

Replication codes for Bisztray, Márta, Miklós Koren and Adam Szeidl: “Learning to Import from Your Peers.” Journal of International Economics. Forthcoming.

Please cite the above paper when using any of these programs.

The codes creating the datasets used for the estimations and preparing the exhibits of the paper use Stata 14 with additional modules `distinct`, `esttab`, `outreg2` and `tabout`.

`master.do` runs all the other codes, preparing the estimation data from raw data and creating all the tables, figures and other calculations presented in the paper.

At the beginning of `master.do` one can set the access route to the input folders with the raw data and to the output folders for the constructed data and for the exhibits.

Once you have the same folder structure as here, with all the necessary datasets in the `data` folder and all the codes as provided in the `code` folder, just use Stata to run the master file from the code folder:

```
cd code
do master.do
```

The sub-folder `data/input` and the folder `results` are placeholders for the necessary inputs and outputs of these scripts. Input data is read from `data` (please see the note on data access below), output data is saved under `data/input`, and exhibits of the paper are saved under `results`.

Workflow

1. Prepare the data from the primary input files. This is done by do-files in `do/prepare_data`.
2. Create list of peers and peer-level statistics. This is done by do-files in `do/define_peers`.
3. Create the exhibits. Create all the tables, figures and calculations presented in the text with `analysis/exhibits.do`.

The detailed workflow is described in `master.do`.

Data used

Our estimates are based on a panel dataset of import and export transactions, balance sheets and earnings statements of Hungarian firms for the period 1992

to 2003, and firm registry data from the same period. Because the data is confidential, we cannot make it available in this replication package.

Researchers interested in replicating our results with this same data, or conducting other academic research on this data, can access the dataset and the necessary data processing scripts at the premises of the Institute of Economics of the Hungarian Academy of Sciences and at Central European University. Please refer to <http://www.mtaki.hu/english/> and <https://economics.ceu.edu/> or contact the corresponding author, Márta Bisztray at bisztray dot marta at krtk dot mta dot hu.

Primary data source

balance_sheet.dta

- *source*: National Tax and Customs Administration of Hungary and Complex firm registry data
- *content*: balance sheet statements of all Hungarian double-entry bookkeeping firms
- *time period*: 1992-2012
- *variables*:
 - **originalid**: firm identifier
 - **year**
 - **emp**: number of employees
 - **sales**: total sales in 1000HUF
 - **export**: export sales in 1000HUF
 - **gdp**: gross value added in 1000HUF (sales + capitalized value of self-manufactured assets - material)
 - **jetok**: total capital in 1000HUF
 - **jetok06**: foreign-owned capital in 1000HUF
 - **ranyag**: material expenditure in 1000HUF
 - **tear03_2d**: 2-digit tear'03 industry code (corresponds to NACE Rev 1.1)
 - **tear_raw**: 4-digit raw industry codes (tear92 in years 1992-1997, tear98 in years 1998-2002, tear03 in years 2003-2007, tear08 in years 2008-2012)
 - **fo2**: dummy for foreign-owned share $\geq 50\%$
 - **so2**: dummy for state-owned share $\geq 50\%$
 - **foundyear**: foundation year
 - **persexp**: payments to personnel in 1000HUF
 - **wbill**: wage bill in 1000HUF
 - **tanass**: tangible assets in 1000HUF

county-nuts-city-ksh-codes.csv

- *source*: Hungarian Central Statistical Office

- *content*: city names with statistical codes, counties and NUTS3 regions
- *variables*:
 - **ksh_code**: statistical code of a settlement
 - **nuts3**: NUTS3 region

deflators.dta

- *source*: Hungarian Central Statistical Office
- *content*: GDP deflator, capital deflator, wage index
- *time period*: 1992-2013
- *variables*:
 - **year**
 - **K_deflator**: capital deflator
 - **wage_index**: wage index

ex91.dta - ex03.dta

- *source*: Hungarian Customs Statistics
- *content*: yearly value of export transactions by firm-country-product category
- *time period*: single year from 1991 to 2003
- *variables*:
 - **e_ft91/92/.../03**: yearly export value in HUF
 - **a1**: firm identifier
 - **szao**: destination country
 - **hs6**: 6-digit HS product category

im91.dta - im03.dta

- *source*: Hungarian Customs Statistics
- *content*: yearly value of import transactions by firm-country-product category
- *time period*: single year from 1991 to 2003
- *variables*:
 - **i_ft91/92/.../03**: yearly import value in HUF
 - **a1**: firm identifier
 - **szao**: source country
 - **hs6**: 6-digit HS product category

PPI.dta

- *source*: Hungarian Central Statistical Office
- *content*: material price index and producer price index by 2-digit industry
- PPI also separately for exports and domestic sales
- *time period*: 1992-2013

- *variables*:
 - **year**
 - **nace2**: 2-digit industry category (NACE Rev. 1.1)
 - **PPI**: producer price index
 - **ex_PPI**: producer price index for export sales
 - **dom_PPI**: produced price index for domestic sales
 - **materialPPI**: material price index

Classifications and correspondance tables

hs6bec.csv

- *content*: correspondence table between HS6 and BEC product categories
- *variables*:
 - **hs6**: 6-digit HS product category
 - **bec**: 2-digit bec product category

unique_nace11_to_teaor08.csv

- *content*: created from NACE Rev 1.1 - NACE Rev 2 correspondence table
- *variables*:
 - **nace**: 4-digit NACE Rev. 1.1
 - **teaor08**: 4-digit TEAOR'08 (corresponds to NACE Rev.2)

Data built from primary sources

1992-2006.csv

- *source*: Complex firm registry data
- *content*: firm-year panel of headquarters (city, street, building)
- *time period*: 1992-2006
- *variables*:
 - **taxid**: firm identifier
 - **year**
 - **cityid**: city name identifier
 - **streetid**: street name identifier
 - **buildingid**: building number

bp_addresses_w_geocoord.dta

- *source*: Complex firm registry data
- *content*: geocoordinates assigned to addresses in Budapest
- *variables*:
 - **cityid**: city identifier
 - **streetid**: street identifier

- **buildingid**: building number
- **lat**: latitude
- **lon**: longitude

country_code.dta

- *source*: Complex firm registry data
- *content*: yearly panel of country codes the firm has owners from
- *time period*: 1992-2003
- *variables*:
 - **tax_id**: firm identifier
 - **year**
 - **country_code**: 2-digit ISO country code

frame.csv

- *source*: Complex firm registry data
- *content*: firms with birth and death dates
- *time period*: 1991-2011 (including earlier birth dates)
- *variables*:
 - **tax_id**: firm identifier
 - **birth_date**: date of incorporation as an 8-digit integer, using the order year month day
 - **death_date**: date of ending operation as an 8-digit integer, using the order year month day

frame_old_format.csv

- *source*: Complex firm registry data
- *content*: firms with birth and death dates
- *time period*: 1991-2014 (including earlier birth dates)
- *variables*:
 - **tax_id**: firm identifier
 - **birth_date**: date of incorporation as an 8-digit integer, using the order year month day
 - **death_date**: date of ending operation as an 8-digit integer, using the order year month day

firm_hqs_yearly_slices.csv

- *source*: Complex firm registry data
- *content*: yearly detailed headquarter addresses of firms (up to floor and door)
- *time period*: 1990-2014
- *variables*:

- **taxid**: firm identifier
- **year**
- **cityid**: city identifier

firmperson_important_liquidator_dummy.csv

- *source*: Complex firm registry data
- *content*: list of firms with connecting people and an indicator for liquidators
 - **tax_id**: firm identifier person_id
 - **felsz**: dummy, 1 if person is ever a liquidator in the firm

location.dta

- *source*: Complex firm registry data
- *content*: firm-year panel of headquarters (city, street, building)
- *time period*: 1992-2006
- *variables*:
 - **tax_id**: firm identifier
 - **year**
 - **cityid**: city identifier
 - **streetid**: street identifier
 - **buildingid**: building number

ownership_mixed_directed.csv

- *source*: Complex firm registry data
- *content*: list of person and firm owners of firms with start and end dates
- *time period*: 1991-2013
- *variables*:
 - **tax_id**: firm identifier
 - **tax_id_owner**: identifier of the owner firm
 - **person_id_owner**: identifier of the owner person
 - **start**: start date of the ownership, string, using the format “yyyy-mm-dd”
 - **end**: end date of the ownership, string, using the format “yyyy-mm-dd”

person_firm_connection_in_all_rovats_using_30d_threshold.csv

- *source*: Complex firm registry data
- *content*: list of firms with connected people having any connection, with start and end dates
- *time period*: 1991-2013
- *variables*:

- **tax_id**: firm identifier
- **person_id**: identifier of the person
- **date_of_start**: start date of the ownership, string, using the format “yyyy-mm-dd”
- **date_of_end**: end date of the ownership, string, using the format “yyyy-mm-dd”

person_firm_connection_by_rovats_using_30d_threshold.csv

- *source*: Complex firm registry data
- *content*: list of firms with connected people, and type of the connection with start and end dates
- *time period*: 1991-2013
- *variables*:
 - **tax_id**: firm identifier
 - **person_id**: identifier of the person
 - **rovat_name**: name of the rovat indicating the type of the relationship, string
 - **date_of_start**: start date of the ownership, string, using the format “yyyy-mm-dd”
 - **date_of_end**: end date of the ownership, string, using the format “yyyy-mm-dd”