

Änderungen am Konzept der Querschnittsgewichtung des Sozio- oekonomischen Panels (SOEP) 1984 - 2001

von Rainer Pischner

Mit der 18. Welle (2001) sind einige Änderungen am Hochrechnungskonzept für die Querschnittsgewichtung vorgenommen worden, die im folgenden kurz dargestellt werden.

Die Berechnung der Längsschnittgewichte sowie die gesonderte Gewichtung der Zuwanderer-Stichprobe bleiben von den Neuerungen unberührt.

Bisheriges Konzept der Querschnittshochrechnung

Für jede neue Stichprobe wurden Designgewichte geschätzt, die als Startwerte für eine Randanpassung verwendet fanden. Diese Randanpassung berücksichtigte relevante Eckzahlen der amtlichen Statistik. Für die Folgewellen wurden die Querschnittsgewichte zunächst mit Hilfe sog.

Bleibwahrscheinlichkeiten fortgeschrieben. Die so ermittelten Gewichte wurden anschließend sparsam an einige Werte des Mikrozensus angepasst. Wesentliches Merkmal dieses Konzept ist die serielle Fortschreibung der Querschnittsgewichte, die sich zwingend aus der wellenspezifischen Abhängigkeit der jährlichen Querschnitte ergibt.

Änderungen am Konzept

1. Für die Startgewichtung der Stichprobe F sind detailliertere Designgewichte als bisher zu Grunde gelegt worden. In den neuen Startgewichten wird auch die Ausschöpfung der Sample-Points berücksichtigt. Zunächst war hierauf verzichtet worden, um eine zu hohe Varianz der Hochrechnungsfaktoren zu vermeiden, zumal die unterschiedliche Ausschöpfung der Sample-Points über die Anpassung an die „Wohnumfeldvariablen“ zumindest teilweise aufgefangen wurde. In Zusammenhang mit der wohl wichtigsten Änderung am Hochrechnungskonzept, die die Werte der Hochrechnungsfaktoren nach oben begrenzt, ist dieser Einwand hinfällig geworden (s.u.).
2. Die Randanpassung an die Altersstruktur ist verfeinert worden. Die Anpassung erfolgt jetzt bis zum 80. Lebensjahr in 5 Jahreskohorten.
3. Über mehrere Wellen – beim SOEP inzwischen bis zu 18 Wellen – führt der serielle Fortschreibungsprozess der Hochrechnungsfaktoren zu Divergenzen und somit zu einer höheren Varianz der Hochrechnungsfaktoren. Dazu tragen vornehmlich temporäre Ausfälle von Haushalten bei, die bei ihrer Rückkehr

in die Stichprobe im allgemeinen deutlich höhere Gewichte erhalten. Im Laufe der Jahre sind einige Haushalte mit sehr hohen Hochrechnungsfaktoren vorgekommen, die vor allem dann größere Probleme bereiten, wenn sie endgültig ausfallen, da ihre Gewichtungsmasse nicht angemessen auf die übrigen Haushalte verteilt werden kann. Um diesem Effekt zu begegnen, wurden maximale Obergrenzen für die Querschnittsgewichte eingeführt. Da gleichzeitig die Bedingungen der Randanpassung erfüllt bleiben müssen, wurde die Limitierung bei den Startgewichten der Randanpassung angesetzt. Für jede Stichprobe wurden Startgewichte, deren Werte das 10-fache ihres stichprobenspezifischen Medians überschritten, gekappt. Die Überschüsse wurden dann ebenfalls stichprobenspezifisch verteilt. Dann erst erfolgte die Randanpassung. In Ausnahmefällen übersteigen die endgültigen Hochrechnungsfaktoren die vorgegebenen Grenzwerte, insgesamt wurde aber eine deutliche Verminderung der Extremwerte und der Varianz der Hochrechnungsfaktoren erreicht.

Eine ausführliche Dokumentation über die Querschnittsgewichtung ist in Vorbereitung.