

Editorial

From Panel Data to Longitudinal Analytical Designs: a Note on Contemporary Research Based on Data from the Socio Economic Panel Study (SOEP)

By Marco Giesselmann, Carsten Schröder,
Johannes Giesecke, John Haisken-DeNew,
Anika Rasner, and Jule Specht

Introduction

Records of contemporary research reveal that panel data is becoming increasingly relevant in social science: In 2014, 32% of all 471 empirical articles in the top-ten economic journals¹ analyzed panel-data. This quota is fairly stable across journals: while only one top-ten Journal has a panel-ratio below 25% (*Journal of Financial Economics*), in the *American Economic Review* and *Econometrica* more than one-third of all empirical articles are based on panel data. Referring to the relevance of panel data in sociology, Young/Johnson (2015) show that 61% of all articles published between 2010 and 2014 in the *Journal of marriage and the family* (JMF), a sociological top-ten journal with about 90 citable articles per year, use longitudinal research designs. Finally, Giesselmann/Windzio (2014) counted 62 articles based on panel data in the two top German sociological journals (*Zeitschrift für Soziologie* and *Kölner Zeitschrift für Soziologie und Sozialpsychologie*) between 2000 and 2009, of which 56% use data from the Socio-Economic Panel Study (SOEP).

This popularity of longitudinal data is primarily driven by a strong focus on causal modeling in contemporary economics (Imbens/Wooldridge, 2009) and sociology (Sobel, 2000). In respective methodological discourses, panel data usually is discussed as a powerful source to validate causal interpretations of coefficients (Giesselmann/Windzio, 2014; Gangl, 2010). These discourses, in turn, are accompanied by a research infrastructure that increasingly produces data in longitudinal form, with numerous long-running and complex panel-datasets available to the research community (Schupp, 2014).

¹ For our inquiry, we used the index provided by IDEAS/RePEc (<https://ideas.repec.org/top/top.journals.all.html>) as basis. We owe thanks to Teresa Backhaus (DIW Berlin) for her help with recording and counting articles.

One of the longest running and most complex longitudinal studies is the German Socio-Economic Panel (SOEP), which released its 30th wave of data to the research community in 2014. The Socio-Economic Panel is a representative survey that collects a variety of objective and subjective variables, aiming at identifying the respondents' economic, demographic, social and psycho-emotional circumstances accurately and exhaustively. The sample is considerably large (24,113 adult persons in 2013), with specific economic and social groups (immigrants, East Germans, wealthy persons) systematically oversampled in order to facilitate precise group-specific analyses (Wagner et al., 2007). The survey uses households as primary sample units and collects information from all household members (Kroh, 2014). Thus, not only individual-level variables, but also household features and partner characteristics can be used to model and explain individual outcomes. The range of collected variables reflects the commitment of the SOEP toward innovative research: in addition to socio-economic data, subjective characteristics (like values, aims and life satisfaction) are elicited from the respondents. Additionally, psychological traits, cognitive skills and information on health status are regularly collected via experiments or scales. The on-going commitment toward pioneering research is emphasized by the 2009 implementation of an innovative module (SOEP-IS), addressing a specific sub-sample, consisting of questions and tests that are proposed by the community and chosen on the basis of a peer-review procedure (Richter/Schupp, 2012).

Since its founding in 1984, the SOEP has served as the empirical basis for path-breaking studies in economics (e.g. Winkelmann/Winkelmann, 1998), sociology (e.g. Gangl, 2004), political science (Zuckermann et al., 2007) and psychology (e.g. Diener et al., 2006). Indeed, the SOEP has become one of the most important data sources for the international social science research community. Given the continuity of basic instruments and recent enhancements of the SOEP, this standing is likely to prevail and to grow.

The 11th International SOEP User Conference (SOEP 2014), which took place from June 30 to July 1, 2014, in Berlin at the Hertie School of Governance and the German Institute for Economic Research (DIW Berlin), attracted more than 120 participants from ten countries. In fourteen parallel sessions over the course of two days, researchers presented around 60 papers and 20 posters. The scientific committee consisted of Johannes Giesecke, Marco Giesselmann, John Haisken-DeNew, Anika Rasner, Carsten Schröder, and Jule Specht. Marco Giesselmann, Christine Kurka, Anika Rasner, and Carsten Schröder were the local organizers. The SOEP User Conference was opened with a welcome address from Marcel Fratzscher, President of the German Institute for Economic Research. He emphasized the role of high-quality micro data for not just economics, but also social sciences in general. He was followed by Jürgen Schupp, Martin Kroh, David Richter, and Marcel Hebinger, who each discussed recent SOEP developments: the family of SOEP longitudinal studies, the

2013/2014 SOEP Migration Sample (Brücker et. al. 2014), the SOEP Innovation Sample (Richter/Schupp, 2012), and DDI on Rails (www.ddionrails.org).

The 11th SOEP User Conference focused on two broad areas: *socio-economic inequalities* and *migration*, reflecting the growing interest of the social sciences in understanding the inter-temporal patterns behind these phenomena and their implications for modern welfare states. As the potentials of longitudinal data in these research domains remain vastly underexplored (Giesselmann/Windzio, 2014; Giesselmann/Goebel, 2013), the conference sought to provide a forum to discuss the potentials of panel data and to promote the possibilities of these data to the research community.

Presentations at the 11th SOEP User Conference

The first keynote lecture of the conference was given by US sociologist Patricia McManus, Associate Professor of Sociology at Indiana University. Her lecture was entitled “The Next Generation: Family Background and Prospects for Immigrant Incorporation in Germany, Great Britain, and the United States.” As McManus summarized, in all three of these countries, immigrant children are disadvantaged as they are living in households that are disproportionately exposed to life risks in the domains of education, employment, income, and health.

The second keynote was given by economist Jacques Silber, Professor Emeritus at Bar-Ilan University. His keynote was entitled “Inequality, Globalization, and Labor Markets”. He noted that, “as far as OECD countries are concerned, the most recent evidence shows clearly that in the past thirty years, wage inequality increased, trade integration spread, technology advanced rapidly and product and labor market institutions and regulations weakened”, and that what seems “to have had the greatest impact on wage inequality in OECD countries is not globalization, but technological change and the weakening of product and labor market institutions”.

While the keynote presentations were outlining desiderata for longitudinal research on social inequality, the common element of all conference presentations was their empirical foundation: German Socio-Economic Panel data. The presentations demonstrated that SOEP data are being used to investigate increasingly complex research questions covering a wide range of topics that include labor markets, intergenerational mobility, education, well-being, and life-course analysis.

Several presentations at the conference dealt with research on the determinants of life satisfaction using the longitudinal dimension of the SOEP data to validate causal claims with quasi-experimental or fixed effects designs. One of the presenters, Frederike Esche (DIW Berlin), reported findings that job loss

and unemployment have short-term effects for the partner of an unemployed person, while another presenter, Jonas Vossmer (University of Mannheim), showed that life-satisfaction converges to the person-specific baseline-level several years after the critical life-event of job-loss. Martin Ehlert (WZB Berlin) discussed the economic consequences of job loss on family members, showing that modern couples in the US and Germany tend to absorb negative economic consequences stemming from job loss by using compensating strategies within the household: the so-called “added worker effect”.

Another benefit of longitudinal data for the field of life-course analysis was explored at the conference: the possibility to jointly model characteristics from different phases of the life course. Using such a design, Holger Lengfeld (University of Leipzig) and Jessica Ordemann (University of Hamburg) showed that job characteristics during the active phase of the life course have a substantive impact on the extent of voluntary work during retirement. Michael Kind (RWI Essen), to cite another example, demonstrated that children of unemployed parents face substantially more difficulties in entering the labor market than children of parents who never experienced unemployment.

Experts in SOEP data management attended all of the conference sessions, giving feedback to presenters, commenting on their research designs, clarifying data issues, and pointing out how unused potentials and capacities of the data could be utilized. As intended, the conference provided a useful channel of communication between the SOEP team and SOEP users. Just as presenters benefited from the input of experts from the SOEP, the SOEP team benefits from exchange with the researchers who use the data: Their input provides a crucial basis for developing the SOEP questionnaires and the data service in general.

Conference presenters also profited from discussions with a number of top researchers who were enlisted as chairs of the parallel sessions, among them David Brady (WZB Berlin), Martina Dieckhoff (WZB Berlin), Jennifer Hunt (Rutgers University), and Thomas Siedler (University of Hamburg). These experts shared their experiences from international top-level research, highlighting gaps in the scientific discourse and encouraging the community to address these using appropriate longitudinal designs.

The closing ceremony began with a speech by Jürgen Schupp, who reviewed the conference as well as the history and past recipients of the SOEP awards. Thereafter, Georg Weizsäcker (Humboldt University Berlin), and Bruce Headey (University of Melbourne), presented the winners of the Felix Büchel Award and the Joachim Frick Award. The awards are made possible by the generous contributions of Vereinigung der Freunde e. V. (VdF).

Jennifer Hunt (Rutgers University) and Thomas Klein (University of Heidelberg) were selected as the recipients of the 2014 Felix Büchel Award. Felix Büchel, who passed away at the age of 47 on July 12, 2004, was not only a

long-time user of SOEP data and active supporter of the SOEP project; he also exhibited all of the qualities that SOEP users should ideally possess. Felix was an excellent scholar and serious policy advisor. His research focused on issues of crucial economic and social importance, and he made effective use of the entire spectrum of possibilities SOEP data has to offer. His research was interdisciplinary, covering topics in education, labor markets, immigration, and income distribution.

Jennifer Hunt, currently on leave from Rutgers University and serving as Deputy Assistant Secretary for Microeconomic Analysis in the U.S. Treasury, was one of the first foreign users of the SOEP. Her publications in top journals, including *The Quarterly Journal of Economics*, *Economica*, *The Journal of Labor Economics*, and *Brookings Papers on Economic Activity* focus on some of the apparent successes of German labor market policies. She has worked on issues relating to immigration, employment and unemployment, wage inequality, transition economics, as well as crime and corruption. Recently she undertook research on the economic crisis of 2008–09 and its effects on Germany (Burda/Hunt, 2011). She has also studied issues of labor supply in the science and engineering workforce in the U.S. and, related to that, innovation in the U.S. economy. She is a distinguished economist and policy practitioner and a very worthy recipient of the Felix Büchel Award.

The co-recipient of the award, Thomas Klein, Professor of Macrosociology and Methods of Empirical Research at the University of Heidelberg and the Director of the Max Weber Institute for Sociology, has been involved with SOEP since its very beginning. In the early 1980s, he was a member of the German Research Foundation's celebrated Collaborative Research Centre, Sfb3, which was set up to develop "The Microanalytical Foundations of Social Policy", or more informally, to develop a set of valid social indicators for Germany. Thomas Klein's record comprises over 50 SOEP-based publications.

The prizes for best paper and poster presented at the SOEP conference are dedicated to our former colleague Joachim R. Frick, who passed away at the age of 49 in 2011. Joachim Frick contributed substantially to shaping the Socio-Economic Panel in its current form. Many aspects of the data that are used in contemporary SOEP-based research and that were presented at the conference can be traced back to his initiative and engagement. Georg Weizsäcker (Humboldt University Berlin) gave the awards speech.

The first Joachim R. Frick Memorial Prize was awarded to Christoph Wunder for his paper "How does the stock market affect subjective expectations of the future? Evidence from linking financial data to survey responses". In his work, Christoph Wunder examines the formation of economic expectations and, in particular, the role of stock market information as determinants of these expectations. Georg Weizsäcker emphasized that the paper was highly innovative, enriching the detailed information already available in the SOEP with

added information from daily stock market data. The paper provides a deeper understanding of the determinants of economic expectations and economic decision-making. In Weiszäcker's words, "the excellent presentation clearly demonstrated that economic expectations respond to short-term stock-market developments, as returns and variability over a 90-day horizon have the highest explanatory power".

The second Joachim R. Frick Memorial Prize was awarded to Wouter Zwysen for his paper "Family background matters for early careers – but not equally". His research deals with the relationship between family background, economic context, and individual economic (labor) outcomes. His study examined the labor market success of individuals over time using the longitudinal features of the SOEP data. His findings show that individual success hinges on a variety of determinants, some people being luckier than others. Although some individuals are advantaged, they may have been unlucky with the economic context at the time when they entered the labor market. For this reason, they may be less successful than initially less advantaged labor market participants. When comparing person A from an advantaged background with person B from a disadvantaged background, person A may be less successful initially than person B, depending on when they entered the labor market. As the economic context worsens and the unemployment rate rises, this finding no longer holds true. Under these adverse conditions, the relationship reverses and the advantaged person has a higher chance statistically of coping with the adverse change in the economic context.

SOEP Data in Practice: A Review on the Contributions to this Volume

This volume covers a selection of extraordinary empirical work presented at the 2014 SOEP conference. The articles were chosen by the program committee, went through careful peer-review processes and are, in turn, an excellent opportunity for us to reflect on how panel data, in general, and SOEP data, in particular, are used in practice.

Among the SOEP features outlined in the introduction, the most significant is, of course, its longitudinal design: persons and households are measured repeatedly on a yearly basis since 1984. According to methodological discourses, the most important benefit of longitudinal data is the option to control for person specific heterogeneity (Gangl, 2010, Wooldridge, 2005, Brüderl/Ludwig, 2015). Controlling for heterogeneity, for example via Fixed Effects regression, hybrid models, or Difference-in-Difference approaches, implicitly holds all unobserved person-specific characteristics constant and consequently increases the validity of causal claims drastically.

In this light, Ehlert (pp. 55–66 in this volume) uses a Difference-in-Difference (DID) approach to measure the impact of the partner's activity status on individual labor earnings. This approach defines changes in the independent variable (here: *partner's activity status*) as an event and controls for heterogeneity by focusing on differences in the dependent variable (here: *labor earnings*) before and after the event. By relating the average event-specific change to regular yearly rates of earnings' increases, the general trend in wages is automatically accurately controlled for by the DID estimator. Similarly, Icardi (pp. 67–82 in this volume) uses the DID technique to measure the impact of workplace training on wages. These two studies emphasize the role of the DID approach as important alternative to the standard Fixed Effects (FE) estimator in modern economics and empirical sociology, if panel data is used to control for unobserved heterogeneity. The DID approach, however, can only be applied to models in which changes in the independent variable constitute genuine events. In such designs, it seems specifically useful if measurements from units who do not experience an event shall be used to control for general trends. Wolbring (pp. 83–96 in this volume) also refers to the statistical benefits of panel data in his contribution. Accordingly, his research question (relating *body weight* to *earnings*) is suitable to be modeled by a causal longitudinal design, at least from a conceptual point of view. However, the author finds insufficient within-variation in his independent variable (*body weight*) to estimate precise coefficients. Consequently, by applying a random effects approach, the author additionally uses cross-sectional covariation to estimate the coefficient of his model, running the risk of having the results biased by unobservable time-constant characteristics (Halaby, 2004, Giesselmann/Windzio, 2014).

In addition to controlling for unobserved heterogeneity, which motivated the use of panel in the three studies discussed above, there are also other important benefits stemming from the use panel data. One is the option to combine prospectively collected characteristics from different time points (or even from different phases of the life course) in one model: Certain economic, social and subjective conditions during young adulthood can, for example, be used to explain later educational and labor market outcomes. These benefits are, on the one hand, sometimes ignored in methodological discourses, but, on the other hand, often used in empirical practice (Giesselmann/Windzio, 2014). Taking a similar perspective on panel data, Zochert (pp. 47–54 in this volume) combines information from different individual measurements in order to construct a valid sample. Here, longitudinal information is not used to control for unobserved heterogeneity, but to identify persons with certain activity patterns. Concretely, persons who experience a job-change are detected by comparing consecutive observations. The valid measurement of such life-course events (like unemployment (Kohler et al., 2012), labor market entry (Giesselmann, 2015), divorce (Andreß et al., 2006), and widowhood (Kohler et al., 2012)) on the basis of repeated measurements is another important, widely used feature of the

SOEP. This feature can be viewed as an infrastructural response to social theorists that put transitional events in the center of ideas on institutional impacts on the emergence of inequalities during the life course (Kohli, 1985). From a related conceptual perspective, Heady (pp. 97–108 in this volume) uses longitudinal information to measure patterns of consecutive observations of life-satisfaction. Once more, panel data is explicitly used to identify and to disentangle life-course dynamics.

The contribution of Vandecasteele and Esche uses panel data in order to perform a discrete life-history analysis. Though this method is not clearly designed to control for unobserved heterogeneity (Giesselmann/Windzio, 2014), it uses longitudinal information to accurately measure transitions, which *de facto* constitute the dependent variable in event-history models. Concretely, the authors are interested in shifts from being employed to being inactive. The probability of experiencing such a labor market exit is predicted on the basis of the partner's socio-economic status. This analysis hints at another important benefit of the SOEP, which is independent of its longitudinal features: the household design allows linking individual measurements on family level. Thus, the characteristics of household and family members (partners, parents, children, sisters) can be used to explain individual outcomes and, consequently, to unravel complex familiarly interdependencies.

The work by Sohn et al. (pp. 13–22 in this volume) treats the data as a pooled cross-section by using it for a trend design – an analytical format, that does not necessarily require panel data. Therefore, for this contribution, not the longitudinal design of the data, but the high quality and precisions of the income measurement were the attractors to use the SOEP. Similarly, Kern (pp. 23–34 in this volume) uses regional context variables in order to explain individual behavior. As in the Paper of Sohn et al., heterogeneity between persons is not intended to be eliminated, but is used to construct the categories of the independent variables and systematically analyzed.

At the end, these two papers remind us that many social phenomena explicitly refer to differences between stable social or demographic groups. Such phenomena, like the distribution of labor incomes across different regional or ethnic groups, accordingly provoke research questions that are inherently cross-sectional. The SOEP, despite its core emphasize on longitudinal research designs, is obviously also strongly committed to such research: besides its longitudinal features, the data gives us very detailed snapshots of the demographic and economic composition of the German population points in time since 1984.

This volume offers three types of longitudinal analyses: *a)* within-designs that allow for controlling unobserved heterogeneity; *b)* designs with longitudinal conceptualizations of the dependent or independent variable; and *c)* models that use information from different time-points of the life-course to construct the sample or to measure a delayed impact of the independent variable. Further-

more, there are designs that do not use the longitudinal aspect of the SOEP, but rather some other genuine features of the data. This offers a good overview on the variety of research designs that can be realized with the data. It explicitly shows that the merits of the SOEP go beyond longitudinal research designs, as other features of the data (like the household design, added experiments and extensive economic measures) constitute unique empirical sources for original research. This implies a clear mission for the providers of SOEP-data: it is not just the provision of panel data that will secure the status of the SOEP as major resource of the international empirical social science community, but its steady commitment to high-quality socio-economic data and the ongoing implementation of innovative features.

Last but not least, we are grateful to our sponsors, Deutsche Forschungsgemeinschaft (DFG), StataCorp, and Dittrich & Partner Consulting (DPC), whose generous contributions made this event possible. Among many other helping hands, the success of the conference was due in large part to the commitment of Christine Kurka. She and her team were in charge of the conference management: the preparation of applications for funding, handling of submissions, setting up a website, organizing rooms, and arranging catering, etc. We also thank Monika Wimmer for her excellent work communicating the conference to the media, and to Gabi Freudenmann for helping to put together this special issue and assisting with the editing process. Our last thank you goes to all the authors and reviewers of this special volume for their contributions.

Berlin, June 2015

References

- Andreß, H. J./Borgloh, B./Brockel, M./Giesselmann, M./Hummelsheim, D. (2006): The Economic Consequences of Partnership Dissolution – A Comparative Analysis of Panel Studies from Belgium, Germany, Great Britain, Italy, and Sweden, European Sociological Review 22 (5), 533 – 60.*
- Brücker, H./Kroh, M./Bartsch, S./Goebel, J./Kühne, S./Liebau, E./Trübswetter, P./Tucci, I./Schupp, J. (2014): The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents, SOEP Survey Paper, Series C 216. Berlin: DIW Berlin.*
- Brüderl, J./Ludwig, V. (2015): Fixed-Effects Panel Regression, in: H. Best/Ch. Wolf (eds.), The Sage handbook of regression analysis and causal inference, Los Angeles.*
- Burda, M. C./Hunt, J. (2011): What Explains the German Labor Market Miracle in the Great Recession?, Institute for the Study of Labor (IZA). IZA Discussion Paper 5800. <http://ideas.repec.org/p/iza/izadps/dp5800.html>.*
- Diener, E./Lucas, R. E./Scollon, Ch. N. (2006): Beyond the hedonic treadmill: Revising the adaptation theory of well-being, American Psychologist 61 (4), 305 – 14.*

- Gangl, M.* (2004): Welfare States and the Scar Effects of Unemployment: A Comparative Analysis of the United States and West Germany, *American Journal of Sociology* 109 (6), 1319–64.
- Gangl, M.* (2010): Causal Inference in Sociological Research, *Annual Review of Sociology* 36, 21–47.
- Giesselmann, M.* (2015): Differences in the Patterns of in-work Poverty in Germany and the UK, *European Societies* 17 (1), 27–46.
- Giesselmann, M./Goebel, J.* (2013): Soziale Ungleichheit in Deutschland in der Längsschnittperspektive. Befunde zur Armutsproblematik auf Basis des Sozio-Oekonomischen Panels (SOEP), *Analyse und Kritik – Zeitschrift für Sozialtheorie* 35 (2), 277–302.
- Giesselmann, M./Windzio, M.* (2012): Studienskripte zur Soziologie, Regressionsmodelle zur Analyse von Paneldaten. Wiesbaden.
- Giesselmann, M./Windzio, M.* (2014): Paneldaten in der Soziologie: Fixed Effects Paradigma und empirische Praxis in Panelregression und Ereignisanalyse, *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 66 (1), 95–113.
- Halaby, Ch. N.* (2004): Panel Models In Sociological Research: Theory into Practice, *Annual Review of Sociology* 30, 507–44.
- Imbens, G. W./Wooldridge, J. M.* (2009): Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature* 47 (1), 5–86.
- Kohler, U./Ehlert, M./Grell, B./Heisig, J. P./Radenacker, A./Wörz, M.* (2012): Verarmungsrisiken nach kritischen Lebensereignissen in Deutschland und den USA, *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64 (2), 223–45.
- Kohli, M.* (1985): Die Institutionalisierung des Lebenslaufs. Historische Befunde und theoretische Argumente, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 37 (1), 1–29.
- Kroh, M.* (2014): Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2012), *SOEP Survey Papers* 177: Series D. Berlin: DIW Berlin.
- Richter, D./Schupp, J.* (2012): SOEP Innovation Sample (SOEP-IS) – Description, structure and documentation, *SOEPpapers* 463. Berlin: DIW Berlin.
- Schupp, J.* (2014): Paneldaten für die Sozialforschung, in: N. Baur/J. Blasius (eds.), *Handbuch Methoden der Empirischen Sozialforschung*, 925–40. Wiesbaden.
- Sobel, M. E.* (2000): Causal Inference in the Social Sciences, *Journal of the American Statistical Association* 95 (450), 647–51.
- Wagner, G. G./Frick, J. R./Schupp, J.* (2007): The German socio-economic panel study (SOEP). Scope, evolution and enhancements, *Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften; Journal of Applied Social Science Studies* 127 (1), 139–69.
- Winkelmann, L./Winkelmann, R.* (1998): Why are the Unemployed So Unhappy? Evidence from Panel Data, *Economica* 65 (257), 1–15. <http://www.jstor.org/stable/2555127>.

- Wooldridge, J. M. (2005):* Introductory econometrics. A modern approach. Mason, Ohio.
- Young, R./Johnson, D. R. (2015):* Handling Missing Values in Longitudinal Panel Data With Multiple Imputation, *Journal of Marriage and Family* 77 (1), 277–94.
- Zuckerman, A. S./Dasovic, J./Fitzgerald, J. (2007):* Cambridge studies in public opinion and political psychology, Partisan families. The social logic of bounded partisanship in Germany and Britain. Cambridge, New York.