

## Replication package for BLM

This git repository contains all the code to replicate the results of **Bonhomme Lamadon and Manresa “A distributional Framework for matched employer-employee data”**, forthcoming at **Econometrica**. The working-paper version is available [here](#). Virtually all code is based on the R platform.

If you are looking for the R package to use the method of the apper, you should use the [rblm package](#). It includes most of the estimators available here, and we keep updating it.

The present replication package is built as an R package that can be easily installed on any system. All package dependencies can be handled using [packrat](#). This option guarantees that results can be reproduced using the exact versions of all the libraries that were used at the time the paper was written. We also provide a [Docker container](#) to ensure full portability. This provides a full linux stack with RStudio and all libraries installed and configured.

Importantly, reproducing the results on Swedish data **requires access to the administrative data from Sweden**. Researchers need to apply to get access to such data. We recommend contacting the [IFAU](#). The institute is hosting this replication package that can be accessed and ran on the data on their servers. The reference name for our project is IFAU-2015-65.

### How do I run this?

The simplest way to use this replication package is to rely on the docker container that we have created as described in solution 1. This will get it running almost instantly.

#### **Solution 1: get it running in less than 10 minutes, run our docker container**

Make sure you have the [docker app](#) installed on your computer. Then run the following command:

```
docker run -d --rm -e PASSWORD=blm -p 8787:8787 tlamadon/blm-replicate
```

This will automatically download our docker container from dockerhub and start it. This will give you access to a fully functioning RStudio with the installed libraries and the code necessary to run the replication code. After completion of the previous command, this Rstudio environment should be available in your browser at <http://localhost:8787>, which points to your local computer. Use login `rstudio` and password `blm`.

From there calling `source("inst/main.R")` will start the full replication, create all necessary intermediate results and generates all figures and tables, saving them in the `tmp` folder. We invite the researcher however to explore the [inst/main.r](#) file.

By default, this will run all of the code using a **synthetic data set**. See below how to get access to Swedish data, and load it into the container.

**Note 1:** make sure the docker app does not limit memory access to less than 16Gb. See [here](#).

**Note 2:** you can stop the container by running `docker stop blm-replicate`. If you want to keep working on the environment, you should not use the `--rm` argument in the original call. Such argument enforces the container to be destroyed upon stopping.

**Note 3:** you can easily move files in and out of a running container using the `docker copy SOURCE DEST` command. Or you can mount a folder from your host computer. See details [here](#).

## Solution 2: install the replication package into your R environment

If you have your own running R system, and you want to run this replication package in your environment, you can directly install the package. In this case we recommend that you make use of the [packrat](#) configuration we are providing.

1. Download the replication package. We recommend to simply clone the github repository, ie: `git clone https://github.com/tlamadon/blm-replicate.git`
2. Start R inside the replication package.

In R, run the following commands:

```
# installing the package locally in your R env.

install.packages("pakcrat") # make sure that packrat is available
install.packages("devtools") # make sure that devtools is available

source("packrat/init.R") # initialize the packrat environment
packrat::restore()       # make sure all is up to date

devtools::install(".")   # build the replication package

source("inst/main.R")    # fire up the replication
```

## Overview of the replication package

The main entry point is [inst/main.r](#). It will **automatically** run all the necessary steps in the other files in order to reproduce all the results of the paper. Note however that this would take a very long time as it will start some bootstraps procedure. The code will generate all figures and tables and put them into folder called `tmp` by default.

We invite researchers to read through [inst/main.r](#) which has explicit calls for each subset of the paper.

## Organization of the code

- All the heavy lifting such as the estimators and simulation codes are in the `R/*.r` folder. This is the usual way to store functions in an R package.
- [inst/server/estimation-static.r](#) contains the code that runs the estimations for the **static** version of the model
- [inst/server/estimation-dynamic.r](#) contains code that runs the different estimations for the **dynamic** version of the model.
- [inst/server/fig-blm.R](#) contains functions that generate all of the **figures and tables** in the paper.

## Replicating the results on Swedish data

### Access to necessary data files

Researchers can directly apply for access to `tmp-2003-static.dat` and `tmp-2003-dynamic.dat` by contacting the [IFAU](#). These two files are the inputs to the replication package and a copy is stored as part of the replication package on the servers at the IFAU. The reference name for our project is IFAU-2015-65.

These files should be placed in the folder `<working-directory>/data-tmp`.

### Re-creating these data files

These two input files are generated using the Swedish employer employee match data. The main data source should be the following list of: `selectedf0educ1.dta`, `selectedf0educ2.dta`, `selectedf0educ3.dta`, `selectedf1educ1.dta`, `selectedf1educ2.dta`, `selectedf1educ3.dta`, `selectedfirms9708.dta`.

The following two scripts use these data sources to construct two data files `tmp-2003-static.dat` and `tmp-2003-dynamic.dat`:

- [inst/server/data-section-static.r](#) contains the code that **processes the data inputs** to prepare the data for the static estimation.

- [inst/server/data-section-dynamic.r](#) contains the code that **processes the data inputs** to prepare the data for the dynamic estimation.

## Using your own data source

This is similar to using the Swedish data. You only need to provide two data sources in the form of a `data.frame`. One should be called `sdata` and contain information on all workers, and one should be called `jdata` and contain information only about the movers. To see the structure that this data.frames should have. The `sdata` and `jdata` frames should be saved into `data-tmp/tmp-2003-dynamic.dat` and `data-tmp/tmp-2003-dynamic.dat` for the static and the dynamic estimation.

We recommend to have a look at the function `generate_simulated_data` in [inst/server/server-utils.R](#). It creates synthetic data simulated from our main specifications and save files to the same format as the actual data. This is your best source to match the structure exactly.

Here is what `sdata` looks like:

	k	y1	y2	j1	j2	j1true	f1	f2	move	birthyear	x	wid	inc
1:	1	9.846396	9.747927	5	1	5	F1335	F1	1	1961	1	W64819	Construction et
2:	2	10.040879	10.075224	5	1	5	F135	F1	1	1963	1	W64807	Retail tra
3:	5	10.638532	10.744525	3	1	3	F143	F1	1	1979	1	W60513	Retail tra
4:	3	8.894678	10.195521	4	1	4	F144	F1	1	1963	1	W62818	Construction et
5:	3	9.718155	9.438086	1	1	1	F181	F1	1	1965	1	W58054	Service
---													
77571:	2	9.983228	10.219231	8	8	8	F998	F998	0	1964	1	W51166	Service
77572:	4	10.471325	10.398645	8	8	8	F998	F998	0	1971	1	W51331	Service
77573:	4	10.331180	10.516750	8	8	8	F998	F998	0	1967	1	W51434	Service
77574:	6	11.375500	11.292524	8	8	8	F998	F998	0	1968	1	W51496	Service
77575:	4	10.399596	10.501993	8	8	8	F998	F998	0	1973	1	W51543	Service
---													
		va1	ind2	va2	educ	size1							
1:		3.3505830	Manufacturing	3.8792234	1	17							
2:		13.7959329	Manufacturing	3.8792234	3	24							
3:		0.2839520	Manufacturing	3.8792234	2	13							
4:		3.0592294	Manufacturing	3.8792234	1	12							
5:		1.0255445	Manufacturing	3.8792234	3	27							
---													
77571:		0.4115116	Services	0.4115116	3	64							
77572:		0.4115116	Services	0.4115116	1	64							
77573:		0.4115116	Services	0.4115116	2	64							
77574:		0.4115116	Services	0.4115116	3	64							
77575:		0.4115116	Services	0.4115116	1	64							