This readme file describes how to replicate the empirical results in Behrens et al. “QUANTIFYING THE GAP BETWEEN EQUILIBRIUM AND OPTIMUM UNDER MONOPOLISTIC COMPETITION” published in the QJE.

Besides this readme file, there are 4 folders (containing various files) and one excel file in this replication package. The excel file “Complete Results and Tables.xlsx” contains a number of detailed results and material needed to reconstruct Tables I, II and III in the paper, as well as Figures I and II, along with some material belonging to the Online Appendix like Tables H-1 to H-3.

In order to replicate our results the first step is to start from the BSD database for the UK and the ESANE database for France. See below for more details about how to access these confidential datasets. In the folder “BSD” you will find two Stata do files: 1) Vai_CES_CARA_Final.do; 2) Vai_CES_CARA_Final_revenue.do. The first one computes the standard deviation of log employment (as well as other stats) for each UK industry in our analysis. The file requires merging a dataset, that you will find in the folder “OECD STAN” (aggregate_ind_R_D_UK.dta), providing info on industry-level total R&D outlays as a % of total revenue computed from the OECD STAN database. The output of this do file is contained, along with other things, in the excel file “Complete Results and Tables.xlsx”. The second file computes the standard deviation of log revenue (as well as other stats) for each UK industry in our analysis. The output of this do file is contained, along with other things, in the excel file “Complete Results and Tables.xlsx”.

In the folder “ESANE” you will find three Stata do files: 1) Vai_CES_CARA_Final.do; 2) Vai_CES_CARA_Final_revenue.do; 3) Vai_CES_CARA_Final_profits.do. The first one computes the standard deviation of log employment (as well as other stats) for each French industry in our analysis. The file requires merging a dataset, that you will find in the folder “OECD STAN” (aggregate_ind_R_D_FR.dta), providing info on industry-level total R&D outlays as a % of total revenue computed from the OECD STAN database. The output of this do file is contained, along with other things, in the excel file “Complete Results and Tables.xlsx”. The second file computes the standard deviation of log revenue (as well as other stats) for each French industry in our analysis. The output of this do file is contained, along with other things, in the excel file “Complete Results and Tables.xlsx”. The third file computes the standard deviation of log employment (as well as other stats) for each French industry in our analysis and it is obtained using a different proxy for fixed costs based on aggregate firm profits. The output of this do file is contained, along with other things, in the excel file “Complete Results and Tables.xlsx”. All 3 files point to merging a correspondence between the 2008 APE industry classification used in the raw ESANE data and our aggregate industry classification based on the NACE rev 1.1 classification. The correspondence is provided in the file “Done2.dta” in the ESANE folder.

The next folder to look at in the one named “Matlab”. This folder contains the Matlab code and data needed to compute a number of key parameters (like the Pareto shape parameters k, the theta, the kappa, the beta, etc.) as well as cutoff and entry distortions. The output of the Matlab code and data is contained, along with other things, in the excel file “Complete Results and Tables.xlsx”. The file “Procedure_UK3_r.m” contains the code related to the case where, for the UK, we consider the standard deviation of log employment and R&D outlays. The file uses the data contained in the file “Data_UK3_r.mat”, which is also available in excel format in the file “Data_UK3_r.xlsx”. The file “Procedure_UK_revenue_r.m” contains the code related to the case where, for the UK, we consider the standard deviation of log revenue. The file uses the data contained in the file “Data_UK_revenue.mat”, which is also available in excel format in the file “Data_UK_revenue.xlsx”. The file “Procedure_FR2_r.m” contains the code related to the case where, for France, we consider the standard deviation of log employment and R&D outlays. The file uses the data contained in the

file “Data_FR2_r.mat”, which is also available in excel format in the file “Data_FR2_r.xlsx”. The file “Procedure_FR_revenue_r.m” contains the code related to the case where, for France, we consider the standard deviation of log revenue. The file uses the data contained in the file “Data_FR_revenue.mat”, which is also available in excel format in the file “Data_FR_revenue.xlsx”. The file “Procedure_FR_emp_prof.m” contains the code related to the case where, for France, we consider the standard deviation of log employment and aggregate profits. The file uses the data contained in the file “Data_FR_emp_prof.mat”, which is also available in excel format in the file “Data_FR_emp_prof.xlsx”. The file “Procedure_UK_FR_one_sect.m” instead contains the code related to the case where, for both France and the UK, we consider the standard deviation of log employment and R&D outlays and one single aggregate sector. Finally, the files “sd_log_rev_int.m”, “sd_loge_int.m” and “eqsystem.m” provide some auxiliary functions used in the above matlab .m files while the file “options_BE.mat” provides the options for the solver.

Last but not least, the folder “Graphs” provides the data and code needed to produce Figures I and II. Both figures are available in the folder as .eps (“Graph_France.eps” and “Graph_UK.eps”) and .pdf (“Graph_France.pdf” and “Graph_UK.pdf”) format. Starting for the Stata data set (“Graph_France.dta” for France and “Graph_UK.dta” for the UK) one simply needs to execute the command in the text file “Graphs_Stata_commands.txt” to produce the relevant figure. The data contained in the Stata .dta files, that is, entry and cutoff distortions for each industry, come straight from the output of the Matlab .m files “Procedure_FR2_r.m” and “Procedure_UK3_r.m”.

Additional details about access to confidential data

The BSD database is accessible to researcher via the Secure Data Service managed by the ONS: https://www.ukdataservice.ac.uk/get-data/how-to-access/accesssecurelab.aspx. The procedure to access the data involves submitting a project to the ONS and is as follows:

1. The purpose of your use/project must provide a public benefit/ serve the public good
2. You must have attended Safe Researcher Training and be an Accredited Researcher (AR). This status is currently provided by the ONS.
3. If you are a postgraduate student you can request access but will need to apply jointly with your supervisor(s)
4. You must check any specific access restrictions for each study in our Data Catalogue that you wish to use, found in the 'Access' tab
5. You must only access data from within the UK (SN 6676 can be accessed by German researchers, subject to meeting other conditions)

The ESANE-FARE database is accessible to researcher via the CASD: https://www.casd.eu/en/

The procedure to access the data is as follows:

In order to access confidential data under the Committee on Statistical Confidentiality (statistical data, tax data, certain administrative data), you must submit a request through the online request management portal CDAP.

You should be careful not to submit a request unless you are sufficiently certain of the objectives of your project and the composition of your team.

Additional requests for changes to the list of sources or persons accessing the data, which would take place within less than two months of an initial request, will only be accepted for demonstrated urgency.

After filling in your file, you will have to send the pdf summary to the producer services concerned for the examination of your request.

You must take into account the availability of the people you are going to request. In order to take this constraint into account, you are invited to send your application **as soon as possible before** the application deadline.

The producer services will examine and assess your application in the light of the conditions laid down in the opinion of the Committee on Statistical Confidentiality. Under the terms of Article 17 of Decree No. 2009-318 of 20 March 2009 on the National Council for Statistical Information, the Statistical Confidentiality Committee and the Committee for the Official Statistics Label, "the Statistical Confidentiality Committee shall deliver its opinion taking into account the nature and relevance of the work for which the application is made, the status of the person or body submitting the application and the guarantees it offers. It checks that **the volume of information requested is not excessive in relation to the work justifying its communication and that this does not lead to excessive prejudice to the interests that the Act** [No. 51-711 of 7 June 1951 as amended on the obligation, coordination and secrecy of statistics] was intended to protect."

In accordance with these provisions, you are invited to assess the quality and completeness of the file you submit. **The use of each source must therefore be justified in relation to the objectives of your project.** If you do not meet these requirements, **the producer services may refuse their agreement to the examination of your file by the Committee**. **On the other hand, this agreement does not presage the final opinion that the Committee will express**.

If your application has received the agreement of all the producer departments concerned, the secretariat will register it for submission to the Committee on Statistical Confidentiality for its opinion. Two cases may arise:

- Your file does not a priori give rise to any particular questions, your file will be submitted to the Committee for its opinion by electronic consultation with its members;
- Your file is likely to provoke questions from the Committee, on its purposes, the use of sources, the respect of the obligations due to statistical secrecy with regard to the missions of your organization, the valorization of the results produced, etc., it will then be examined in the face of the Committee and **you will be asked to present your project**.

If the Committee's opinion is favourable, you will then be able to access the requested data after the formalities for signature by the competent authorities and, for sources disseminated via the CASD (Centre for Secure Access to Data), after contractualisation between the CASD and your institution and a registration session (registration). **Although the Committee makes its best efforts to ensure that these formalities, which are essential to guarantee the legal security of the communication of data, are completed as quickly as possible, they may nevertheless take several weeks**. You are invited to take this into account in the organization of the timetable for your project.

For more information: see the page On-line submission of applications

Obligations relating to the processing of personal data

A personal data is a data relating to an identified or identifiable natural person, which in practice applies to all data accessible through the Committee on Statistical Confidentiality, with the rare exceptions of sources that would only relate to legal persons. **The processing of personal data is subject to conditions laid down in the General Regulation on Data Protection (GDPR) and the Data Protection Act**.

Applicants are warned that the favourable opinion granted by the Committee **applies only to access to the data, and in no way prejudges the formalities they must complete to comply**.

Applicants' attention is particularly drawn to the fact that the processing of **"sensitive" data**, as defined in Article 9 of the DPMR (e.g. health, ethnic origins, religious practices, sexual practices) is subject to specific conditions. **An impact assessment** may also be necessary if the processing is likely to cause a high risk to the rights and freedoms of natural persons. This is the case in particular if a processing operation combines at least two criteria, including the scale of the processing operation (which applies when exhaustive sources on a population, as in the case of administrative sources, are used), the use of « sensitive » data and the link between several data sources.

Applicants are invited to contact the **Data Protection Officer** of their organisation for further information. Information is also available on the CNIL website.