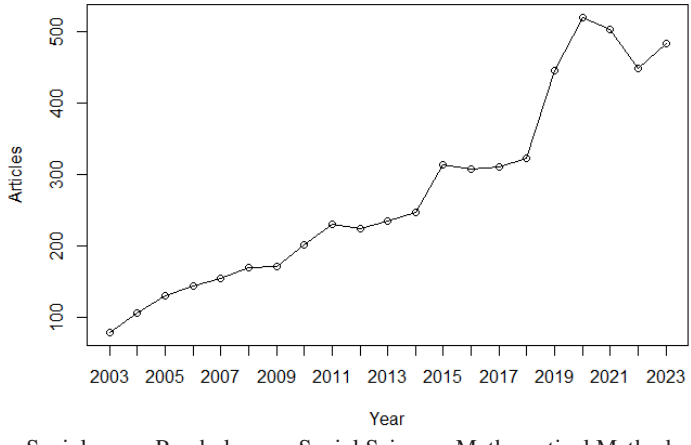
methods, data, analyses | Vol. 18(1), 2024, pp. 3-6

Editorial
Recent Developments and Current Approaches to the Analysis of Panel Data

Henrik Kenneth Andersen[1], *Jochen Mayerl*[1] & *Elmar Schlüter*[2]

[1] *Chemnitz University of Technology*
[2] *University of Giessen*

Panel data refer to repeated observations of the same units over time. Due to the growing interest in causal inference in the social sciences, and the increasing feasibility of collecting (intensive) longitudinal data, interest in panel data has grown steadily in the social sciences (Rohrer & Murayama, 2023). Figure 1 shows the number of articles containing the term "panel data" published just in the fields of Sociology, Psychology, and Social Sciences Mathematical Methods over the last 20 years (according to Web of Science, as of January 2024).



Categories: Sociology or Psychology or Social Sciences Mathematical Methods

Figure 1 Articles featuring keywords "panel data" (all fields), Web of Science years 2003-2023

Panel data offer a wide variety of advantages over cross-sectional data or even other types of longitudinal data. For one, they are valuable for the purposes of causal inference, that is, drawing causal conclusions from observational (rather than experimental) data. Indeed, as Hamaker (2012) notes, most social science theories are implicitly formulated at the *within-person* level. And the potential outcomes framework always begins with formulating a unit-specific causal effect: a contrast between realized and counterfactual states at the individual level (Rohrer & Murayama, 2023). For example, when we think of the relationship between typing speed and typing errors, most of us would probably expect the effect to be positive: the faster one types, the more mistakes she or he makes (Hamaker, 2012). This is exactly because we are thinking at the *within-person* level rather than the between-person level: if an individual increases her or his typing speed (holding all else constant), she or he is likely to make more errors. Panel data allows us to get closer to this ideal. By comparing the same individuals over time, we can be sure that we're holding constant all the things that don't change for a given individual, such as place and time of birth, upbringing, and potentially even psychological traits.

With panel data, researchers can respect the fact that processes and effects "unfold over time" (Hamaker & Wichers, 2017). Thus, social change over time can be analyzed at the individual rather than aggregate level, avoiding ecological fallacies. As technology evolves to make (intensive) longitudinal data collection more feasible, and as causal inference becomes the focus of many social science studies (e.g. fixed effects panel regressions), panel data are becoming increasingly important (Rohrer & Murayama, 2023).

The field of panel data research is still growing, addressing the need for research on innovative panel data collection methods as well as panel data analysis techniques. On the methodological side, the quality of panel data collection is challenged by issues such as panel conditioning (e.g., learning effects), the question of optimal lags for identifying causal effects, and high attrition rates that require missing value treatment techniques or weighting procedures. To further improve panel data analysis, research is needed on issues such as dealing with violations of the parallel assumption and heterogeneous growth, comparing different statistical approaches to panel data analysis, mediation analysis based on panel data, estimation of treatment effect dynamics and dealing with negative weighting bias, the challenges of dynamic panel models and the inclusion of bidirectional effects and lagged dependent variables, and continuous versus discrete time modeling, to name just a few current research issues.

This special issue contains applications to methodological issues and statistical problems in panel data analysis in a variety of content-related areas:

The contribution from *Dominik Becker*, entitled "Many Roads to Mediation: A Methodological and Empirical Comparison of Different Approaches to Statistical Mediation", examines the use of panel data to investigate social mechanisms in the form of mediation analyses. While mediation analysis is often done using cross-sectional data, the use of panel data has several interesting advantages. For one, mediation analysis with panel data allows for drawing causal inference under less strict assumptions. If confounders of the effects of interest are stable within individuals over time, then the broad category of panel fixed effects panel models can eliminate unobserved time-invariant heterogeneity. Second, panel data allow researchers to empirically establish the theoretical causal order of cause, mediator, and outcome. In particular, the specification of lagged effects between variables helps to rule out reverse causality. The article constructs a simulation study and compares a variety of modeling techniques with respect to their ability to recover the true parameter values, and provides researchers with valuable recommendations for approaching questions of causal mechanisms with panel data.

Judith Lehmann contributes an article entitled "Analyzing the Causal Effect of Obesity on Socioeconomic Status – the Case for Using Difference-in-Differences Estimates in Addition to Fixed Effects Models" in which she compares Difference-in-Differences (DiD) with Fixed Effects (FE) models to investigate the empirically well-established obesity penalty with respect to labor market outcomes. Like other articles in this issue, this one also combines strong substantive and methodological components. Substantively, the author finds no effects of obesity on socioeconomic status in either the FE or the DiD model. However, the DiD estimator explicitly models the development of the control group, providing a deeper understanding of the relationships. Namely, the non-obese individuals in the analysis showed stronger socioeconomic development over time compared to the group of obese individuals.

Manuel Holz and Jochen Mayerl compare health outcomes of migrants and native Germans over time in a contribution entitled "Migrant health inequalities or unequal measurements? Testing for cross-cultural and longitudinal measurement invariance of subjective physical and mental health". The so-called healthy migrant effect describes both the self-selection of comparatively healthy individuals to migrate from their home countries and the greater decline in health among migrants compared to the native population. The paper draws attention to an aspect of cross-cultural comparisons of health outcomes that has been overlooked in the previous research: to make valid comparisons of (especially) subjective measures of health, one must establish that components of the measurement instrument have the same meaning and importance across cultures and time. Thus, this article compares the trajectories of subjective health (SF-12 for physical and mental health)

of migrants and native-born Germans, testing for measurement invariance across groups and over time.

Christina Beckord tackles an interesting methodological topic in her contribution entitled "Challenges in Assigning Panel Data with Cryptographic Self-generated Codes – Between Anonymity, Data Protection and Loss of Empirical Information". The article examines the difficulties of linking data across 13 survey waves of the "Crime in the Moden City" (CrimoC) study and details a unique strategy for dealing with ambiguous user-generated codes. The author describes a meticulous, error-tolerant matching process, involving manual handwriting comparison, to merge individual data over time. The matching process resulted in 3,589 filled missing units.

The final contribution by *Jost Reinecke, Anke Erdmann, & Manuel Voelkle* entitled "Continuous Time Modeling with Criminological Panel Data: An Application to the Longitudinal Association between Victimization and Offending" re-examines the well-known victim-offender overlap – that offenders tend to have been victimized themselves – with novel panel data from the Crime in the Modern City (CrimoC) study. Methodologically, this paper adds to the new but growing literature on so-called continuous time panel models. Unlike the more commonly applied discrete time models (e.g., cross-lagged panel models, latent growth curves), continuous time model recognize that panel data provide multiple discrete snapshots of constructs over time. Yet effects between constructs over time are highly sensitive to the time interval between these snapshots, which is often chosen arbitrarily (e.g., one panel wave per year) or set based on time and budget constraints. The article discusses the results of the continuous time models, explains how researchers can transform continuous parameters into discrete parameters and visualizes the dynamic effects of constructs on each other (and themselves) as time unfolds.

References

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), Handbook of research methods for studying daily life (pp. 43–61). The Guilford Press.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. Current Directions in Psychological Science, 26(1), 10-15.

Rohrer, J. M., & Murayama, K. (2023). These are not the effects you are looking for: causality and the within-/between-persons distinction in longitudinal data analysis. Advances in methods and practices in psychological science, 6(1), 25152459221140842.