A Global Foreign Direct Investment Stocks Database

E. Bekkers; C. Lakatos; R. So; W. J. Yuan; F. Traverso

World Trade Organization

June 24, 2021

Background

Our project is an update of [Gouel et al., 2012] from CEPII:

- Fully balanced, three-dimensional (bilateral & by sector) FDI stock and flow database
- Created for the GTAP-FDI module
- Consistent with GTAP Version 7 – 113 regions, 57 sectors, baseline year of 2004[1]

[1] Regional aggregation and sectoral disaggregation were performed by [Lakatos and Walmsley, 2011]

Motivation

The original FDI database was adapted into later GTAP versions by assuming a proportionate growth of FDI to the host country's GDP

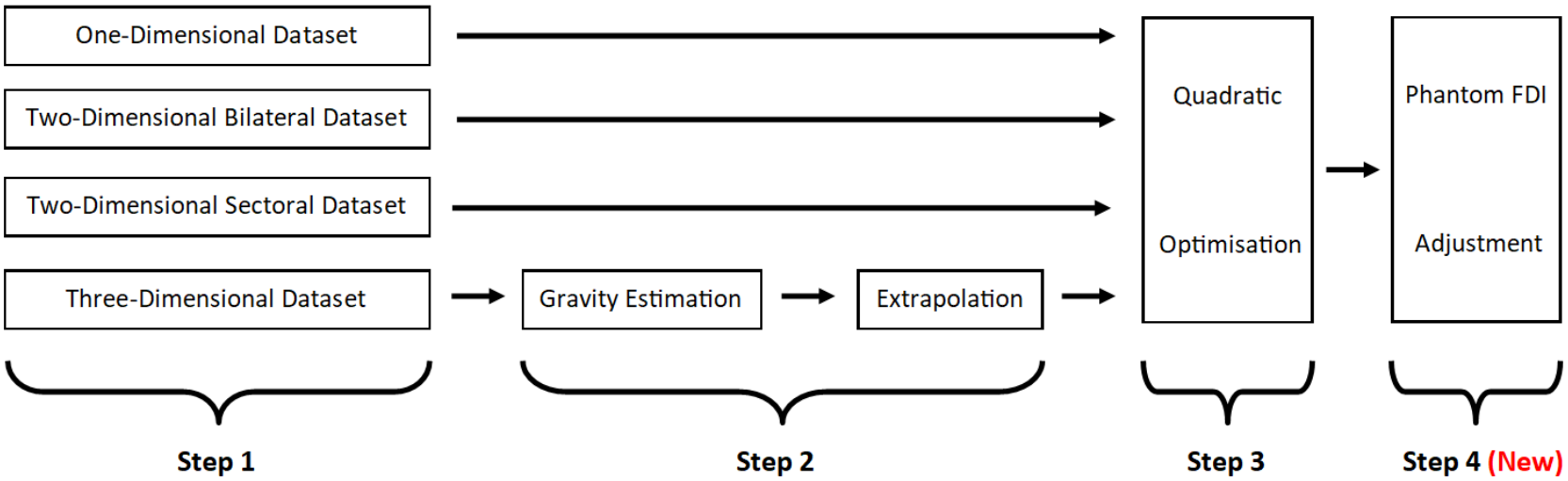
Reasons for update:

- Rise of North-South and South-South FDI
- Expanded list of regions and new sector concordance in GTAP
- Better quality and greater availability of raw FDI data
- Synergy with existing work on expanding the GTAP-FDI module and the update on Foreign Affliate Sales (FAS) database

Note that we are only updating the FDI stock database

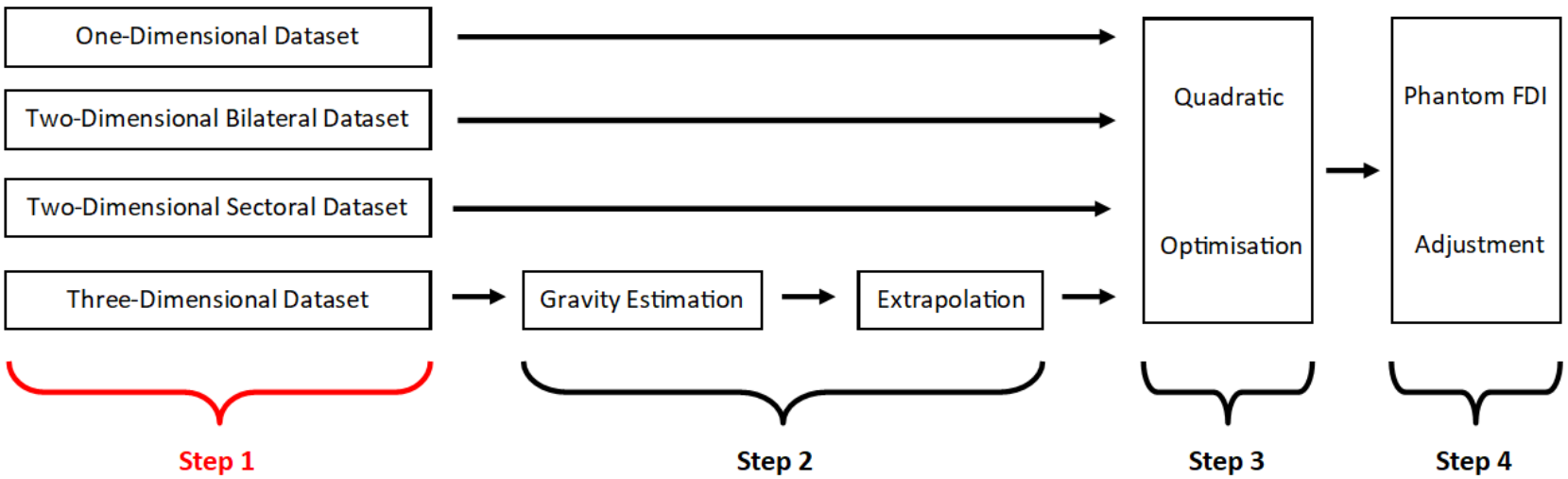
Overview

An overview of the database construction procedures:



Output: FDI Stock databases that are fully consistent with GTAP Version 10 and 11 (baseline year of 2014 & 2017)

Overview


One-Dimensional Dataset
Two-Dimensional Bilateral Dataset
Two-Dimensional Sectoral Dataset
Three-Dimensional Dataset
Gravity Estimation
Extrapolation
Quadratic
Optimisation
Phantom FDI
Adjustment
Step 1
Step 2
Step 3
Step 4

Introduction

Raw FDI stock data from different sources are available across multiple dimensions:

- One-dimensional Dataset – total FDI stock from World
- Two-dimensional Bilateral Dataset – FDI stock with partners, no sector breakdown
- Two-dimensional Sectoral dataset – total FDI with World, with sector breakdown
- Three-dimensional dataset – FDI stock with partners, and by sector

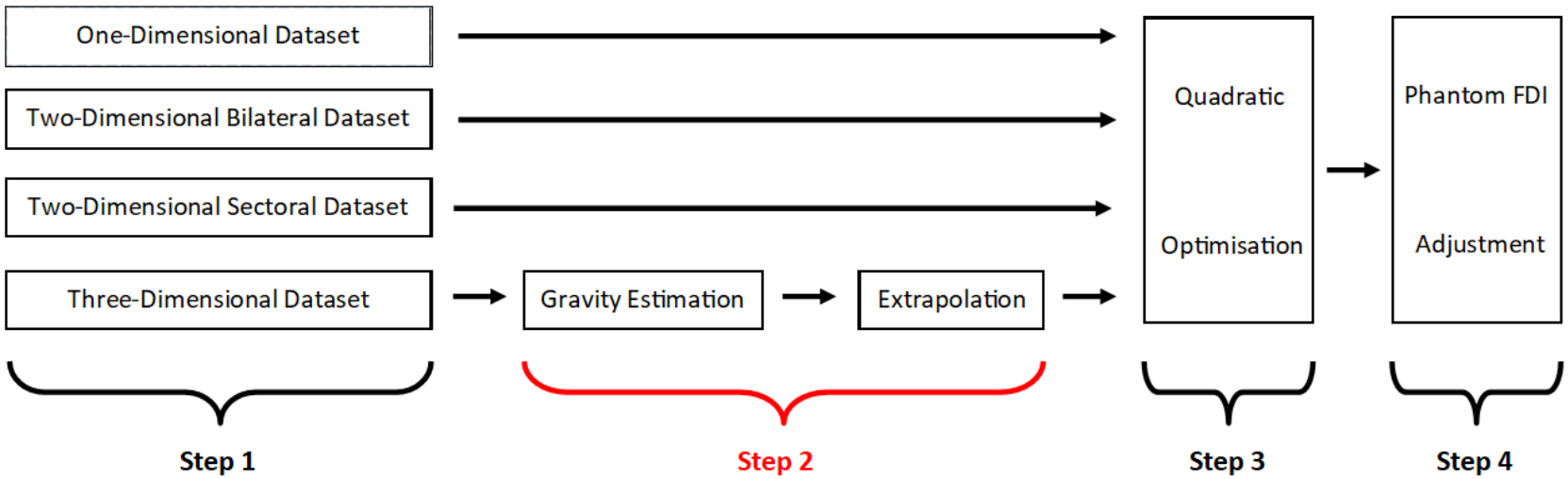
Summary Table

Dataset	Data Sources	Outward Values	Year
1D	IMF, UNCTAD	Discarded	2014, 2017
2D Bilateral	IMF	Mirroring	2014, 2017
2D Sectoral	OECD, Eurostat, ITC	Discarded	2014, 2017
3D	Eurostat, ITC	Mirroring	2013 – 2018

- Only inward values are collected for the 1D and 2D Sectoral dataset
- In the 2D Bilateral and 3D dataset, inward FDI stocks are preferred over mirrored values

Dataset	Country Coverage	Sectoral Coverage
1D	180	N/A
2D Bilateral	120 + 125	34
2D Sectoral	63	N/A
3D	50 + 164	34

Overview


One-Dimensional Dataset
Two-Dimensional Bilateral Dataset
Two-Dimensional Sectoral Dataset
Three-Dimensional Dataset
Gravity Estimation
Extrapolation
Quadratic
Optimisation
Phantom FDI
Adjustment
Step 1
Step 2
Step 3
Step 4

Methodology:

- Pseudo Poisson Maximum Likelihood (PPML) estimator to account for zero observations and heteroskedasticity
- Regressions are run separately for each of the 34 sectors
- Year fixed effects only – does not control for multilateral resistance terms

Model Specification:

$$\begin{aligned} FDI_{ijt,k} = \exp(&\beta_1 + \beta_2 \ln(GDP_{ti}) + \beta_3 \ln(GDP_{tj}) + \beta_4 \ln(GDPpc_{ti}) \\ &+ \beta_5 \ln(GDPpc_{tj}) + \beta_6 \ln(DIST_{ij}) + \beta_7 COMLG_{ij} + \beta_8 COL_{ij} \\ &+ \beta_9 DEV_i + \beta_{10} DEV_j + \beta_{11} LUX_{ij} + \beta_{12} NLD_{ij} + \beta_{13} HKG_{ij} \\ &+ \eta_{t,k} + \epsilon_{ijt,k}) \end{aligned}$$

Estimation Results

VARIABLES	(1) Original†	(2) Replication	(3) Country Dummies
ln_DIST	-0.756***	-0.588***	-0.428***
	(0.01)	(0.0411)	(0.0468)
ln_GDP_host	0.858***	0.531***	0.673***
	(0.009)	(0.0270)	(0.0275)
ln_GDP_source	0.775***	0.615***	0.827***
	(0.009)	(0.0435)	(0.0435)
ln_GDPpc_host	2.417***	1.195***	0.611***
	(0.061)	(0.0892)	(0.0877)
ln_GDPpc_source	2.995***	1.718***	1.211***
	(0.061)	(0.132)	(0.109)
Developing Country_host	1.558***	1.282***	0.887***
	(0.086)	(0.122)	(0.139)
Developing Country_source	-0.345	1.533***	1.255***
	(0.215)	(0.223)	(0.239)
Common Language	0.683***	0.0569	0.694***
	(0.029)	(0.130)	(0.0715)
Colonial Relationship	0.136***	0.744***	0.0793
	(0.033)	(0.181)	(0.109)
LUX Dummy	-	-	2.467***
			(0.121)
NLD Dummy	-	-	2.074***
			(0.0872)
HKG Dummy	-	-	2.272***
			(0.259)
Observations	68,230	307,621	307,621
Year F.E.	Yes	Yes	Yes
Sector F.E.	Yes	Yes	Yes
Pseudo R-squared	-	0.573	0.688

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

† Source: Directly extracted from Gouel et al. (2012)

- Replication of [Gouel et al., 2012] yielded similar results
- Improved goodness of fit (R-squared) following the introduction of 3 country dummies
- Results remain robust under sector-by-sector estimations

Out-of-sample validation:

- 70% of sample as trainning set, 30% as testing set
- Gravity estimation is run on the training set and extrapolate to the testing set to acquire mean absolute error (MAE)
- Repeat 50 times

Model Specification	Average of MAE (mil. USD)
Original (pooled estimation)	357.43
Original + Country Dummies	310.74
Sector-by-Sector + Country Dummies	273.71

Robustness Checks:

- Extra Variables (CEPII, treaties variables, policy indices)

Extrapolation

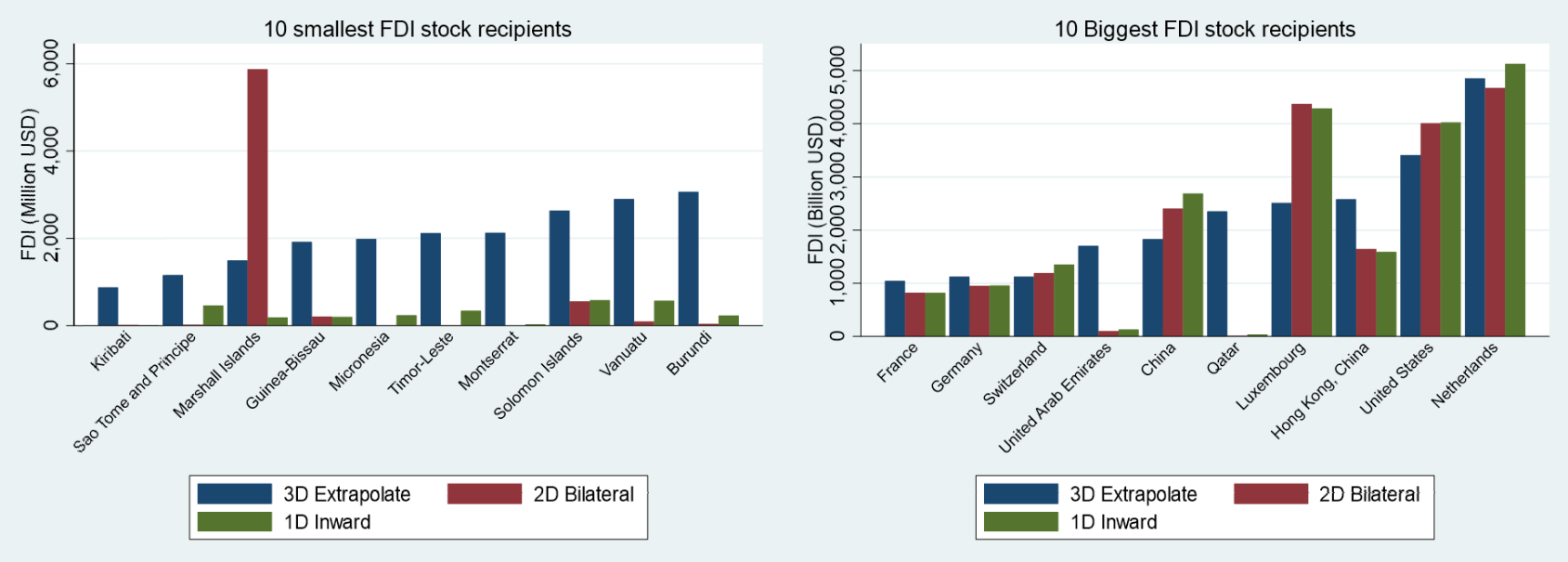
Missing values are extrapolated using coefficients acquired from gravity estimation

- Dataset dimension: $204 \times 203 \times 34$ for each year
- 2.8 million total observations, only 110k (4%) are reported values from the three-dimensional dataset

	1-D Inward	2-D Bilateral	3-D Reported	3-D Extrapolate
2014	30.1	30.8	12.5	38.1
2017	36.9	37.4	16.1	42.3

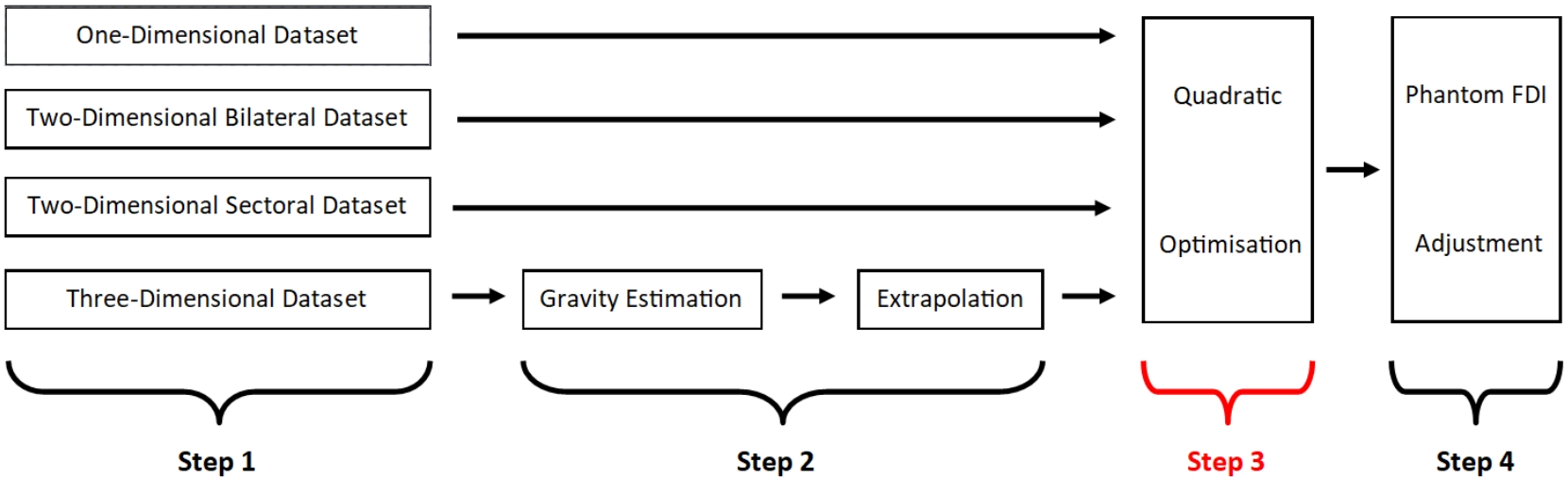
- Slight overshoot (unit: trillion USD)

Extrapolation (Continued)



- 10 smallest and biggest total Inward FDI ranked according to the extrapolated dataset (year: 2017)
- Overshooting for small and resource-rich economies

Overview


One-Dimensional Dataset
Two-Dimensional Bilateral Dataset
Two-Dimensional Sectoral Dataset
Three-Dimensional Dataset
Gravity Estimation
Extrapolation
Quadratic
Optimisation
Phantom FDI
Adjustment
Step 1
Step 2
Step 3
Step 4

Pre-optimisation Adjustment

1. Fixing FDI values at the bilateral dimension:
   - Two-dimensional bilateral dataset constructed with data from the IMF
   - Disaggregate to three-dimensional values based on sectoral share from the extrapolated dataset
   - Extrapolated values are adopted for missing country-pair values

2. Aggregating to GTAP region (120 economies plus 21 residual regions)

Quadratic Optimisation

Minimise:

$$\sum_{kij}(FDI^1_{kij} - FDI^0_{kij})^2/w_{kij} + \sum_{ki}(FDI^1_{ki} - FDI^0_{ki})^2/w_{ki} + \sum_{i}(FDI^1_{i} - FDI^0_{i})^2/w_{i}$$

Subject to:

$$\sum_{kij} FDI^1_{kij} = FDI_i^{\text{1D\_Inward}}$$

$$\sum_{kij} FDI^1_{kij} = FDI_j^{\text{1D\_Outward}}$$

$$\sum_{k} FDI^1_{kij} = FDI_{ij}^{\text{Bilateral}}$$

$$\sum_{j} FDI^1_{kij} = FDI_{ki}^{\text{Sectoral\_Inward}}$$

$$\sum_{i} FDI^1_{kij} = FDI_{kj}^{\text{Sectoral\_Outward}}$$

Country Comparison

	Original (Year 2004)		Current (Year 2017)	
Country/Regions	By Host	By Source	By Host	By Source
U.S.	19.6%	21.7%	10.6%	14.4%
China	3.7%	0.4%	6.1%	1.9%
EU_27	47.8%	45%	45.9%	43.2%
Japan	0.8%	3.2%	0.5%	4.1%
Britain	6.5%	11.6%	4.1%	7.6%
India	0.4%	0.08%	1.3%	0.3%
Canada	2.6%	3.1%	1.8%	2.9%
Russia	1.0%	0.9%	1.1%	0.8%
Brazil	1.3%	0.6%	1.7%	0.2%
ROW	16.3%	13.4%	26.9%	24.6%

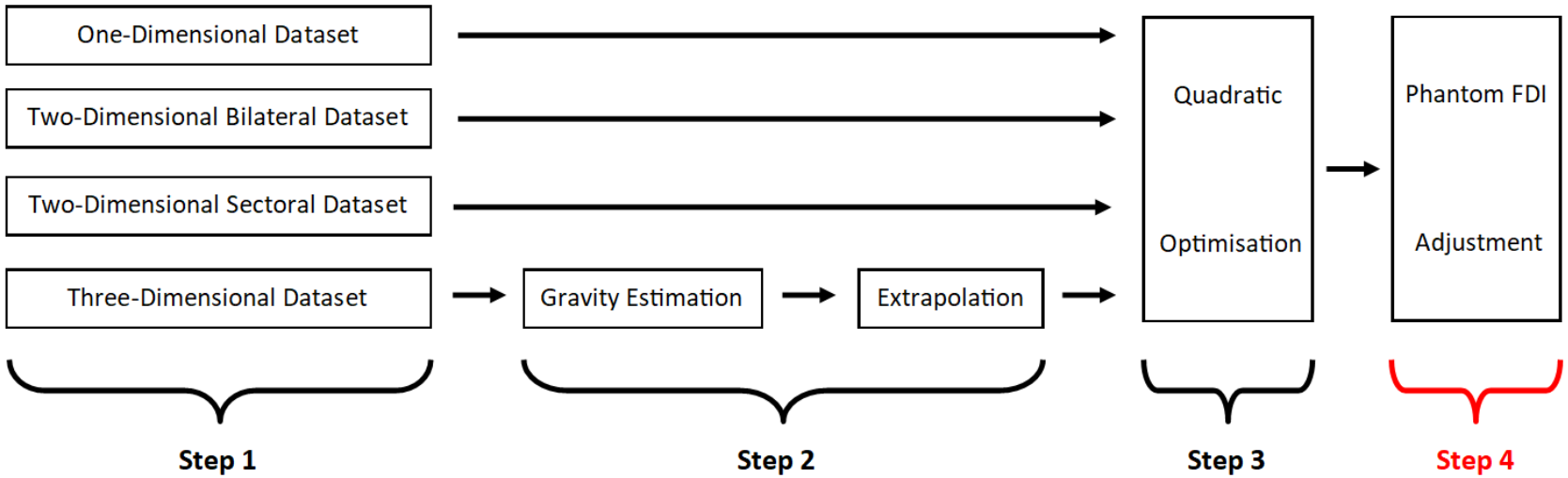
Sectoral Comparison

Sector	Original (Year 2004)	Current (Year 2017)
Agriculture & Extraction	0.37%	7.0%
Manufacturing	19.5%	14.1%
Services	76.7%	78.9%

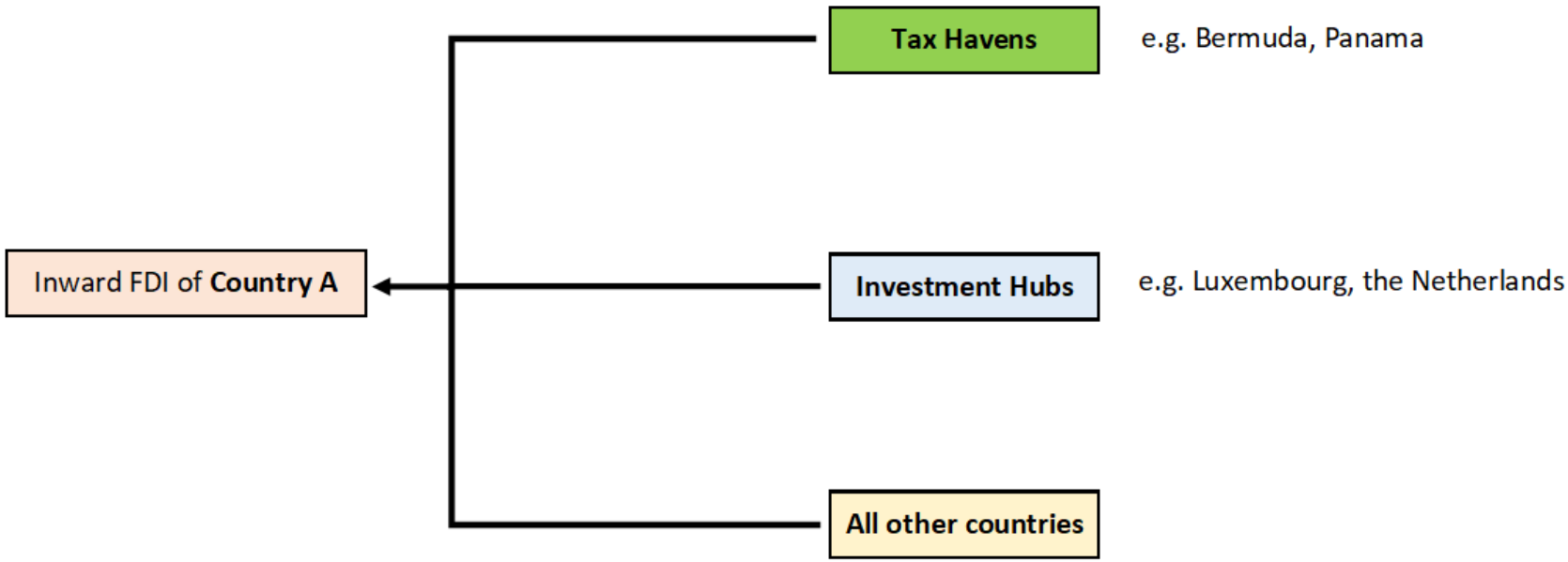
- Estimations from [UNCTAD, 2017] suggests that service sectors account for 63.5% of global FDI stock in 2015. For primary and manufacturing, they are 7.8% and 27.3% respectively

- Not final, subject to changes

Overview


One-Dimensional Dataset
Two-Dimensional Bilateral Dataset
Two-Dimensional Sectoral Dataset
Three-Dimensional Dataset
Gravity Estimation
Extrapolation
Quadratic
Optimisation
Phantom FDI
Adjustment
Step 1
Step 2
Step 3
Step 4

Introducing Phantom FDI

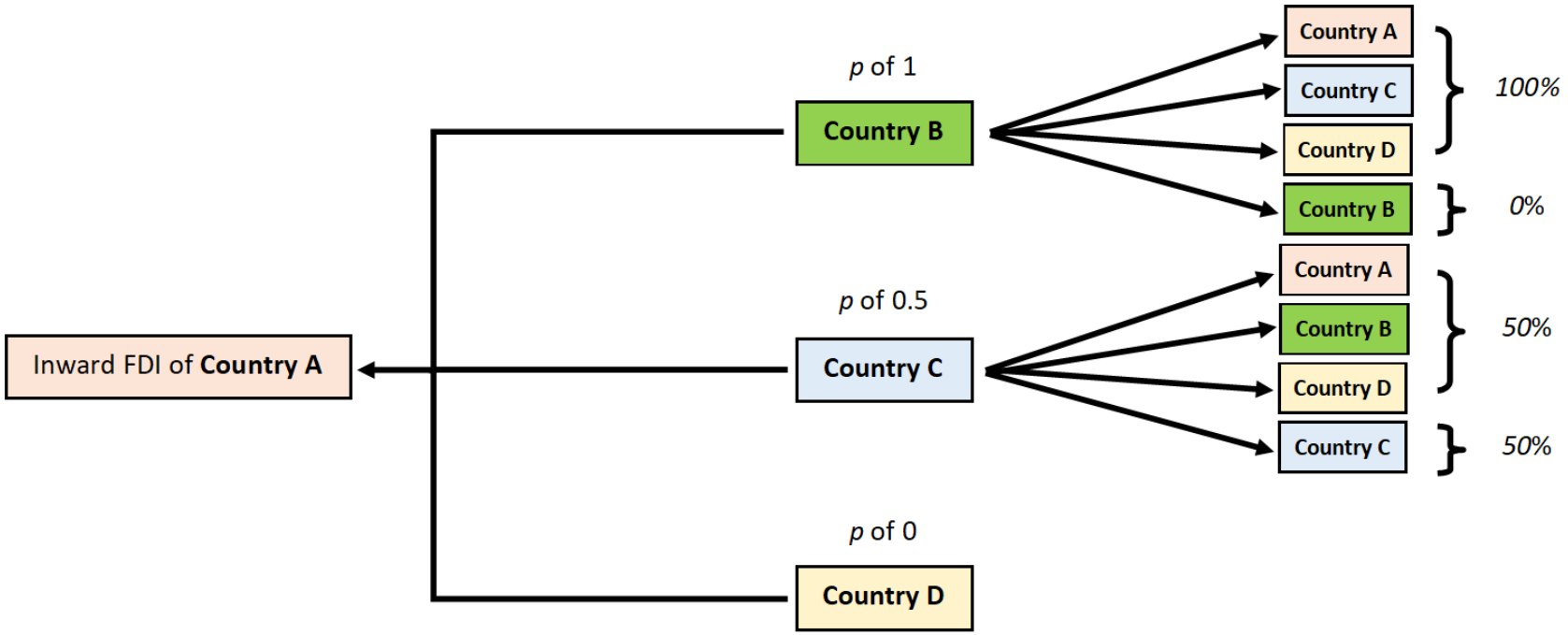


- As shown above, FDI statistics are heavily distorted by the presence of tax havens/offshore financial centres

Methodology

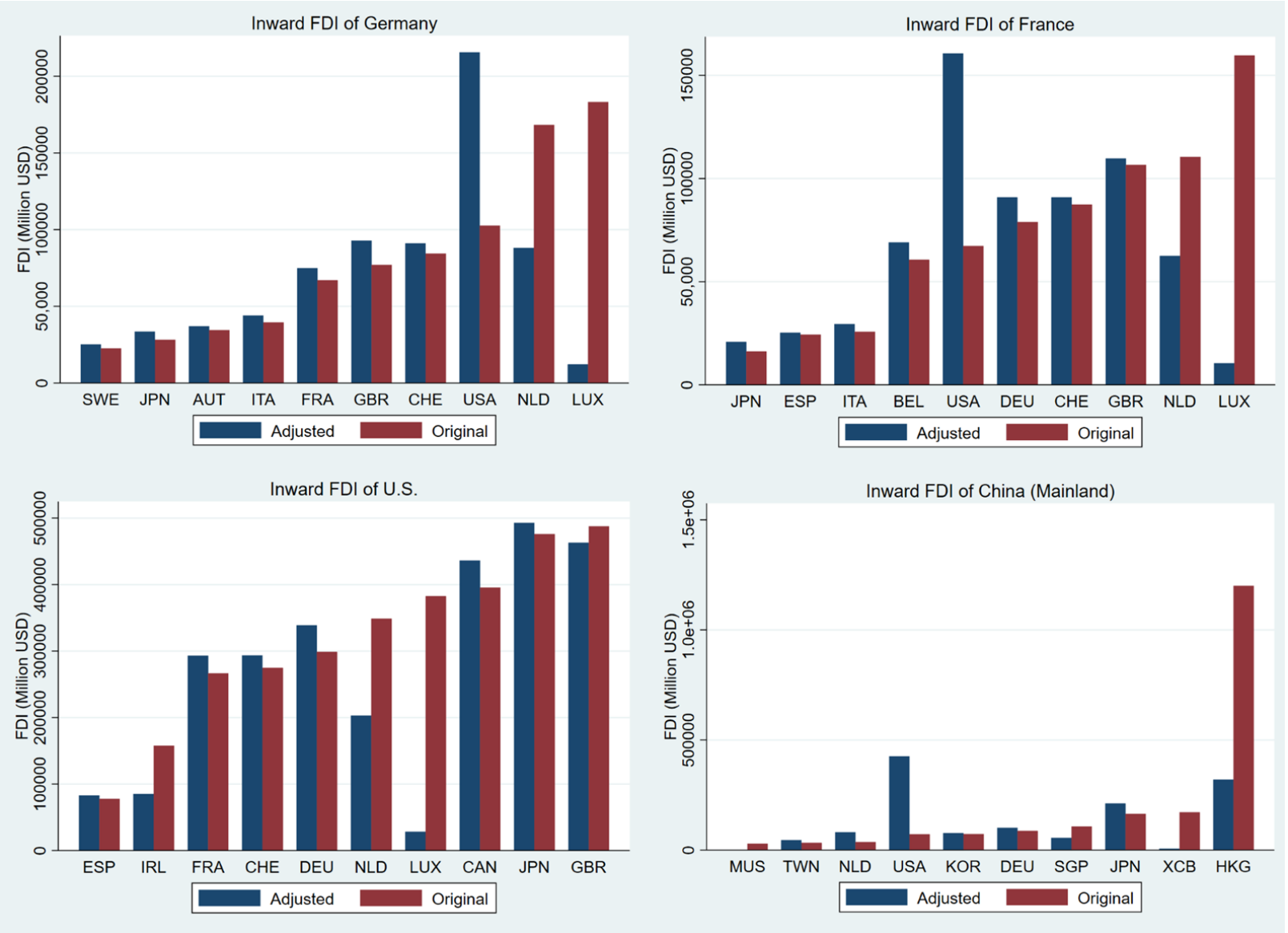
- Exact replication of [Casella, 2019] from UNCTAD
- Proposed solution: a probabilistic approach (absorbing markov chain) to estimate conduit/phantom FDI and then redistribute to their ultimate investors
- Key parameter: the conduit probability of a country which is imposed exogenously
  - 38 tax havens defined by the OECD: p of 1
  - Self-reporter or estimated investment hubs: $p \in$ [0-1]
  - All other regions: p of 0

Methodology (Continued)



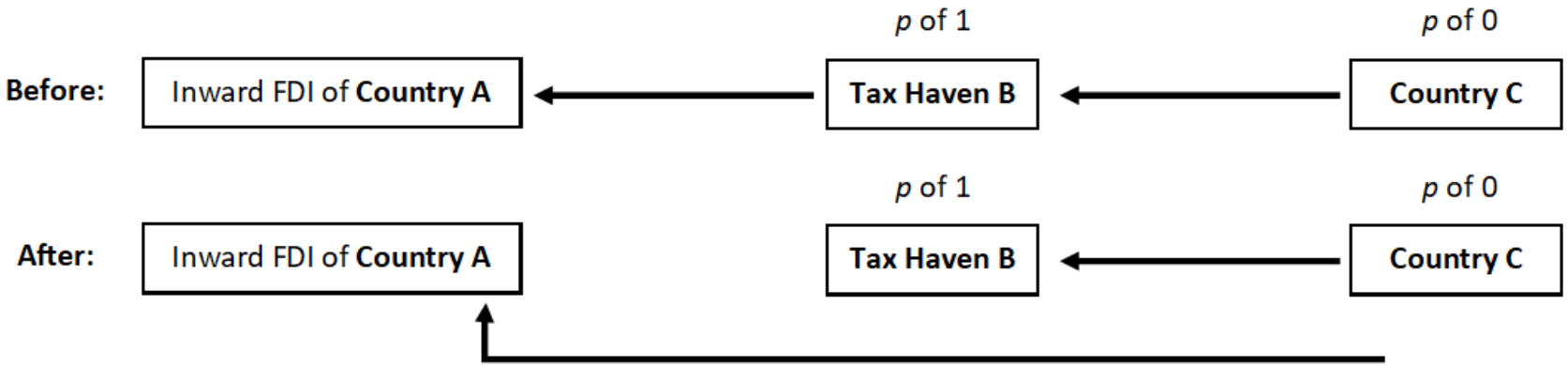
- Adjustment takes place at the bilateral level and then disaggregated to the three-dimensional level using the same sectoral share

Post-adjustment Comparison



Limitations

- Conduit probability is a country-specific parameter, it is not bilateral (e.g. US – Ireland) or sectoral (e.g. finance)
- Total world FDI and total inward FDI of each country remain the same – does not account for double accounting



- Other limitations – see [Casella, 2019]

Conclusion

Final Output:

- Fully consistent with GTAP version 10 & version 11
- Four datasets in total for 2014 & 2017, with & without phantom FDI adjustment
- 141 regions and 34 sectors
- Further sector disaggregation and reconciliation may be required – see [Lakatos and Walmsley, 2011]

Comparing to [Gouel et al., 2012]:

- Data collection: preference is given to inward FDI reporting
- Gravity: sector-by-sector estimation + 3 country dummies
- Quadratic optimisation: fixing FDI values at the bilateral level
- Additional phantom FDI adjustment

Thank You!

[Casella, 2019] Casella, B. (2019).
Looking through conduit FDI in search of ultimate investors âĂŞ a probabilistic approach.
Transnational Corporations, 26(1):109–149.

[Gouel et al., 2012] Gouel, C., Guimbard, H., and Laborde, D. (2012).
A Foreign Direct Investment database for Global CGE models.
CEPII Working Paper, No. 2012-08:35.

[Lakatos and Walmsley, 2011] Lakatos, C. and Walmsley, T. (2011).
A Global Multi-sector Multi-region Foreign Direct Investment Database for GTAP.
GTAP Research Memoranda, No. 18.

[UNCTAD, 2017] UNCTAD (2017).
World Investment Report 2017: Investment and the Digital Economy.
United Nations Conference on Trade and Development, Geneva.