Replication Data for
"On the Estimation of Cross-Firm Productivity Spillovers with an Application to FDI"
by Emir Malikov & Shunan Zhao, Review of Economics and Statistics, 2021

---

* Programs and OS

All analysis was carried out using RStudio version 1.4.1106 (R version 4.0.3) on a Windows 10 machine

* Data

The "*Ind39.dta*" file is in the Stata format and contains data used in the analysis which come from the Chinese Industrial Enterprises Database survey conducted by China's National Bureau of Statistics. The variable definitions are as follows (also see the labels of variables in the data file).

Variable	Definition
Firm	firm id
Year	year
city	firm location: city
Province	firm location: province
sic2	2-digit level industry code
sic4	4-digit industry code
F	foreign equity share
y	log output
k	log capital
l	log labor
m	log material
sm	material-output ratio
ppi	producer price index
ppii	price index for intermediate inputs
pifa	price index for fixed assets
gdpd	GDP deflator
age	firm age
TotalAssets	total assets
TotalLiab	total liability
TotalProfit	profit
Lev	leverage
ROA	return on assets
Employment	labor employed
SOE	state share of equity
Export	export value
ExpR	export intensity (export share of output)
Subsidy	subsidies received
coast	coastal province dummy
East	East Coast province dummy
ExY	ratio of the sum of imports and exports to GDP at the province level
T	time trend

The data file also includes the lagged variables. The naming convention is that, for any variable **x**, its first lag is denoted as **x1**, the second lag is denoted as **x2**, etc. For example, the variables **F1** and **F2** are the first and second lags of the foreign equity share variable **F**, respectively.

F1	first lag of F
F2	first lag of F1 (second lag of F)
dF	first difference of F

k1	first lag of k
l1	first lag of l
m1	first lag of m
ppi1	first lag of ppi
ppii1	first lag of ppii
ExY1	first lag of ExY

The other four files (*bou2_4p.dbf*, *bou2_4p.shp*, *bou2_4p.shx* and *capital location.csv*) are auxiliary and contain spatial information that some code/program files call for when generating maps.

* Code/program files

There are forty-two R code/program files. Each file is named in accordance with a table/figure (including those in the online appendix) that it replicates. For example, the file "*Table H4 Figure H1 H2.R*" produces Table H4 and Figures H1 and H2 in the appendix. For some tables with bootstrap confidence intervals, such as Table 6, bootstrap and jackknife code files should be run first before the final tabular code file.