DOCUMENTATION ON DATA ACCESS AND REPLICATION PACKAGE OF THE PAPER:

***EXPLOITING GROWTH OPPORTUNITIES: THE ROLE OF INTERNAL LABOR MARKETS***

Giacinta Cestone (Bayes Business School, City University of London and ECGI)

Chiara Fumagalli (Università Bocconi (Department of Economics), IGIER and CEPR)

Francis Kramarz (CREST (ENSAE), Institut Polytechnique de Paris, Uppsala University)

Giovanni Pica (Università della Svizzera Italiana (USI), CSEF, Centro Luca D'Agliano)

The analysis in the paper makes use of confidential firm-level and matched employer-employee data gathered by the French statistical institute, INSEE (*Institut National de la Statistique et des Études Économiques*). Accordingly, access to the datasets used in our paper is subject to prior authorization, and we are not allowed to disseminate these data. However, access is not exclusive. To obtain access to the data, one needs to follow a standardized procedure that applies to every research project: all interested researchers can present a project to the French *Comité du Secret Statistique* and, upon authorization, are allowed to work on the data sources through the CASD ([www.casd.eu](http://www.casd.eu)), as described in more details below.

**Data Availability and Provenance Statements**

We combine the following data sets: the “All employees data base (BTS)”, firm balance sheets (FICUS-FARE), firm-level financial links between firms (LIFI).

* The matched employer-employee data (BTS - Base Tous Salariés) is related to the annual declarations of French firms (Annual declaration of social data – DADS in French). We actually use the Job Position data (BTS fichier POSTES). A description of the source file is available here: <https://www.casd.eu/source/base-tous-salaries-fichier-postes/>. Each observation in BTS corresponds to a unique individual-plant combination in a given year, with detailed information about the plant-individual relationship, including the number of days during the calendar year that individual worked in that plant, the (gross and net) wage, the type of occupation (classified according to socio-professional categories), the full time/part time status of the employee. Moreover, it provides the fiscal identifier of the firm that owns the plant, the geographical location of both the employing plant and firm, as well as the industry classification of the activity undertaken by the plant/firm.
* The FICUS-FARE contains information on firms' balance sheets and income statements. Each observation in FICUS-FARE corresponds to a firm in a given year. It is constructed from administrative fiscal data, based on mandatory reporting to tax authorities for all French tax schemes, and it covers the universe of French firms, with about 2.2 million firms per year. FICUS-FARE provides accounting information, including firm's assets, EBITDA, Value Added, sales, capital expenditures, cash flows and interest payments. A description of the source file is available here: <https://www.casd.eu/source/statistique-structurelle-annuelle-dentreprises-issue-du-dispositif-esane/> (FARE) and <https://www.casd.eu/source/statistique-structurelle-annuelle-dentreprises-issue-du-dispositif-suse/> (FICUS).
* The LIFI collects information on direct financial links between firms, but it also accounts for indirect stakes and cross-ownerships. This allows INSEE to precisely identify the group structure even in the presence of pyramids. More precisely, LIFI defines a group as a set of firms controlled, directly or indirectly, by the same entity (the head of the group). The survey relies on a formal definition of direct control, requiring that a firm holds at least 50% of the voting rights in another firm's general assembly. Each observation in LIFI corresponds to a firm in a given year. A description of the source file is available here: <https://www.casd.eu/en/source/financial-links-between-enterprises-survey/>.

The data span the period 2002-2010

By following the procedure described here, any researcher can remotely access the data stored on a secured server, no need to be a French citizen. This service is provided by the CASD (*Centre d’Accès Sécurisé aux Données*). The data is stored on a server of the French National Statistical Institute and researchers work remotely using a box secured by finger-print. All details provided on the dedicated portal here: <https://www.casd.eu/en/> and here <https://www.casd.eu/en/your-project/procedures-dhabilitation/>. In order to access the data, any interested researcher need to follow the steps below:

* Demands are filled on line. The researcher firstly creates an account here (click on “langue de contact anglais” at the bottom of the page just above the captcha): <https://cdap.casd.eu/creation-compte>. Then the researcher logs-in and fills the demand here: <https://cdap.casd.eu/>
* Select the data sources needed for the project and describe the project

The researcher gets in touch with the administration producing the statistics to double-check that the requested file has the right format: definition of variables, years, identifiers. The CASD procedure is designed in such a way that it automatically routes the researcher to the statistician in charge of any data set the researcher is interested in.

* Apply to the next session of the "*Comité du Secret Statistique*"
* The " *Comité du Secret Statistique* " of the CNIS ultimately provides the authorization. It can be contacted as follows:

Insee, Secrétariat du Comité du secret statistique

Timbre C050, 18 boulevard Adolphe Pinard

75675 Paris cedex 14 (France)

Tel : (33).1.41.17.67.17

Mail : [comite-secret@cnis.fr](mailto:comite-secret@cnis.fr)

* The process may take a bit of time to be granted access because, depending on the data sources (for most of those used in the project but for FICUS-FARE, in particular), access will require a signed agreement (by the researcher and the producers, also by the French Archives) and satisfy some location requirements (the DGFIP data – FARE -- cannot be accessed outside the EU and must be accessed using a CASD box, see below).
* Once every authorization is in place to access the data, the interested researcher will have to contact the CASD to obtain a box through which every step of the work will have to be performed. This is described in full detail in the enclosed document “*casd\_user\_guide-5.pdf*” which is also available at: https://www.casd.eu/en/useful-documents.

**Computational requirements**

The codes were last run on a server with an 8 vCPUs, (Intel processor), on STATA version 16.1 for 64-bit windows and SAS version 9.4 TS Level 1M0. Approximate time needed to reproduce the analysis on a standard machine is, for almost all codes, between 10 and 60 minutes. Some codes (those described in section A.1.4 computing the mean excess probability of within-group firm-to-firm transitions for the (conditional) descriptive evidence on ILM activity) may take up to 10 hours.

**Guide to replication package**

The replication package provides all the codes needed to extract and merge the raw data, and conduct the analysis that allows one to reproduce all the results presented in the paper.

The folder “**REPLICATIONPACKAGE**” contains three subfolders:

1. “**Codes**”
2. “**Dta**”
3. “**Output**”

The folder “**Codes**” contains STATA and SAS codes; the folder “**Dta**” contains a number of subfolders where the datasets generated by the codes will be saved; .the folder “**Output**” contains the output files (graphs and tables) generated by the codes.

We start describing the folder “**Codes**”

1. **Folder “Codes”**

The folder “**Codes**” contains two subfolders:

1. “**DoFiles**”
2. “**SASFiles**”
   1. **Subfolder “Codes\DoFiles”**

The folder “**DoFiles**”, in turn contains four subfolders:

1. “**data\_construction**”
2. “**Descriptives**”
3. “**event\_study**”
4. “**gammas**”

and six do-files:

1. “**1\_shell\_FICUSLIFI\_construction.do”**
2. “**2\_shell\_DADS\_BilateralFlows\_construction.do**”
3. “**3\_shell\_DADS\_BilateralFlowsByQualif\_construction.do**”
4. “**4\_shell\_descriptives.do**”
5. “**5\_shell\_event\_study.do**”
6. “**6\_shell\_gammas.do**”

Each of the six do-files runs additional do-files contained in the four subfolders listed above and provides detailed instructions on when to run the SAS codes contained in the folder “**SASFiles**”. We describe the six do-files below.

* + 1. **Do-file “Codes\DoFiles\1\_shell\_FICUSLIFI\_construction.do”**

This do-file builds a number of firm-level datasets based on FICUS-FARE and LIFI.

First, the do-file defines a number of global variables that identify the paths that contain both the data and additional do-files to be run.

**Step 1**: use raw LIFI data to build a firm-level panel identifying BG-affiliated firms and the head of each group:

* ***(LINE 40)***: converts raw yearly .csv files (located at YOURPATH\LIFI) into STATA format .dta files and append them all, by running the file “**dataassembling.do**", contained in the subfolder “**codes\DoFiles\data\_construction\FICUSLIFI**”.

**do "$home\\$dofile\dataassembling.do**";

* ***(LINE 43)***: next, the file “**cleaning Ent\_Tg.do**” does some cleaning. Most notably, it harmonizes sectors, i.e. it converts NAF REV 2 in years 2008-2010 into NAF REV 1.

**do "$home\\$dofile\cleaning Ent\_Tg.do".**

* output: **ent\_tg\_panel\_clean.dta**.

**Step 2**: build a panel of firm-level balance sheets and merge it with **ent\_tg\_panel\_clean.dta** to identify BG-affiliated firms.

* need to first run the SAS code “**ficus\_extraction.sas**”, contained in the subfolder “**codes\SASFiles\FICUSLIFI**”. This code converts the yearly raw FICUS-FARE data from SAS to STATA format:

"**$home\\$sasfile\ficuslifi\ficus\_extraction.sas**”;

* ***(LINE 76)***: the file “**ficus\_2002\_2010\_bflows.do"** contained in the subfolder “**codes\DoFiles\data\_construction\FICUSLIFI**” appends the yearly dataset and harmonizes sectors codes:

**do "$home\\$dofile\ficus\_2002\_2010\_bflows.do"**

* Output: **ficus\_2002\_2010\_bflows.dta**

**Step 3**: build firm-level dataset for event study regressions in which the (sector-level) shock is the closure of a large competitor

**Substep 3.1**: identify firms’ closures from BTS POSTES. To do so run the following SAS codes contained in the subfolder “**codes\** **SASFiles\CLOSURES**”:

"**$home\\$sasfile\CLOSURES\closures\_0208.sas**”

"**$home\\$sasfile\CLOSURES\closures\_0809.sas**”

"**$home\\$sasfile\CLOSURES\closures\_0910.sas**”

"**$home\\$sasfile\CLOSURES\closures\_0208etp.sas**”

"**$home\\$sasfile\CLOSURES\closures\_0809etp.sas**”

"**$home\\$sasfile\CLOSURES\closures\_0910etp.sas**”

"**$home\\$sasfile\CLOSURES\appendplants.sas**”

The files **CLOSURES\_xxyyetp.sas**: identify closures (and false closures) looking at firm exits from the dataset or a 90% drop of the employment stock from year xx to year yy. The employment stock is computed using headcount in **CLOSURES\_xxyy.sas** and full-time equivalent (ETP) adjustments in **CLOSURES\_xxyyetp.sas**. False closures are defined as closures in which 70% or more of the firms’ workforce is directed towards a unique other firm.

**SIDE NOTE**: the files **closures\_xxyy.sas** also create ancillary output files used as input codes in **closures\_xxyyetp.sas**.

**Example:** closure\_0809 generates the files siren09/siren09etp that identifies firms that are observed last in 2008. In addition it creates the datasets prefinal\_l08 and prefinal\_f09

The output files are:

- siren02/siren02etp: firms observed last in 2002- 2007

- siren09/siren09etp: firms observed last in 2008

- siren10/siren10etp: firms observed last in 2009

- prefinal\_lxx/\_fyy datasets: ancillary output files.

NOTE1: year of closure in the datasets is reported as the first year of absence.

NOTE2: prefinal\_lxx and prefinal\_fyy are datasets of stocks of workers from the BTSyy.

Finally, **appendplants.sas** appends the closure datasets of the different years and saves the resulting list of firm-level identifiers with the associated closure year in STATA format.

* Output: **sirenclosuretp.dta**

**Substep 3.2**: identify the subset of big closures:

* ***(LINE 130)***: **to reduce computing time** the following do-file (in “**codes\DoFiles\data\_construction\FICUSLIFI**”) creates a lighter (few vars no LIFI merged) FICUS (compared to **ficus\_2002\_2010\_bflows.do**) to be used in the next do file

**do "$home\\$dofile\ficus\_1995\_2010\_bflows.do"**

* Output: **ficus\_1995\_2010\_bflows.dta**
* ***(LINE 131)***: the following do-file (in “**codes\DoFiles\data\_construction\FICUSLIFI**”) builds a list of shocked sectors (i.e. 4-digit sectors where large closures happen) and a dataset of shocked firms (i.e. firms in sectors in which large closures happen)

**do "$home\\$dofile\DataBuilding\_v2.do"**

* Output**: ficus152naf4.dta** (dataset of shocked firms)

**152xsectionnaf4ape.xlsx** (dataset of shocked SECTORS)

**Step 4**: use the merged FICUS-LIFI to build firm-level and group-level variables

* ***(LINE 147)***:rundo-files in “**codes\DoFiles\data\_construction\FICUSLIFI**”

**do "$home\\$dofile\groupsize.do"**

group size

**do "$home\\$dofile\vapw data preparation.do"**

group hhi (using employment)

rest-of-group EBITDA

rest-of-group value added per worker

rest-of-group sales

....

**do "$home\\$dofile\siren\_l\_f\_chars\_distr.do"**

Distribution of firm-level variables

* + 1. **Do-file “Codes\DoFiles\2\_shell\_DADS\_BilateralFlows\_construction.do”**

This do-file builds a dataset in which the unit of observation is a pair of firms *jk* in a given year, in which the destination *j* is a shocked BG-firm and the origin *k* is a labor market partner from which the shocked firm may hire workers.

First, the do-file defines a number of global variables that identify the paths that contain both the data and additional do-files to be run.

**Step 0**: each of the files “**DADSPOSTESxxyy\_nm.sas**” contained in the folder “**codes\** **SASFiles\ DADS\EXTRACTION**” extracts the job-to-job flows from year **xx** to year **yy** from the region-year raw BTS POSTES:

“**$home\\$sasfile\EXTRACTION\DadsPostes0205\_nm.sas**”

“**$home\\$sasfile\EXTRACTION\DadsPostes0508\_nm.sas**"

“**$home\\$sasfile\EXTRACTION\DadsPostes0809\_nm.sas**"

“**$home\\$sasfile\EXTRACTION\DadsPostes0910\_nm.sas**"

Only the observations with non missing gross wage, region and ident\_s (worker ID) are selected.

Other cleaning: removal of professional categories strictly related to the public function (cs='45' or cs='33' or cs='52') or unidentified categories (cs='00'); removal of particular firm identifiers (i.e. sirens: the ones starting with letters F, S or P; this correction is particularly effective for years from 2008 onwards)

In order to identify a univocal flow, for each worker it is selected only the best contract, defined as the longest or the one with the highest professional attainment (highest qualification).

* Output: **totflowsyy** (SAS): job to job movements from year *yy-1* to year *yy*.

The file “**DadsLifi0210\_merge\_14\_10\_14.sas**”merges totflowsyy with LIFI in order to identify business groups heads for the affiliated firms in BTS data. In addition, it organizes the data so that each row corresponds to the transition of a given worker and appends the different years:

“**$home\\$sasfile\EXTRACTION\DadsLifi0210\_merge\_14\_10\_14.sas**"

* Output: **jtotj1lineyy** (SAS and STATA) Worker-level dataset of job-to-job movements (starting in year *yy*).

**Step 1**: build a dataset of bilateral worker flows from any firm in the economy to firms operating in shocked sectors

**Substep 1.1**: the SAS file “**00\_dads\_apen\_tot.sas**” extracts a list of firms and the associated sector from DADS

“**$home\\$sasfile\BuildingBilateralFlows\00\_dads\_apen\_tot.sas**"

This sas code extracts the sector from the BTS POSTES for the whole economy, looking both at the forward wave and the lag wave. Ideally, the info extracted looking at wave “f” of time x and at the wave “l” of time x+1 should be the same. Thus, we look at dads x and 1) extract all info on siren\_f and record them as for siren\_f at year x, 2) extract all info on siren\_l and record them as for siren\_l at year x-1.

**Substep 1.2**: the SAS file “**01\_bilateralflows\_noqual.sas**” extracts the subset of firms operating in shocked sectors and merges it with the pre-existing dataset of bilateral job flows (see step 0). Here flows are assumed to be equal to zero (and sector temporarily left missing) for missing pairs when both firms exist in the data (this may happen if existing firms exchange zero workers).

“**$home\\$sasfile\BuildingBilateralFlows\01\_bilateralflows\_noqual.sas**”

**Substep 1.3**: the SAS file “**02\_dads\_apen\_lalista\_first.sas**” extracts sector information for (some zero-flow) cases in which the sector was missing because could not be retrieved from step 1.2 (e.g. when the firm does not appear in the BTS in a given year because it has no worker), using further away lags and leads of the firm

“**$home\\$sasfile\BuildingBilateralFlows\02\_dads\_apen\_lalista\_first.sas**”

**Substep 1.4**: the SAS file “**03\_add\_apen\_to\_bilateralflows.sas**” adds the sector of the firm for the cases described in step 1.3 or for missing pairs (described in step 1.2) for which flows were assumed to be equal to zero and sector temporarily left missing

“**$home\\$sasfile\BuildingBilateralFlows\03\_add\_apen\_to\_bilateralflows.sas**”

* Output: **inflows\_ps\_noqual\_final.dta**

**Substep 1.5**: creates a list of firms, which, at least in one year, appear in a sector that is NOT one of the shocked sectors. For these firms/year observations there are no flows because substep 1.2 only extracts flows for the firm/year observations in which the receiving firm belongs to shocked sectors

* ***(LINE 95)***:

**do "$home\\$dofile\04\_build\_flussi\_recuperare.do"**;

* Output: **flussi\_recuperare.xlsx**

**Substep 1.5.1**: uses the above list created in the previous step to extract the job flows of the subset of problematic firms

“**$home\\$sasfile\BuildingBilateralFlows\05\_recupero\_flussi\5.1\_bilateralflows\_noqual\_recupero.sas**”

**Substep 1.5.2**: same as step 1.3 (only for firms listed in **flussi\_recuperare.xlsx**):

**“$home\\$sasfile\BuildingBilateralFlows\05\_recupero\_flussi\5.2\_dads\_apen\_lalista\_second.sas”**

**Substep 1.5.3**: same as step 1.4 (only for firms listed in **flussi\_recuperare.xlsx**):

**“$home\\$sasfile\BuildingBilateralFlows\05\_recupero\_flussi\5.3\_add\_apen\_to\_bilateralflows.sas”**

**Substep 1.6**:

* ***(LINE 121)***: next, the file “**06\_worker\_flows.do**” builds a complete dataset putting together the results from steps 1.1-1.4 and the results from step 1.5.

**do "$home\\$dofile\06\_worker\_flows.do"**

* output: **worker\_flows.dta**.

**Substep 1.7**:

* ***(LINE 127)***: next, the file “**bilateral\_cflows\_pl\_2cluster.do**” fixes BG status at *t*=-1; eliminates both firms of origin and destination that appear only after the shock; eliminates both firms of origin and destination that appear only before the shock

**Do “$home\\$dofile\ bilateral\_cflows\_pl\_2cluster.do”**

* output: **sampleforestimates\_FIXEDCLEAN.dta**.
  + 1. **Do-file “Codes\DoFiles\3\_shell\_DADS\_BilateralFlowsByQualif\_construction.do”**

This do-file builds a dataset in which the unit of observation, in a given year, is a triplet *jko*, where *j* is a shocked BG-firm and *k* is a labor market partner from which the shocked firm may hire workers of the specific occupational category *o*.

**Step 1**: build a dataset of bilateral worker flows by occupational category from any firm in the economy to firms operating in shocked sectors

**Substep 1.1**: the SAS file “**01\_bilateralflows\_qual.sas**” extracts the subset of firm operating in shocked sectors and merges it with the pre-existing worker-level datasets (**jtotj1lineyy**) of bilateral flows. Again, flows are assumed to be equal to zero (and sector left temporarily missing) for missing pairs when both firms exist in the data (this may happen if existing firms exchange zero workers).

“**$home\\$sasfile\BuildingBilateralFlowsByQualif\01\_bilateralflows\_qual.sas**”

**Substep 1.2**: the SAS file “**03\_add\_apen\_to\_bilateralflows.sas**” adds the sector of the firms with missing sector or for missing pairs for which flows were assumed to be equal to zero and sector left temporarily missing.

“**$home\\$sasfile\BuildingBilateralFlowsByQualif\03\_add\_apen\_to\_bilateralflows.sas**”

* Output: **inflows\_ps\_qualif\_final.dta.** dataset of job-to-job flows by occupational category from any firm in the economy to firms operating in shocked sectors

**Substep 1.3**: uses the list in “**flussi\_recuperare.xlsx**” to extract the job flows of the subset of problematic firms

“**$home\\$sasfile\BuildingBilateralFlowsByQualif\05\_recupero\_flussi\5.1\_bilateralflows\_qual\_f\_recupero.sas**”

**Substep 1.4**: same as step 1.2 (only for firms listed in **flussi\_recuperare.xlsx**):

**“$home\\$sasfile\BuildingBilateralFlowsByQualif\05\_recupero\_flussi\5.3\_add\_apen\_to\_bilateralflows\_f\_recupero.sas”**

**Substep 1.5**:

* ***(LINE 77)***: next, the file “**06\_worker\_flows.do**” builds a complete dataset putting together the results from steps 1.1-1.4 and the results from step 1.5.

**do “$home\\$dofile\BuildingBilateralFlowsbyQualif\06\_worker\_flows.do**”

* Output: **worker\_flows\_byqualif.dta**.

**Substep 1.5**:

* ***(LINE 83)***: next, the file “**bilateral\_cflows\_pl\_2cluster.do**” fixes BG status at *t*=-1; eliminates both firms of origin and destination that appear only after the shock; eliminates both firms of origin and destination that appear only before *the shock*.

**do "$home\\$dofile\BuildingBilateralFlowsbyQualif\bilateral\_cflows\_pl.do**

* Output: **sampleforestimates\_BQ\_FIXEDCLEAN.dta**.
  + 1. **Do-file “Codes\DoFiles\4\_shell\_descriptives.do”**

This do-file runs the sequence of codes that produce

* FIGURE 2
* FIGURE 3
* TABLE 2
* TABLE 3

**FIGURE 2 AND FIGURE 3**:

* ***(LINE 36)***: the file “**groupsinfrance.do**” produces Figures 2 and 3.

**do** "**$home\\$dofile\TABLE1\groupsinfrance.do**"

* Output in **REPLICATIONPACKAGE\Output\Descriptives\figure2**”:
  + **groupsinfrance1.png** (panel a)
  + **empl.png** (panel b)
* Output in “**REPLICATIONPACKAGE\Output\Descriptives\figure3**”:
  + **sirlifi.png** (panel a)
  + **av\_empl.png** (panel b)
  + **contasect.png** (panel c)
  + **hhi.png** (panel d)
  + **contareg.png** (panel e)
  + **hhi\_reg.png** (panel f)

**TABLE 2**

**Step 0**:

The SAS file “**dadsflowscart\_csdeptcorrection.sas**” computes the mean excess probability of within-group firm-to-firm transitions for the (conditional) descriptive evidence on ILM activity. We call these probabilities "gammas" based on the methodology discussed in online Appendix A.2.

"**$home\\$sasfile\GAMMAS\TABLE2\dadsflowscart\_csdeptcorrection.sas**” (conditional on the French Departments in which the firms of origin and destination are located).

**Step 1**

* ***(LINE 57)***: compute group-level diversification measures:

**do "$home\\$dofile\TABLE2\0\_concentration\_dataset.do"**

**Step 2**

* ***(LINE 61)***: merge gammas with LIFI and with group-level diversification measures

**do "$home\\$dofile\TABLE2\1A\_gammalififirmmerge.do"**

**Step 3**

* ***(LINE 65)***: run regressions to produce table 2

**do "$home\\$dofile\TABLE2\1B\_lifi\_gammafirm.do"**

* Output in “**REPLICATIONPACKAGE\Output\Descriptives\TABLE2**”:
  + **Table2.out**

**TABLE 3**

The do-file “**TABLE3.do**” produces the evidence shown in Table 3.

**do "$home\\$dofile\TABLE3\TABLE3.do"**

* Output in “**REPLICATIONPACKAGE\Output\Descriptives\TABLE3**”:
  + **TABLE3.log**
    1. **Do-file “Codes\DoFiles\5\_shell\_event\_study.do”**

This do-file runs the sequence of codes that produce results on:

Firm level outcomes

* FIGURE 4
* FIGURE 5
* FIGURE 6
* FIGURE 7
* FIGURE 8
* FIGURE 9
* FIGURE 10
* FIGURE 13

Bilateral flows

* FIGURE 11
* FIGURE 12
* FIGURE 14

Together with the associated tables reporting the coefficients, standard errors and number of observations.

**FIRM LEVEL OUTCOMES**

**FIGURE 4-7-8-9**:

* ***(LINE 45)***: the file “**FIGURES 4 7 8 9.do**” produces Figures 4, 7, 8 and 9.

**do "$home\\$dofile\FIGURES 4 7 8 9.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 4**”:
  + **empl\_FIXEDBGCLEAN.png** (panel a); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.10)
  + **Dempl0\_FIXEDBGCLEAN.png** (panel b); **Dempl0\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.10)
  + **invcorp\_FIXEDBGCLEAN.png** (panel c); **invcorp\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.10)
  + **salesshare\_FIXEDBGCLEAN.png** (panel d); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.10)
  + **stats.log, line 111-126** (number of observations by distance from the shock reported in Table A.10, column 6. NB: this file is located in folder “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURES 7 8 9**”)
* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\** **FIGURES 7 8 9**”:
  + “**panel a\empl\_FIXEDBGCLEAN.png**” (Figure 7, panel a); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.13)
  + “**panel b\empl\_FIXEDBGCLEAN.png**” (Figure 7, panel b); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.13)
  + “**panel c\empl\_FIXEDBGCLEAN.png**” (Figure 7, panel c); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.13)
  + “**panel d\empl\_FIXEDBGCLEAN.png**” (Figure 7, panel d); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.13)
  + **stats.log, lines 127-211** (number of observations by distance from the shock reported in Table A.13, columns 3-5-8-11-14)
  + “**panel a\invcorp\_FIXEDBGCLEAN.png**” (Figure 8, panel a); **invcorp\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.14)
  + “**panel b\invcorp\_FIXEDBGCLEAN.png**” (Figure 8, panel b); **invcorp\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.14)
  + “**panel c\invcorp\_FIXEDBGCLEAN.png**” (Figure 8, panel c); **invcorp\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.14)
  + “**panel d\invcorp\_FIXEDBGCLEAN.png**” (Figure 8, panel d); **invcorp\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.14)
  + “**panel a\salesshare\_FIXEDBGCLEAN.png**” (Figure 9, panel a); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.15)
  + “**panel b\salesshare\_FIXEDBGCLEAN.png**” (Figure 9, panel b); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.15)
  + “**panel c\salesshare\_FIXEDBGCLEAN.png**” (Figure 9, panel c); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.15)
  + “**panel d\salesshare\_FIXEDBGCLEAN.png**” (Figure 9, panel d); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.15)

**FIGURE 5**:

* ***(LINE 49)***: the file “**FIGURE 5.do**” produces Figure 5.

**do "$home\\$dofile\FIGURES 5.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 5**”:
  + **shareint\_FIXEDBGCLEAN.png**; **shareint\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.10, column 7)
  + **stats.log, lines 31-51** (number of observations by distance from the shock reported in Table A.10, column 8)

**FIGURE 6**:

* ***(LINE 53)***: the file “**FIGURE 6a.do**” produces Figure 6, panel a.

**do "$home\\$dofile\FIGURE 6a.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 6a**”:
  + **share\_int\_by\_ILM.png**; **share\_int\_by\_ILM.dta** (associated coefficients reported in Table A.12, columns 1 and 2)
  + **stats\_share\_int\_by\_ILM.log, line 132-162** (number of observations by distance from the shock reported in Table A.12 in columns next to columns 1 and 2)
* ***(LINE 57)***: the file “**FIGURE 6b.do**” produces Figure 6, panel b.

**do "$home\\$dofile\FIGURE 6b.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 6b**”:
  + **zemp\_nump\_UNC\_allext3W\_FIXEDBGCLEAN.png**; **zemp\_nump\_UNC\_allext3W\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.12, columns 3 and 4)
  + **stats\_nump.log, line 50-64** (number of observations by distance from the shock reported in Table A.12 in columns next to columns 3 and 4)

**FIGURE 10**:

* ***(LINE 61)***: the file “**FIGURES 10a 10b.do**” produces Figure 10, panels a and b.

**do "$home\\$dofile\FIGURES 10a 10b.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 10\Panels a b**”:
  + **empl\_FIXEDBGCLEAN.png** (panel a); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.16, columns 3 and 4)
  + **salesshare\_FIXEDBGCLEAN.png** (panel b); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.16, columns 1 and 2)
* ***(LINE 65)***: the file “**FIGURES 10c 10d.do**” produces Figure 10, panels c and d.

**do "$home\\$dofile\FIGURES 10c 10d.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 10\Panels c d**”:
  + **empl\_FIXEDBGCLEAN\_HILM.png** (panel c); **empl\_FIXEDBGCLEAN\_above.dta** (associated coefficients reported in Table A.17, columns 3 and 4)
  + **salesshare\_FIXEDBGCLEAN\_HILM.png** (panel d); **salesshare\_FIXEDBGCLEAN\_above.dta** (associated coefficients reported in Table A.17, columns 1 and 2)

**FIGURE 13**:

* ***(LINE 69)***: the file “**FIGURE 13.do**” produces Figure 13, panels a and b.

**do "$home\\$dofile\** **FIGURE 13.do"**

* Output in “**REPLICATIONPACKAGE\Output\FirmLevelOutcomes\FIGURE 13**”:
  + **empl\_FIXEDBGCLEAN.png** (panel a); **empl\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.23, columns 3 and 4)
  + **salesshare\_FIXEDBGCLEAN.png** (panel b); **salesshare\_FIXEDBGCLEAN.dta** (associated coefficients reported in Table A.23, columns 1 and 2)

**BILATERAL FLOWS**

**FIGURE 11**:

* ***(LINE 89)***: the file “**FIGURE 11a.do**” produces Figure 11, panel a.

**do "$home\\$dofile\** **FIGURE 11a.do"**

* Output in “**REPLICATIONPACKAGE\Output\BilateralFlows\figure 11a**”:
  + **“figure 11a.png**”(panel a); “**figures for table A20.dta**”(associated coefficients reported in Table A.20, columns 3 and 4)
  + “**NumobsByDistance for Table A20.log**”**, lines 9-22** (number of observations by distance from the shock reported in Table A.20 in columns next to columns 1 and 2)
* ***(LINE 94)***: the file “**FIGURE 11b.do**” produces Figure 11, panel b.

**do "$home\\$dofile\** **FIGURE 11b.do"**

* Output in “**REPLICATIONPACKAGE\Output\BilateralFlows\figure 11b**”:
  + **“figure 11b.png**”(panel b); “**figures for table A21 cols 1-4.dta**”(associated coefficients reported in Table A.21, columns 1-4)
* ***(LINE 99)***: the file “**TABLE A21 (COLUMNS 5-8).do**” completes Table A.21.

**do "$home\\$dofile\** **TABLE A21 (COLUMNS 5-8).do"**

* Output in “**REPLICATIONPACKAGE\Output\BilateralFlows\TABLE A21 (COLUMNS 5-8)**”:
  + “**figures for table A21 cols 5-8.dta**”(coefficients reported in Table A.21, columns 5-8)

**FIGURE 12**:

* ***(LINE 104)***: the file “**FIGURE 12.do**” produces Figure 12.

**do "$home\\$dofile\FIGURE 12.do"**

* Output in “**REPLICATIONPACKAGE\Output\BilateralFlows\FIGURE 12**”:
  + **“figure12a.png**”(panel a); “**figure12a.dta**”(associated coefficients reported in Table A.22, columns 1-2)
  + **“figure12b.png**”(panel b); “**figure12b.dta**”(associated coefficients reported in Table A.22, columns 3-4)

**FIGURE 14**:

* ***(LINE 109)***: the file “**FIGURE 14.do**” produces Figure 14.

**do "$home\\$dofile\FIGURE 14.do"**

* Output in “**REPLICATIONPACKAGE\Output\BilateralFlows\FIGURE 14**”:
  + “**figure14a\_bluecollars.png**”(panel a);
  + “**figure14b\_whitecollars.png**”(panel b);
  + “**figure14c\_intprof. png**”(panel c);
  + “**figure14d\_mgmt.png**” (panel d);
  + “**figure14.dta**” (associated coefficients of the 4 panels reported in Table A.24)
    1. **Do-file “Codes\DoFiles\6\_shell\_gammas.do”**

This do-file runs the sequence of codes that produce Table 1.

**TABLE 1**

**Step 0**:

The following SAS files compute the mean excess probability of within-group firm-to-firm transitions for the (conditional) descriptive evidence on ILM activity. We call these probabilities "gammas" based on the methodology discussed in online Appendix A.2.

"**$home\\$sasfile\** **dadsflowscart\_uncon.sas**” (unconditional)

"**$home\\$sasfile\dadsflowscart\_cs\_cat.sas**” (conditional on the occupation of the worker in the firms of origin and destination)

"**$home\\$sasfile\dadsflowscart\_ze.sas**” (conditional on the local labour market where the firms of origin and destination are located)

"**$home\\$sasfile\dadsflowscart\_csze.sas**” (conditional both on the occupation of the worker in the firms of origin and destination, and on the local labour market where the firms of origin and destination are located)

**Step 1**:

The next SAS file extracts the set of firms for which the gammas conditional on occupation pairs and origin/destination local labour market are computed; this is the “common sample” for final gammas computed below

"**$home\$sasfile\extract\_siren\_sets.sas**"

**Step 2**:

The SAS file below compute the gammas on the common sample

"**$home\\$sasfile\dadsflowscart\_uncon\_cs.sas**” (unconditional)

"**$home\\$sasfile\dadsflowscart\_occ\_cs.sas**” (conditional on the occupation of the worker in the firms of origin and destination)

"**$home\\$sasfile\dadsflowscart\_ze\_cs.sas**” (conditional on the local labour market where the firms of origin and destination are located)

"**$home\\$sasfile\dadsflowscart\_occze\_cs.sas**” (conditional both on the occupation of the worker in the firms of origin and destination, and on the local labour market where the firms of origin and destination are located)

"$**home\$sasfile\yearappend\_rev2.sas**" (append years)

**Step 3**:

* ***(LINE 78)***: compute group-level diversification measures:

**do "$home\\$dofile\gamma\_TABLE1.do"**

* Output in “**REPLICATIONPACKAGE\Output\Descriptives\TABLE1**”:
  + **Table1.txt**;(Table 1)
  + **apptableA2A3.tex**; Tables A2 and A3 in online appendix)

**Data References**

* INSEE (Institut national de la statistique et des études économiques)[producer]. 2010. "*Enquête LIFI : Enquête Liaisons financières entre sociétés, 2002-2010*." CASD (Centre d'accès sécurisé aux donnés) [distributor].  <https://www.casd.eu/en/source/financial-links-between-enterprises-survey/>  DOI listed in the Data Appendix.
* INSEE (Institut national de la statistique et des études économiques)[producer]. 2010. "*BTS - Postes: Base Tous Salariés: fichier Postes, 2002-2010*." CASD (Centre d'accès sécurisé aux donnés) [distributor].  <https://www.casd.eu/source/base-tous-salaries-fichier-postes/>.  DOI listed in the Data Appendix.
* INSEE (Institut national de la statistique et des études économiques) & DGFiP (Ministère des Finances) [producer]. 2010. "*FARE: Statistique structurelle annuelle d'entreprises issue du dispositif ESANE, 2008-2010*." CASD (Centre d'accès sécurisé aux donnés) [distributor].  <https://www.casd.eu/source/statistique-structurelle-annuelle-dentreprises-issue-du-dispositif-esane/>.  DOI listed in the Data Appendix.
* INSEE (Institut national de la statistique et des études économiques) & DGFiP (Ministère des Finances)[producer]. 2007. "*FICUS: Statistique structurelle annuelle d'entreprises issue du dispositif SUSE, 2002-2007*." CASD (Centre d'accès sécurisé aux donnés) [distributor]. <https://www.casd.eu/source/statistique-structurelle-annuelle-dentreprises-issue-du-dispositif-suse/>.  DOI listed in the Data Appendix.

**Data Appendix**

**LIFI:**

2010 : <https://doi.org/10.34724/CASD.272.164.V1>

2009 : <https://doi.org/10.34724/CASD.272.150.V1>

2008 : <https://doi.org/10.34724/CASD.272.149.V1>

2007 : <https://doi.org/10.34724/CASD.272.148.V1>

2006 : <https://doi.org/10.34724/CASD.272.147.V1>

2005 : <https://doi.org/10.34724/CASD.272.146.V1>

2004 : <https://doi.org/10.34724/CASD.272.145.V1>

2003 : <https://doi.org/10.34724/CASD.272.144.V1>

2002 : <https://doi.org/10.34724/CASD.272.143.V1>

**BTS Postes:**

2010 : <https://doi.org/10.34724/CASD.21.97.V1>

2009 : <https://doi.org/10.34724/CASD.21.54.V1>

2008 : <https://doi.org/10.34724/CASD.21.53.V1>

2007 : <https://doi.org/10.34724/CASD.21.52.V1>

2006 : <https://doi.org/10.34724/CASD.21.51.V1>

2005 : <https://doi.org/10.34724/CASD.21.50.V1>

2004 : <https://doi.org/10.34724/CASD.21.49.V1>

2003 : <https://doi.org/10.34724/CASD.21.48.V1>

2002 : <https://doi.org/10.34724/CASD.21.47.V1>

**FARE**

2010 : <https://doi.org/10.34724/CASD.42.124.V1>

2009 : <https://doi.org/10.34724/CASD.42.552.V1>

2008 : <https://doi.org/10.34724/CASD.42.551.V1>

**FICUS**

2007 : <https://doi.org/10.34724/CASD.68.557.V1>

2006 : <https://doi.org/10.34724/CASD.68.556.V1>

2005 : <https://doi.org/10.34724/CASD.68.555.V1>

2004 : <https://doi.org/10.34724/CASD.68.554.V1>

2003 : <https://doi.org/10.34724/CASD.68.553.V1>

2002 : <https://doi.org/10.34724/CASD.68.779.V1>