

Forschungsdatenzentrum

der Bundesagentur für Arbeit  
im Institut für Arbeitsmarkt-  
und Berufsforschung

FDZ

# FDZ-Methodenreport

Methodische Aspekte zu Arbeitsmarktdaten

12/2017

DE

## Zusammenführen der Wellen des IAB-Betriebspanels

Ein Do-File für die grundlegende Aufbereitung eines  
Paneldatensatzes in Stata

Matthias Umkehrer



Bundesagentur für Arbeit

# Zusammenführen der Wellen des IAB-Betriebspanels

Ein Do-File für die grundlegende Aufbereitung eines Paneldatensatzes in Stata

Matthias Umkehrer (Institut für Arbeitsmarkt- und Berufsforschung)

Die FDZ-Methodenreporte befassen sich mit den methodischen Aspekten der Daten des FDZ und helfen somit Nutzerinnen und Nutzern bei der Analyse der Daten. Nutzerinnen und Nutzer können hierzu in dieser Reihe zitationsfähig publizieren und stellen sich der öffentlichen Diskussion.

FDZ-Methodenreporte (FDZ method reports) deal with methodical aspects of FDZ data and help users in the analysis of these data. In addition, users can publish their results in a citable manner and present them for public discussion.

## **Inhaltsverzeichnis**

Zusammenfassung	3
Abstract	3
1 Ausgangslage	4
2 Grundlagen der Datenaufbereitung	5
3 Der Aufbau des Do-files	7
4 Wie das Do-file zu bedienen ist	9
5 Anmerkungen zu aufbereiteten Panelvariablen	11
Literaturverzeichnis	13
Appendix	13

### Tabellenverzeichnis

Tabelle 1 - Beispiel für die Variablen-Zeitpunkte-Matrix.....	5
---	---

### Abbildungsverzeichnis

Abbildung 1 - Panelstruktur - vom Format "Wide" zum Format "Long".....	6
Abbildung 2 - Aufbau des Do-files (level 1).....	7
Abbildung 3 - Aufbau des Do-files (level 2).....	8

## Zusammenfassung

Dieser FDZ-Methodenreport stellt das Do-file *BP\_panelgen\_93JJ\_v1* vor, welches alle Wellen des IAB-Betriebspanels von 1993 bis zum aktuellsten Jahr zu einem Paneldatensatz in Stata zusammenfügt. Er erläutert die Ausgangslage, skizziert den Aufbau des Do-files, beschreibt den Paneldatensatz und bietet Bedienhinweise. Die Variablen im generierten Paneldatensatz sind konsistent benannt, kodiert und beschriftet, um eine standardisierte Grundlage für Panel-Analysen zu schaffen. Der Programmcode ist einfach und transparent gehalten, sodass Nutzerinnen oder Nutzer bei Bedarf leicht Anpassungen vornehmen können. Allerdings gilt es zu beachten, dass sich die generierten Panelvariablen, zum Beispiel aufgrund von Änderungen der Fragestellung, noch immer in ihrem Inhalt über die Wellen hinweg unterscheiden können.

## Abstract

This FDZ-Methodenreport introduces the do-file *BP\_panelgen\_93YY\_v1\_en* for combining all waves of the IAB Establishment Panel from 1993 to the most recent year to one panel data set in Stata. It explains the initial problem, outlines the structure of the do-file, describes the panel data set and provides operating instructions. The variables in the resulting panel data set are consistently named, coded and labelled to provide a standardized basis for panel analyses. The program code is kept simple and transparent so that users can easily adjust it if necessary. However, caution has to be taken as generated panel variables might still differ in content across waves due to changes in the questionnaire, for instance.

**Keywords:** IAB-Betriebspanel, Datenaufbereitung, Paneldaten, Datenmanagement

## Anmerkung und Danksagung

Für die Inhalte der FDZ-Daten- und Methodenreporte sind die jeweiligen Autoren verantwortlich. Matthias Umkehrer hat den Kern der Stata Syntax konzipiert und den FDZ-Methodenreport verfasst. Sandra Dummert hat die Fertigstellung der Syntax unterstützt. Besonderer Dank kommt Jochen Späth für die Bereitstellung seiner Aufbereiter-routinen zu. Weiterer Dank gebührt Stefanie Wolter, Dana Müller, Jens Stegmaier und Tina Hinz für wertvolle Anmerkungen und Kommentare, ebenso wie Jasmin Werle, Falk Voigt, Lisa Schmidlein, Florian Heinritz, Marie Sophie Jestadt und Stefan Scholz für ihre unterstützenden Arbeiten.

## Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich. Nähere Informationen dazu sind auf der Internetseite: <http://fdz.iab.de/> unter der Rubrik „Betriebsdaten“ zu finden.

# 1 Ausgangslage

Das IAB-Betriebspanel ist eine repräsentative Befragung von Betrieben, die seit 1993 jährlich durchgeführt wird. Es stellt die zentrale Datenbasis für Analysen der Arbeitsnachfrage in Deutschland dar, indem es detaillierte Informationen zu einer Vielzahl an arbeitsmarktpolitisch relevanten Themen bietet, wie beispielweise Personalentwicklung, Beschäftigtenstruktur, Geschäftspolitik, Investitionstätigkeit, öffentliche Förderung und betriebliche Ausbildung. Eine detaillierte Beschreibung zur Erhebung des IAB-Betriebspanels findet sich in Fischer et al. (2008); weiterführende Informationen zu Datenzugang, Methodik und Arbeitshilfen finden sich auf der Internetseite des FDZ: <http://fdz.iab.de/>.

Das IAB-Betriebspanel ist als Langzeitbefragung ausgelegt, d.h. zum einen nimmt eine Vielzahl der Betriebe wiederholt an der Befragung teil. Zum anderen sind die Fragebogeninhalte über die Jahre zu großen Teilen konstant. Aufgrund der langfristigen Auslegung wird es möglich Entwicklungen über die Zeit nicht nur anhand von Querschnittvergleichen sondern auch innerhalb der Betriebe zu untersuchen.

Jedoch ist das IAB-Betriebspanel nicht als Paneldatensatz aufbereitet, vielmehr stehen die einzelnen Querschnittsdaten zur Verfügung. Eine Harmonisierung von Querschnittvariablen über die Zeit kann mühsam sein und ist nicht immer unproblematisch. Die Variablennamen sind wellenspezifisch angelegt. Sie setzen sich aus einem Kleinbuchstaben, der die Welle kodiert, gefolgt von der Nummer der Frage im jeweiligen Fragebogen zusammen. So bezeichnet a07 die siebte Frage im Fragebogen des Startjahres 1993. Zudem treten Änderungen des Wortlauts einer Frage, der Antwortkategorien, der Kodierung einzelner Werte oder der Messeinheit und des Bezugszeitraums auf.

Um die Nutzerinnen und Nutzer beim Zusammenführen der Wellen des IAB-Betriebspanels zu unterstützen, bietet das FDZ das Stata Do-File *BP\_panelgen\_93JJ\_v1* an.<sup>1</sup> Dieses Do-File erstellt einen Paneldatensatz für alle Jahre von 1993 bis zum aktuellsten Jahr und für alle Variablen im IAB-Betriebspanel. Die resultierenden Panelvariablen sind über alle Wellen hinweg konsistent benannt, kodiert und gelabelt. Soweit möglich wird die ursprüngliche Information beibehalten, und es wird auch keinerlei künstlich generierte Information hinzugefügt, beispielsweise durch vor- und zurückschreiben.

Das Ziel der Datenaufbereitung ist es eine einheitliche Grundlage für Panelanalysen mit dem IAB-Betriebspanel zu schaffen. Allerdings gilt es zu beachten, dass einige Panelvariablen, aufgrund der zuvor genannten Probleme, noch immer in ihrem Inhalt über die Wellen hinweg abweichen können. Es liegt in der Verantwortung der Nutzerin oder des Nutzers sicherzustellen, dass die generierten Daten für das jeweilige Forschungsvorhaben geeignet sind. Insbesondere sollte geprüft werden, ob die zugrunde liegenden Fragen und/oder Umkodierungen über die Wellen hinweg hinreichend konsistent für die Beantwortung der wissenschaftlichen

---

<sup>1</sup> Das IAB-Betriebspanel ist auch Bestandteil der verschiedenen Modelle der Linked-Employer-Employee-Daten des IAB (LIAB), siehe Klosterhuber et al. (2016) zum Querschnittmodell und Heining et al. (2016) zum Längsschnittmodell, ebenso wie von WeLL und LPP. Nutzerinnen und Nutzer dieser Daten können das Do-file in exakt derselben Weise wie in diesem Methodenreport beschrieben anwenden.

Fragestellung sind. Aus diesem Grund empfiehlt das FDZ ausdrücklich neben der Arbeit mit dem Do-file auch immer die Fragebögen sowie das Excel-File „Variablenliste“, verfügbar auf der Internetseite des FDZs, heranzuziehen.

Der folgende Abschnitt beschreibt die Veränderung der Datenstruktur beim Zusammenführen der Wellen zu einem Paneldatensatz sowie grundlegende Schritte der Harmonisierung. Anschließend werden der Aufbau des Do-Files erläutert und Bedienhinweise gegeben. Der letzte Abschnitt beinhaltet einige Anmerkungen zur Aufbereitung bestimmter Panelvariablen.

## 2 Grundlagen der Datenaufbereitung

Um einen Überblick über die Variablen zu schaffen, die in mindestens zwei Wellen des IAB-Betriebspanels mit hinreichender Übereinstimmung erhoben wurden, bietet das FDZ eine Variablen-Zeitpunkte-Matrix an.<sup>2</sup> Ein erläuterndes Beispiel für diese Matrix wird in Tabelle 1 präsentiert.

**Tabelle 1 - Beispiel für die Variablen-Zeitpunkte-Matrix**

Variablen-Zeitpunkte-Matrix	Modul	1993	1994	...	20JJ	Panelvariable	Anm. zum Panel
	(A) bis (Q), sensible Merkmale,						
Variablen Label zu "panvar1"	Zusatzinformationen	a1	b1	...	X1	panvar1	-
Identifikationsnummer	Zusatzinformationen	idnum	idnum	...	idnum	idnum	-
Jahr der Befragung	-	-	-	...	-	year	-

Jede Zeile der Variablen-Zeitpunkte-Matrix beinhaltet die originären Namen der zusammenhängenden Variablen aus den einzelnen Querschnitten, beispielsweise a1 bis X1 als hypothetisches Beispiel. Die erste Spalte bietet eine knappe Beschreibung des Inhalts der jeweiligen Querschnittvariablen, welche auch als Variablenlabel der neu generierten Panelvariable dient. Die zweite Spalte umfasst die thematischen Module, denen jede Gruppe an Variablen zugeordnet ist. Organisationsvariablen, wie die idnum, sind beispielsweise Teil des Moduls „Zusatzinformationen“.<sup>3</sup> Die Variable year, hingegen, ist nicht in den einzelnen Querschnitten enthalten, sondern wird während der Aufbereitung des Panels erstellt. Die Namen der Panelvariablen sind in der vorletzten Spalte enthalten, wie beispielsweise panvar1 in diesem hypothetischen Beispiel. Schließlich beinhaltet die letzte Spalte noch einige Anmerkungen zu bestimmten Panelvariablen, siehe Abschnitt 5 für Details.

Die Original-Querschnitte sind als separate Datensätze für jede Welle im Format „wide“ abgelegt, d.h. jeder Datensatz enthält eine Datenzeile pro erfassten Betrieb. Die Tabellen im oberen Teil von Abbildung 1 bieten ein fiktives Beispiel für dieses Format: Die Zeilen beinhalten alle Variablen-Ausprägungen, die einem jeden Betrieb zugeordnet werden können, wie etwa den Betriebs-Identifikator idnum und sämtliche Variablen-Ausprägungen  $wq_j$ , die in der aktuellen Welle erhoben wurden, mit  $j = [1; N]$  für den Betrieb,  $w = [a; X]$  für die Befragungswelle und  $q$

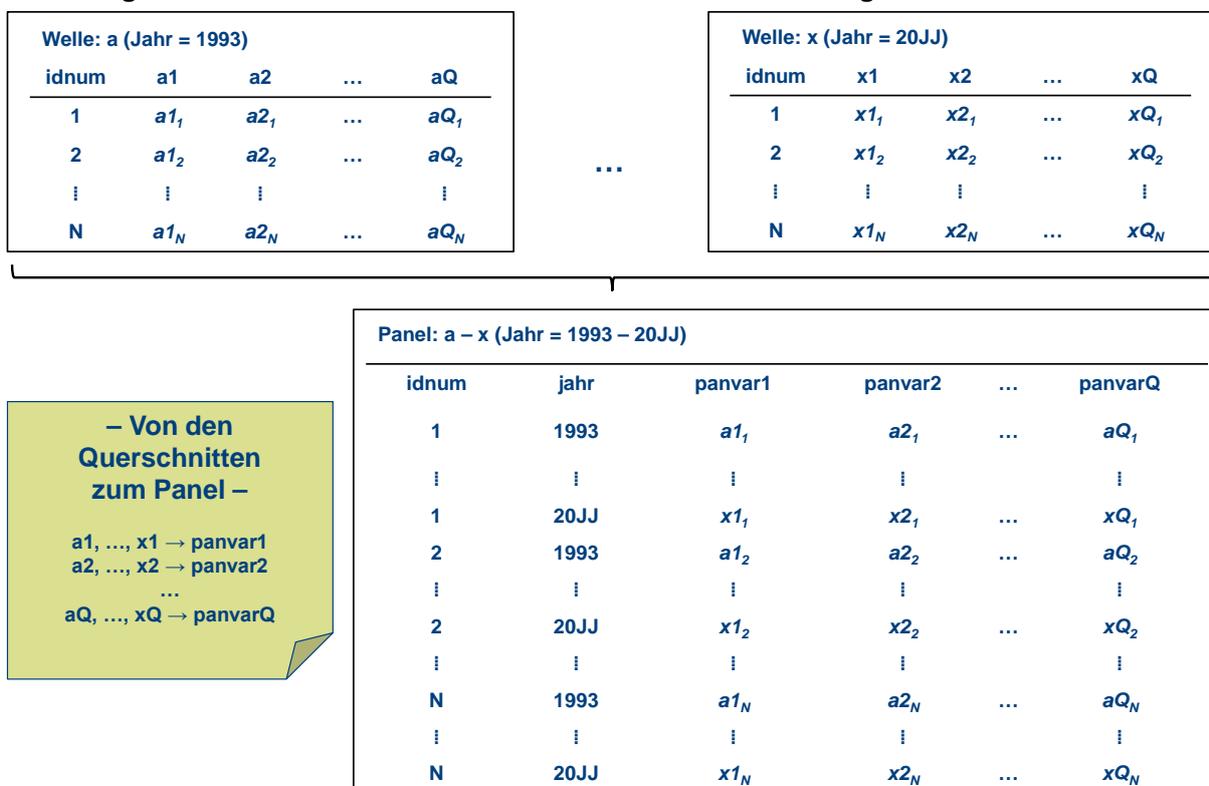
<sup>2</sup> Diese Matrix ist im Tabellenblatt „Panelvariablen“ des Excel-Files „Variablenliste“ zu finden. Das Tabellenblatt „Zusatzvariablen“ hingegen listet diejenigen Variablen, die lediglich in einer einzelnen Welle erhoben wurden, üblicherweise als Teil eines wellenspezifischen Schwerpunktthemas.

<sup>3</sup> Abschnitt 4 erläutert wie die Nutzerin oder der Nutzer die Erstellung des Paneldatensatzes auf einzelne Module oder Modulgruppen beschränken kann.

= [1; Q] für die zugrunde liegende Frage (unter der vereinfachenden Annahme, dass Fragen  $q$  identisch für alle Wellen sind).

Die Aufbereitung zu einem Paneldatensatz ändert das Datenformat nun von „wide“ auf „long“, wie im unteren Teil von Abbildung 1 dargestellt. Dies impliziert, dass die Querschnitte so zusammengespielt werden, dass der sich ergebende Datensatz eine Zeile für jeden Betrieb und für jedes Jahr, in dem der Betrieb erfasst wurde, beinhaltet. Gleichzeitig werden die einzelnen Querschnittvariablen  $a_q$  bis  $x_q$  zu einer Panelvariable zusammen geführt. In diesem Beispiel trägt die Panelvariable den Namen  $panvar_q$ . Folglich enthalten die Spalten des Paneldatensatzes, neben  $idnum$  und  $year$ , die neuen Panelvariablen, welche wiederum alle zusammenhängenden Querschnittvariablen umfassen (ursprünglich die Einträge in den Zeilen der Variablen-Zeitpunkte-Matrix).

**Abbildung 1 - Panelstruktur - vom Format "Wide" zum Format "Long"**



Im Rahmen der Panelaufbereitung nimmt das Do-file einige grundlegende Schritte vor, um die Querschnittvariablen zwischen den Wellen anzugleichen. Dazu gehören

- Zuweisung eindeutiger Namen zu den Panelvariablen (Spalte „Panelvariable“ der Variablen-Zeitpunkte-Matrix),
- Zuweisung von Variablen-Labels zu den Panelvariablen (erste Spalte der Variablen-Zeitpunkte-Matrix),
- konsistente Beschriftung der Kategorien, die in allen relevanten Wellen vorkommen (Vereinheitlichung von Wertelabels),
- Hinzufügen einer Kategorie „nein“ / „0“ zu Variablen, wenn lediglich die Kategorie „ja“ / „1“ in den Ursprungsdaten kodiert wurde,

- Zusammenfassen unterschiedlicher Missing-Kategorien zu einer Missing-Kategorie mit Ausprägung „9“ und Label „n.s.“ (nicht spezifiziert) – wurde eine Variable nicht erfasst wird der standard-Wert „.“ für fehlende Werte kodiert – und
- diverse fallabhängige Anpassungen, wie in Abschnitt 5 beschrieben.

### 3 Der Aufbau des Do-files

Der Programmcode von *BP\_panelgen\_93JJ\_v1* ist einfach und transparent gehalten, sodass Nutzerinnen oder Nutzer bei Bedarf leicht Anpassungen vornehmen können. Es gibt nur ein einziges Do-file, dessen grundlegender Aufbau in Abbildung 2 und Abbildung 3 beschrieben ist.

Abbildung 2 - Aufbau des Do-files (level 1)

```

BP_panelgen_9316_v1_red.do* Untitled.do
* Zweck: Erstellung eines Panel-Datensatzes aus dem IAB-Betriebspanel *
*** 1. Grundlagen

*** 2. Wahl der zu erstellenden Module

** (A) Allgemeine Angaben zum Betrieb
local A_gen = 0

** (B) Altersstruktur und Fachkräftebedarf
local B_gen = 0
...
** (Q) Personalstruktur
local Q_gen = 0

** (ZZ) Zusatzinformationen
local zuinf = 0

** Daten aus Einzelerhebungen den entsprechenden Modulen hinzuspielen
local ezinf = 0

** Sensible Merkmale, sofern vorhanden, bitte separat zuzuspielen

*** 3. Generierung der Panel-Daten

if `A gen' == 1 {
else if `A gen' == 0 {

if `B gen' == 1 {
else if `B gen' == 0 {
...
if `Q gen' == 1 {
else if `Q gen' == 0 {

if `zuinf' == 1 {
else if `zuinf' == 0 {

*** 4. Zusammenspielen und schliessen

```

Abbildung 2 zeigt, dass das Do-file in vier Abschnitte untergliedert ist. Der erste Abschnitt konfiguriert grundlegende Einstellungen, wie die Version von Stata, Verzeichnisse und die Benennung von Log-/ und Daten-Files.

Im zweiten Abschnitt sollen die für das Forschungsprojekt relevanten Module von (A) bis (ZZ) ausgewählt werden, die dann in den Paneldatensatz aufgenommen werden. Hierzu wird das entsprechende lokale Makro von „0“ auf „1“ gesetzt. Standardmäßig werden nur die Variablen in den Paneldatensatz aufgenommen, die in mehreren Wellen erhoben wurden. Jedoch besteht die Möglichkeit, sogenannte „Zusatzvariablen“, also Variablen die nur einmal erhoben wurden, an die gewählten Module anzuspielen, indem das lokale Makro ezinf von „0“ auf „1“ gesetzt wird.

Abbildung 3 - Aufbau des Do-files (level 2)

```

BP_panelgen_9316_v1_red2.do  Untitled.do
*** 3. Generierung der Panelvariablen

if `A_gen' == 1 {

    di " (A) Allgemeine Angaben zum Betrieb"

    * Lade Variablen des gewählten Moduls
    local A_var_1993 = "a26 a73 a74 a75 a76a a76b"
    local A_var_1994 = "b65 bz02 bz03 bz04 bz05 bz06 bz07a bz07b"
    ...
    local A_var_2016 = "x80 x77a x77d x77e x77g x77b x77c x77f ..."

    gen year = .
    foreach YEAR of numlist 1993(1)2016 {
        append using ${orig}\iabbbp_`YEAR'.dta, keep(idnum `A_var_`YEAR')
        replace year = `YEAR' if year >= .
    }

    *-----*
    ** Panelvariablen erstellen und speichern
    ...
    *-----*

    ** Zusatzvariablen anspielen, sofern aktiviert
    if `ezinf' == 1 {
    }

    ** Leeren Datensatz speichern, sofern nicht aktiviert
    else if `A_gen' == 0 {

```

In einem dritten Schritt bereitet das Programm temporär Teildatensätze für jedes Modul auf. Ist ein Modul nicht aktiviert wird ein leerer Datensatz erstellt. Ist das Modul aktiviert erfolgt die Aufbereitung wie in Abbildung 3 dargestellt: Erstens werden alle zusammenhängenden Variablen des jeweiligen Moduls mittels lokaler Makros wellenweise eingelesen, also diejenigen Variablen, die in den Zeilen der Variablen-Zeitpunkte-Matrix abgebildet sind. Zweitens werden die Daten Welle für Welle untereinander angespielt, wobei auch die Variable year für die Zeitdimension des Panels generiert wird. Daraus ergibt sich ein zusammengespielter Datensatz, der eine Spalte pro Querschnittvariable enthält. Sind keine Informationen in einem Jahr vorhanden werden fehlende Werte „.“ gesetzt. Drittens werden die Querschnittinformationen im zusammengespielten Datensatz mit Hilfe der egen-Funktion rowmin() zu einer neuen Variablen verdichtet. Dies funktioniert da die Querschnittvariablen eindeutig benannt und somit nach

dem Zusammenspielen nur für das relevante Jahr gefüllt sind. Diese neue Variable wird beibehalten und zur Panelvariable aufbereitet, wie zu Ende des vorherigen Abschnittes sowie im letzten Abschnitt beschrieben.

In einem vierten Schritt werden nun die modulspezifischen Teildatensätze über `idnum` und `year` zu einem Paneldatensatz zusammengeführt. Der Name des Paneldatensatzes entspricht dabei dem Namen des Do-Files, durch welches er generiert wurde, also hier *BP\_panelgen\_93JJ\_v1*.

Das Do-file wurde mit der Datenversion IABBP\_9316\_v1 und unter Stata 14.1 getestet. Das FDZ übernimmt für die Korrektheit der erstellten Inhalte keine Garantie. Hier ist von der Nutzerin oder dem Nutzer zu prüfen, ob die generierten Merkmale richtig und für die eigene Forschungsfrage geeignet sind. Des Weiteren können wir Fehler im Programmcode und in der Variablen-Zeitpunkte-Matrix nicht vollständig ausschließen. Sollten Sie einen Fehler finden, informieren Sie uns bitte per E-Mail an die FDZ-Box ([iab.fdz@iab.de](mailto:iab.fdz@iab.de)). Als allgemeine Einführung zur Datenanalyse mit Stata empfehlen wir Kohler und Kreuter (2012a und 2012b).

#### 4 Wie das Do-file zu bedienen ist

Folgende Punkte sollten im Umgang mit dem Do-file *BP\_panelgen\_93JJ\_v1* berücksichtigt werden:

- Download des Do-files

Das Do-File kann von der Internetseite des FDZ heruntergeladen werden (siehe auch den Link im Appendix). Außerdem ist es im Gastaufenthalt verfügbar.

- Anforderung an die Daten

Die Routine benötigt alle Wellen des IAB-Betriebspanels, wie sie im Namen des Do-Files angezeigt werden, also beispielsweise die Wellen 1993 bis 2016 im Fall von *BP\_panelgen\_9316\_v1*. Aktuell gibt es keine Option zur Wahl bestimmter Wellen. Wellen, die nicht relevant für das Forschungsvorhaben sind, können nach dem Durchlauf von der Nutzerin oder dem Nutzer selbst gelöscht werden.

- Nutzung innerhalb der FDZ Umgebung

Wird das Do-File im Gastaufenthalt oder im Rahmen der Datenfernverarbeitung mittels JoSuA verwendet müssen keine Pfadangaben spezifiziert werden. Das Programm nimmt automatisch an, dass eine Ordnerstruktur entsprechend derjenigen im FDZ-Gästenetz/JoSuA vorliegt und definiert die entsprechenden Makros.

Um das Programm zur Panelaufbereitung im Gastaufenthalt oder über JoSuA auszuführen, sollte das Do-File mit

```
do "$prog\BP_panelgen_93JJ_v1.do"
```

(mit Ersetzen von „JJ“ durch das aktuellste Jahr in der genutzten Version) in das *master.do* eingebunden werden. Für die Datenfernverarbeitung müssen sowohl *master.do* als auch *BP\_panelgen\_93JJ\_v1.do* in JoSuA hochgeladen werden.

- Nutzung mit den Testdaten

Das Programm läuft auch unter Verwendung der Testdaten für das IAB-Betriebspanel. Die Nutzerinnen oder Nutzer müssen hierzu eine Ordnerstruktur entsprechend derjenigen in der FDZ Umgebung anlegen. Auf die folgenden Ordner greift das Panel-Do-File zu:

- Hauptverzeichnis – die jeweilige Projektablage (z.B. "C:\fdzXXXX")
  - prog – hier liegen alle Do-Files
  - orig – hier liegen die Ausgangsdaten/Testdaten
  - data – hier wird der erstellte Paneldatensatz abgespeichert
  - log – hier wird das zugehörige Log-File abgespeichert

Zusätzlich sind die zugehörigen Makros zu definieren. Der entsprechende Code ist bereits im Do-File unter dem Punkt „\*\* Pfade“ enthalten. In der Grundeinstellung ist dieser Code jedoch auskommentiert. Sobald er aktiviert ist muss nur noch das globale Makro main durch die Nutzerin oder den Nutzer mittels Eingabe des Pfades des Hauptverzeichnisses definiert werden. Anschließend funktioniert das Do-File in derselben Weise wie im vorherigen Punkt beschrieben.

- Wahl der Module, die im Paneldatensatz enthalten sein sollen

Die Module, die im Paneldatensatz enthalten sein sollen, können in Abschnitt 2 des Do-Files durch die Nutzerin oder den Nutzer festgelegt werden. Dies erfolgt indem die zugehörigen lokalen Makros von „0“ auf „1“ gesetzt werden. Aus Gründen der Datensparsamkeit und Speicherplatzbeschränkung ist die Zahl der aktivierten Module so gering wie möglich zu halten. In der Grundeinstellung werden hierbei nur Variablen berücksichtigt, die in mindestens zwei Wellen erhoben wurden.

- Hinzufügen von „Zusatzvariablen“ zu den gewählten Modulen

Um Variablen die nur in einer einzigen Welle erhoben wurden an die gewählten Module anzuspielden, muss das lokale Makro ezinf von „0“ auf „1“ gesetzt werden.

- Hinzufügen sensibler Merkmale

Sensible Merkmale finden bei der Datenaufbereitung keine Berücksichtigung. Die Zuspiegelung muss von der Nutzerin oder dem Nutzer selbst vorgenommen werden, indem eindeutig nach den Variablen `idnum` und `year` „gemerged“ wird. Hierbei ist zu beachten, dass die Variable `year` nicht in den jeweiligen Ausgangsdateien enthalten ist. Sie muss vor dem Zuspiegeln erst definiert werden.

- Erstellte Dateien

Das Do-File speichert ein Log-File zur Dokumentation der gewählten Module im Ordner „log“ und ein `.dta`-File des Paneldatensatzes im Ordner „data“ ab. Standardmäßig werden sowohl das Log-/ also auch das Datenfile genauso benannt wie das generierende Do-File.

## 5 Anmerkungen zu aufbereiteten Panelvariablen

Die letzte Spalte der Variablen-Zeitpunkte-Matrix beinhaltet einige Anmerkungen zu spezifischen Panelvariablen, die sich entweder auf die Datenaufbereitung oder die Zeitkonsistenz der zugrunde liegenden Querschnittinformation beziehen. Zu beachten ist, dass Variablen in Zeilen ohne jeglichen Kommentar nicht zwingend problemfrei sind. Beispiele für solche Anmerkungen umfassen:

- *Integriert* – Fehlende Werte wurden in der Panelvariable aufgefüllt, wenn diese in separaten Querschnittvariablen abgelegt wurden.
- *Informationen in einigen Jahren vergrößert* – Einige Ausprägungen wurden in mindestens einem Jahr bei der Erstellung der Panelvariable gruppiert. Dies impliziert, dass die Information im Querschnitt in mindestens einer Welle detaillierter ist als im Panel.
- *Zeilen-übergreifend generiert* – Variablen aus unterschiedlichen Zeilen der Variablen-Zeitpunkte-Matrix wurden bei der Erstellung der Panelvariable kombiniert.
- *Klassifikation abhängig vom Zeitraum* – Mehrere Panelvariablen wurden erstellt, da Abweichungen in der Klassifikation über die Zeit zu schwerwiegend für eine Harmonisierung sind.
- *bis einschl. 2001 durch 1.95583 geteilt* – Werte in DM wurden in EUR umgerechnet.
- *Frage weicht ab: JJJJ* – Die Frage im gegebenen Jahr weicht von verwandten Fragen aus anderen Jahren ab.
- *Frage ändert sich: ab JJJJ* – Der Wortlaut oder Inhalt verwandter Fragen ändert sich ab dem gegebenen Jahr.
- *Zeilen markiert mit einem Sternchen* – Deskriptionen der Panelvariable haben deutliche Abweichungen in mindestens einem Jahr gezeigt; dies bedeutet jedoch nicht zwingend dass eine Inkonsistenz vorliegt.

Letztlich zeigt die letzte Stelle des Namens einer jeden Panelvariable an, ob die Variable

- \*\_b – binär; üblicherweise „ja/nein“,
- \*\_d – diskret oder
- \*\_c – stetig ist.

## Literaturverzeichnis

Alda, Holger; Dundler, Agnes; Müller, Dana; Spengler, Anja (2006): *Aufbereitung eines Panel-datensatzes aus den Querschnittsdaten des IAB-Betriebspanels*. FDZ-Datenreport, 02/2006

Fischer, Gabriele; Janik, Florian; Müller, Dana; Schmucker, Alexandra (2008): *Das IAB-Betriebspanel - von der Stichprobe über die Erhebung bis zur Hochrechnung*. FDZ-Methodenreport, 01/2008

Heining, Jörg; Klosterhuber, Wolfram; Lehnert, Patrick; Seth, Stefan (2016): *Linked-Employer-Employee-Daten des IAB: LIAB-Längsschnittmodell 1993 – 2014 (LIAB LM 9314)*. FDZ-Datenreport, 10/2016

Klosterhuber, Wolfram; Lehnert, Patrick; Seth, Stefan (2016): *Linked-Employer-Employee-Daten des IAB: LIAB-Querschnittmodell 2 1993 – 2014 (LIAB QM2 9314)*. FDZ-Datenreport, 05/2016

Kohler, Ulrich; Kreuter, Frauke (2012a): *Datenanalyse mit Stata \* allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung*. 4. Auflage. München: Oldenbourg Wissenschaftsverlag

Kohler, Ulrich; Kreuter, Frauke (2012b): *Data Analysis Using Stata*. Third Edition. Stata Press

## Appendix

Download des Stata Do-files:

- [http://doku.iab.de/fdz/iabb/BP\\_panelgen.zip](http://doku.iab.de/fdz/iabb/BP_panelgen.zip)

## Impressum

FDZ-Methodenreport 12/2017

### Herausgeber

Forschungsdatenzentrum (FDZ)  
der Bundesagentur für Arbeit  
im Institut für Arbeitsmarkt- und Berufsforschung  
Regensburger Str. 104  
90478 Nürnberg

### Redaktion

Dana Müller, Dagmar Theune

### Technische Herstellung

Dagmar Theune

### Rechte

Nachdruck - auch auszugsweise - nur mit  
Genehmigung des FDZ gestattet

### Bezugsmöglichkeit

[http://doku.iab.de/fdz/reporte/2017/MR\\_12-17.pdf](http://doku.iab.de/fdz/reporte/2017/MR_12-17.pdf)

### Internet

<http://fdz.iab.de/>

### Rückfragen zum Inhalt an:

Matthias Umkehrer  
Institut für Arbeitsmarkt- und  
Berufsforschung (IAB)  
Forschungsdatenzentrum  
Regensburger Str. 104  
90478 Nürnberg  
Telefon: 0911 / 179-6211  
E-Mail: [Matthias.Umkehrer@iab.de](mailto:Matthias.Umkehrer@iab.de)