deprived areas hold true as well for non-poor or not deprived areas. (It is like studying a gentrifying area without controlling for upgrading processes in other areas of the city.)

The third question seems to be of specific interest, since the context hypothesis is that poor will fare better in non-poor than in poor neighborhoods. It is this hypothesis that underlies the northamerican Gautreaux and more specifically the MTO-program of relocating poor residents from poor to non-poor areas. Boldly generalizing the results from the MTO program, the studies suggest more positive than negative effects.

3. Research Strategies

To arrive at a research strategy for the study of neighborhood effects, I will first review the options and restrictions of three major large data sets available in Germany, then discuss the potential of combining existing data sets, finally explore the costs of new surveys and their implications for the methodology of neighborhood study designs.

3.1 Large Data Sources

The major large (or mass) data sets available in Germany are the ALLBUS (General Social Survey), the Microcensus., and the Socio-economic Panel (SOEP). For the purposes of studying neighborhood effects, these data sets have to be regionalized and cumulated. Unfortunately, this is only to a limited extent possible.

A single wave of the ALLBUS is by no means sufficient for the regionalization required. If several waves are combined, as in the Cumulative ALLBUS, due to different sampling we have different residents in the regional units and thus residential change as an error.

The Microcensus can be regionalized, but the Public Use Files comprises only a 70 percent sample of the total sample; further, the spatial units have changed after 1990. From 1990 onwards, each sample unit comprises an average of nine dwellings – which is by far too small for our purposes.

Therefore, the only candidate from the large surveys – except for a new survey – is the Socio-economic Panel. Due to data protection, the SOEP data presently cannot be regionalized, although it would be possible to allocate respondents by their residence to urban districts. However, efforts to regionalize SEOP data are under way, an example being the study of total poverty quotas and those of persons aged 65 and older for the 97 Regional Policy Regions in Germany (Knies / Krause, 2005). Moreover, data are now regionalized down to the level of zip codes (Knies / Spiess 2007), and presently work is under way to link SOEP data with data from microm Micromarketing on a household basis (Goebel et al., 2006).

Such regionalized SOEP datasets would allow to test hypotheses derived from the typology in Tables 1 and 2. If this does not result in a sufficient number of cases, we could then aggregate urban districts for several cities having similar characteristics assumed to be related to the individual outcome, such as GDP, poverty rates or percent ethnic minorities. In both strategies, the individual residences do not have to be disclosed.

Nonetheless, two drawbacks remain: First, the number of cases per urban district is to small (e.g., for each of the 85 Cologne districts) to conduct contextual analyses in the tradition of neighbourhood effect studies; further, zip codes according to our experiences cannot be aggregated to urban districts. Second, this large data set comprises only some of the dependent variables relevant for assessing neighborhood effects, e.g. income and educational achievement, but no variables of deviant behaviour or social capital (collective efficacy), which are crucial for an international cumulative and comparative study.

It should be noted, however, that only panel data allow us to do research truly testing the *causal* effects that most scholars of neighbourhood effects

have proposed. In contrast, the majority of the studies which posit effects of the neighbourhood on behaviour of residents use cross-sectional survey data for one point in time. Finally, the SOEP panel data would allow us to examine changes in the composition of urban areas by specifying the socio-economic characteristics of those moving in vs. those moving out.

3.2 Combining Existing Datasets

A second strategy is to combine official statistical data from different sources. In Cologne, for instance, we combined data for the urban districts from the Statistical Office with data from the Department of Health on a survey of children aged 4 to 6 years on their health, physical and mental conditions, these tests serving as a criterion of the ability to enter elementary school. Even at the risk of stating the obvious, I suggest to search for such data "hidden" in urban departments or other institutions.

3.3 New Surveys

An evident third strategy is to conduct new surveys. In the following section, I attempt to give examples for the calculation of costs of fieldwork. They serve to more specify the relationship between methodology of neighborhood effects studies if costs are taken into account.

Which cities? The study cannot compare too many cities, if survey data for all districts or neighborhoods are to be collected. Example 1: Imagine, we wish to draw a probability sample of residents in the 85 Cologne districts with only 150 residents in the final sample, this would require 12,750 completed interviews. If we calculate € 25 Euro per face-to-face interview, the costs amount to € 318,750 – excluding all costs of questionnaire construction, pretesting, data correction and data entry. If total field costs are calculated, a commercial institute will charge € 100 Euro per interview (including a corrected SPSS file), with the total amount rising to € 1,127,000.

Example 2: If, as proposed in Table 1 and 2, we use two neighborhoods per four cells by 150 interviews at € 100 Euro each, costs are 120,000 per city for the field work only.

Example 3: Since we wish to study several cities, selected e.g. by different extents of income inequality or different GDP's, the calculation in example 2 and categorizing the independent variable into three categories (e.g., low – medium – high income) and taking two cities per category, the total costs of fieldwork would be $6 \times$ € 120,000 = 720,000.

It is easy to vary the calculations given in the three examples, for instance by doubling the number of interviews per cell. Of course, a high amount of

funding is feasible if project costs are shared by several countries. But under which conditions is such a study justified? We may think of two major legitimate reasons. First, the study is innovative and makes a crucial contribution to our knowledge of neighborhood effects, taking the PISA studies as a model. This given, we arrive at a second reason: The policy implications. Results should allow for programs and the implementation of measures to alleviate living conditions of the poor segment of the population. One of the conclusions may pertain to the question of whether area-based or people-based strategies are advisable.

4. Conclusions: Towards a "Puzzle Strategy"

It is evident, that comparative studies of neighborhood effects require large data sets, with data for the aggregate level of the neighborhood and on the individual level. To cope with the data problems outlined, we have to carefully design a study combining elements from all three strategies, based on the existing theoretical reasoning and empirical evidence on neighborhood effects. I therefore suggest a "puzzle strategy".

The design of a comparative study, encompassing several countries and cities, should include several of those dependent variables listed in the introductory section. Further, large data sets available might lend themselves to different parts of the total design, e.g. the SOEP data of the U.K. data set on wards. We therefore have to specify which effect can be approached by a specific data set. Even further, we have to explore for each city if so far not used individual data from institutional sources exist. Based on this evidence of available data, we can determine those parts of the total design which are not captured. The final step would then require a larger new survey in few cities filling the methodological and data gaps. We assemble the different data a like parts of a puzzle. As stated above, both for scholarly and policy reasons, we need a European study.

If such new surveys are conducted, the data should be available, e.g., at the DIW in Berlin or the Central Archive for Empirical Social Research, Cologne, Germany, to enable scholars from different countries to make use of the data und perform secondary analyses, as the widely used SOEP convincingly demonstrates. Further, the design of the study might as well serve as guideline for additional studies in other cities.

References

Blalock, H. M. (1984): Contextual Effects Models: Theoretical and Methodological Issues, Annual Review of Sociology 10, 353 – 372.

Blasius, J. / *Friedrichs,* J. (2007): Internal Heterogeneity in a Deprived Urban Area and its Impact on Resident's Perceptions of Deviance, Housing Studies 22 (5), 753 – 780.

Blau, P. M. (1994): Structural Contexts of Opportunities. Chicago (IL).

Boehm, Th. P. / *Schlottmann,* A. M. (1999): Does Home Ownership by Parents Have an Economic Impact on Their Children?, Journal of Housing Economics 8, 217 – 232.

Boehm, Th. P. / *Schlottmann,* A. M. (2002): Housing and Wealth Accumulation: Intergenerational Impacts, in: N. P. Retsinas / E. S. Belsky (eds.): Low-Income Homeownership. Examining the Unexamined Goal, Cambridge (MA): Joint Center for Housing Studies und Washington, DC, 407 – 426.

Brown, C. (1982): The Nazi Vote. A National Ecological Study, American Political Science Review 76, 285 – 301.

Coleman, J. S. (1987): Microfoundations and Macrosocial Behavior, in: J. C. Alexander et al. (eds.): The Micro-Macro Link, Berkeley (CA), 153 – 173.

Dietz, R. (2000): Estimation of Neighborhood Effects in the Social Sciences, URAI Working Paper No. 00 – 03, Columbus, OH: Ohio State University, Urban and Regional Analysis Initiative.

DiPasquale, D. / *Glaeser,* E. L. (1999): Incentives and Social Capital: Are Homeowners Better Citizens?, Journal of Urban Economics 45, 354 – 384.

Falter, J. (1991): Hitlers Wähler (Hitler's Voters). München.

Friedrichs, J. / *Blasius,* J. (2006): Attitudes of Owners and Renters in a Deprived Neighborhood, Paper, presented at the ENHR Conference, Workshop "Poverty Neighborhoods", Ljubljana July 10, 2006.

Friedrichs, J. / *Galster,* G. / *Musterd,* S. (2003): Neighborhood Effects on Social Opportunities, The European and American Research and Policy Context, Housing Studies 18, 797 – 806.

Galster, G. (2003): Investigating Behavioral Impacts of Poor Neighborhoods: Towards New Data and Analytic Strategies, Housing Studies 18, 863 – 914.

Galster, G. / *Quercia,* R. G. / *Cortes,* A. (2000): Identifying Neighborhood Thresholds: An Empirical Investigation, Housing Policy Debate 11, 701 – 732.

Goebel, J. / *Knies,* G. / *Krause,* P. / *Spiess,* C. K. / *Scheike,* I. (2006): Verknüpfung des SOEP mit microgeo-graphischen Daten: Neue Möglichkeiten für Wissenschaft und Forschung, Vortrag, gehalten auf der Statistischen Woche 2006 der Deutschen Statistischen Gesellschaft, Dresden.

Green, R. K. / *White,* M. J. (1997): Measuring the Benefits of Homeowning: Effects on Children, Journal of Urban Economics 41, 441 – 461.

Harkness, J. / *Newman,* S. J. (2002): Homeownership for the Poor in Distressed Neighborhoods: Does This Make Sense?, Housing Policy Debate 13, 597 – 630.

Haynie, D. L. (2001): Delinquent Peers Revisited: Does Network Structure Matter?, American Journal of Sociology 106, 1013 – 1057.

Hernes, G. (1989): The Logic of the Protestant Ethic, Rationality and Society 1, 123 – 162.

Hox, J. J. (1996): Applied Multilevel Analysis, www.fss.uu.nl / ms / jh / publist / amaboek.pdf.

Jargowsky, P. A. (1996): Poverty and Place, New York.

Jencks, Ch. / *Mayer,* S. E. (1990): The Social Consequences of Growing up in a Poor Neighborhood, in: L. E. Lynn / M. G. H. McGeary (eds.): Inner-City-Poverty in the United States, Washington (DC.), 111 – 186.

Knies, G. / *Krause,* P. (2005): Armut und Alter, in: Nationalatlas Bundesrepublik Deutschland, Vol. 7: Arbeit und Lebensstandard, Heidelberg / Berlin, 118 – 199.

Knies, G. / *Spiess,* C. K. (2007): Regional Data in the German Socio-Economic Panel Study (SOEP), Data Documentation 17. Berlin: Deutsches Institut für Wirtschaftsforschung.

Kristen, C. (2002): Hauptschule, Realschule oder Gymnasium? Ethnische Unterschiede am ersten Bildungsübergang, Kölner Zeitschrift für Soziologie und Sozialpsychologie 54, 534 – 552.

Leventhal, T. / *Brooks-Gunn,* J. (2000): The Neighborhoods They Live in: The Effects of Neighborhood Residence on Child and Adolescent Outcomes, Psychological Bulletin 126, 309 – 337.

Oberwittler, D. (2003): Die Messung und Qualitätskontrolle kontextbezogener Befragungsdaten mithilfe der Mehrebenenanalyse – am Beispiel des Sozialkapitals von Stadtvierteln, ZA-Information 53, 11 – 41.

Oberwittler, D. (2004): Sozialstruktur, Freundeskreise und Delinquenz. Eine Mehrebenenanalyse zu sozialökologischen Kontexteffekten auf schwere Jugenddelinquenz, in: D. Oberwittler / S. Karstedt (Hg.): Soziologie der Kriminalität, Wiesbaden, 135 – 170. (Sonderheft 43 der Kölner Zeitschrift für Soziologie und Sozialpsychologie).

Sampson, R. J. / *Morenoff,* J. D. / *Gannon-Rowley,* Th. (2002): Assessing "Neighborhood Effects": Social Processes and New Directions in Research, Annual Review of Sociology 28, 443 – 478.

Sampson, R. J. / *Raudenbush,* S. W. / *Earls,* F. J. (1997): Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy, Science 277, 918 – 924.

Sampson, R. J. / *Raudenbush,* S. W. / *Earls,* F. J. (1999): Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children, American Sociological Review 64, 633 – 660.

Snijders, T. A. B. / *Bosker,* R. (1999): Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. London.

van Tubergen, F. / *Maas,* I. / *Flap,* H. (2004): The Economic Incorporation of Immigrants in 18 Western Societies: Origin, Destination, and Community Effects, American Sociological Review 69, 704 – 727.

Vartanian, Th. P. (1999): Adolescent Neighborhood Effects on Labor Market and Economic Outcomes, Social Service Review 73, 142 – 167.

Weicher, J. C. (1990): How Poverty Neighborhoods are Changing, in: L. E. Lynn / M. G. H. McGeary (eds.): Inner-City Poverty in the United States, Washington (DC.), 68 – 110.