Globalization and Prosperity Lab

Computational Model: Data

July 30, 2021

Agenda

- Data, Principles and Definitions
- Missing Production
- Overlay of Final Uses
- Mutual Imputation of Missing Flows
- Disaggregating Use Activities to Use Industries

Agenda

- Data, Principles and Definitions
- Missing Production
- Overlay of Final Uses
- Mutual Imputation of Missing Flows
- Disaggregating Use Activities to Use Industries

Data

- WIOT for trade, input-output flows and uses in 2014
- ITPD-E for proportions of bilateral trade by industry
  Mean of years 2013-15
- WIOT flow data W_{sdab}
  with source country s, destination d, supply activity $a(i)$, use $b(j)$
- ITPD-E trade and self-trade data X_{sdi}
  with source country s, destination country d, supply industry i

Bilateral Trade Flows by Supply Industry and Use

- X_{sdij}: Trade or self-trade flow from source country s and supply industry i to destination country d (including $d = s$) and use j

- General principles of data handling:

  1. use WIOD for b uses within sdi and totals to preserve value added: 38 aggregate activities $b(j)$, 5 final uses b

  2. use ITPD for proportional imputations of sdi: 170 industries i

  3. use flows reported in either data set to complete the other

  4. take production function of industries to be identical within activity

Definitions

- **Countries**: 43 and ROW (rest of world).
  Mutual consistency between ITPD and WIOD

- **Industries**: 170 ITPD-based industries.
  Aggregation for WIOD consistency, plus non-traded sector.
  Called GPL industries (Globalization and Prosperity Lab)

- **Activities**: 38 aggregate WIOD activities (including HH and international organization activities), plus 5 final uses (total 43): final consumption expenditure C_{sdi}^{HH} by households and non-profit organisations serving households (NPISH) C_{sdi}^{NPISH}, final expenditure by government G_{sdi}, gross fixed capital formation I_{sdi}^{K}, and changes in inventories and valuables I_{sdi}^{IV}.

GPL Country Groups

- Country group $\mathbb{G}_c$ based on World Development Indicators (WDI): High, Upper middle, Lower middle, Low income

- Regroup WDI countries for consistent per-capita GDP ranking 2014
  Top Upper middle income country Equatorial Guinea at $19.4k GDP p.c.; reassign High-income countries with less GDP p.c. to Upper middle; similarly for Lower middle (top Iran with $5.6k) and Low income (top Yemen with $1.7k GDP)

- WIOD sample countries (43): pop weighted mean GDP p.c. $14.9k
  ROW (173 WDI countries): pop wgt'd mean $4.3k (**Lower middle**)
  WIOD: 28 High, 12 Upper middle, 2 Lower middle, 1 Low income
  India single low-income country in WIOD, grouped with lower middle income
  ROW: 30 High, 43 Upper middle, **49 Lower middle**, 43 Low income

Aggregate Variables

- **Production** (output, not value added): $Y_{si} = \sum_d \sum_j X_{sdij} = X_{s \cdot i \cdot}$
  self trade included ($s = d$ in sum)

- **Market size** (expenditure): $X_{di} = \sum_s \sum_j X_{sdij} = X_{\cdot di \cdot}$
  self trade included ($s = d$ in sum)

- **Exports** $EX_{si} = \sum_{d \neq s} \sum_j X_{sdij}$
  Imports $IM_{di} = \sum_{s \neq d} \sum_j X_{sdij}$

- **Trade balance** country c: $TB_{ci} = EX_{ci} - IM_{ci}$

Value Added (JUNYUAN TO CHECK)

- **Production**: $Y_{ci} = \sum_d \sum_j X_{cdij} = X_{c \cdot i \cdot}$
  self trade included ($c = d$ in sum)

- **Intermediate inputs**: $M_{dj} = \sum_s \sum_i X_{sdij} = X_{\cdot d \cdot j}$
  self trade included ($s = d$ in sum)
  Restated: $M_{ci} = \sum_s \sum_k X_{scki} = X_{\cdot c \cdot i}$

- **Value added**: $V_{ci} = Y_{ci} - M_{ci}$
  Underlying X_{sdij} combines ITPD and WIOD information
  after disaggregation (extending Lindner, Legault & Guan 2012)
  Problem cases: XXX

Capital and Labor Shares (JUNYUAN TO COMPLETE))

- SEA: Results from comparison value-added shares WIOD and SEA
- Winsorize the extremely low VA shares
- ROW as mean of sector-specific world average for lower middle income countries
- XXX

Current Account Balance (FOR LATER)

- **Current Account balance** country c: $CA_c = TB_{c\cdot} + r_c \cdot B_c$
  (cross-border foreign net wealth position B_c, no cross-border labor movements)

- To be calibrated. Proposal:
  - Use CPIS (Coordinated Portfolio Investment Survey) 2014 to find B_c, use WDI 2014 to find $r_c B_c$, and compute $CA_c = TB_{c\cdot} + r_c B_c$ using WDI and $r_c = (CA_c - TB_{c\cdot}) / B_c$ using WDI and CPIS
  - Problem to be addressed: United States $B_{US} < 0$ but $r_{US} B_{US} > 0$ because interest rate on gross foreign credit exceeds interest rate on gross foreign debt

Activity Aggregation

- Aggregate activities a from WIOD and GPL industries i defined so that strict nesting $a(i)$

- Set of industries within activity a: $i \in \mathbb{I}_a = \{i : a(i)\}$, $j \in \mathbb{J}_b = \mathbb{I}_a$. Universe of industries: $\bigcup_a \mathbb{I}_a$

- **Production** (self trade included): $Y_{sa} = \sum_{i \in \mathbb{I}_a} X_{s \cdot i \cdot}$
  Market size (self trade included): $X_{da} = \sum_{i \in \mathbb{I}_a} X_{\cdot di \cdot}$

- **Intermediate inputs**: $M_{db} = \sum_{j \in \mathbb{J}_b} X_{\cdot d \cdot j}$
  Value added: $V_{sa} = Y_{sa} - M_{sa}$

Agenda

- Data, Principles and Definitions
- Missing Production
- Overlay of Final Uses
- Mutual Imputation of Missing Flows
- Disaggregating Use Activities to Use Industries

Agenda

- Data, Principles and Definitions
- Missing Production
- Overlay of Final Uses
- Mutual Imputation of Missing Flows
- Disaggregating Use Activities to Use Industries

Agenda

- Data, Principles and Definitions
- Missing Production
- Overlay of Final Uses
- Mutual Imputation of Missing Flows
- Disaggregating Use Activities to Use Industries

Agenda

- Data, Principles and Definitions
- Missing Production
- Overlay of Final Uses
- Mutual Imputation of Missing Flows
- Disaggregating Use Activities to Use Industries

Data With Non-zero Production After Mutual Imputations

- Combined ITPD 2013-15 and WIOD 2014 data X_{sdib}
  10,848,350 observations (corresponding to 269,277 at X_{sdi} level)

- In ITPD 2013-15 (mean): 256,985 X_{sdi} observations (some self trade)
  Appended non-zero production: 133 X_{ssi} observations
  Resulting total: 257,118 observations

- Unmatched ITPD 2013-15 X_{sdi} in WIOD 2014 after imputations: 226 X_{sdi} observations (less than 1/1,000 of 1 percent of total trade)

- Imputation: Append 9,115 X_{sdi} observations, resulting: 269,277 X_{sdi}

Disaggregating Use Activities to Use Industries (JUNYUAN)

- Wolsky (1984) and Lindner, Legault & Guan (2012):
  Disaggregate X_{sdab} into X_{sdij} under multiplicity of possibilities
  $\left(X_{sdab} = \gamma_{sdab} \cdot \sum_{i \in \mathbb{I}_a} X_{sdi}^{ITPD}\right)$

- We know: X_{sdib}
  Single consistent disaggregation of $b(j)$ into j given observed i
  $\left(X_{sdib} = \gamma_{sdab} \cdot X_{sdi}^{ITPD}\right)$

- Rationale, Main assumptions, Plausibility of assumptions

APPENDIX

Reverse Imputations as Alternative to Main Approach

- X_{sdij}: Trade or self-trade flow from source country s and supply industry i to destination country d (including $d = s$) and use j

- Alternative principles of data handling:
  1. use ITPD for sdi totals: 170 industries i
  2. use WIOD for proportional imputations, especially of j within sdi: 38 aggregate activities $b(j)$, 5 final uses b
  3. use flows reported in either data set to complete the other

- **Problem**: Pervasive cases of negative value added
  Likely cause: ITPD flows out of proportion for input-output relations

Missing Production: All Industries Produce in All Countries

- ITPD data incomplete for production $Y_{si} = \sum_d \sum_j X_{sdij} = X_{s \cdot i.}$
  (WIOD suggests 1 no-production case of 133 only: media reproduction in Russia)
  - 7,480 square observations expected: 44 source countries (including ROW) $\times$ 170 supply industries
  - 44 missing: nontraded sector missing from ITPD for 44 countries
  - 20 missing: 5 WIOD aggregate activities missing for some countries (e.g. rental services in 11 countries, publishing in 3)
  - 69 missing: GPL industries within WIOD aggregates missing (e.g. rice in Taiwan, live cattle in Japan, prepared vegetables in Canada)

Imputation of Missing Production *within* Activity Level

- In 69 cases of missing production, some ITPD industries i within WIOD activity a for a source country s non-missing but others missing

- From ITPD, define production share of industry by activity:

$$\beta_{si|a}^{ITPD,\mathbb{G}} = \frac{Y_{\cdot i}^{ITPD,\mathbb{G}}}{\sum_{i \in \mathbb{I}_a} Y_{\cdot i}^{ITPD,\mathbb{G}}}, \quad \text{where } Y_{\cdot i}^{ITPD,\mathbb{G}} = \sum_{s \in \mathbb{G}_s} Y_{si}$$

  for three groups $\mathbb{G}_s$ by income p.c. 2014 (WDI)

- Imputation: $Y_{si}^{ITPD} = \beta_{si|a}^{ITPD,\mathbb{G}} Y_{sa}^{ITPD} = \beta_{si|a}^{ITPD,\mathbb{G}} \beta_{sa}^{WIOD} Y_{s\cdot}^{ITPD}$,
  where $Y_{sa}^{ITPD} = \beta_{sa}^{WIOD} Y_{s\cdot}^{ITPD}$ given below.

Imputation of Missing Production *at* Activity Level

- From WIOD, supply activity share in total (38 activities) by country:
$$\beta_{sa}^{WIOD} = Y_{sa}^{WIOD} / Y_{s\cdot}^{WIOD}.$$

- Intended imputation: $Y_{sa}^{ITPD} = \beta_{sa}^{WIOD} Y_{s\cdot}^{ITPD}$, but $Y_{s\cdot}^{ITPD}$ unobserved

- Known: $\check{Y}_{s\cdot}^{ITPD} = \left(1 - \sum_{i \in \mathbb{M}_s} \beta_{si|a}^{ITPD,\mathbb{G}} \beta_{sa}^{WIOD}\right) Y_{s\cdot}^{ITPD}$
where $\mathbb{M}_s$ is set of industries missing from ITPD by country s

- Complete **imputation**: $Y_{si}^{ITPD} = \dfrac{\beta_{si|a}^{ITPD,\mathbb{G}} \beta_{sa}^{WIOD}}{1 - \sum_{i \in \mathbb{M}_s} \beta_{si|a}^{ITPD,\mathbb{G}} \beta_{sa}^{WIOD}} \check{Y}_{s\cdot}^{ITPD}$

Manual Flow Corrections: Reset zero to $100k

- In ITPD data (2013-2015 mean), 25,172 zero trade flows. Zero kept
- Of the 133 imputations into ITPD, WIOD suggests one zero trade flow: Russian self-trade in reproduced media, set to $100k
  133 i industries imputed into ITPD, corresponding to 72 a activities
- Some imputations may be implausibly large under above method, so manually reduce to $100k: Soybeans in Cyprus and Norway; Tobacco leaves and cigarettes in Finland; Cocoa and cocoa products in Cyprus, Ireland, Latvia, Malta and Slovakia; Cotton in Malta; Live swine in Turkey.

Final Uses in WIOD 2014

- In WIOD 2014, there are 73,568 $W_{sda.}$ flow observations (with non-zero or zero values)
  - 48 w/ negative investment (gross fixed capital formation $I^K_{sda} < 0$)
  - 1,516 w/ negative inventory and valuables change ($I^{IV}_{sda} < 0$)
  - no other use or final use categories exhibit negative values
- To prevent negative changes from inducing implausible flows, we use time aggregation for steady-state final-use shares in total

Time Aggregation of Final Uses in WIOD 2000-2014

- WIOD 2000-14 to measure steady-state share of final uses in total

  W_{sdabt}: 43 countries plus ROW s, d; 38 activities $a(i), b(j)$; $t = 2000, \ldots, 2014$

- In steady state, gross investment or inventory change non-negative

  – But time and use aggregation $\sum_t \sum_b W_{sdabt} = W_{sda\cdot\cdot} < 0$ in 14 cases (with ROW destination) and $\sum_t \sum_{b \in \mathbb{F}} W_{sdabt} < 0$ in 110 cases (where $\mathbb{F}$ is set of final uses): impacts of negative inventory changes (and five cases of negative investment), mostly for ROW destination, publishing and basic metals activities

- For final use shares, we set final use share **to zero when otherwise negative** (positive total) and **to unity when otherwise in excess of unity** (negative total).

  – 292 zero final use shares (from 190), 73 unitary final use shares (from 65)

Use Activity Shares from WIOD in ITPD

- Combine WIOD activities with ITPD flows by industry i (for $a(i)$).

  38 supply activities $a(i)$, 38+5 use activities plus final uses $b(j)$

- Final use share: $\bar{\gamma}_{sda}^{finl} = \frac{\sum_t \sum_{b \in \mathbb{F}} W_{sdabt}}{\sum_t \sum_b W_{sdabt}}$.

  Time aggregation of WIOD 2000-14, where $\mathbb{F}$ is set of final uses

- For 38 use activities $b \notin \mathbb{F}$: $\gamma_{sdab} = \left(1 - \bar{\gamma}_{sda}^{finl}\right) \frac{W_{sdab,2014}}{\sum_{b \notin \mathbb{F}} W_{sdab,2014}}$

  For 5 final uses $b \in \mathbb{F}$: $\gamma_{sdab} = \bar{\gamma}_{sda}^{finl} \frac{W_{sdab,2014}}{\sum_{b \in \mathbb{F}} W_{sdab,2014}}$

  Set $W_{sdab,2014} = 0$ if < 1 cent of \$1 (20,389 of 3,236,992 obs.)

  Cap I_{sdi}^{K}, I_{sdi}^{IV} below so $\sum_{b \in \mathbb{F}} W_{sdab,2014} >= 0$ (815 of 3,236,992 obs.)

- **Overlay uses on bilateral flows by industry**: $X_{sdib} = \gamma_{sdab} \cdot X_{sdi}^{ITPD}$

Excess and Unaccounted Trade Flows in WIOD 2014

- In 2014 (final sample year), WIOD reports 70,802 non-zero trade flows $W_{sda.}$ by source country s, destination d, supply activity $a(i)$
  Set $W_{sdab} = 0$ if < 1 cent of \$1 (20,389 of 3,236,992 obs.), then aggregate uses b.

- In 2014, ITPD reports 55,939 non-zero flows $X_{sda} = \sum_{i \in \mathbb{I}_a} X_{sdi.}$
  by source-country s, destination d, supply activity $a(i)$
  after imputing non-zero production for all activities and countries

- 16,325 *excess* trade flows $W_{sda.}$ in WIOD (not in ITPD)

- 1,462 *unaccounted* trade flows X_{sda} in ITPD (not in WIOD)

Treating *Excess* WIOD-2014 Flows

- 16,325 $W_{sda.}$ flows (617 self-trade $s = d$) in WIOD but not in ITPD
  Excess self-trade (617) accounts for 40.1% of self trade and 35.0% of total in WIOD
- Proportion missing d to all non-missing: $\zeta_{sdab}^{WIOD} = \dfrac{W_{sdab}}{\sum_{c \in \mathbb{G}_{ITPD}} W_{scab}}$

  $\mathbb{G}_{ITPD}$: non-missing destinations in WIOD *and* ITPD (cover all but 23 of 16,325)
  Remove if $I_{sdi}^{IV} < 0$ dominant so that no $\sum_{c \in \mathbb{G}_{ITPD}} W_{scab} < 0$ (drop 15 of 16,325)

  Share of industry i in activity a by country group: $\gamma_{sdi|a}^{ITPD,\mathbb{G}}$ (next page)

- **Impute excess WIOD flows into ITPD** (16,287 successes)

$$X_{sdab} = \zeta_{sdab}^{WIOD} X_{s \cdot a}^{ITPD} \quad \text{and} \quad X_{sdib} = \gamma_{sdi|a}^{ITPD,\mathbb{G}} X_{sdab},$$

where $X_{s \cdot a}^{ITPD} = \sum_{i \in \mathbb{I}_a} \sum_{c \in \mathbb{G}_{ITPD}} X_{sci}^{ITPD}$ is observed ITPD flow

Treating *Excess* WIOD Flows: Industry Shares 2013-15

- For 7,172 of 16,287 observations industry shares directly observed. Share of industry i in activity a by country pairs and $i \in \mathbb{I}_a$:

$$\gamma_{sdi|a}^{ITPD,\mathbb{G}} = \frac{X_{sdi}^{ITPD,13\text{-}15}}{\sum_{i \in \mathbb{I}_a} X_{sdi}^{ITPD,13\text{-}15}}$$

- For 8,257 of 16,287 observations industry shares require imputation. Share of industry i in activity a by country **group** pairs and $i \in \mathbb{I}_a$:

$$\gamma_{sdi|a}^{ITPD,\mathbb{G}} = \frac{\sum_{s \in \mathbb{G}_s} \sum_{d \in \mathbb{G}_d} X_{sdi}^{ITPD,13\text{-}15}}{\sum_{s \in \mathbb{G}_s} \sum_{d \in \mathbb{G}_d} \sum_{i \in \mathbb{I}_a} X_{sdi}^{ITPD,13\text{-}15}}$$

- For 868 of 16,287 observations no pair flows, **reuse** $\beta_{si|a}^{ITPD,\mathbb{G}}$ above

Treating *Unaccounted* ITPD Trade Flows

- 1,462 *unaccounted* X_{sda} flows in ITPD (not in WIOD) concentrated: mostly printing, publishing, basic metals, and repair and installation

- To preserve WIOD I-O structure, assume flows only to 5 final uses.
  For 5 final uses $b \in \mathbb{F}$: $\gamma^{\mathbb{F},\mathbb{G}}_{sdab} = \frac{\sum_{s\in\mathbb{G}_s}\sum_{d\in\mathbb{G}_d}\sum_{i\in\mathbb{I}_a} X_{sdib}}{\sum_{s\in\mathbb{G}_s}\sum_{d\in\mathbb{G}_d}\sum_{i\in\mathbb{I}_a}\sum_{b\in\mathbb{F}} X_{sdib}}$,
  for three groups $\mathbb{G}_c$ by income p.c. 2014 (WDI) and X_{sdib} as imputed
  Some industries show no cases of cross-group shipments into final uses
  34 of 1,462 flows therefore not imputable

- **Impute** $X_{sdab} = \gamma^{\mathbb{F},\mathbb{G}}_{sdab} X_{sda}$ (1,428 successes)

Data Quality Assessments (JUNYUAN)

- Value added comparisons between WIOD, SEA and COMPLETE DATA after imputation

- XXX