DMP title

Project Name DMP "Marriage markets, household consumption and individual welfare" - DMP title

Project Identifier DMP_L3_2022

Grant Title D-2022-1377

Principal Investigator / Researcher Laurens Cherchye, Bram De Rock and Frederic Vermeulen

Description The ultimate goal of the research project is to develop a comprehensive toolkit for empirical revealed preference analysis of the relationship between marital matching patterns, intrahousehold allocations of consumption and time, and individual welfare. In addition to theoretical contributions, we will demonstrate the practical usefulness of our novel methodology through various empirical applications to data describing household decisions regarding the intrahousehold allocation of time and money.

Institution KU Leuven

1. General Information

Name of the project lead (PI)

Laurens Cherchye

Bram De Rock

Frederic Vermeulen

Internal Funds Project number & title

C14/21/010 "Marriage markets, household consumption and individual welfare".

2. Data description

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

- Reuse existing data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

MEqIn	Cross-sectional microdata with detailed socioeconomic and demographic information on a representative sample of Belgian households.	.dta	Less than 1 GB	Computer assisted personal interviewing + drop off questionaire.
LISS	Panel microdata with detailed socioeconomic and demographic information on a representative sample of Dutch households.	.dta	Depending on the modules used less than 1 GB	Representative online panel.
PSID	Panel microdata with detailed socioeconomic and demographic information on a representative sample of American households.	.dta	Depending on the modules used less than 1 GB	Variety of interview modi over the years: e.g. face to face, computer assisted telephone interview.

3. Ethical and legal issues

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.

Data are anonymous social survey data.

3.2. 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.

3.3. Does your research 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.

3.4. 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 regarding reuse and sharing are in place?

The three data sources mentioned can be used for academic research without any restrictions (conditional upon registration).

4. Documentation and metadata

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

1. MEqIn. General documentation on the original data can be found on https://sites.google.com/view/meqin/data. Our empirical applications will likely make use of data drawn from the original dataset. All steps from the original data to the sample used for our studies will be carefully documented in a Stata-do-file and/or a readme file.

2. LISS. General documentation on the original data can be found on https://www.lissdata.nl/. Our empirical applications will likely make use of data drawn from the original dataset. All steps from the original data to the sample used for our studies will be carefully documented in a Stata-do-file and/or a readme file.

3. PSID. General documentation on the original data can be found on https://psidonline.isr.umich.edu/. Our empirical applications will likely make use of data drawn from the original dataset. All steps from the original data to the sample used for our studies will be carefully documented in a Stata-do-file and/or a readme file.

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, 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 project

5.1. Where will the data be stored?

The original MEqIn, LISS and PSID data are stored by the three institutions that organized the data. The specific data sets that will used by our research team are stored on internal KU Leuven servers (Onedrive) and on Dropbox (given its widespread use among academics). In addition, local copies can be kept on personal devices.

5.2. How will the data be backed up?

The data will be stored on Onedrive (via KU Leuven servers) and Dropbox. The latter has permanent back-up facilities. The university's central server also has automatic back-up procedures.

5.3. 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. Each dataset that we will use in our empirical applications is relatively small in size (less than 0.1 GB). This does not imply any problem for storage on standard local devices, nor for the university servers and Dropbox.

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

The data storage on the university's servers is free. The three main investigators also have a subscription to Dropbox. Per investigator, the total costs are about 500 euro over the course of the project. The costs are covered through the project's funding. We will keep our subscription to Dropbox after the project given that it is a widely used tool to share data and documents.

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

The original data from MEqIn, LISS and PSID are stored through a secure environment. We refer to the respective organizers for details.

6. Data preservation after the end of the project

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

All data we intend to use can be stored for at least 10 years.

6.2. Where will these data be archived (= stored for the long term)?

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Given the small size data sets, we do not expect any substantial costs for the storage of the data in Dropbox and the university's servers. The expected costs can be easily covered through project funding.

7. Data sharing and re-use

7.1. 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 or because of IP potential)?

The three data sources we intend to use can be freely used by academic researchers for noncommercial use. The only condition is that they need to register. Once this is done, access to the data is granted. From there on, our Stata do files and/or readme files can be consulted to go from the original data to the specific data used in the various applications.

More and more academic journals require that the data sets used for published papers are sent to these journals for replication purposes (examples in economics are American Economic Review, Economic Journal, Review of Economics & Statistics). In addition, these journals require researchers to send their programs. Of course, we will carefully follow these steps in the publication process.

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

The final datasets of our respective empirical applications will be made available in Stata (.dta) format. In addition do-files, readme-files and programs will be made available.

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

- Upon request by mail

7.4. When will the data be made available?

- Upon publication of the research results

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

Academic and policy researchers without commercial motives.

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

Negligible costs through Dropbox or mail sharing.

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

Laurens Cherchye, Bram De Rock and Frederic Vermeulen have the final responsibility.

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

Laurens Cherchye, Bram De Rock and Frederic Vermeulen have the final responsibility.

8.3. Who will be responsible for ensuring data preservation and sharing?

Laurens Cherchye, Bram De Rock and Frederic Vermeulen have the final responsibility.

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

Laurens Cherchye, Bram De Rock and Frederic Vermeulen have the final responsibility.