

Warum Panelstudien? (Version 1.1)

Pfarr, Klaus; Schröder, Jette

Erstveröffentlichung / Primary Publication

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Pfarr, K., & Schröder, J. (2015). *Warum Panelstudien? (Version 1.1)*. (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/gesis-sg_008

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

GESIS Survey Guidelines

Warum Panelstudien?

Klaus Pforr & Jette Schröder

Zusammenfassung

Das Papier gibt einen Überblick über die methodischen Vor- und Nachteile von Panelstudien und über die grundlegenden Überlegungen bei der Entscheidung für oder gegen ein Panel-Designs.

Zitierung

Pfarr, Klaus und Schröder, Jette (2015). Warum Panelstudien. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_008

1. Vorteile von Panelstudien

Panelstudien sind ein Erhebungsdesign mit langer Tradition in den Sozialwissenschaften (Lazarsfeld, Berelson, & Gaudet, 1944; Lazarsfeld & Fiske, 1938), bei dem die Beobachtungseinheiten über die Zeit hinweg mehrfach befragt werden. Panelstudien können somit Längsschnittdaten auf Individualebene liefern, die gegenüber Querschnittdaten große Vorteile für die Datenanalyse bieten (Allison, 2009; Brüderl, 2010; Giesselmann & Windzio, 2014; Halaby, 2004).

Längsschnittdaten auf Individualebene haben mehrere Stärken für die Beantwortung sozialwissenschaftlicher Fragestellungen: Ausgehend vom Ziel der Hypothesenprüfung, ist eine notwendige Bedingung für das Vorliegen eines Kausalzusammenhangs, dass die unabhängige Variable der abhängigen Variable zeitlich vorausgeht. Bei der Analyse mit Querschnittdaten kann diese Bedingung aber empirisch nicht überprüft werden, die zeitliche Struktur zwischen den Variablen muss also angenommen bzw. theoretisch begründet werden (Opp, 2010). Längsschnittdaten können hingegen sicherstellen, dass die angenommene zeitliche Struktur zwischen den Variablen auch tatsächlich gegeben ist. Ein besonderes Potenzial von Paneldaten für die sozialwissenschaftliche Analyse kommt dann zum Tragen, wenn den Befragten die gleichen Fragen über die Zeit immer wieder gestellt werden. Dies ermöglicht es einerseits auf deskriptiver Ebene, Veränderungen und Entwicklungsverläufe der Beobachtungseinheiten über die Zeit darzustellen. Stellt man beispielsweise die Erwerbsbeteiligungen von Frauen nach dem Alter mit Querschnittdaten dar, so bleibt unklar, ob eine abnehmende Erwerbsbeteiligung mit zunehmendem Alter auf einen Kohorteneffekt zurückzuführen ist, d. h., ob Frauen der älteren Kohorten grundsätzlich eine geringere Erwerbsbeteiligung haben als Frauen der jüngeren Kohorten, oder ob es sich um einen Lebenszykluseffekt handelt, d.h. ob die Erwerbsbeteiligung von Frauen mit zunehmendem Alter sinkt (Andreß, Golsch, & Schmidt, 2013, p. 5).

Entscheidende Vorteile von Paneldaten ergeben sich aber insbesondere bei der Überprüfung von Hypothesen mittels multivariater Analysemethoden. Die Ergebnisse von Analysen mit Querschnittdaten sind verzerrt, wenn Faktoren, die mit der unabhängigen Variable und dem Fehlerterm korrelieren, nicht kontrolliert werden. Da meist nicht alle diese Faktoren erhoben werden bzw. erhoben werden können, ist die Gefahr groß, dass Querschnittsanalysen unter unbeobachteter Heterogenität bzw. Endogenität leiden (Arellano, 2003; Baltagi, 2005; Cameron & Trivedi, 2009; Cameron & Trivedi, 2010; Hsiao, 2003; Mátyás & Sevestre, 1996; Wooldridge, 2010). Findet man beispielsweise in Querschnittsanalysen einen negativen Effekt von Kindern auf das Einkommen von Frauen, so kann dieser Zusammenhang darauf zurückzuführen sein, dass Frauen, die Kinder bekommen, ohnehin z. B. eine geringere Karriereorientierung haben als Frauen, die keine Kinder bekommen. Wenn die Karriereorientierung in der Querschnittsanalyse nicht kontrolliert wird, wird der negative Effekt von Kindern auf das Einkommen überschätzt. Untersucht man den Effekt jedoch mit Längsschnittdaten, so kann man den Effekt mittels Fixed-Effects-Regression zumindest dann konsistent schätzen, wenn die unbeobachteten Faktoren zeitkonstant sind. Hier werden zur Schätzung nämlich nicht wie in der Querschnittsanalyse Frauen mit und ohne Kinder miteinander verglichen („between“-Vergleich), sondern man vergleicht das Einkommen der einzelnen Frauen vor und nach der Geburt von Kindern („within“-Vergleich).

Längsschnittdaten lassen sich zwar prinzipiell auch mittels Retrospektivfragen in Querschnittsstudien erheben, allerdings ist dies nur eingeschränkt möglich, weil Retrospektivfragen mit starken Messfehlern behaftet sind. Schon bei der Erfassung vergangener Ereignisse leidet die Datenqualität je nach Untersuchungsgegenstand in unterschiedlichem Ausmaß, weil z. B. unerwünschte Ereignisse verdrängt, subjektiv unwichtige bzw. seltene Ereignisse vergessen oder fehldatiert werden und gegenwärtige Umstände als verzerrender Anker verwendet werden (Belli, Stafford, & Alwin, 2009). Bei subjektiven

Merkmale wie z.B. Einstellungen, Kompetenzen und psychische Skalen, ist dieser Messfehler so groß, dass eine Erfassung mittels Retrospektivfragen unmöglich ist.

Beispiele für deutsche Panelstudien, die darauf ausgerichtet sind, die Schätzung von Kausaleffekten durch die Erhebung geeigneter Längsschnittinformationen in der dargestellten Weise zu verbessern, sind SOEP, pairfam, SHARE, PASS, NEPS, das Mikrozensus-Panel und die Nationale Kohorte. Die Wiederholung der Messung der gleichen Sachverhalte bzw. die Erfassung von Variablen, die der abhängigen Variablen zeitlich vorgelagert sind, ist bei diesen Studien zentral.

Neben einem besseren Analysepotential im Vergleich zu Querschnittsstudien gibt es jedoch auch andere Gründe, die zur Entscheidung für ein Panel-Design führen können. So ist ein möglicher Grund, dass man von den gleichen Individuen so viele unterschiedliche Informationen benötigt, dass die Dauer des Interviews zu lang für eine einzige Befragung wäre. Um die Belastung der Befragten zu reduzieren, kann man die Inhalte auf mehrere Befragungswellen verteilen. Im Vorfeld muss allerdings sorgfältig überlegt werden, ob es für die angestrebten (Querschnitts-)Analysen nicht problematisch ist, dass einige Variablen zu einem früheren Zeitpunkt erhoben werden als andere. Dies gilt insbesondere für zeitveränderliche Variablen.

Ein anderer Grund für die Anwendung eines Panel-Designs ist es, Forschern einen schnellen und kostengünstigen Zugang zu Befragungspersonen zu ermöglichen. So kann es günstiger sein, eine Stichprobe von Personen über die Zeit wiederholt zu befragen, anstatt immer wieder neue Stichproben von Befragungspersonen aus der gleichen Grundgesamtheit zu ziehen. Zusätzlich ermöglicht ein bestehendes Panel es dem Forscher, viel schneller mit einer Befragung auf aktuelle Themen in Politik und Gesellschaft zu reagieren als dies der Fall wäre, wenn vorher eine Stichprobe von Befragungspersonen gezogen werden müsste. Bei Panels, die aus diesen Gründen eingerichtet werden, dominiert die Erhebung von Querschnittsdaten, die wiederholte Messung des gleichen Sachverhalts zur Verbesserung der Kausalanalyse steht hingegen im Hintergrund. Häufig werden Studien mit diesem Ansatz auch als Access-Panels bezeichnet (Engel, Bartsch, Schnabel, & Vehre, 2012).

2. Nachteile von Panelstudien

Neben den obengenannten Vorteilen der Längsschnittdaten sind Panelstudien auch mit bestimmten erhebungsmethodischen Problemen verbunden. Zum einen kann der Ausfall der Befragungseinheiten über die Zeit hinweg (Panelattrition) selektiv verlaufen, d.h. die Stichprobe wird mit zukünftigen Wellen weniger repräsentativ für die Grundgesamtheit (Kasprzyk, Duncan, Kalton, & Singh, 1989; Lynn, 2009; Menard, 2008). Je nach Analyseplan ist dieses Problem mehr oder weniger schwerwiegend. Es kann erstens durch Maßnahmen bei der Studiendurchführung, wie z.B. Panelpflege, Interviewertraining, Kontakt- und Refusal-Conversion-Prozedere, und Incentives abgeschwächt werden. Zweitens kann man das Problem nach der Befragung durch statistische Verfahren bearbeitet werden. Beispiele hierfür sind Gewichtung, Heckman-Modelle, die den Ausfall explizit modellieren und Imputationsverfahren. Allerdings setzen diese Verfahren voraus, dass die Einflussgrößen, die den selektiven Ausfall abbilden, auch erhoben wurden.

Eine weitere methodische Besonderheit von Panelstudien ist, dass die Befragung in vergangenen Wellen die Befragung in zukünftigen Wellen beeinflussen kann (panel conditioning) (Warren & Halpern-Manners, 2012). So kann sich einerseits der tatsächliche Wert einer Variable in Folge des Interviews ändern, weil das Verhalten oder die Einstellung des Befragten durch die Befragung beeinflusst wird. Eine andere Variante des Panel-Conditionings ist, dass sich zwar der wahre Wert einer Variablen nicht verändert, jedoch das Antwortverhalten. Die bisherige Forschung zeigt, dass Panel-Conditioning je nach Forschungsgegenstand und Frage auftreten kann, aber nicht auftreten muss. Es gibt bisher nur wenige

Erkenntnisse und Hypothesen, bei welcher Art von Fragen Panel-Conditioning auftritt, etablierte Lösungsansätze gibt es nicht.

3. Entscheidung für oder gegen ein Panel-Design

Wie oben dargestellt, sind – abgesehen von Experimenten – Panelstudien Querschnittsstudien in der Regel für die Kausalanalyse überlegen. Sind für eine Analyse hingegen nur Querschnittsdaten notwendig, wie dies z. B. bei deskriptiven Analysen zur Verteilung bestimmter Merkmale in der Bevölkerung der Fall ist, so sind umgekehrt Querschnittsstudien gegenüber Panelstudien zu bevorzugen. Dies gilt prinzipiell auch für Trendanalysen, bei denen der zeitliche Verlauf von Merkmalsverteilungen untersucht wird. Hier sind Trenddaten auf mehreren Querschnittsstudien Paneldaten aus Panelstudien in der Regel vorzuziehen. In beiden Fällen bringen Paneldaten keinen Analysegewinn, sind aber, wie eben schon erwähnt, mit möglichen Abstrichen in der Datenqualität durch selektive Panelattrition und Panel-Conditioning verbunden. Bei lange laufenden Panels kommt hinzu, dass die Stichprobe zwar die Bevölkerungszusammensetzung zum Zeitpunkt der Stichprobenziehung repräsentiert, aber unter Umständen (z. B. durch Migration) nicht der aktuellen Bevölkerungszusammensetzung entspricht.

Die Abwägung der Vor- und Nachteile eines Panel-Designs hängt entsprechend den obigen Ausführungen von der Ausrichtung des Panels ab. Ist ein Panel geplant, das ein besseres Analysepotential durch die Erhebung von individuellen Längsschnittdaten anstrebt, so sind die Vor- und Nachteile dieses Designs mit einer (unter Umständen retrospektiven) Querschnittsstudie bzw. einer Trendstudie zu vergleichen, die die Daten für den gleichen Analysezweck erhebt. Die Panelstudie ist durch die zusätzlichen Kosten der Mehrfacherhebung und der Panelpflege teurer und mit zusätzlichem zeitlichem Aufwand und Wartezeit verbunden, liefert aber andererseits bessere Daten zur Überprüfung von kausalen Zusammenhängen oder zur deskriptiven Darstellung von individuellen Entwicklungen über die Zeit. Sowohl die mit dem Panel-Design verbundenen Vorteile hinsichtlich der Datenanalyse als auch die Kosten bzw. die Realisierbarkeit einer Panelstudie variieren hier stark mit der Forschungsfrage. Entsprechend müssen die Vor- und Nachteile im Vergleich zu alternativen Erhebungsplänen immer für den Einzelfall abgewogen werden. Ist hingegen – als anderes Extrembeispiel – ein Accesspanel geplant, mit dem in wiederholten Befragungen lediglich unterschiedliche Querschnittsdaten erhoben werden sollen, so ist dieses Design mit der Alternative von wiederholten Querschnittsbefragungen mit jeweils eigener Stichprobenziehung zu vergleichen. Beim Accesspanel ergibt sich gegenüber der wiederholten Querschnittsbefragung unter Umständen ein Kostenersparnis durch die einmalige Stichprobenziehung und es ist möglich, schneller mit einer Befragung auf aktuelle Themen in Politik und Gesellschaft zu reagieren. Dem stehen mögliche Einbußen in der Datenqualität durch selektive Panelattrition und Panel-Conditioning gegenüber. Auch hier ist die Abwägung immer vom Einzelfall abhängig.

Literaturverzeichnis

Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles: SAGE.

Andreß, H.-J., Golsch, K., & Schmidt, A. S. (2013). *Applied panel data analysis for economic and social surveys*. Berlin, London: Springer.

Arellano, M. (2003). *Panel data econometrics*. Oxford, New York, NY: Oxford University Press.

Baltagi, B. H. (2005). *Econometric analysis of panel data* (3rd ed.). Chichester, Hoboken, NJ: John Wiley & Sons.

- Brüderl, J. (2010). Kausalanalyse mit Paneldaten. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 963–994). Wiesbaden: VS Verlag.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics: Methods and applications*. Cambridge, New York, NY: Cambridge University Press.
- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata*. College Station, TX: Stata Press.
- Engel, U., Bartsch, S., Schnabel, C., & Vehre, H. (2012). *Wissenschaftliche Umfragen: Methoden und Fehlerquellen*. Frankfurt am Main: Campus.
- Giesselmann, M., & Windzio, M. (2014). Paneldaten in der Soziologie: Fixed Effects Paradigma und empirische Praxis in Panelregression und Ereignisanalyse. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(1), 95–113. doi:10.1007/s11577-013-0248-z
- Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, 30, 507–544.
- Hsiao, C. (2003). *Analysis of panel data* (2nd ed.). Cambridge, New York, NY: Cambridge University Press.
- Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M. P. (Eds.). (1989). *Panel surveys*. New York, NY: Wiley.
- Lazarsfeld, P., & Fiske, M. (1938). The "panel" as a new tool for measuring opinion. *Public Opinion Quarterly*, 2(4), 596–612.
- Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). *The people's choice: How the voter makes up his mind in a presidential campaign*. New York, NY: Duell, Sloan and Pearce.
- Lynn, P. (Ed.). (2009). *Methodology of longitudinal surveys*. Chichester: Wiley.
- Mátyás, L., & Sevestre, P. (Eds.). (1996). *The econometrics of panel data: Handbook of theory and applications*. Dordrecht: Kluwer Academic.
- Menard, S. (Ed.). (2008). *Handbook of longitudinal research: Design, measurement, and analysis*. Amsterdam, Boston, MA: Elsevier.
- Opp, K.-D. (2010). Kausalität als Gegenstand der Sozialwissenschaften und der multivariaten Statistik. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 9–38). Wiesbaden: VS Verlag.
- Warren, J. R., & Halpern-Manners, A. (2012). Panel Conditioning in Longitudinal Social Science Surveys. *Sociological Methods & Research*, 41(4), 491–534. doi:10.1177/0049124112460374
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.