

DMP title

Project Name To work or not to work: constructing structural labour market models for meaningful policy simulation. - DMP title

Project Identifier 11I6522N

Grant Title 11I6522N

Principal Investigator / Researcher Thibault Deneus

Project Data Contact thibault.deneus@kuleuven.be

Description This project aims to construct structural models that allows for labour supply policy simulation. These models will be estimated on socio-economic survey data. The survey data that will be used is requested by STATBEL and Deutsches Institut für Wirtschaftsforschung (DIW). The data is already anonymised.

Institution KU Leuven

1. General Information

Name applicant

Thibault Deneus

FWO Project Number & Title

11I6522N

To work or not to work: constructing structural labour market models for meaningful policy simulation.

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

I will use two panel datasets.

1) BE-SILC

The BE-SILC dataset consists of socio-economic panel survey data for Belgium. The dataformat is .sas7bdat and it has a volume of 279 MB. It is created by Belstat.

2) GSOEP

The GSOEP dataset consists of socio-economic panel survey data for Germany. The dataformat is .dta and it has a volume of 14.3 GB. It is created by DIW.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

Privacy Registry Reference: G-2022-4923

Short description of the kind of personal data that will be used: Socio-economic panel data for Belgium and Germany.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- Yes

BE-SILC:

The data remains exclusive property of statbel. The Data Recipient agrees to follow the General Data Protection Regulation of 30 July 2018. Only prof. Erwin Ooghe and Thibault Deneus have access to the data. The Data Recipient has to notify Statbel for every change w.r.t. the individuals who can access the data. The Data Recipient is forbidden to share the data, also partial datasets are forbidden to share.

GSOEP:

1 DIW Berlin grants the Data Recipient the non-exclusive right of use of the Socio-Economic Panel (SOEP) microdata (referred to below as "SOEP data").

2 The following are the restrictions to this right of use:

2.1 The Data Recipient agrees not to give or make the data available to any person or institution other than those persons listed in Section 2.3 who are working on the research project stated there and have agreed to the data protection regulations. The same applies to any modified data.

2.2 The data may only be used for the Data Recipient's own research. Use of the data for commercial or other scientific purposes is strictly forbidden. Permission may be granted in particular cases upon submission of a request for a separate contract. The use of the complete dataset in teaching is also forbidden. Only the version of the SOEP data provided for use in teaching, which contains 50% of the number of cases in the full dataset, may be used for this purpose. The teaching version of the dataset is provided to Data Recipients on specific request.

2.3 Use of the SOEP data is allowed only in the following research projects: To work or not to work: constructing structural labour market models for meaningful policy simulations. Other use of the data than for this project is not permitted. The data are to be used only by the persons who are employed at the Data Recipient's institution and who are responsible for working on the aforementioned research project. These are:

Prof Erwin Ooghe

Thibault Deneus

The Data Recipient is to ensure that the aforementioned persons who are permitted to use the data adhere to the regulations to which they are subject under this contract, and that they treat the SOEP data in accordance with the legal provisions on the protection of personal data.

2.4 De-anonymization measures (identifying individuals in the dataset) are not permitted.

Publication of individual datasets is prohibited. Merging of the data with other non-anonymized SOEP data is not permitted. Linking of the data with other individual or household data (for instance, using the procedure of statistical matching) is only permitted with written approval from the SOEP Research Data Center. DIW Berlin – Deutsches Institut für Wirtschaftsforschung e.V. Mohrenstraße 58, 10117 Berlin Postanschrift: DIW Berlin SOEP Service und Dokumentation 10108 Berlin T +49 30 897 89 -0 F +49 30 897 89 -200 www.diw.de Vorstand Prof. Marcel Fratzscher, Ph.D. (Präsident) Angelica E. Röhr (Geschäftsführerin) Prof. Dr. Stefan Liebig Kuratorium Prof. Dr. Axel A. Weber (Vorsitzender) Rechtsform Eingetragener Verein Sitz in Berlin Amtsgericht Charlottenburg 95 VR 136 NZ DIW Berlin Deutsches Institut für Wirtschaftsforschung Mohrenstraße 58 10117 Berlin Name of institution 2/2

2.5 The Data Recipient is responsible for taking the necessary technical and organizational measures needed to ensure the protection of personal data in accordance with German data protection law as well as the respective national data protection laws. DIW Berlin can provide further information if needed. In the event of any legal disputes, German data protection laws apply.

2.6 Upon completion of the research project stated in Section 2.3, the SOEP data provided to the Data Recipient as well as any back-up copies, extracts, and help files must be deleted. DIW Berlin is to be informed in writing of the completion of the project, including the date and manner of deletion.

2.7 The Data Recipient agrees to make all findings based on SOEP data and all publications of these findings available to DIW Berlin at no charge for the purpose of inclusion in the SOEPlit database.

2.8 When publishing their research findings, both the Data Recipient and all persons listed in Section 2.3 must strictly ensure that no information is released that allows for identification of individual persons. If publishers/journals refuse to publish the submitted paper unless the underlying SOEP data are made publicly available, it should be kept in mind that passing on the dataset provided by us is strictly forbidden. All users may, however, archive their datasets at the SOEP Research Data Center for reanalysis.

2.9 The Data Recipient hereby certifies that he/she has obtained permission from the project head stated in Section 2.3 for DIW Berlin to publish project head's name, research institute/department, and project title to appear in the next issue of the SOEP newsletter in the section "New data users."

2.10 The Data Recipient agrees to cite the SOEP data in all publications using of the SOEP data by stating the source as "SOEP" with the corresponding DOI.

3 DIW Berlin agrees provide the data at no charge. No fees will be charged for data transfer or documentation.

4 DIW Berlin reserves the right to a final decision in cases where disagreements arise over the right to use the SOEP data.

5 The right of each of the persons listed in Section 2.3 to use the data ends if and when he/she leaves the Data Recipient's institution or if and when that institution is dissolved, taken over by new management, or re-established as a new institution. This is also the case if the same persons continue the project at a different institution. In this case, the persons are only permitted to continue using the SOEP data after a new data distribution contract has been signed between the new institution and the SOEP. In all such cases, the data as well as any back-up copies, extracts, and help files must be deleted. DIW Berlin is to be informed of any of the changes covered in this Section. If this does not occur, DIW Berlin reserves the right to revoke SOEP data use privileges at any time. In this case as well, the Data Recipient is required to delete the SOEP data. DIW Berlin is to be informed of all deletions carried out by the Data Recipient as stipulated in Section 2.6.

6 In cases of dispute or discrepancy, the German version of this contract is binding. Any changes or amendments to this contract must be made in writing

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1) For the variables already present in the data-sets, the insitution (STATBEL and DIW) already have an exhaustive codebook.

2) The procedure that I will use to merge and clean the data will be documented. The R- and STATA-files that are used to clean the data will also be saved on KU Leuven servers.

3) The procedure that I will use to generate new variables will be documented. The R-, STATA and Matlab files that are used to generate these variables will be saved on KU Leuven servers.

2) and 3) ensure that everyone will be able to replicate the sample size and variables I use in this project given that they have acces to the initial data-sets.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

5. Data storage and backup during the FWO project

Where will the data be stored?

The source data will be stored on the one-drive of the KU Leuven. The code used to process the data will also be stored on the one-drive.

How is backup of the data provided?

Onedrive provides automatic back-ups.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

KU Leuven provides 2000GB capacity via Onedrive. This vastly exceeds the less than 20GB data this project needs.

What are the expected costs for data storage and back up during the project? How will these costs be covered?

The data and codes will not exceed the standard capacity of Onedrive KU Leuven that is offered for free to KU Leuven staff. There are therefore no additional costs.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data is already anonymised. It is also in a map only available to Erwin Ooghe (promotor) and myself.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

The source files (BE-SILC and GSOEP) have to be deleted at the end of the projects conform with the contractual agreement with STATBEL and DIW. All data are however publicly available for research institutions and can be granted through a motivated data request. Code will be kept for at least 10 years, in line with KU Leuven policies.

Where will the data be archived (= stored for the longer term)?

Code will be kept for at least 10 years, in line with KU Leuven policies. This will be stored on the university's central servers (with automatic back-up procedures).

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

The code scripts will not exceed 1GB and therefore the price will be low?

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

The source data (GSOEP and BE-SILC) are not allowed to be shared due to contractual agreement.

Which data will be made available after the end of the project?

The code-scripts of the data-processing will be available.

Where/how will the data be made available for reuse?

- Other (specify):

We are not allowed to share the source data due to contractual agreement. The code scripts for the data processing will be public available via ???

When will the data be made available?

All source data are publicly available for research institutions and can be granted through a motivated data request by DIW and STATBEL. The code scripts for the data processing will be public available.

Who will be able to access the data and under what conditions?

All source data are publicly available for research institutions and can be granted through a motivated data request by DIW and STATBEL. The code scripts for the data processing will be public available.

What are the expected costs for data sharing? How will the costs be covered?

The only data I share will be less than 1GB, the costs will be negligible.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Thibault Deneus

Who will be responsible for data storage & back up during the project?

Erwin Ooghe

Who will be responsible for ensuring data preservation and reuse ?

Erwin Ooghe

Who bears the end responsibility for updating & implementing this DMP?

Erwin Ooghe bears the end responsibility of updating & implementing this DMP.