Plan Overview

A Data Management Plan created using DMPonline.be

Title: New methods to control for unobserved heterogeneity in nonlinear panel data models

Creator: Geert Dhaene

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

ID: 212356

Last modified: 23-01-2025

New methods to control for unobserved heterogeneity in nonlinear panel data models
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
R package	Software code for the methods	Generate new data	Digital	Software	.R	<100MB	
National Longitudinal Survey of Youth (NLSY) data	Data for an application	Re-use existing data	Digital	Observational	.csv	<100MB	
Current Popula-tion Survey (CPS)	Data for an application	Re-use existing data	Digital	Observational	.csv	<100MB	
Panel Study of Income Dynamics (PSID)	Data for an application	Re-use existing data	Digital	Observational	.csv	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

National Longitudinal Survey of Youth (NLSY): https://www.bls.gov/nls/home.html
Current Population Survey (CPS): https://www.bls.gov/cps/
Panel Study of Income Dynamics (PSID): https://psidonline.isr.umich.edu/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

We will use README.txt files for the observational data. For the R package, we will make an application guide with detailed instructions available on GitHub. The package will also be uploaded on CRAN.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

The R package will be uploaded on CRAN at https://cran.r-project.org/ , allowing the latest version to be easily retrieved.

3. Data storage & back-up during the research project

Where will the data be stored?

During the project, we will use the range of storage solutions provided by KU Leuven: a KUL-managed computer, desktop file storage, KU Leuven OneDrive. These storage types are encrypted by a personal password.

How will the data be backed up?

Automatic backup by our IT department.

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

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Only the PI has access to the repositories, which are protected by strong passwords.

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

There are no costs as KU Leuven already provides us with the necessary storage.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

Data is stored on KU Leuven servers for at least 10 years.

Where will these data be archived (stored and curated for the long-term)?

On KU Leuven servers.

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

There are no costs, since we will use KU Leuven servers.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

If access is restricted, please specify who will be able to access the data and under what conditions.

Not applicable.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The R package will appear on CRAN and GitHub.

When will the data be made available?

We will make the data available as soon as possible.

Which data usage licenses are you going to provide? If none, please explain why.

We will use an MIT license from https://opensource.org/license/mit

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

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

There are no costs for data sharing. We only use free means of sharing the data.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Both PIs: Geert Dhaene and Jad Beyhum

Who will manage data storage and backup during the research project?

Both PIs: Geert Dhaene and Jad Beyhum

Who will manage data preservation and sharing?

Both PIs: Geert Dhaene and Jad Beyhum

Who will update and implement this DMP?

Both PIs: Geert Dhaene and Jad Beyhum