
Gender inequality in the labor market: The role of flexible jobs

A Data Management Plan created using DMPonline.be

Creator: Alexander Wintzéus

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 11P2J24N

ID: 206768

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Over the last half-century, gender disparities in the labor market have narrowed considerably. However, gender gaps in labor market outcomes still persist today. Despite an extensive amount of economic research aimed at identifying the factors underlying these gaps, traditional factors have not been able to capture their full extent. Recent evidence indicates that gender differences in preferences for flexible jobs could be a key factor driving the persistent wage gap between men and women. Importantly, further research suggests that childbirth and child-rearing responsibilities play a role in shaping these differences in preferences.

In light of the rising prevalence of employees working in flexible jobs, it becomes increasingly important for policy makers to understand how preferences for flexible jobs interact with other work and family-related decisions in shaping gender gaps in labor market outcomes. The aim of my research is to contribute to this understanding. To this end, I start by examining the role of job flexibility in the decision-making process of couples with children. Next, I turn to quantify how men and women value flexible jobs and investigate how the presence of children in the household affects these valuations. Finally, I study the importance of job flexibility and fertility preferences in driving the life-cycle gender wage gap.

Last modified: 25-04-2024

Gender inequality in the labor market: The role of flexible jobs

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
Household survey data (Bilendi)	Microeconomic household survey dataset. This dataset contains information on the allocation of time and expenditure in the household for Dutch couples with children.	(N) Generate new data	(D) Digital	<ul style="list-style-type: none"> • Observational • Text 	.sav .dta	<1GB	
Household panel data (GSOEP)	Microeconomic survey dataset with panel structure. This dataset contains information on wages and worker characteristics for German workers.	(E) Reuse existing data	(D) Digital	<ul style="list-style-type: none"> • Observational • Compiled/Aggregated 	.dta	<10GB	
Administrative data (I)	Administrative microeconomic data provided by Statistics Netherlands (CBS). This data includes employment histories for the entire Dutch workforce.	(E) Reuse existing data	(D) Digital	<ul style="list-style-type: none"> • Compiled/Aggregated 	.sav	<10GB	

Administrative data (II)	Linked employer-employee data for Germany provided by the German Institute for Employment Research (IAB).	(E) Reuse existing data	(D) Digital	<ul style="list-style-type: none"> • Compiled/Aggregated 	.sav	<10GB	
Computer code (I)	Computer code for cleaning <i>raw</i> datasets.	(N) Generate new data	(D) Digital	<ul style="list-style-type: none"> • Software 	.do (Stata code) .r (R code)	<1GB	
Computer code (II)	Computer code for numerically solving and estimating structural models.	(N) Generate new data	(D) Digital	<ul style="list-style-type: none"> • Software 	.do (Stata code) .jl (Julia code) .py (Python code)	<1GB	
Results	Empirical results: Estimation results and policy counterfactuals.	(N) Generate new data	(D) Digital	<ul style="list-style-type: none"> • Numerical Simulation Data • Text • Images 	.csv .png	<1GB	

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:

I use microeconomic datasets that are available either publicly or upon agreement:

1. **Household panel data (GSOEP)** can be obtained upon request from DIW - SOEP from: https://www.diw.de/en/diw_01.c.678568.en/research_data_center_soep.html
2. **Administrative data (I)** can be consulted after an application procedure from Statistics Netherlands (CBS) through Remote Access (RA) technology: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen>
3. **Administrative data (II)** can be consulted after an application procedure from the German Institute for Employment Research (IAB): <https://fdz.iab.de/en/our-data-products/overview-of-available-data/data-sources/>

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.

- Yes

Household survey data (Bilendi) is a microeconomic dataset that is constructed on the basis of household surveys. To this end, my supervisors have collaborated with the Belgian division of the market research company Bilendi. This company has a pool of pre-recruited participants at its disposal that are regularly invited to partake in online surveys. Any personal data is processed on their side and has been rendered **anonymous** before being made available for research purposes. Only the anonymized data is

reusable for replication.

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.

- Yes

1. **Household panel data (GSOEP):** Confidentiality agreements must be respected in order to be able to access the data.
2. **Administrative data (I):** This data from Statistics Netherlands (CBS) are only accessible on-site or through a remote desktop connection (Remote Access; RA). Access to the data (e.g., for replication purposes) must be done with the agreement of CBS.
3. **Administrative data (II):** This data from the German Institute for Employment Research (IAB) can only be consulted on-site or through remote access. Data and program codes can only be obtained with the approval of IAB.

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).

1. **Household survey data (Bilendi):** Original data files will be saved and accompanied with metadata created by myself and my supervisors.
2. **Household panel data (GSOEP):** Original data files (including metadata provided by the host organization DIW-GSOEP) will be saved together with the download date.
3. **Administrative data (I):** Data files and metadata can be obtained after approval from Statistics Netherlands (CBS).
4. **Administrative data (II):** Data files and metadata can be obtained after approval from the German Institute for Employment Research (IAB).

Any data manipulations of original data files to obtain the datasets to be used in the empirical analyses will be clearly documented.

Computer code used to manipulate the original data files -- **Computer code (I)** -- will contain extensive comments to justify any manipulations and make it understandable.

Similarly, computer code for performing empirical analyses -- **Computer code (II)** -- will carry comments for comprehensibility.

Finally, I will include a README.txt or README.md file that aims to explain how to obtain the results of the analyses. The empirical results follow as an automatic output of **computer code (II)**. All output will be described in the README file.

It is, however, important to stress that any program code and files related to the **administrative datasets** can solely be obtained upon approval of the associated hosting organizations.

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.

- Yes

The **DataCite Metadata Schema** will be used for the creation of coherent metadata to **Household survey data (Bilendi)**.

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

Where will the data be stored?

In general, data will be stored in the following locations:

- OneDrive (KU Leuven)
- Personal drives at FEB KU Leuven (Y drive).

For collaboration with other researchers across different research units, DropBox and GitHub will be employed for the active use of data files and program codes during the project.

However, any computer code and data files related to the **administrative datasets** will be stored on the servers of the respective host organizations (CBS and IAB).

How will the data be backed up?

Data will be backed up automatically by KU Leuven FEB ICTS for the storage solutions mentioned above (**OneDrive, Personal drives**). They are managed according to the latest security standards to prevent loss of data.

In addition, I intend to back up any files distributed between collaborators on **DropBox and GitHub**.

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

The estimated volume of all data files should not exceed 40GB. The storage capacity provided by the **KU Leuven OneDrive and Personal drive** solutions should well exceed this estimated amount.

Moreover, both **DropBox and GitHub**, which are to be employed to facilitate the sharing and active use of data files and program codes between collaborators, offer ample storage possibilities.

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

For data stored on the **FEB KU Leuven servers**, i.e.,

- **Household survey data (Bilendi)**
- **Household panel data (GSOEP)**
- **Computer code (I & II)**
- **Estimation results**

standard security measures put in place by the KU Leuven will be followed (including MFA). These servers are managed according to the latest security standards and data storage can be encrypted. Data can only be accessed by authorized researchers.

For collaboration with other researchers across different research units, **DropBox and GitHub** will be employed for the active use of data during the project. Any files shared on **DropBox and GitHub** can only be accessed by collaborators (e.g., by hosting private repositories).

Administrative data I and II can only be accessed on site or through a remote desktop connection. These data are stored on the local servers of the respective host organizations (i.e., CBS and IAB) and are subject to the security measures put in place by these organizations. Furthermore, any program code manipulating either administrative dataset will also be securely stored on the local servers of the host organizations.

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

The core storage facilities (**One Drive, Personal drive**) are readily available and free of use for staff at the FEB.

To ensure sufficient storage capacity on **DropBox**, I will use the FWO bench fee (DropBox Plus subscription; 2TB storage; €120/year).

The free storage capacity on **GitHub** is limited, but should suffice for the intended collaboration through this service.

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 will be retained for **10 years** following the **KU Leuven RDM policy**. Certain data, however, are subject to restrictions:

- **Household panel data (GSOEP):** This data can be stored locally on the servers of the Faculty of Economics and Business (FEB) KU Leuven. However, in order to access the data confidentiality agreements must be respected.
- **Administrative data I:** This data is only accessible on-site or through a remote desktop connection. They will not be stored locally on the servers of the FEB. Access to the data for replication purposes can only be obtained after agreement of Statistics Netherlands (CBS).
- **Administrative data II:** This data can only be accessed on-site or through a remote desktop connection. Data and program code can be consulted with approval from the German Institute for Employment Research (IAB).

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

Data will be archived in the **KU Leuven Research Data Directory**.

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

No additional costs for preservation during the expected retention period should occur.

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 a restricted access repository (after approval, institutional access only, ...)

In general, datasets and program code will be publicly accessible through journals, the authors' websites, in public GitHub repositories, or on request.

However, data and program code related to **administrative data I and II** can only be accessed upon approval from the respective organizations from which they are obtained (CBS and IAB).

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

- **Administrative data (I):** Data is obtainable from Statistics Netherlands (CBS). Access is either on-site or via a remote desktop connection. Access to the data for replication purposes must be done with the agreement of CBS. Approval can only be obtained if the applicant under consideration is employed at an approved organization and intends to use the data for research (or replication) purposes only.
- **Administrative data (II):** Data is obtainable from the German Institute for Employment Research (IAB). Data can be consulted either on-site or via a remote desktop connection. Access to the data must be done with the approval of IAB which can only be obtained by applicants employed at approved organizations and for research purposes only.

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.

- Yes, Other

Data and program code related to **administrative data I and II** can only be accessed upon approval from the respective organizations from which they are obtained (CBS and IAB).

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

- KU Leuven RDR (Research Data Repository)
- GitHub

When will the data be made available?

Upon publication of research results.

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

- CC-BY 4.0 (data)
- Data Transfer Agreement (restricted data)

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.

- Yes

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

No (additional) costs for data sharing are to be expected.

6. Responsibilities

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

Alexander Wintzéus (PhD fellow)

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

Alexander Wintzéus (PhD fellow)

Who will manage data preservation and sharing?

Alexander Wintzéus (PhD fellow), Frederic Vermeulen (supervisor), and Iris Kesternich (co-supervisor)

Who will update and implement this DMP?

Alexander Wintzéus (PhD fellow)