

SOEP Survey Papers

Series G - General Issues and Teaching Materials

SOEP – The German Socio-Economic Panel Study at DIW Berlin

2014

Job submission instructions for the SOEPremote System at DIW Berlin – Update 2014

Jan Goebel

Running since 1984, the German Socio-Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentations (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at
<http://www.diw.de/soepsurveypapers>

Editors:

Prof. Dr. Gert G. Wagner, DIW Berlin and Technische Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Jan Goebel. 2014. Job submission instructions for the SOEPremote System at DIW Berlin – Update 2014. SOEP Survey Papers 195: Series G. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soepapers@diw.de

JAN GOEBEL

**JOB SUBMISSION INSTRUCTIONS FOR THE
SOEPRREMOTE SYSTEM AT DIW BERLIN –
UPDATE 2014**

Berlin, 2014

Please note that we assume that you have a basic knowledge of the statistical package Stata, and that we are not a helpdesk for Stata.

Contents

1	BASIC INFORMATION	2
1.1	How to access the <i>SOEPremote</i> system	2
1.2	Some remarks on job submission for the <i>SOEPremote</i> system .	2
1.3	Privacy or confidentiality rules	2
1.4	Dataset \$kreise: Additional Regional Codes at County Level . .	3
2	VERY IMPORTANT! Remarks about your e-mail client	4
3	Structure of the <i>SOEPremote</i> job submissions	4
3.1	Stata job syntax	4
3.1.1	<i>SOEPremote</i> Email Header	4
3.1.2	Loading a dataset	5
3.1.3	Other Syntax issues	5
3.2	Stata commands your are not allowed to use with <i>SOEPremote</i>	6
3.3	Use of external data with <i>SOEPremote</i>	7
4	Output from <i>SOEPremote</i>	7
5	Some recommendations and advices	9
6	Addresses for further information	9

List of Tables

1	Accessing different SOEP versions with <i>SOEPremote</i>	5
2	Disallowed Stata Commands within <i>SOEPremote</i>	6

1 BASIC INFORMATION

1.1 How to access the *SOEP*remote system

The databases of *SOEP*remote can be accessed only by preparing a STATA job request in the specific *SOEP*remote format. This job request follows the normal STATA batch conventions (STATA 8.0, for Linux) with a few exceptions and modifications that are needed for the *SOEP*remote system to process your request. These exceptions, conventions, and other details concerning job submission are described in this document. You must submit your job request by sending an e-mail to: **soepremote [at] diw.de** (without spaces). *Please note that the database cannot be downloaded. Direct access to the microdata is not possible, with the exception for users visiting DIW in Berlin (for further details please contact Jan Goebel (jgoebel@diw.de)).* In order for the *SOEP*remote processing system to accept a job request, you must register with the DIW to obtain a user identification code and a password and sign a pledge stating that you will abide by the projects rules for accessing data and publishing results. You can send your request for registration or any questions to the *SOEP*Hotline (soepmail@diw.de)

1.2 Some remarks on job submission for the *SOEP*remote system

The *SOEP*remote system resides on a Linux/Unix Network at the DIW Berlin. The microdata are housed on a secured data server on the network in an area that is secure and can not be accessed from outside of the DIW. Jobs are received at an internet node located at the DIW and are forwarded automatically to the *SOEP*remote operating system. *SOEP*remote will automatically process your jobs and return the standard STATA listings to the e-mail address you gave us when you became a user (you can always ask us to temporarily or permanently change it). Processing time varies and depends on the number of users that are using the system at the time you submit your jobs and, of course, the complexity of your jobs. As a rule of thumb, you may expect your results to be back within one hour. In case the number of jobs in our system is low, you can receive your listings back within minutes. Maintenance on our system will cause longer delays. Remember that the *SOEP*-Team does not guarantee a fixed maximum delay, nor do we accept responsibility in case your listings have not been returned within the normal expected timeperiod.

1.3 Privacy or confidentiality rules

Printing or listing out individual data is prohibited

Do not try to print out (parts of) datasets in any way. Please abstain from commands like LIST or TAB on continuous variables such as PERSNR, the weights, or income variables. Such commands will be detected by the *SOEP*remote system, and those listings are not returned to the user,

but are directed to our review. The review is a security provision where all long (and thus suspect) listings are sent.

Long listings

If you combine many commands in the same job the length of the resulting listing will grow accordingly. Even though all commands themselves may be strictly allowed, the fact that the listing exceeds a certain length will cause it to be directed to our review. As these listings have to be examined manually, this not only means an additional burden for our team, it also causes longer delays for you to receive your results. Therefore we advise you also in your own interest, to split such programs over several separate jobs and to use some of the options for specific statistical packages suggested in this document (at this time only for stata).

SOEPrmote is a true workaholic and works seven days a week and 24 hours a day. Unfortunately computers are not perfect and it can happen that **SOEPrmote** takes some unscheduled time off. We try to detect and solve these problems as quickly as possible. However, when they occur at night (CET time), during the weekend or during other holidays, we won't be able to fix it until the next working day.

1.4 Dataset \$kreise: Additional Regional Codes at County Level

Additional codes indicating the county for each household and year is available in a dataset. This dataset is available in two “flavors”:

1. In wide format \$kreise
2. In long format \$kreise_l

As **county identifier** we use the first five digits (as numeric and hence without leading zero) of the official Community Identification Number¹. We provide two versions of this information, the first one is the “original” one (variable `bkkz` - `bdkkz` or just `kkz` in the long version), coded by the field institute at the end of the survey year (using mostly the territorial definition from december of one year before). Because this territorial definition is changing over time² we provide a recoded version of this regional codes describing a territorial definition at one point in time (variables with suffix “_rek”). The territorial definition is the most recent used by the [BBSR](#) for the [INKAR CD](#) at the time the specific **SOEP** version was first available to our users.

You can get an updated description of the variables in these datasets by using SOEPrmote, sending an ‘describe’ or ‘codebook’ statement to the server, i.e. ‘describe using \$soep29/kreise_l’.

¹An online description in english can be found at [wikipedia](#), the official keys can be found at [Destatis](#), the Federal Statistical Office.

²For example in 1990 Germany was divided in 533 counties, nowadays Germany has only 403 counties. An excellent overview can be found at [wikipedia](#), however only in German language.

2 VERY IMPORTANT! Remarks about your e-mail client

1. Your jobs should be written inside the body of an e-mail message and not as an attachment to such a message.
2. It is important that you disable some of the more advanced settings of your E-mail editor:
 - DO NOT use html formatted E-mails!
 - DO NOT use colour, bold cases, etc.
 - DO NOT use electronic business cards, auto-confirmation (which informs you whether the person that you sent an E-mail message to has opened your message), etc.
 - Our system needs to receive your job as **plain ASCII e-mail message**.
 - DO NOT use special characters like accents (' ^), nor country-specific characters like German umlaut (ü ß), Scandinavian vowels (æ Å), Spanish letters (ç ñ) etc. Our system will translate non-ASCII settings into ASCII characters (such as '3D', 'E20', etc), which will make your job unreadable. In all cases where you fail to follow these rules, the **SOEPremote** system will send you the following error message: WRONG HEADER (see below in section 4).

3 Structure of the *SOEPremote* job submissions

3.1 Stata job syntax

3.1.1 SOEPremote Email Header

The first 4 lines of your job will allow you to be recognised as a *SOEP* registered user and will thus give you access to the *SOEP* database. They are essential for ALL jobs and for Stata, they should be as follows:

```
** user = your username
** password = your password
** package = STATA
** project = GSOEP
```

After these first lines you can write your normal stata program lines, bearing in mind the following points, which are specific to *SOEPremote*:

3.1.2 Loading a dataset

To load a datasets into Stata you have to make use of aliases defined in global macros. To access the directory where all datasets from one data distribution are stored you have to use the pre defined globals for the file path.

If you want to use the dataset `ppfad.dta` from version 29 of the SOEP you have to write `$soep29/ppfad` (please note the forward slash “/”).

For the defined path are shown in table 1.

Table 1: Accessing different SOEP versions with *SOEPremote*

<i>SOEP</i> version	<i>SOEPremote</i>
SOEPlong (version 29)	<code>\$soep291/</code>
SOEP-core (version 29), incl. kreise dataset wide and long	<code>\$soep29/</code>
SOEP-core (version 28), incl. kreise dataset wide and long	<code>\$soep28/</code>
SOEPlong (version 28)	<code>\$soep281/</code>
SOEP-core (version 27), incl. kreise dataset wide and long	<code>\$soep27/</code>

Be careful with typing mismatches: wrong abbreviations will result in a “file not found” message in your listing. Remember: all the datasets have to be referred to in lowercase (Stata is case sensitive).

3.1.3 Other Syntax issues

Delimiter: The delimiter is not mandatory in *SOEPremote*/stata, but we strongly recommend to introduce one in order to avoid problems of multiple or long lines.

Intermediate files: You may need to save intermediate datasets for later use (not necessarily in the same job), for example to append or merge two datasets. When doing so, please always save your intermediate files with the following command:

```
save $mydata/myfile, replace ;
```

The macro `$mydata` indicates the path where your temporary intermediate datasets will be saved. Substitute `myfile` with a unique name you want to give to your dataset.

Privacy or confidentiality rules: Do not use the commands screened out as data protection problems by *SOEPremote*, such as `list` (see above Privacy or confidentiality rules).

Automatic set-up: There are things *SOEPremote* does automatically when processing a stata job request. For this reason you should never use the command: `set mem Xm` (as it is automatically set to the maximum) and never `exit stata` after you finish your commands.

Stata "ado" files: We have installed the official (stb) Stata ado files that are commonly used by researchers. If you have an ado file that you would like us to add, just send an email to [Jan Goebel \(jgoebel@diw.de\)](mailto:jgoebel@diw.de) and we will do so if possible.

Note on long listings in STATA: To avoid too long output using, for instance, logit or probit functions, the option `", nolog"` can be put at the end of the command. And you can also make use of the commands `"quietly"` or `"qby"`. For further information please have a look at the corresponding help files of these commands.

3.2 Stata commands your are not allowed to use with *SOEP*remote

Because of privacy and confidentiality rules some Stata commands are not allowed within a *SOEP*remote job, for a list of these commands see table 2.

Table 2: Disallowed Stata Commands within *SOEP*remote

Stata Command	Description
!	Temporarily invoke operating system
_getfilename	Command for handling path-filename specifications
cd	Change directory
copy	Copy file from disk or URL
cscript	Begin certification script
dir	Display filenames
do	Execute commands from a file (other than your email)
drop	Eliminate variables or observations (see below)
file	Read and write ASCII text and binary files
infile	Reading ASCII data into Stata
keep	Keep variables or observations (see below)
list	List values of variables
log	Echo copy of session to file or device
macro	Macro definition and manipulation
mkdir	Create directory
net	Using the Internet
print	Printing output
set_defaults	Reset system parameters
shell	Temporarily invoke operating system
ssc	Install and uninstall packages from SSC
sysuse	Use Stata's shipped datasets
webuse	Use dataset from web (!!!)
type	Display contents of files

The *SOEP*remote equivalent for drop and keep

Instead of using Stata commands to select observations and variables (`drop` and `keep`) you are bound to use the equivalent *SOEP*remote ado files `soepdrop` and `soepkeep`. The commands have exactly the same syntax as the original commands and are working exactly the same way. The only difference is, that your selection statement will be checked whether at least three cases are left in memory.

Description of the command syntax from the Stata online help:

```
-----  
help for drop, keep, clear                manual:  [R] drop  
                                           dialog: drop/keep  
-----
```

Eliminate variables or observations

```
[soep]drop varlist  
[soep]drop if exp  
[soep]drop in range [if exp]
```

```
[soep]keep varlist  
[soep]keep if exp  
[soep]keep in range [if exp]
```

3.3 Use of external data with *SOEP*remote

An analysis of data at county level often necessitates additional use of external information and merging this data with the *SOEP*. However, we **do not** expect any datasets as attachment in your mail. The only way to use external data is to setup your external dataset within your do file. You can do this with the command `input`, as you can see in the following example:

```
input id var1 var2  
    4.58 55  
    1.86 60  
    1.61 .  
end
```

4 Output from *SOEP*remote

*SOEP*remote acts very polite, whatever message you send to *SOEP*remote, you will always receive an answer. The e-mail you receive will contain either of the following:

1. The listing with the output of the job execution if you followed all the **SOEPrermote** rules above. **SOEPrermote** will automatically forward you the listing containing the output of the job execution. Please note that this will be within the text of the e-mail message, and NOT as attachment as a log file. This will be returned to the address that you provide when you register with the **SOEP** staff (*and NOT to the address from which you sent the e-mail!*).
2. A warning because of long listing. There are two scenarios dealing with long listings:
 - (a) **Review queue:** in case the job is fully executed but the size of its listing is larger than a given limit, the job is automatically put in the review queue and a message is automatically sent to you saying that your job has to be manually reviewed by the staff. In this case, you can expect that within two working days you will either receive the listing with the output of the job execution, if no confidentiality rules were broken, or receive nothing in case the listing contained confidential data. Both the warning and the listing (if applicable) will be returned to the address that you provide when you register with the **SOEP** staff (and NOT to the address from which you sent the email!).
 - (b) **Trash-can:** in case the job is exceptionally long, it is automatically put to the trash-can and **SOEPrermote** will immediately inform you accordingly (you will receive an e-mail to the address you provide when you register with the **SOEP** staff).
3. An error message, which can be of three different types:
 - (a) **Wrong package/project:** in case you specify a non-existing package or project in the first 4 lines of your job submission (for the Syntax of these first lines, see section 3.1), the job cannot be executed, and you will receive a message telling you clearly what was wrong. The error message will be returned to the address that you provide when you register with the **SOEP** staff (and NOT to the address from which you sent the e-mail!).
 - (b) **Wrong userid/password:** in case you do not specify a valid user-id and password, **SOEPrermote** cannot recognise you as a registered **SOEPrermote** user and will thus not execute the job. Please note that in this case the error message will be sent back to the e-mail address from where you sent the job (and NOT the one you provided when you registered!).
 - (c) **Wrong header:** in case **SOEPrermote** cannot find the first 4 lines (the header), and can thus not recognise you as a registered **SOEPrermote** user, it will automatically and immediately inform you accordingly, with a message to the e-mail address from where

you sent the job (and NOT the one that you provided when you registered!). Please note that this can happen both if you omit (or mistype/misplace) the first 4 lines, or if there was a non-ascii or rich text character in the text of your e-mail message (see the very Important remark about your E-mail client above, section 2).

5 Some recommendations and advices

1. It is advisable that you give a subject to your e-mail so that you can easily distinguish the listings from one another in your out and inbox.
2. It is advisable that you keep the code of all the (important) jobs you send to **SOEPremote**, as there is no way of obtaining them again from **SOEPremote**.
3. It is strongly recommended that you carefully check the *Documentation* on-line at the **SOEP**-homepage (in particular the *Desktop Companion*).
4. We strongly recommend that all statistical operations performed with the **SOEP** database make use of the appropriate weighting variables.

6 Addresses for further information

If you still have any problems concerning the use of **SOEP** data via **SOEPremote**, please contact us at the addresses below.

Administrative questions:	<i>SOEPHotline (soepmail@diw.de)</i>
Geocode data and access:	<i>Jan Goebel (jgoebel@diw.de)</i>
SOEPremote questions:	<i>Jan Goebel (jgoebel@diw.de)</i>