

## **Hinweise zur Generierung von Oesch-Klassen mit dem SOEP (V40<sup>1</sup>)**

Der vorliegende SPSS-Job generiert die Oesch16-Klassen aus den Daten der SOEP Long-Files (V40) für den gesamten Erhebungszeitraum seit 1984.

Für die Aufbereitung der Selbstständigen wird auf einen Stata Job von Oliver Lipps auf der Oesch-Website zurückgegriffen: SOEP - CNEF equivalent file, all available waves. Der Job wurde an die SPSS-Syntax bzw. die veränderten Variablen in den SOEP-Long-Files angepasst. Für die Generierung der Klassen bei den abhängig Beschäftigten wurden die auf der Website angebotenen Excel-Files genutzt (Excel for occupations measured with ISCO-88, 4-digits bzw. Excel for occupations measured with ISCO-08, 4-digits). Da für die Gruppe der Soldaten und der Auszubildenden bzw. Praktikanten keine ISCO-Zuordnungen vorlagen, wurden diese ausgeschlossen.

Bei der Generierung stellen sich drei wesentliche Probleme:

### **1. ISCO-Daten im SOEP**

Im SOEP werden die aktuellen Berufe auf Basis von ISCO88 bzw. ISCO08 klassifiziert. Dementsprechend werden mit dem Job zwei verschiedene Variable generiert:

- oesch16\_88 für 1984-2017
- oesch16\_08 für 2013-2023

Der Überlappungsbereich ermöglicht eine Abschätzung des Einflusses der verschiedenen ISCO-Standards. Er ist m.E. angesichts anderer Unschärfen nicht erheblich.

### **2. Kodierung der Selbstständigen**

Bei den Selbstständigen liegen die Probleme in den sich verändernden Kategorisierungen der verschiedenen Selbstständigengruppen. So können zwischen 1984 und 1999 die Soloselbstständigen im SOEP nicht abgegrenzt werden. D.h. die Klasse 4 (Small business owners without employees) besteht bis 1999 nur aus den Mithelfenden Familienangehörigen. Folglich kommt es zwischen 1999 und 2000 zu einem Sprung in der Zeitreihe der Selbstständigen.

### **3. Fehlende Werte bei den abhängig Beschäftigten**

Bei den abhängig Beschäftigten stellt sich das Problem fehlender Werte bei den Angaben zu den ISCO-Berufen. Für alle Befragten liegen qualifikationsdifferenzierte Angaben zur „Stellung im Beruf“ vor; bei den ISCO-88 bzw. ISCO-08 Angaben finden sich jedoch nicht wenige fehlende Werte. Diese sind in den Variablen oesch16\_88 und

---

<sup>1</sup> Wegen eines Datenfehlers in der Datei pl.sav wurde dieser Job bislang nur mit der V39 getestet. Nach Sichtung der Variablenstruktur müsste der Job aber auch mit der Version 40 laufen.

oesch16\_08 als fehlende Werte erkennbar. Für sehr wenige Fälle liegen Angaben aus der KldB1992, KldB2012 bzw. ISCO-68 vor. Diese Informationen wurden angesichts der geringen Zahl der Fälle jedoch nicht berücksichtigt, da sie nicht zu einer erheblichen Verbesserung der Klassifizierung beitragen.

Um das Problem der fehlenden ISCO-Daten abzumildern, wurden in den Variable oesch16\_88\_a bzw. oesch16\_08\_a bei fehlenden ISCO-Angaben die Oesch-Klassen aus den qualifikatorischen Angaben zur Stellung im Beruf und aus den Angaben zur Branche (NACE-2-stellig) generiert; das ist jedoch nur eine Annäherung, da die Angabe zur Branche nur betriebs- und nicht personenspezifisch erfolgt.

Leider liegen für immer hin 13,7% der oesch16\_88\_a und 16,8% der oesch16\_08\_a Generierungen weder ISCO- noch NACE-Angaben vor. Für eine Abschätzung der daraus erwachsenden Verzerrungen wurde geprüft, ob sich die qualifikatorische Struktur der erfolgreich kodierten Klassen von denen der fehlenden Werte unterscheidet. Leider gibt es hier nicht geringe Abweichungen, wie die folgende Tabelle zeigt.

ISCO-88a	erfolgreich kodiert	fehlende Werte	ISCO-08a	erfolgreich kodiert	fehlende Werte
low skilled	16,9%	31,4%	low skilled	17,5%	29,6%
skilled	41,6%	43,9%	skilled	35,4%	41,9%
semiprofessional	24,2%	19,2%	semiprofessional	26,7%	23,8%
professional	17,3%	5,4%	professional	20,8%	4,7%

D.h. die fehlenden Angaben führen tendenziell zu einer Unterschätzung der untersten und einer Überschätzung der obersten Klassen der verschiedenen worklogics. Da sich dieses Problem nur bei den fehlenden Werten (13,7% bzw. 16,8%) stellt, kann man mit der Generierung oesch16\_88\_a bzw. oesch16\_08\_a durchaus arbeiten, wenn man diese Verzerrungen benennt. Wer die Unschärfen der Näherungslösung über Stellung im Beruf und Branche vermeiden möchte, sollte mit den Variablen oesch16\_88 bzw. oesch16\_08 arbeiten, die Zahl der Missings und die oben skizzierten Unschärfen werden dann jedoch größer sein.

Christoph Weischer