

**Replication package for
“Surveillance and Self-Control”**

**by Deborah A. Cobb-Clark, Sarah C. Dahmann, Daniel A. Kamhöfer,
and Hannah Schildberg-Hörisch,
Economic Journal**

June 2023

The main data source of this paper are survey data taken from the Innovation Sample of the German Socio-Economic Panel Study (SOEP-IS). For parts of the analysis, we add data on the number of unofficial informers of the GDR’s Stasi as surveillance measure, provided by Lichter et al. (JEEA Replication Package, 2021), to the SOEP-IS survey data. We also compare our SOEP-IS extract with the Core data of the SOEP to assess whether the representativeness of the Core data holds true for our analysis. Based on these data sources, the empirical analysis is divided into four blocks (folder in the replication package/block names in parenthesis): (i) the general determinants of trait self-control in the survey data (“SOEP-IS”), (ii) the combination of SOEP-IS and surveillance data on county level using remote processing (“SOEPremote”), (iii) the comparison of our SOEP-IS extract with the SOEP-Core data (“SOEP”), and (iv) the analysis of the surveillance data, independent of the survey data (“Unofficial informers”).

Blocks (i), (iii), and (iv) can be run in Stata on the local processor. Block (ii) requires remote-access to the SOEP-IS data and needs to be run remotely by the team of the SOEP Department of the DIW Berlin. The replication package includes a master do-file (called “MASTER.do”) that calls all do-files that can be run locally and produces all the empirical results presented in the tables and figures. The master do-file also sets the paths to the various datasets and to the folder that contains the empirical results. Users need to adjust these paths before the master file runs through. Apart from folders for the four analysis blocks, containing do-files for each block, the replication package includes a data folder that includes the data we are allowed distribute (this is only data on the number of unofficial informers) and a results folder.

The following table maps the empirical results shown in the paper into the four analysis blocks. All results are stored in the folder “Stored results” of the replication package. The structure follows the master do-file:

Analysis step	Do-file	Input file(s)	Output file
(i) SOEP-IS			
Data preparation	data_prep.do	pgen.dta hgen.dta bio.dta bioparen.dta inno.dta ppfad.dta biobirth.dta	det_data.dta
Table 1	table_1.do	det_data.dta	table_1.dta
Table 2	table_2.do	det_data.dta	table_2.dta
Figure 1	figure_1.do	det_data.dta	figure_1.pdf
Figure 2	figure_2.do	det_data.dta	figure_2.pdf
Table A3	table_A3.do*	det_data.dta	table_A3.txt
Figure A1	figure_A1.do	det_data.dta	figure_A1.pdf
Information in text	text_SOEP_IS.do	det_data.dta	<i>Stata window</i>
(ii) SOEPremote			
Prepare Stasi data for remote server	prep_Stasi_data.do	Stasi_Data.csv	informers.dta
Run analysis on remote server (send do-file via email to soepremote@diw.de)	SOEPremote_analysis.do (inserts "informers.dta" using Stata's <code>input</code> command)	inno.dta hgen.dta ppfad.dta regionl.dta	<i>output email</i>
Figure 3	figure_3.do	<i>output email</i>	figure_3.pdf
Information in text	Step 10 in SOEPremote_analysis.do		<i>output email</i>
(iii) SOEP			
Data preparation	prep_SOEP_core.do	ppfad.dta bgpgan.dta bghgen.dta biobirth.dta bioparen.dta phrf.dta	det_core.dta
Table A2	table_A2.do	det_data.dta det_core.dta	table_A2.dta
(iv) Unofficial informers			
Information in text	text_unofficial_informers.do	Stasi_Data.csv	<i>Stata window</i>

*Requires the user-written program `mkcorr`, version 2.0.1.

Table A1 (variable definitions) does not present empirical results. Figure A2 (map showing the density of unofficial informer on county level) is taken from Lichter et al. (2021b).

In the following we elaborate on each step of the analysis:

Block 1: SOEP-IS

While the Innovation Sample is part of the SOEP data infrastructure (see SOEP-IS, 2018; Richter and Schupp, 2015), it does not contain the same respondents as the more well-known SOEP (the SOEP-Core data, see Goebel et al., 2019). The application for data access is the same for the SOEP-Core data and SOEP-IS data.

The SOEP data infrastructure is subject to data protection legislation and we are prohibited to distribute the data ourselves. Interested researchers can obtain data access free of charge by signing a data distribution contract with the German Institute for Economic Research (DIW Berlin):

Contact information: The German Socio-Economic Panel
DIW Berlin
Mail address: 10108 Berlin, Germany
Visitors address: Mohrenstraße 58, 10117 Berlin, Germany
Tel: +49 30 89789 292
Email: soepmail@diw.de

Further information on how to apply for data access may be found here:

https://www.diw.de/en/diw_01.c.601584.en/data_access.html

After signing the data distribution contract, users can choose which survey and wave of the SOEP infrastructure they would like to get access to. All do-files in this block run on scientific use files (SUF) of the 2018 Innovation Sample (doi:10.5684/soep.is.2018) that can be locally stored on the user's device. The do-file "MASTER.do" executes the do-file "data_prep.do" that creates the final estimation sample based on the SOEP-IS data files and runs all do-files for the tables and figures that only require SOEP-IS data, see the table above for a list.

Block 2: SOEPremote

Figure 3 relates self-control (assessed in the SOEP-IS survey) with the number of unofficial informers of the GDR's Stasi on county level. This block consists of three do-files. The first do-file, "prep_Stati_data.do", can be run locally and prepares the Stasi information. The second do-file, "SOEPremote_analysis.do", runs on the SOEPremote server, see below. The third do-file, "figure_3.do", can be run locally and produces Figure 3 based on data obtained by the second do-files.

The Stasi information

The number of unofficial informers is taken from the Replication Package (Lichter et al., 2021b, doi:10.7910/DVN/WMVC4U) of the 2021 Journal of the European Economic Association article of Lichter et al. (2021a) that can be accessed freely through the Harvard Dataverse repository:

<https://doi.org/10.7910/DVN/WMVC4U>

The data are originally collected by the Stasi Records Agency (Stasi-Unterlagen-Archiv, part of the German Federal Archives), a government agency that restored Stasi files that were shredded after the collapse of the GDR, see

<https://www.stasi-unterlagen-archiv.de/en/>

Lichter et al. (2021a, b) use the data version compiled by Müller-Enbergs (2008). The file “County_Data.csv” in the Replication Package (Lichter et al., 2021b) includes the number of unofficial informers relative to the population in the GDR counties. To match this information with the SOEP-IS respondents, we searched the current county identifiers referring to the GDR counties by hand (the counties’ German Wikipedia entries usually include a history of the county, including county reforms). The dataset “Stasi_Data.csv” in this replication package of our study includes the information of Lichter et al. (2021b) as well as the current county identifier. That is, the file “Stasi_Data.csv” is a version of the file “County_Data.csv” provided by Lichter et al (2021b) that is augmented for our purposes. The file “Stasi_Data.csv” is in the data folder of our replication package.

The do-file “prep_Stasi_data.do”

The do-file “prep_Stasi_data.do” collapses the Stasi information in the file “Stasi_Data.csv” on the county level. This is the level needed in order to merge the information to SOEP-IS respondents. The number of unofficial informers to the Stasi per county is stored in the dataset “informers.dta” in the results folder. Note, this file is only for record-keeping.

Access to SOEPrmote

Because of the confidentiality of the information, the SUF of the SOEP-IS used in Block 1 does not include respondent’s county of residence. Analyzing the surveillance density, therefore, requires accessing the SOEP-IS data via an access mode called SOEPrmote. SOEPrmote allows researchers to analyze sensitive data via email. Researchers can send their code in an email to soeprremote@diw.de and members of the SOEP team at the DIW run the code on a server at the DIW. An email that contains the Stata output is sent back to the researcher’s email address. This output email is basically a log-file.

For a manual on how to use SOEPrmote, see Goebel (2014):

https://www.diw.de/documents/publikationen/73/diw_01.c.570638.de/diw_ssp0195.pdf

Accessing SOEPrmote requires regular access to the SOEP infrastructure (see Block 1). Researchers who have access to SOEP data can get remote access by sending an informal request to soepmail@diw.de. Researchers should specifically mention that they require access to the Innovation Sample.

The do-file “SOEPrmote_analysis.do”

It is not necessary to merge the number of unofficial informers (prepared by the do-file “prep_Stasi_data.do” and stored in “informers.dta”) directly to the confidential SOEP-IS infor-

mation. The do-file “SOEPremote_analysis.do” that is sent via to the SOEPremote server creates a new dataset with the information on the server using Stata’s `input` command in step 1 of the do-file. The inserted data points are the data stored in “infomers.dta”.

Steps 2 to 8 in the do-file “SOEPremote_analysis.do” prepare the SOEP-IS and surveillance data for analysis.

Step 9 of the do-file “SOEPremote_analysis.do” reports the average self-control of the SOEP-IS respondents per bin of unofficial informers (both adjusted for covariates). This is the information required to produce Figure 3 later on.

Step 10 of the do-file “SOEPremote_analysis.do” conducts additional analyses we refer to in the text of the paper.

The email containing the Stata output we have received from the DIW after running the do-file “SOEPremote_analysis.do” on the SOEPremote server is contained in the “Stored Results” folder. This email is basically a log-file of the analysis run on the SOEPremote server.

The do-file “figure_3.do”

In the do-file “figure_3.do”, that runs locally, the data points reported by step 9 in the do-file “SOEPremote_analysis.do” on the SOEPremote server are used to create a new dataset using Stata’s `input` command. This dataset is then used to create Figure 3.

Block 3: SOEP

Block 3 runs the balancing analysis, comparing our SOEP-IS extract with the SOEP-Core data (SOEP, 2016; Goebel et al., 2019) in Table A2. Researchers wishing to run this analysis should order the SOEP-Core data in addition to the SOEP-IS data when applying for access as described in Block 1. The do-file in this block runs on SUF of SOEP-Core, wave 33 (doi:10.5684/soep.v33.1) that can be locally stored on the user’s device. The do-file “prep_SOEP_core.do” prepares the SOEP-Core data. The do-file “table_A2.do” merges the SOEP-Core data to our estimation sample (“det_data.dta”, generated in Block 1) and conducts the balancing analysis.

Block 4: Unofficial informers

This block is only required to calculate the correlation between the number of unofficial Stasi informers per county with the number of arrests, reported in section 3.2 of the paper. This block only consists of the do-file “information_in_text_unofficial_informers.do” and it only requires the data “Stasi_Data.csv”.

General information

The locally-run analyses (all do-files except for “SOEPremote_analysis.do”) is conducted using Stata/SE, version 14.2 on a Microsoft Windows 10 Intel Core i7-7500U CPU, 64-Bit OS processor. The estimated running time of the locally-run analyses is less than 3 minutes. For Table A3, Stata users require the user-written program `mkcorr`, version 2.0.1.

Data references

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics* 239 (2), 345-360.

Goebel, J. (2014). Job submission instructions for the SOEPremote System at DIW Berlin – Update 2014. SOEP Survey Papers 195: Series G. German Institute for Economic Research (DIW Berlin).

Lichter, A., M. Löffler, and S. Sieglöcher (2021a). The Long-Term Costs of Government Surveillance: Insights from Stasi Spying in East Germany. *Journal of the European Economic Association* 19(2), 741-789.

Müller-Enbergs, H. (2008). *Inoffizielle Mitarbeiter des Ministeriums für Staatssicherheit*. Teil 3: Statistik. Berlin: Ch. Links Verlag.

Richter, D. and J. Schupp (2015). The SOEP Innovation Sample (SOEP IS). *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften* 135 (3), 389-400.

Data DOIs

Lichter, A., M. Löffler, and S. Sieglöcher (2021b). Replication Files for: The Long-Term Costs of Government Surveillance: Insights from Stasi Spying in East Germany. Harvard Dataverse: doi:10.7910/DVN/WMVC4U.

SOEP (2016). Socio-Economic Panel Study, wave 33. German Institute for Economic Research (DIW Berlin). doi:10.5684/soep.v33.1.

SOEP-IS (2018). Socio-Economic Panel Study Innovation Sample, wave 2018. German Institute for Economic Research (DIW Berlin). doi:10.5684/soep.is.2018.