CHELEM
International Trade

Building Methods of the CEPII Database

Alix de SAINT VAULRY

November 2013

Content of the Database

- International trade flows (exporter, importer, product, year)

- Yearly data stretching back to 1967

- Trade values in millions of US current dollars

- World coverage: 95 countries/zones (83 countries and 12 zones, and a heading for not elsewhere specified countries); 4 nested levels

- 3 product classifications: CHELEM (71 categories), GTAP (43 categories), ISIC (147 categories)

- Multiple sectoral aggregations (production chains, stages in the production process, technological levels…)

- Multiple indicators: Trade balances, Coverage rates, Revealed comparative advantages…, some of them being linked with CHELEM-GDP (populations, GDP in value, volume or PPP, exchange rates) or CHELEM-BOP (balances of payments)

A harmonized database

- For <u>each flow from country *i* to country *j*</u>, there can be either:

  - <u>1 reporting</u> from the exporting country *i* or from the importing country *j*

  - <u>2 reportings</u> from both *i* and *j*: they <u>must be harmonized</u>

  - <u>no reporting</u> at all

- A <u>harmonized bilateral trade database</u> gives:

  - <u>consistent</u> data over both <u>products</u> and <u>countries</u>

  - a <u>world coverage</u> with the maximum detail possible to keep good quality data

  - <u>exhaustive</u> data (missing flows are estimated)

Sources

- Main source: <u>COMTRADE</u> database from the United Nations

- <u>IMF</u> (Direction of Trade Statistics)

- Other <u>international sources</u> (OECD, UNCTAD, World Bank)

- <u>National sources</u> (Hong Kong for its reexports by origin, Taiwan, Luxembourg from 1993 to 1998…)

Building process

8 steps:

– A. First tests, aggregations, consistency tests

– B. Reexports and reimports

– C. Missing data for usually reporting countries

– D. Removing freight and insurance costs from *cif* imports

– E. Geographical aggregations

– F. Comparison of mirror flows and corrections

– G. Harmonization

– H. Estimation of countries issued from former USSR, former Czechoslovakia and former Yugoslavia, Belgium and Luxembourg before 1992 or 1993, Bangladesh separated from Pakistan from 1967 to 1971

A. First tests, aggregations, consistency tests

- **First tests**: from source data, check the formats, units, currency, geographical and sectoral classifications (and make the necessary updates with new items)

- **Aggregations**: products are aggregated in an intermediary CHELEM classification which allows 3 possible aggregations in 71 CHELEM, 43 GTAP or 147 ISIC classifications

- **Consistency tests**:

- For confidentiality reasons, some countries do not report their exchanges for some products, or do not break down some flows between all partner countries.

- Thus there is a difference between the reported « total » of the exports or imports with a given partner and the sum of the reported flows by product.

- If positive, this difference is attributed to the « not elsewhere specified » product code (or exceptionnally to crude petroleum for petrol exporters, for instance).

- If negative, the « not elsewhere specified » product code is set to zero for this partner.

B. Reexports and reimports (1): definitions

- **Reexports** (COMTRADE's definition): foreign goods exported from any part of the *economic* territory of a country in the same state as previously imported (or which underwent processing that did not change their origin)

- **Reimports** (COMTRADE's definition): domestic goods in the same state as previously exported (or having undergone processing that did not change their origin) which reenter any part of the *economic* territory of their country of origin

- In **COMTRADE**, reexports and reimports are included in exports and imports

- In **CHELEM**, reexports and reimports are excluded from exports and imports and special processings occur for Hong Kong (whose reexports reach 3% of international trade) and Singapore (whose reexports, as reimports, are near 1% of international trade)

B. Reexports and reimports (2): Hong Kong

- **<u>Hong Kong exports</u>**: reexports are removed from COMTRADE's exports reported by Hong Kong, as for other countries reporting reexports.

- **<u>Hong Kong imports</u>**:

  To measure correctly <u>Hong Kong domestic imports</u>, reexports by country of origin and product (source: Hong Kong Census and Statistics Department) are removed from COMTRADE's imports reported by Hong Kong, as far as possible.

  For a given product, if reexport from a given partner to Hong Kong exceeds Hong Kong import from this partner, Hong Kong import of this product from this partner is set to zero.

- **<u>Other countries' exports to Hong Kong finally destinated to others</u>**:

  Some countries, like <u>China, USA, Japan, South Korea, Taiwan, Malaysia, Thailand and India</u>, report big amounts of exports to Hong Kong, including goods later reexported by Hong Kong to other countries. Their exports <u>transiting via Hong Kong</u> are reaffected to their true final countries of destination.

B. Reexports and reimports (3): Singapore

For Singapore, reexports and reimports, not reported in COMTRADE, have to be estimated.

- Total reexports and reimports from 1970 to 1993 are edited by the IMF.

- From 1994 up to now, only total reexports (RX) are available on www.singstat.gov.sg so total reimports (RM) are estimated by CEPII: RM = 0.8 * RX (in 1993, RM/RX=0.8).

- **Reexports**: 1991 ratios of reexports to exports detailed by product k (source: Singapore Trade Statistics December 1991, at SITC level) are multiplied by the exports of product k to partner j and a multiplier in order to keep the total reexports:

$$RX_{jk}^t = Min\,(X_{jk}^t, \alpha_k^{1991}.X_{jk}^t \cdot \frac{RX^t}{\sum_k \alpha_k^{1991}.X_k^t}) \qquad \text{with} \quad \alpha_k^{1991} = \frac{\sum_j RX_{jk}^{1991}}{\sum_j X_{jk}^{1991}}$$

- **Reimports**: there are calculated accordingly to reexports. They must not exceed imports, as reexports must not exceed exports.

$$RM_{jk}^t = Min\,(M_{jk}^t, \quad M_{jk}^t \cdot \frac{RX_k^t}{M_k^t} \cdot \frac{RM^t}{RX^t})$$

- Reexports and reimports calculated this way are removed respectively from Singapore exports and imports reported in COMTRADE to estimate Singapore domestic exports and imports. Beginning in 1993, special procedures are applied to electronic components, computer equipment and refined petroleum products (half of Singapore exports).

C. Missing data for usually reporting countries (1)

- When a usually reporting country does not report its exports and/or imports for a given year n, they have to be **estimated**. If no detailed data is available, the principal methods are the calculations by **structure** or by **interpolation** and the use of **mirror statistics** (reported by the **partner countries**).

- **Total exports and imports** may be found in the Direction of Trade Statistics (DOTS) of the IMF or in the Monthly Bulletin of Statistics (MBS) of the United Nations (or in the UNCTAD statistics or in national statistics). Otherwise they can be estimated by trend (variation of total exports or imports of the zone or continent to which the country belongs or of the world).

- **By structure**: the geographical and product crossed breakdown of the exports (resp. imports) of an available year is applied to the total exports (resp. imports) of the missing year. The reference year must be as near as possible to the estimated year. If the total exports (resp. imports) of the missing year is unknown, the variation of the total exports (resp. imports) of the zone or continent to which the country belongs (or those of the world if there are not available) is used.

- **By interpolation**: assuming that the geographical and product crossed breaking down of the exports (resp. imports) are available for two distant years *t0* and *tn* but not for the years *t* in between, the data for these missing years are calculated by interpolation. Two cases are possible:

- The total exports (resp. imports) of the missing year are known:

$$X_{ijk}^t = X_i^t \cdot \left(\frac{tn - t}{tn - t0} \cdot \frac{X_{ijk}^{t0}}{X_i^{t0}} + \frac{t - t0}{tn - t0} \cdot \frac{X_{ijk}^{tn}}{X_i^{tn}} \right)$$

- The total exports (resp. imports) of the missing year are not known:

$$X_{ijk}^t = \frac{tn - t}{tn - t0} \cdot X_{ijk}^{t0} + \frac{t - t0}{tn - t0} \cdot X_{ijk}^{tn}$$

Exports are usually reported *fob* (free on bord) whereas imports are reported *cif* (cost, insurance and freight to the frontier of the importing country included), following the recommendations of the United Nations (see reference on last slide). In order to get comparable (*fob*) values of exports and imports, freight and insurance costs are removed from *cif* imports.

$$Mfob_{ijk}^{t} = \frac{Mcif_{ijk}^{t}}{1 + freight_{ijk}^{t}}$$

with $$freight_{ijk}^{t} = rate_{IJK} * (1 + \frac{1}{9}) * coef_{k} * ind_{k}^{t}$$

i, j, k, t   : exporting country, importing country, product, year
I   : zone where country *i* belongs
J   : zone where country *j* belongs
K   : reference product for *k*
rate $_{IJK}$: freight rate for reference product *K* from zone *I* to zone *J* in 1969
1 + 1/9   : to take into account added insurance cost, estimated to 1/9 of freight cost
coef $_{k}$: multiplicative coefficient linking *k* to *K*
ind_{k}^{t} : evolution index of product *k* (depending on the product, it is calculated with liner
      or petroleum tankers, bulk or trip freight indices and unit value indices;
      so it takes into account the distance and the value/weight ratio)

E. Geographical aggregations

- **Geographical aggregations**:

- As it is difficult to get satisfying exhaustive data on trade broken down by products for every country in the world to every country in the world every year, only 83 countries are kept in the CHELEM - International Trade database. The other countries are aggregated in 12 zones: others in America, Caucasus, other CIS, Serbia/Montenegro, Southafrican Union, African LDCs, Africa (others), Gulf nes, Middle East no OPEC, Cambodia and Lao PDR, East Asian LDCs, East Asia nes.

- Besides, some geographical aggregations are necessary to keep consistent series:

USA includes US Virgin islands and Porto Rico; France includes overseas departments and Monaco... Intra-trade within each of these aggregations is removed.

- These aggregations may occur before and/or after the comparisons and corrections.

13

Steps A to D give two sets of data: reports of domestic exports and reports of domestic imports, both *fob*.

Comparison of mirror flows: those *mirror* flows reported respectively by the exporting country i (X_{ij}^k) and by the importing country j (M_{ij}^k) for the product k and the year t are compared, relatively to (i) the total of international trade W and to (ii) the total of the trade of the product k reported by all reporting countries:

(i) $$Q1 = \frac{X_{ij}^k - M_{ij}^k}{W}$$

(ii) $$Q2 = \frac{X_{ij}^k - M_{ij}^k}{\sum (X_{i'j'}^k + M_{i'j'}^k)/2}$$

i' and j' describing
all reporting countries

The biggest ratios are analysed and then some reportings are corrected before the final harmonization. Iterations may occur: comparison, corrections, then new comparison, new corrections, and so on.

14

For large Q1 and Q2, various **corrections** are possible:

• Replace the missing reports by the reports of the partner. Be careful, the country which reports a bigger amount is not always the most reliable. For instance, Germany reports the total value of co-produced Airbus planes at each crossing of the French frontier, for both exports and imports. CHELEM chooses the French reports of added value only. The balances are roughly the same.

• Reaffect « not elsewhere specified » items to the correct products (known by another year report or by the reports of the partner(s)).

• Reaffect « not elsewhere specified » partners to the correct partners (known by another year report or by the reports of the partner(s)).

• Sub-sectoral breaking down: for instance between vehicles and parts of vehicles, the sub-total (vehicles + parts) is kept but with the partner's sub-breaking down.

• Sub-geographical breaking down: for instance a set of european countries for US exports. The total exports reported by the US to Western Europe for a product is kept with the geographical breaking down reported by european countries.

G. Harmonization (1): the four categories

- **Two sets of matrices** are now available: reports of <u>exporters</u> and reports of <u>importers</u>. A given flow from country i to country j can be reported by none of the two countries, by only one country (the exporter or the importer) or by both exporter and importer, generally with different amounts. In this last case the data have to be harmonized. If there is no report the data have to be estimated.

- The **countries are sorted** in four categories according to the quality of their reports, then the best reporting ones have priority over the others and different harmonization processes are implemented in order to get a unique value for each elementary flow.

- The **four categories** are the following (see the CEPII website for details):
  - **alpha 1: more reliable countries** (North America, Western Europe, Turkey, Israel, Japan, Asian NICs, Australia, New Zealand): more than 80% of international trade
  - **alpha 2: less reliable countries** (or reporting with delay), as China when reporting, Mexico, Brazil, India, Russia, Malaysia...
  - **beta 1: zones containing rather good reporters** (miscellaneous countries in Latin America, miscellaneous countries in Asia and Oceania)
  - **beta 2: countries or zones with no good reporter** (or in which good reporters do not represent a large enough part of the zone's exports and imports): Libya, Middle-East, LDCs in Africa, LDCs in Asia and Oceania, Cambodia... (and former USSR…) 16

G. Harmonization (2): the different processes

IMPORTERS

	Alpha 1	Alpha 2	Beta 1	Beta 2
Alpha 1	**Q11** harm. RAS	**Q12** report. X	**Q21** report. X	**Q22** report. X
Alpha 2	**Q13** report. M	**Q14** max (report. M, report. X)	**Q23** report. X	**Q24** report. X
Beta 1	**Q31** report. M	**Q32** report. M	**Q41** estim.	**Q42** estim.
Beta 2	**Q33** report. M	**Q34** report. M	**Q43** estim.	**Q44** estim.

EXPORTERS

G. Harmonization (3): Q11
trade between best reporters (alpha 1)

• **Why harmonize?** The alpha 1 countries report regularly and reliably. But the exporter and the importer do not report the same value for the same flow, even after removing the freight from *cif* imports and doing some corrections. The two reported values are harmonized with the RAS iterative method (see Stone and alii (1963)).

• **Initial matrix and aimed totals**: First of all, the matrix (report. X + 2 report. M)/3 at the aggregate level of 18 product categories is calculated. Indeed the importers know better the origin of imported products than the exporters know the destination of the exported products. But the exporters know the values better, so the total of exports of each category is kept. The « aimed » totals are calculated, for each line (total of exports of a product from an alpha 1 country to all alpha 1 countries) and each column (total of imports).

• **Iterative multiplications**: Then successive multiplications (by the aimed total of the line (resp. column) divided by the current calculated total) are made on lines and on columns in alternance until the matrix converges: when the total of each line and the total of each column are less than 0,1% different from the corresponding aimed totals, the interlocked multipliers of lines and columns for the aggregate category are applied for each detailed CHELEM product category.

18

G. Harmonization (4): Q14
trade between second best-reporters (alpha 2)

- For trade between second-best reporters (means of less than 5% of international trade since 1992), the method is very simple: the maximum of export and import reportings:

$$H_{ijk} = \max(X_{ijk}, M_{ijk})$$

with:

X_{ijk} : exports of product k from country i to country j reported by country i

M_{ijk} : exports of product k from country i to country j reported by country j

H_{ijk} : harmonized exports of product k from country i to country j

- For the « not elsewhere specified » product code $(k = nes)$, it is the opposite:

$$H_{ijnes} = \min(X_{ijnes}, M_{ijnes})$$

G. Harmonization (5): Q41
trade between zones with few reporters (beta 1)

In each beta 1 zone there are good reporters, for instance Costa Rica, Guatemala, Jamaica and Trinidad for « Others in America ». Each exporting zone *I* (and each importing zone *J*) is divided in two parts: {*i*} and {*j*} for good reporters (ex: Costa Rica + Guatemala + Jamaica + Trinidad) and {*I-i*} and {*J-j*} for the other countries.

For each beta 1 exporting zone *I*, each beta 1 importing zone *J* and each product *k*, the harmonized trade from zone *I* to zone *J* for product *k* is calculated as follows:

$$H_{IJ}^k = \frac{(\max(X_{\{i\}\{j\}}^k, M_{\{i\}\{j\}}^k) + X_{\{i\}\{J-j\}}^k + M_{\{I-i\}\{j\}}^k)}{(\sum_{k'}\max(X_{\{i\}\{j\}}^{k'}, M_{\{i\}\{j\}}^{k'}) + X_{\{i\}\{J-j\}}^{\bullet} + M_{\{I-i\}\{j\}}^{\bullet})} * \max(X_{IJ}^{\bullet}, M_{IJ}^{\bullet})$$

{*i*} : good reporters in zone *I*

{*I-i*} : other countries in zone *I*

$X_{\{i\}\{j\}}^k$: exports of product *k* from {*i*} to {*j*} reported by {*i*}

$M_{\{I-i\}\{j\}}^k$: imports of product *k* coming from {*I-i*} reported by {*j*}

$M_{\{I-i\}\{j\}}^{\bullet}$: total imports from {*I-i*} reported by {*j*}

$X_{IJ}^{\bullet}$: total exports from zone *I* to zone *J* (source: DOTS of the IMF or MBS of the UN…)

The breaking down by product reported by good reporting countries of beta 1 zones is applied to bilateral totals found in IMF and UN yearbooks in order to calculate harmonized data.

For <u>exports from beta 1 zones I to beta 2 zones J (Q42)</u>:

$$H_{IJ}^{k} = \frac{X_{iJ}^{k}}{X_{iJ}^{\bullet}} * X_{IJ}^{\bullet}$$

For <u>exports from beta 2 zones I to beta 1 zones J (Q43)</u>:

$$H_{IJ}^{k} = \frac{M_{Ij}^{k}}{M_{Ij}^{\bullet}} * X_{IJ}^{\bullet}$$

i : good reporters in zone I

j : good reporters in zone J

X_{iJ}^{k} : exports of product k from i to J reported by i

$X_{iJ}^{\bullet}$: total exports from i to J reported by i

M_{Ij}^{k} : imports of product k from I to j reported by j

$X_{IJ}^{\bullet}$: total exports from zone I to zone J (source: DOTS of the IMF or MBS of the UN)

21

For each exporting beta 2 zone *I*, the product breakdown of the imports of all importing countries but beta 2 zones (alpha 1, alpha 2 and beta 1 zones) coming from the beta 2 zone *J* is applied to bilateral totals found in IMF and UN yearbooks in order to calculate harmonized data.

$$H_{IJ}^k = \frac{\displaystyle\sum_{\substack{j\,describing\,alpha\,countries\\ and\,beta1\,zones}} M_{Ij}^k}{\displaystyle\sum_{\substack{j\,describing\,alpha\,countries\\ and\,beta1\,zones}} M_{Ij}^\bullet} * X_{IJ}^\bullet$$

j : countries alpha 1 and alpha 2 and zones beta 1

M_{Ij}^k : imports of product k from I to j reported by j

$M_{Ij}^\bullet$: total imports from I to j reported by j

$X_{IJ}^\bullet$: total exports from zone I to zone J (source: DOTS of the IMF or MBS of the UN)

22

H. Estimation of countries issued from former USSR, Yugoslavia, Czechoslovakia, Belgium and Luxembourg before 1992 or 1993 (1)

Motivation:

Countries issued from former USSR, Yugoslavia and Czechoslovakia were recognized as trade partners (and reporters) by the UN only from 1992 on (1993 for Czech and Slovak republics). In order to keep the same nomenclature over the whole period, the data of these countries stretching back to 1967 had to be estimated.

The same issue occurs for Belgium and Luxembourg from 1967 to 1992.
Belgium and Luxembourg began to report their trade separately to the UN in 1999. As Luxembourg compiled its exports and imports since 1993 and sent these data to the CEPII, data for Belgium from 1993 to 1998 could be estimated by removing Luxembourg data from data sent by the Belgium and Luxembourg Economic Union (BLEU) to the UN.

In the same way, the trade of Bangladesh, which was part of Pakistan before 1972, is removed from pakistanese data from 1967 to 1971.

The « intra » trade (within countries issued from former USSR, or within former Yugoslavia, before 1992, or within former Czechoslovakia, or within BLEU, before 1993, or between Bangladesh and Pakistan, before 1972) is removed in the estimations.

23

Calculation: for missing years, for each zone I, partner J and product k, the 1993 (or 1992, or 1972) breakdown of the « zone » I into countries i issued from I is applied:

$$H_{iJk}^{t} = H_{IJk}^{t} * \frac{H_{iJk}^{1993}}{H_{IJk}^{1993}} \quad \text{and} \quad H_{Jik}^{t} = H_{JIk}^{t} * \frac{H_{Jik}^{1993}}{H_{JIk}^{1993}} , \quad \text{with always} \quad H_{IIk}^{t} = 0$$

I : former USSR, former Yugoslavia, former Czechoslovakia, former Pakistan, or BLEU without intra trade
i : Belgium and Luxembourg if I = BLEU, Czech or Slovak republic if I = former Czechoslovakia, …
J : partner country

If partner J is one of the 4 other zones, $\quad H_{ijk}^{t} = H_{IJk}^{t} * \dfrac{H_{ijk}^{1993}}{H_{IJk}^{1993}}$

If $\quad H_{IJk}^{1993} = 0 \quad$ or $\quad H_{JIk}^{1993} = 0$, the breakdown for k = total or J = world is applied.

For rare remaining cases, either the breakdown for k = total <u>and</u> J = world is applied, or the total amount is affected to the <u>principal country of the zone</u> (Russia, Serbia/Montenegro, Czech republic, Belgium or Pakistan).

References

• de SAINT VAULRY, A. (2008), "Base de données CHELEM - commerce international du CEPII", CEPII Working Paper 2008-09, June 2008

• http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp

• United Nations (2004), "International Merchandise Trade Statistics: Compiler Manual", UN Statistics Division (UNSD), Department of Economic and Social Affairs, Series F, N° 87   (http://unstats.un.org/unsd/trade/methodology imts.htm)

• Stone, R., Bates, J., Bacharach, M. (1963), "Input-Output Relationships, 1954-1966", in Stone, R., "A Programme for Growth", Vol. 3, Chapman & Hall, London