

Kurzanleitung für die kontrollierte Datenfernverarbeitung (KDFV) im FDSZ

1 Eignungskriterien für die Nutzung der kontrollierten Datenfernverarbeitung (KDFV)

Die KDFV ist für Forscher geöffnet, die mit einem aktuellen oder vorherigen Forschungsvorhaben eine ausreichende Erfahrung mit den jeweiligen Mikrodaten unter Beweis gestellt haben.

2 Verfügbare Datensätze für die KDFV

Eine aktuelle Liste mit den für die KDFV verfügbaren Datensätzen kann auf unserer Homepage (www.bundesbank.de/fdsz) eingesehen werden. Momentan steht lediglich die Mikrodatenbank Direktinvestitionen für die KDFV zur Verfügung.

3 Analysesoftware

Zurzeit ist STATA (aktuelle Version: 14) für die Analyse im Rahmen der KDFV verfügbar.

4 Datenstrukturfile und Mustercode

Auf Anfrage sendet das FDSZ Forschern ein Datenstrukturfile zu, welches dem Originaldatensatz strukturell gleicht, jedoch keine realen Werte enthält. Zusammen mit dem vom FDSZ bereitgestellten Musterprogrammcode kann das Datenstrukturfile dazu genutzt werden, einen eigenen STATA Programmcode zu schreiben

Forscher werden gebeten sich an der Struktur des Mustercodes zu orientieren, da dieser den gängigen Programmierstandards im FDSZ entspricht. Der Mustercode ist in vier Teile unterteilt:

1. Ein Masterfile, in dem Pfade und Globals definiert werden (0_master.do). In diesem File werden ebenfalls alle anderen Programmcodes aufgerufen.
2. Ein beispielhafter Programmcode für die Datenaufbereitung (1_data_preparation.do).
3. Ein beispielhafter Programmcode für deskriptive Analysen (2_descriptive_analysis.do).
4. Ein beispielhafter Programmcode für Regressionen (3_regressions.do).

Die Nutzung des Masterfiles ist verpflichtend. Die beispielhaft zur Verfügung gestellten Programmcodes können verändert werden.

5 Analyseprozess im Rahmen der KDFV

Nachdem der Forscher den Programmcode geschrieben hat, ist er verpflichtet, seinen Code am Datenstrukturfile zu testen. Erst wenn der Programmcode erfolgreich ausgeführt werden kann, kann er per Mail (fdsz-data@bundesbank.de) an das FDSZ übermittelt werden. Hierbei ist die Projektnummer in der Betreffzeile der Mail anzugeben.

Im FDSZ wird der Programmcode archiviert, an das technische Umfeld angepasst und ausgeführt. Wenn der Programmcode nicht erfolgreich ausgeführt werden kann, wird der Fehlercode in eine Mail kopiert und ohne weiteren Kommentar an den Forscher geschickt. Dieser Prozess wird so lange wiederholt, bis eine erfolgreiche Ausführung erreicht wird.

6 Outputkontrolle

Nachdem der Programmcode erfolgreich ausgeführt wurde und Analyseergebnisse erzeugt hat, werden diese Ergebnisse durch die FDSZ Mitarbeiter im Hinblick auf die statistische Geheimhaltung geprüft. Wenn die erzeugten Ergebnisse nicht die im Vertrag definierten Bedingungen erfüllt, wird der Programmcode an den Forscher zurückgegeben und er / sie wird gebeten, seinen Programmcode anzupassen.

Die Regeln für die Outputkontrolle können im Vertrag eingesehen werden. Der Forscher ist verpflichtet, den Aufwand, der bei den FDSZ Mitarbeitern im Rahmen der Outputkontrolle anfällt, zu minimieren. Wenn der Output ein vertretbares Volumen überschreitet, wird der Programmcode an den Forscher zurückgegeben und er / sie wird gebeten, seinen Programmcode anzupassen.

Erst nachdem die generierten Ergebnisse die im Vertrag definierten Bedingungen und Kriterien erfüllt, werden sie dem Forscher für die weitere Verwendung zur Verfügung gestellt.

7 Publikationskontrolle

Werden die generierten Ergebnisse verwendet, um einen Artikel, Buchbeitrag oder Aufsatz in einer wissenschaftlichen Zeitschrift zu veröffentlichen, muss diese Veröffentlichung von den FDSZ Mitarbeitern überprüft werden. Die Dauer der Prüfung der beabsichtigten Publikation wird höchstens 5 Werktage dauern. Jede Änderung, die das FDSZ vom Forscher verlangt, erneuert die Dauer der Prüfung.

8 Zitationsvorgaben

Jedes Diagramm und/oder Tabelle müssen nach den folgenden Zitationsvorgaben zitiert werden:

Quelle: Research Data and Service Centre (RDSC) of the Deutsche Bundesbank, <name of the micro dataset used from the standard RDSC list (where appropriate with the common abbreviation)>, <period during which the microdata were collected>, own calculations.

Beispiel: Research Data and Service Centre (RDSC) of the Deutsche Bundesbank, Microdatabase Direct investment (MiDi), 1999-2012, own calculations.

Im Nachgang der Publikation muss dem FDSZ eine elektronische Kopie (und die korrekte Zitiervorgabe der Publikation) kostenfrei zur Verfügung gestellt werden. Diese kann per Mail an das FDSZ gesendet werden.