DMP title

Project Name Motherhood and maternity leave: effects and responses from workers and firms in the presence of informal labor markets - DMP title

Project Identifier G0B8122N

Grant Title G0B8122N

Principal Investigator / Researcher Mariana Zerpa Reisch

Project Data Contact mariana.zerpa@kuleuven.be

Description Despite important progress in the past 50 years, there remain large gaps in employment and earnings between men and women. Recent literature has shown that part of these gaps start after the birth of the first child, with large effects on short and long run employment and earnings of women. While some of this motherhood effect can be explained by intrahousehold labor supply decisions, less is known how firms respond to maternity leave. In the short run, firms may have to adapt their workload across workers and increase labor demand. In addition, in contexts with high labor market informality (as is the case in most developing and transition economies), firms could also use unregistered workers and underreported hours of work as margins of response. In the medium run, firms could update their expectations and potentially adapt their hiring practices to avoid these costs, which could harm womenâ€™s employment opportunities. This project aims to study the impact of childbearing on the labor market, both from the point of view of workers and firms, in the presence of labor market informality. It leverages detailed linked employer-employee administrative data, household surveys and a project-specific manager survey from Uruguay. We use dynamic difference-in-differences research designs to take advantage of the different timing of births across individuals and firms, to study how the trajectories of worker and firm level outcomes are shifted after experiencing a maternity leave.

Institution KU Leuven

1. General Information

Name applicant

Mariana Zerpa Reisch

FWO Project Number & Title

G0B8122N

Motherhood and maternity leave: effects and responses from workers and firms in the presence of informal labor markets

Affiliation

- KU Leuven

2. Data description

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

- Generate new 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).

Type of data	Format	Volume	How created
Social security administrative records of sample of workers	.txt	4GB	Confidential social security administrative records obtained from Banco de Previsión Social del Uruguay
Social security administrative records of sample of firms	.txt	20GB	Confidential social security administrative records obtained from Banco de Previsión Social del Uruguay
Household surveys	.dta	6GB	Public use survey data from Instituto Nacional de Estadística processed by UDELAR
Manager survey	.dta	0.5 to 5GB	Survey of firm managers in Uruguay to be conducted as part of the project (through email, phone or face-to-face)

In addition to these source datasets, the project will create processed data files, i.e. intermediate customized datasets that contain combinations of existing datasets but also models generated as part of the project. These datasets will be used for descriptive and econometric analysis using Stata (.dta), generating tables and graphs with descriptive statistics and model parameter estimates.

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

First, we use two confidential datasets obtained from Uruguay's social security administration (Banco de Previsión Social--BPS). These are monthly records from 1996 to 2016 for (a) a sample of workers and (b) a sample of firms. The datasets are anonymized by BPS before giving them to us to protect the confidentiality of the records. Second, we use publicly available household survey data, which is already anonymized and published by Uruguay's statistics office (Instituto Nacional de Estadística). In both of these cases, we do not use personal data because we receive the data anonymized and the information does not allow one to identify a person.

Third, we plan for a survey of firms to collect information on informal employment practices in Uruguay. We will contract a local survey company with the experience and resources to collect such data according to the highest standards. We will gather basic demographic information about the respondent, information about managenerial practices and information about the firm.

We will anonymize or pseudonimize the data or request the contracted local survey firm to anonymize it for us.

PRET application: G-2022-5456 (in process of elaboration, to be submitted during year 1 of the project)

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

We use two confidential datasets obtained from Uruguay's social security administration (Banco de Previsión Social--BPS). These are monthly records from 1996 to 2016 for (a) a sample of workers and (b) a sample of firms. We have access to the data under a data use agreement with BPS. The agreement prevents us from sharing the data with third parties.

4. Documentation and metadata

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

At the end of the project, we will create a document detailing the sources of data used and instructions on how to obtain the data, including how to apply for the use of confidential data. We will also make available all the code we use to process the data and for the statistical and econometric analysis. The detailed code will enable any researcher with access to the original dataset to replicate our analysis.

In the case of our own survey, we will make the anonymized dataset available to other researchers by request. The information on how to request this data will be included in the data documentation, as well as in my professional webpage. We will create a sharable dataset, together with a data dictionary and descriptive 'read me' document.

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

We will carefully document the data creation, management, and statistical and econometric analysis.

For the creation of the documents, and for creating the metadata of our survey of managers, we will use the standards of the leading Economics journals to allow any researcher with access to the original dataset to replicate our analysis. To make our survey data available, we will use the data repository and standards used by the journal(s) where our research is published. If the journal(s) does not have a repository, we will use a repository used by one of the leading journals in Economics.

5. Data storage and backup during the FWO project

Where will the data be stored?

The master copy of the data will be kept on my personal computer and backed up in Dropbox. Copies can be made and kept on personal devices. Since I will collaborate with two coauthors,

copies can be make and kept on our personal devices, and we will use Dropbox for active use, collaboration and real time back-up of the data during the project.

How is backup of the data provided?
The original data will be stored in the PI's personal computer with back up on Dropbox. Intermediate analysis datasets will by kept in the PI's computer and the computers of her collaborators, with automatic real-time back up on Dropbox.

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

Yes, there is currently sufficient storage and backup capacity in both my server at KU Leuven and my Dropbox account where the back-ups will be made. The funds in the project allow for maintenance of the servers, purchasing a computer for a student assistant, and purchasing more storage and back-up capacity if needed.

What are the expected costs for data storage and back up during the project? How will these costs be covered?
For IT-material, we foresee 2* 2.500€ for powerful computers for the principal investigator and the research assistant to work on, including software (incl Stata licenses, purchase of credits for the use of KU Leuven servers, Dropbox extended storage and version history subscription).

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?
Since we will be working with confidential datasets (although all data will be anonymized before we obtain it), we will take security measures by keeping the data only in password protected personal use devices and using password protected cloud back up through a reputable service (Dropbox). Servers and PCs are managed by ICT department of FEB, according to latest security norms. Data sets can only be accessed by researchers who obtained permission.

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, ...).
In line with KU Leuven policy, the original raw data files will be preserved for at least 10 years. In addition to these source datasets, the project will create processed data files, i.e. intermediate customized datasets that contain combinations of existing datasets but also models generated as part of the project. These datasets will be used for descriptive and econometric analysis using Stata (.dta), generating tables and graphs with descriptive statistics and model parameter estimates, and will not be preserved after the project. Instead, the code that creates these datasets and all project output from the original raw data, as well as any other data processing, will be stored and publicly available.

Where will the data be archived (= stored for the longer term)?
The data at KU Leuven will remain stored for 10 years under the custody of the principal investigator, under her own service and online backup services.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
Because of the size and the current capacity of the servers, the marginal costs of preserving the data are small. The strategy of keeping the data in different forms (hard disks drives and online backup services) guarantees the preservation of the data at a cost of up to 5000 euro under the budget of the project.

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:

We use two confidential datasets obtained from Uruguay's social security administration (Banco

de Previsión Social--BPS). These are monthly records from 1996 to 2016 for (a) a sample of workers and (b) a sample of firms. We have access to the data under a data use agreement with BPS, which prevents us from sharing the data with third parties.

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

We will make available the data documentation, including instructions on how to request the proprietary data and how to obtain the public use household survey data, and the code we use to process the data and obtain the project's research output.

We will also make the data from our project's survey available to other researchers upon request.

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

- In an Open Access repository
- In a restricted access repository

To make all data documentation and code available, we will use the data repository and standards used by the journal(s) where our research is published. If the journal(s) does not have a repository, we will use a repository used by one of the leading journals in Economics. We will make our survey data available upon request through a restricted access repository. The procedure for requesting access to data will be available on the principal investigator's website.

When will the data be made available?

- Upon publication of the research results

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

Any interested party will be able to request access to our survey data. Access will be considered after a request is submitted explaining the planned reuse. Only uses for research purposes will be allowed and commercial reuse will be excluded.

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

There is no additional cost of data sharing our data documentation and our own survey data, besides the time of the researchers and the data storage costs already mentioned above.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Mariana Zerpa (PI) will have overall responsibility. She will instruct collaborators how to document the data and program codes relating to this project.

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

Mariana Zerpa (PI) will be responsible for the data storage at KU Leuven.

Who will be responsible for ensuring data preservation and reuse ?

Mariana Zerpa (PI) will be responsible for data preservation and reuse.

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

The PI bears the end responsibility of updating & implementing this DMP.