

Documentation for the CEPII trade prices dataset

For internal use only, please do not circulate

Pierre Cotterlaz, Guillaume Gaulier, Aude Sztulman, Deniz Unal

26 septembre 2023

On calcule des indices de Törnqvist. Ceux-ci utilisent les variations de valeur unitaire (valeur/quantité) au sein de chaque catégorie de produit. On note $d\ln(UV_{i,j,k,t})$ la variation du logarithme des valeurs unitaires : $d\ln(UV_{i,j,k,t}) = \ln(UV_{i,j,k,t}) - \ln(UV_{i,j,k,t-1})$.

On calcule les indices de prix par groupe de produits, un groupe de produits pouvant correspondre à une branche (ISIC 2 chiffres), à un stade de production, ou à l'intersection des deux dimensions. L'indice de prix d'un groupe de produits, $T_{K,t}$, est une moyenne pondérée des variations de valeur unitaire des produits de la branche :

$$d\ln(T_{K,t}) = \sum_{i,j,k \in K} w_{i,j,k,t} d\ln(UV_{i,j,k,t})$$

Le poids reflète l'importance du flux au sein des échanges du groupe de produits, c'est une moyenne de la part du flux en t et de sa part en $t-1$.

$$w_{i,j,k,t} = \frac{1}{2} \left(\frac{v_{i,j,k,t}}{v_{K,t}} + \frac{v_{i,j,k,t-1}}{v_{K,t-1}} \right)$$

Au niveau de chaque produit, les variations de valeur unitaire aberrantes sont remplacées par des valeurs moins aberrantes. Plus précisément, on calcule pour chaque produit la distribution des variations de valeurs unitaires et l'on remplace les variations se situant en deça du 5e centile par la valeur au 5e centile, et les observations au delà du 95e centile par la valeur au 95e centile.

Dans un premier temps, nous avons exploré 4 méthodes :

1. Pas de retrait des valeurs aberrantes.
2. Retrait des valeurs aberrantes sur la base d'un filtrage non pondéré (10% des observations).
3. Retrait des valeurs aberrantes sur la base d'un filtrage pondéré (10% du w).
4. Filtrage non pondéré, puis remplacement des $d\ln(UV_{i,j,k,t})$ par leur médiane au niveau HS 4 chiffres, i, j, t , ou plus agrégé encore si besoin.

NB : l'imputation permet de faire apparaître des $d\ln(UV_{i,j,k,t})$ pour des i, j, k, t avec v_t ou v_{t-1} manquant mais on ne les conserve pas.

La troisième méthode a été retenue (filtrage pondéré). Se pose ensuite la question de la source des données : BACI ou WTFC ?

L'indice de prix "tous produits" se rapproche davantage de l'indice de référence (CPB) lorsqu'il est calculé à partir des données de WTFC que lorsqu'il est calculé à partir des données de BACI. On opte donc pour WTFC. Cependant, la base WTFC n'ayant plus vocation à être actualisée, on utilise pour les années les plus récentes (à partir de 2020) des séries d'indices de prix issues de BACI.

Nomenclature sectorielle : Il s'agit de la nomenclature ISIC à deux chiffres. Pour le moment, la table de passage SH-ISIC est obtenue via une table de passage auparavant mise à disposition par Ramon, mais plus disponible. Un cheminement alternatif serait d'utiliser les tables de passage de l'UNSD. HS96 - CPC1.0 puis CPC1.0 - CPC1.1 puis CPC1.1 - ISIC3.1.

Seuls les secteurs représentant au moins 1% des flux mondiaux sur l'ensemble de la période étudiée (en proportion du nombre d'observations) sont conservés, les autres secteurs sont regroupés en un résidu ("NEC").

Stades de production : Les stades de production sont déterminés à partir de la BEC 4.0. Le passage SH - BEC se fait à partir des tables de passage de l'UNSD.

Obtention des variables finales Le taux de croissance annuel des prix s'obtient par l'exponentielle de la variation du log de l'indice puisque :

$$d \ln(T_{K,t}) = \ln \left(\frac{T_{K,t}}{T_{K,t-1}} \right)$$

donc

$$\exp(d \ln(T_{K,t})) = \frac{T_{K,t}}{T_{K,t-1}}$$

D'autre part :

$$\frac{T_{K,t}}{T_{K,2000}} = \frac{T_{K,t}}{T_{K,t-1}} \cdots \frac{T_{K,2001}}{T_{K,2000}}$$

donc en prenant l'exponentielle de la somme cumulée en t des $d \ln(T_{K,t})$ à partir de 2000 on obtient l'indice de prix en base 2000.

C'est ainsi que l'on crée la variable **price_index_base_100**.

A partir de BACI, on obtient également le commerce total au niveau K, t , que l'on peut utiliser pour calculer un indice **trade_value_base_100** au niveau K, t .

On obtient le commerce en volume en divisant l'indice de commerce en valeur par l'indice de prix.