prepd.sh

bulk_download.R

Bulk download annual trade data from Comtrade. Remove redundant data. Update download info.

%-comtrade/download.log

bulk_download.log

%-comtrade/#Y.csv.bz2

HK_#Y_rx.csv.bz2, Swiss_mx_71-72.csv

Updated Years

<= 2

> 2

mirror_match-node.R

mirror_match.R

Preprocessing, pair import and export declarations; separate matched, orphaned, lost

%-mirror-analysis/@-#Y-M|X-junked-HK-adj.csv.bz2

%-mirror-analysis/mirror_match.csv, mirror_match.log

%-mirror-analysis/@-#Y-M_orphaned|lost.csv.bz2

%-mirror-analysis/@-#Y-M_matched.csv.bz2

pub_note.R

Download and clean Comtrade publication note; extract exchange rates and fob info.

%-comtrade/mfob.csv

%-comtrade/exr.csv

Using global data to build the model that estimates the gap between CIF and FOB.

input_adj.R

CEPII_GeoDist.csv, EIA.csv.bz2

bridge.csv

Subset matched data and add attributes for building and applying the model.

cifob_model.R

/efs/unct/@-#Y-input.csv.bz2

mol_cty_agg.R

%-work/cifob_model.rds.bz2

./data/trains.csv, wto.csv

Clean tariff rate data from WITS.

trf_effective.R

Identify import & export, aggregate commodities. input: k=2, all_trade=T&F, cty, years=years.

# of Countries

Applying the model to country data and calculating weights to calculate TM. Input: years=all

cty=null

cty <55

gaps+tm. Input: cty=cty, k=6

/efs/work/mm|o|l_m|x_k2T|F.csv.bz2

gaps.R

%-work/summ_prd.csv, gaps.log

%-work/@-Y#-gaps.csv.bz2

gaps_tm.R

/efs/work/trf.csv

Identify flows, aggregate commodity. input: cty, all_trade=T&F, years=years.

tm_cty_agg.R

k=6

/efs/work/tm_k6T|F.csv.bz2

k=2

/efs/work/tm_k2T|F.csv.bz2

revenue_loss.R

add tariff, aggregate commodities. no change of k.

./data/rl#cty.csv

Legend:

Data (Stored on S3)

Program (single computing unit)

Data (Stored locally or on EFS)

Program (multiple computing units)

By-product Outputs

Another Process

IFFs worldwide reports

IFFs national reports

implicit inputs

#: a number, #Y: year
%: gfi, @: Comtrade, a|b: a or b
./: folder of the program
files in ~/UN-Comtrade/ , files in ~/WB-WITS/

1. Every cloud-side program produces a log file of the same name as itself.

2. To launch multiple computing units, run 'spark-master.sh' in the master node and 'spark-worker.sh' in the worker nodes. 3. For single node program, attach swap memory and run 'spacious.sh'.