Readme file for:

"Knocking on Tax Haven's Door: Multinational Firms and Transfer Pricing"
By Ron Davies, Julien Martin, Mathieu Parenti, and Farid Toubal
Review of Economics and Statistics (accepted for publication in Dec. 2016)

The archive contains the do-files necessary for replicating the figures and tables of the paper and the online appendix.

Do-files: 29 do-files are used for the construction of the data and the replication of the results.

The file master_dmpt.do is the main do-file that details all steps needed, with corresponding intermediary do-files.

To conduct the empirical analysis, we build a dataset using these four do-files:

- 01_TradeDataset.do
- 02_CountryInfo.do
- 03_FinalData
- 04_DataMgmt.do

Then, there are 23 do-files replicating the 20 tables and the 3 figures of the paper (including those in the online appendix). The names of these do-files are simply "table*#ofthefigure*.do" or "figure*#ofthefigure*.do" as in the manuscript.

An additional do-file (figure.do) is used as an auxiliary do-file to generate some figures.

These do-files have been run on Stata MP 13.1 and 14.0.

Data:

The archive file also contains country-level and sector-level data used in the paper.

Access to the raw firm-level data is restricted. To obtain the data, one must apply for access through the committee in charge of the diffusion of French administrative data (the "CNIS"). The documentation (in French) can be found at

http://www.cnis.fr/cms/Accueil/activites/_trois_comites/Comite_du_secret_statistique/Les_procedures_dacces_aux_donnees_confidentielles

Researchers who would like to access the data should follow this procedure:

- Download and fill in the application form (*Access to confidential data*)
- Contact the data producers (See this list of contacts):
  - M. Patrice Dumont (DGFIP) for the customs data
  - M. Fabienne Rosenwald for the LIFI dataset (Insee)
  - M. Beatrice Maubras (INSEE) for INSEE the firm-level survey on the foreign activities of French multinational firms (called : Enquête Mondialisation)
- Send the form via email to the "Secrétariat du Comité du secret statistique" : comite-secret@cnis.fr

All the datasets and do files used for the paper are stored on a CASD server for replication purposes (https://casd.eu/en/home). Replication requires getting access to the secure data center after through procedure described above.

Alternatively, a request can be sent to the corresponding author for the program provided (or modifications of it) to be run and output to be sent back to the researcher asking for it. All generated results are subject to disclosure control and may be used in publications following their approval.

We commit to respond quickly to future requests.

Here is the list of datasets used in the paper:

	Source	**Description**
Individual data		
cadre-b-c.dta	EIIG INSEE	Intra-firm exports
douanes_1999_12.txt	French Customs	French exports 1999
Lifi	LIFI INSEE	Ownership status
country_hq	Own computation based on LIFI	Country head-quarter
country_affiliates	Own computation based on LIFI	Countries of affiliates
Country data		
updates_loretz_data_with_Haven	Loretz (2008)	EATR and EMTR
dist_cepii	CEPII	Distances
population_worldbank	World Bank	Population
gdp_worldbank	World Bank	GDP
iso_pays	Own computation	Correspondance iso2-iso3
cname_pays	CEPII	Name of countries
tax_rates	OECD	Corporate tax rates
FSI2011	Tax Justice Network	Financial Secrecy Index
dum_tax	Own computation	Tax haven dummy
DataJobID-351903_351903_france1999	TRAINS	Tariffs
geo_cepii	CEPII	Country info
ER1999	World Bank	Exchange rate Franc /ROW
rta_FRA99	CEPII	Dummy Regional Trade Agreement
ppp1999	World Bank	Price level ratio
market_potential	CEPII	Mayer's measure of market potential
gini	World Bank	Gini index
remot	Own computation	Geographic remotness of countries
Sector data		
hs1996ToIsicRev2	UNSTAT	Correspondence HS6-96 to ISIC rev2
rauch	Rauch (1999)	Product differentiation
contract_intensity_ISIC_1997	Nunn (2007)	Input specificity
ladder_isic	Khandelwal (2010)	Quality ladder