
Nomenclature de produits et concordance dans le temps. Procédure de correction et analyses de sensibilité

Cécile Le Roy, Monique Harel, Karine Latouche, Carl Gagné et Stéphane Turolla



Édition électronique

URL : <http://journals.openedition.org/economierurale/4466>

DOI : 10.4000/economierurale.4466

ISSN : 2105-2581

Éditeur

Société Française d'Économie Rurale (SFER)

Édition imprimée

Date de publication : 15 septembre 2014

Pagination : 87-102

ISSN : 0013-0559

Référence électronique

Cécile Le Roy, Monique Harel, Karine Latouche, Carl Gagné et Stéphane Turolla, « Nomenclature de produits et concordance dans le temps. Procédure de correction et analyses de sensibilité », *Économie rurale* [En ligne], 343 | septembre-octobre 2014, mis en ligne le 01 janvier 2016, consulté le 19 avril 2019. URL : <http://journals.openedition.org/economierurale/4466> ; DOI : 10.4000/economierurale.4466

Nomenclature de produits et concordance dans le temps

Procédure de correction et analyses de sensibilité

Cécile LE ROY • Monique HAREL • Karine LATOUCHE

INRA, UR1134 LERECO, Nantes

Carl GAIGNÉ • Stéphane TUROLLA

INRA, UMR1302 SMART, Rennes

Dans cet article, les auteurs présentent un algorithme permettant un suivi des statistiques de production ou d'échange d'un produit donné, malgré les actualisations fréquentes des nomenclatures officielles. L'intérêt de cette procédure est démontré à travers une application sur les données des douanes françaises sur les exportations et importations des produits agroalimentaires. Sans l'application de cette procédure, les gains aux échanges liés à l'introduction de nouvelles variétés peuvent être surestimés.

L'essor des recherches empiriques et théoriques en commerce international, observé au cours de ces dernières années, doit beaucoup à un accès facilité aux données statistiques de production et d'échanges. L'utilisation de ces données et les comparaisons qui en sont faites reposent sur l'existence de systèmes de classification, aussi appelés nomenclatures. Les nomenclatures telles que PRODCOM (classification des produits de la Communauté européenne) ou la nomenclature combinée (NC) permettent d'obtenir un niveau d'observation très fin.

Toutefois les nomenclatures sont soumises à de nombreuses révisions chaque année, ce qui rend difficile le suivi dans le temps des données statistiques. Bien que les données détaillées sur les catégories de biens¹ produits et échangés par les firmes soient collectées par les agences nationales de statistique, leur suivi dans le temps est rendu extrêmement complexe en raison de la difficulté à faire correspondre les

codes de produits d'une année sur l'autre². Chaque année, les nomenclatures sont actualisées de façon à assurer une représentation fidèle des produits échangés sur les marchés (innovation, obsolescence ou apparition de segments nouveaux, par exemple, viennent modifier sensiblement l'ensemble des produits offerts), de façon à garantir une harmonisation entre les différents systèmes de nomenclature de par le monde. Ainsi, pour un niveau d'agrégation donné, une catégorie de biens peut d'une année sur l'autre être remplacée, agrégée avec une ou plusieurs catégories existantes, scindée entre plusieurs nouvelles catégories (ou encore subir une combinaison de ces deux opérations) ou peut tout simplement disparaître, ce qui complexifie sensiblement le suivi dans le temps des produits destinés au marché national ou à l'export³.

1. Les termes « catégorie de biens » et « produit » sont employés indifféremment dans la suite.

2. La problématique est la même si l'on se positionne au niveau des échanges commerciaux *via* la nomenclature combinée ou au niveau des biens produits en sortie d'établissements *via* la Classification des produits française (CPF).

3. De plus, pour compliquer davantage, un numéro de code peut être réutilisé une année suivante pour désigner un produit différent.

Ces révisions de nomenclature sont loin d'être anecdotiques. À titre d'exemple, elles concernent près de 70 % des codes à huit chiffres de la nomenclature combinée pour la période 1995-2010.

En conséquence, et étant donné l'état actuel des statistiques publiques, un certain nombre de questions d'économie internationale ne peuvent être traitées convenablement. En particulier, les problématiques centrées sur la mesure des facteurs d'ajustement dits à la marge extensive (évolution dans le temps du nombre de produits échangés ; voir pour un survol de la littérature Krugman, Obstfeld, Melitz, 2011⁴) et intensive pâtissent de l'absence d'une information fiable dans le temps sur la nature des produits échangés. Si de nombreux travaux théoriques ont montré que les variations des échanges entre pays résultent essentiellement des ajustements de la marge extensive, peu de travaux empiriques ont pu valider ces prédictions à un niveau de détail satisfaisant. Les quelques exceptions reposent sur une analyse des échanges commerciaux au niveau des pays, avec pour problématique centrale la mesure d'un *effet variété*⁵ (Feenstra, 1994⁶ ; Hummels, Klenow, 2005⁷ ; Broda, Weinstein, 2006⁸ ; Goldberg

et al., 2009⁹). Bien que moins sensibles par nature au phénomène d'entrée/sortie des produits, du fait d'un niveau d'agrégation (pays), ces études n'en restent pas moins soumises au biais discuté ici. À l'inverse, très peu d'analyses empiriques conduites au niveau de la firme et utilisant des données plus précises que le niveau du système harmonisé à six chiffres (SH6) ont pu être réalisées du fait des contraintes imposées par les révisions de nomenclatures au niveau des codes à huit chiffres¹⁰.

Dans cet article, nous présentons un algorithme permettant de suivre dans le temps les codes associés aux catégories d'une nomenclature à un niveau d'information très proche du niveau le plus fin disponible. Le principe de cette procédure repose sur la construction de familles de produits *homogènes*, pour une période donnée. La traduction en algorithme des tables d'évolution des nomenclatures est un enjeu important dans de nombreux pays. À notre connaissance, seuls Pierce et Schott (2011)¹¹ ont entrepris un travail similaire sur données américaines. Confrontés au même problème de révision des nomenclatures de produits au niveau de désagrégation le plus élevé, ces auteurs proposent un algorithme permettant d'assurer la concordance dans le temps des nomenclatures de produits. Si nos deux travaux s'accordent sur la définition d'une famille

4. Krugman P. R., Melitz M. J., Obstfeld M. (2011). *International Economics: Theory and Policy*. Prentice Hall (9th Ed.)

5. L'effet variété (i.e., l'accroissement du nombre du couple produits-pays échangés) est un des principaux arguments avancés en faveur de la libéralisation du commerce international. En effet, l'intégration internationale des marchés induirait une plus grande variété de produits disponibles au niveau des économies nationales.

6. Feenstra R. C. (1994). New Product Varieties and the Measurement of International Prices. *American Economic Review*, vol. 84, n° 1, p. 157-177.

7. Hummels D., Klenow P. J. (2005). The Variety and Quality of a Nation's Exports. *American Economic Review*, vol. 95, n° 3, p. 704-723.

8. Broda C., Weinstein D. E. (2006). Globalization and the Gains from Variety. *Quarterly Journal of Economics*, vol. 121, n° 2, p. 541-85.

9. Goldberg P. K., Khandelwal A. K., Pavcnik N., Topalova P. (2009). Trade Liberalization and New Imported Inputs. *American Economic Review*, vol. 99, n° 2, p. 494-500.

10. Les rares exceptions utilisent la procédure de Pierce et Schott (2011) discutée ci-après (e.g., Bernard. A. B., Redding S. J., Schott P. K. (2011). Multiproduct Firms and Trade Liberalization. *The Quarterly Journal of Economics*, vol. 126, n° 3, p. 1271-1318 ; Martin J., Méjean I. (2011). *Low-Wage Countries Competition, Reallocation Across Firms and the Quality Content of Exports*. CEPR discussion paper 8231).

11. Pierce J. R., Schott P. K. (2011). *Concording U.S. Harmonized System Categories Over Time*. NBER Working Paper 14837.

de produits, ils diffèrent sur l'algorithme permettant d'identifier ces familles de produits, et donc sur le degré d'agrégation de la codification générée. Nous détaillerons par la suite les différences entre les deux approches.

Notre algorithme peut s'appliquer à toute nomenclature composée d'une codification chiffrée et sujette à des révisions régulières. Dans cet article, nous appliquons notre algorithme sur la NC et son système de codification à huit chiffres propre aux pays de l'Union européenne. Nous regardons ensuite les effets de cette procédure sur la NC appliquée aux données des douanes françaises pour les produits agroalimentaires échangés sur la période 1995-2010. Nous montrons que l'emploi de notre procédure sur la NC permet un suivi dans le temps des échanges des produits au niveau de la firme proche de celui obtenu au niveau NC8 (et bien plus détaillé que le niveau d'agrégation NC6).

Cet article est organisé de la façon suivante. Nous décrivons tout d'abord la nomenclature utilisée pour le commerce extérieur européen et les problèmes posés par les révisions régulières de cette nomenclature. Ensuite, nous présentons notre procédure permettant d'identifier les produits sur une période longue. Nous discutons également ses principaux avantages et la manière de gérer ses principales limites. Puis, nous mesurons les biais corrigés par notre procédure en l'appliquant à des données d'échanges de la France concernant le secteur agroalimentaire. Enfin, nous concluons.

La nomenclature combinée

En 1988, la NC a été créée par la Commission européenne pour satisfaire aux besoins des statistiques du commerce extérieur et à l'application du tarif douanier commun. Elle repose sur la nomenclature du système harmonisé (SH – nomenclature

internationale polyvalente élaborée sous l'égide de l'Organisation mondiale des douanes) et comporte des subdivisions supplémentaires créées au niveau communautaire. À chacune des sous-positions de la nomenclature combinée correspondent un code numérique à huit chiffres et un libellé. Les six premiers chiffres indiquent les positions et sous-positions de la nomenclature du système harmonisé. Les septième et huitième chiffres identifient les sous-positions de la nomenclature combinée¹². Plus précisément, un code est créé avec une date de début de validité (année de création) et une date de fin de validité forfaitaire (2500). Si un code est supprimé, la date de fin de validité est mise à jour avec l'année de fin de validité.

L'évolution de la NC est principalement liée aux mises à jour du SH afin de tenir compte des modifications arrêtées soit par l'Organisation mondiale des douanes, soit lors des négociations commerciales (OMC). Ainsi, les années 1996, 2002, 2007, années de modifications importantes, correspondent à des phases d'adaptation de la NC au SH. Par ailleurs, la NC est révisée chaque année pour tenir compte des évolutions structurelles des produits dans les échanges commerciaux (apparition de nouveaux produits ou obsolescence de certains d'entre eux, par exemple), mais aussi pour affiner les besoins de l'Union européenne en matière de politique commerciale. La NC permettant l'application des droits de douane, les révisions proposées retranscrivent l'évolution des politiques commerciales en termes de protection

12. Par exemple, la sous-position de la NC identifiée sous le code à huit chiffres 07070005 correspond aux « concombres ». Si l'on se limite aux six premiers chiffres du code (070700), on retombe sur la sous-position du SH correspondant aux « concombres et cornichons, à l'état frais ou réfrigéré » et dont le chapitre d'appartenance « légumes, plantes, racines et tubercules alimentaires » est identifié par les deux premiers chiffres (07).

FAITS ET CHIFFRES

des marchés européens. L'actualisation annuelle de la NC est publiée sous forme d'un règlement de la Commission, au journal officiel de l'Union européenne.

Au niveau NC8, trois mécanismes élémentaires conduisent à la suppression d'un code.

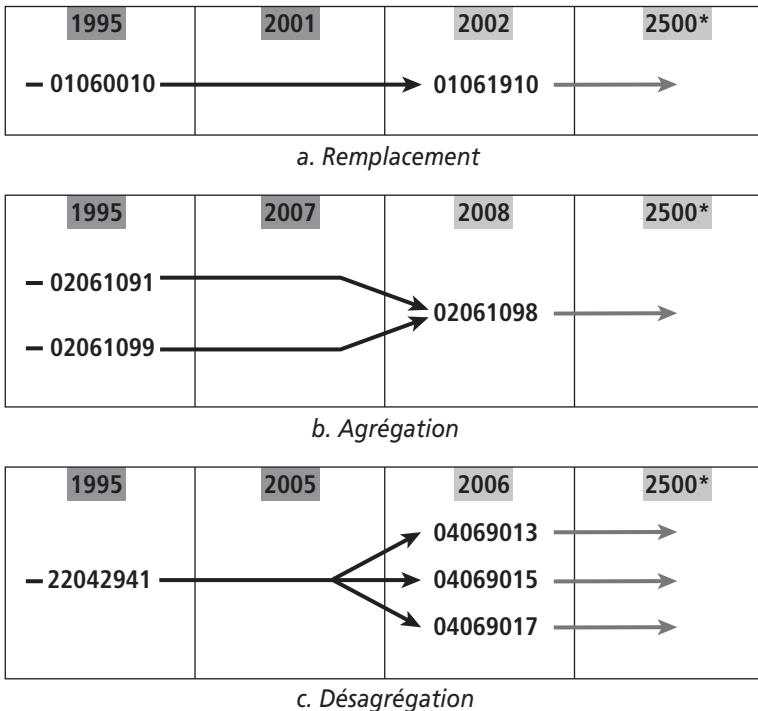
- (i) Remplacement : le produit a un nouveau numéro de code. Par exemple, le code 01060010 est remplacé en 2002 par le code 01061910 (*figure 1a*). C'est le cas le plus simple, car il implique une correspondance unique entre un code pour une année t et son remplaçant pour l'année $t+1$.
- (ii) Agrégation : deux codes sont fusionnés. Par exemple, les codes 02061091 et 02061099 sont remplacés, en 2008, par un code unique 02061098 (voir *figure 1b*).

- (iii) Désagrégation : un code produit est désagrégé en plusieurs nouveaux codes produits (voir *figure 1c*).

Ainsi, un code peut être remplacé par un ou plusieurs codes. Dans ce cas, il devient « prédécesseur de », tandis que celui ou ceux qui lui succèdent seront dits « successeurs de ».

En plus de ces trois situations s'ajoutent des combinaisons de celles-ci, par exemple la combinaison d'agrégations et de désagrégations entre deux périodes ou sur plusieurs périodes. Le produit « viandes et abats comestibles de l'espèce porcine » permet d'illustrer ce cas (*figure 2*). On constate que ce produit, défini en 2004 sous le code 02063000, est le résultat d'une succession d'agrégations et désagrégations de codes NC8 opérées depuis 1988. Autrement dit, c'est l'arbre de ses correspondances (ou

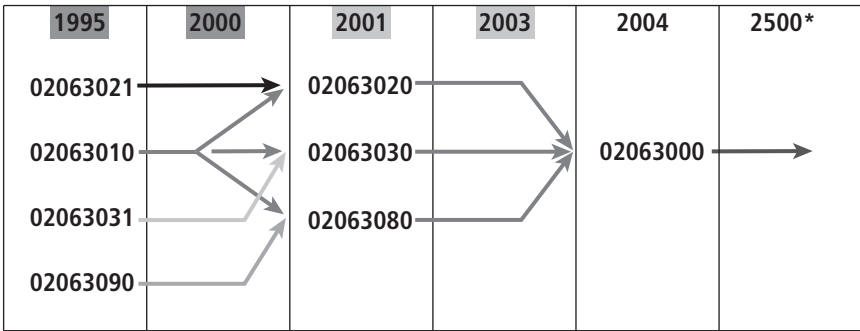
Figure 1. Exemples de changements de codes



Note : * 2500 est une date de fin de validité forfaitaire.

Source : les auteurs.

Figure 2. Cas mixte



Note : * 2500 est une date de fin de validité forfaitaire.

Source : les auteurs.

famille homogène) qui définit le produit tel qu’il est codifié en 2004.

Il apparaît donc complexe de suivre un produit dans le temps. En effet, il faut établir la ou les relations qui existent entre le ou les codes NC8 qui définissent le produit sur la période étudiée. Cette mise en relation crée un lien entre les codes NC8 et c’est ce lien qui va permettre de définir le produit sur l’ensemble de la période. Les évolutions des codes NC8 présentées précédemment sont des reconstitutions manuelles, au cas par cas. Pour généraliser cette approche, une procédure automatisée est proposée.

Procédure

L’objectif de la procédure est de créer un identifiant que l’on appellera *index* pour chaque famille de produits. Cet index identifie une famille homogène de produits en associant le ou les codes NC8 qui la définissent sur la période étudiée. Dans la suite de l’article, on nommera *produit générique*, une famille homogène de produits (i.e. l’ensemble du ou des codes NC8 liés dans le temps) auxquels on affecte un identifiant numérique, que l’on nomme *index*. Autrement dit, un produit générique est composé soit d’un unique code NC8 (aussi appelé *produit mono-NC8*, et donc jamais concerné par une révision de

nomenclature) soit de plusieurs codes NC8 (aussi appelé *produit multi-NC8*).

La première difficulté repose sur l’identification des liens entre les codes NC8 d’un produit générique. Reprenons l’exemple de la figure 2. Le produit « viandes et abats comestibles de l’espèce porcine » est représenté dans le temps par huit codes NC8. Ces codes NC8 sont accompagnés d’une date de début de validité (dénote par *ddeb*) et d’une date de fin de validité (dénote par *dfin*). Par exemple, pour le code 02063021 (*ddeb* = 1988 et *dfin* = 2000), pour le code 02063030 (*ddeb* = 2001 et *dfin* = 2003) et pour le code 02063000 (*ddeb* = 2004 et *dfin* = 2500, c’est-à-dire la date définie par défaut).

La procédure qui va être décrite conduit à la création d’un identifiant commun (que l’on nomme *index*) qui sera affecté à tous les triplets (code NC8/*ddeb*/*dfin*) définissant un même produit générique. Pour l’application de la procédure aux données des douanes françaises, nous retiendrons l’année 1995 comme date de début de construction de l’identifiant.

1. Algorithme

Étape 1 : création de la table prédécesseur-successeur

La table de référence construite recense l’ensemble des changements de codes sur

Tableau 1. Extrait de la table prédécesseur-successeur

code_nc8_prédécesseur	ddeb_pre	dfin_pre	code_nc8_successeur	ddeb_suc	dfin_suc
02063010	1988	2000	02063020	2001	2003
02063010	1988	2000	02063030	2001	2003
02063010	1988	2000	02063080	2001	2003
02063021	1988	2000	02063020	2001	2003
02063031	1988	2000	02063030	2001	2003
02063090	1988	2000	02063080	2001	2003
02063020	2001	2003	02063000	2004	2500
02063030	2001	2003	02063000	2004	2500
02063080	2001	2003	02063000	2004	2500

Source : les auteurs.

la période 1995-2010. Elle est constituée de la liste exhaustive des codes NC8 pour l'année 2010, ainsi que de l'ensemble des correspondances de la NC sur la période¹³. Dans cette table, chaque ligne est formée du code NC8 qui est supprimé (*code_NC8_prédécesseur*) et du code NC8 (*code_NC8_successeur*) qui le remplace. Chaque code NC8 est complété par sa date de début de validité (*ddeb*) et sa date de fin de validité (*dfin*). La date de fin de validité est renseignée si le code a évolué sur la période ou est forfaitaire (*dfin* = 2500) si le code est toujours valide en fin de période. Ceci est illustré dans le *tableau 1*. On peut donc avoir plusieurs lignes avec le même triplet prédécesseur si le code NC8 est désagrégé. De même, on peut avoir plusieurs

lignes avec le même triplet successeur si le code NC8 est issu d'une agrégation¹⁴.

Par ailleurs, les codes NC8 non concernés par une évolution de la NC au cours de la période se voient affectés un enregistrement propre. Pour ces enregistrements, les modalités des variables précédentes sont renseignées comme suit : *code_nc8_prédécesseur* = code NC8, *ddeb_pre* = date de début de validité, *dfin_pre* = 2500, et les champs *code_nc8_successeur*, *ddeb_suc* et *dfin_suc* sont renseignés par défaut avec la valeur « null ».

Étape 2 : création de pré-index

La table prédécesseur-successeur est lue séquentiellement dans l'ordre croissant des dates de début de validité des codes NC8

13. Ces données sont disponibles sur le site de la Commission européenne aux adresses suivantes : http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?sort=1&file=comext%2F201401%2Ftext%2Fnc_successor_predecessor.txt et http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?sort=1&file=comext%2F201401%2Ftext%2Ffrancais%2Fnc.txt.

14. Dans l'extrait reproduit, il apparaît, en 2001, que le code 2063010 est désagrégé en trois codes NC8 (2063020, 2063030 et 2063080). En même temps, ces trois nouveaux codes sont le résultat des agrégations suivantes : 2063020 agrège 2063010 et 2063021, 2063030 agrège 2063010 et 2063031, 2063080 agrège 2063010 et 2063090. Cette configuration correspond donc au cas mixte combinant les cas agrégation et désagrégation. Puis en 2004, les nouveaux codes sont agrégés pour former un unique code, le 2063000.

Figure 3. Exemple de création de pré-index

1995		2000	2001		2003	2004		2500
Code_nc8	pre_index		Code_nc8	pre_index		Code_nc8	pre_index	
02063021	19882734							
			02063020	19882734,19882733				
02063010	19882733							
			02063030	19882733,19882735		02063000	19882734,19882733, 19882735,19882736	
02063031	19882735		02063080	19882733 19882736				
02063090	19882736							

Source : les auteurs.

prédécesseur. Lors de la première occurrence du triplet prédécesseur (code_NC8_pre, ddeb_pre, dfn_pre), on lui affecte un pré-index. Le pré-index est une valeur numérique correspondant à la concaténation de la date de début du prédécesseur et d'un numéro qui s'incrémente par pas de 1. On affecte ce même pré-index à chaque triplet successeur de ce triplet prédécesseur.

En conséquence, à la fin de la lecture de la table, on obtient une nouvelle table dans laquelle tous les triplets (code NC8/date début/date fin) sont affectés d'un ou plusieurs pré-index. Cette étape est illustrée dans la figure 3 à partir de l'exemple de la figure 2. On a, dans ce cas précis, quatre triplets prédécesseurs distincts, d'où la création de quatre pré-index en 2004¹⁵.

Étape 3 : création des index

À partir de cette table nouvellement créée, composée de l'ensemble des triplets (code NC8/date début/date fin) et de leur(s) pré-index, on affecte le même index à tous les triplets qui ont un ou des pré-index identiques. En conséquence, chaque produit est affecté d'un index qui rassemble le ou les codes NC8 définissant ce produit générique dans le temps.

Si l'on poursuit notre exemple, représenté par la figure 3, on affecte au triplet le plus récent (02063000/2004/2500) un index (dans notre cas présent index = 29). Puis, on affecte à tous les triplets qui ont le ou les mêmes pré-index que lui le même index = 29. Ainsi, cet index permet d'identifier tous les codes NC8 qui définissent un produit générique pour une période donnée¹⁶.

Enfin, il peut arriver que deux triplets ayant un même code NC8 (mais des dates de début de validité et des dates de fin de validité différentes, autrement dit deux triplets différents) n'aient pas le même index. Ce cas apparaît lorsqu'aucun lien n'existe entre ces triplets lors des révisions (donc dans la table prédécesseur-successeur).

2. Avantages et limites

Cette procédure permet non seulement une cohérence dans le temps du suivi des produits échangés, mais elle a aussi pour avantages (i) d'être applicable quelle que soit la date de début et de fin et (ii) de détecter l'apparition et la disparition de produits pendant la période définie. Enfin, cette procédure produit exactement les mêmes résultats, quel que soit l'ordre chronologique de lecture des tables de

15. Les programmes sont écrits en PHP et les bases de données sont gérées sous PostgreSQL. Le programme STATA de la procédure, plus coûteux en temps de calcul, est téléchargeable sur la page internet des auteurs (<http://www6.rennes.inra.fr/smart/Membres>).

16. Toujours dans notre exemple, notons que si la période d'étude porte sur 1995-2001, les triplets générés en 2001 sont liés par un pré-index commun ; ce qui conduit à la création d'un seul produit générique.

FAITS ET CHIFFRES

correspondance (*i.e.* de l'année la plus ancienne vers la plus récente ou inversement).

Cependant, cette procédure peut potentiellement biaiser les résultats dans les deux cas suivants. Tout d'abord, l'apparition d'un nouveau code NC8 peut ne pas être détectée si, quelques années après, il est agrégé avec un code existant depuis le début de la période. Plus la période est longue, plus cette probabilité est élevée. De la même façon, la disparition d'un code NC8 peut ne pas être observée si celui-ci appartient à un index composé de plusieurs codes. Pour illustrer ce problème, prenons l'exemple suivant où l'on a six périodes (T1 à T6), deux codes à la date de départ T1 ([A] et [B]) et deux codes à la date d'arrivée T6 ([A2] et [BC]). Les six périodes sont dans l'ordre d'apparition T1, T2, T3, T4, T5 et T6. En T2, un nouveau code apparaît avec le code [C] et en T3, le produit [A] est désagrégé en code [A1] et en code [A2]. En T4, un nouveau code [BC] est créé par la fusion de [B] et [C] et en T5 le code [A1] disparaît. En T6, il reste donc deux codes : [A2] et [BC]. Ceci est illustré dans la *figure 4*.

Si la période d'analyse débute en T1 et s'arrête en T6, alors notre procédure aboutit à deux index, et donc deux produits génériques. On conclurait à tort que le nombre de produits n'a pas évolué dans le temps. Si la période d'analyse commence en T1 et se termine à la fin de T3,

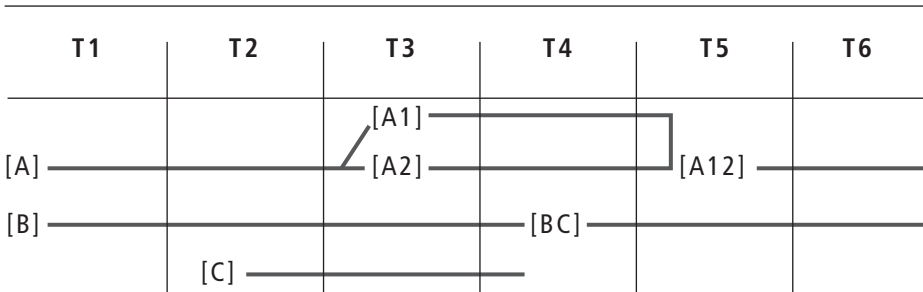
alors notre procédure aboutit à trois index, signifiant une hausse du nombre de produits. De même, si l'analyse va de T4 à T6, le nombre d'index est égal à 3 en début de période puis seulement de 2 en fin de période. On conclut donc à une baisse du nombre de produits.

Il est donc nécessaire de tester la sensibilité du nombre de produits génériques identifiés à des changements des années de départ et d'arrivée.

3. Tests de sensibilité

Nous proposons dans cette sous-section de tester la robustesse de notre procédure au regard de l'évolution du nombre de produits dans le secteur agroalimentaire. Les produits issus des secteurs de l'agriculture et de l'agroalimentaire sont représentés par les 24 premiers chapitres de la NC. Dans la liste exhaustive 2010, on recense 3 825 codes pour le secteur agroalimentaire, dont 2 251 sont en cours de validité et 1 151 ont expiré en raison d'une date de fin de validité antérieure à 2010. La mise en œuvre de notre procédure sur cette liste exhaustive a permis de générer 2 020 index pour la période 1995-2010 (voir *tableau 2*). Le niveau d'agrégation reste donc faible si l'on compare aux 2 251 codes NC8 recensés en 2010. En moyenne, on dénombre 1,6 codes NC8 par produit générique. Mais il est à noter que près de 72 % des produits génériques sont représentés par un seul code NC8 (produits

Figure 4. Sensibilité des résultats à la période retenue



Source : les auteurs.

Tableau 2. Nombre de produits génériques selon la période initiale

	1988-2010	1995-2010	2000-2010
Nb de produits génériques	1 749	2 020	2 089
dont % de produit mono-NC8	63	72	75
Nb de codes NC8/produit générique	2.08	1.6	1.4

Source : les auteurs.

mono-NC8). En revanche, 1,2 % des produits génériques sont constitués de 8 à 31 codes NC8 (produits multi-NC8).

Si l'analyse porte sur la période 2000-2010, on obtient un nombre légèrement supérieur de produits génériques (2 089 index ont été créés) et plus de 75 % des produits génériques sont dits mono-NC8. Cependant la moyenne de codes NC8 par produit générique reste comparable (1,4) et 1 % des produits génériques sont constitués de 6 à 28 codes NC8. Comme cela est attendu, plus la période est courte, plus le nombre d'index augmente et plus la moyenne de codes NC8 par index est faible. Toutefois, même en raccourcissant la période de cinq ans (11 ans au lieu de 16 ans), les résultats changent peu. À l'inverse, si la procédure porte sur la période 1988-2010 (soit un allongement de sept ans), le nombre de produits génériques baisse (1 749) et la moyenne de codes NC8 par produit générique augmente (2,08). Le nombre de produits génériques demeure cependant relativement élevé même si la période s'étale sur 22 ans. Notons que seul un nouveau produit apparaît sur l'ensemble de la période 1995-2010 (l'Oléorésine de vanille, un extrait naturel de vanille).

En complément de l'analyse de sensibilité menée sur la date de début de la procédure, nous regardons si nos résultats ne sont pas liés à des spécificités propres au secteur agroalimentaire, pour lequel la nomenclature est très détaillée. Pour ce faire, nous avons appliqué notre procédure sur l'ensemble de la NC (chapitres 1

à 99), toujours sur la période 1995-2010. Au total, nous recensons 17 503 codes NC8, dont 9 443 sont toujours valides en 2010 et 8 060 ont expiré. Une fois notre procédure employée, nous comptabilisons 7 780 produits génériques, dont 70 % (soient 5 475 produits génériques) sont mono-NC8. Les autres produits génériques sont composés de 2 à 359 codes NC8. En moyenne, les produits génériques sont composés de 1,99 codes NC8. Malgré des changements structurels plus fréquents dans le secteur manufacturier (nombreuses apparitions ou disparitions de produits, changements de codification à des niveaux plus agrégés), ce qui pouvait laisser présager un nombre plus faible de produits mono-NC8, nous obtenons un pourcentage de produits mono-NC8 sensiblement équivalent au cas spécifique des 24 premiers chapitres de la NC.

4. Comparaison avec la procédure de Pierce et Schott (2011)

Nous discutons à présent des différences entre notre méthodologie et celle proposée par Pierce et Schott (2011), à partir d'une comparaison sur les 24 premiers chapitres de la NC définie sur la période 1995-2010.

Nous reportons en annexes le *tableau 5* qui comptabilise la fréquence de produits génériques définis par le nombre de codes NC8 les composant. Globalement, les deux procédures donnent des résultats sensiblement identiques bien que des différences existent. La procédure de Pierce et Schott génère 2 017 produits génériques, tandis que nous obtenons 2 020 produits

avec notre procédure. En comparant plus précisément les codes NC8 composant les produits génériques, il apparaît que plus de 99 % des contenus sont identiques. Le tableau 5 en annexes précise le détail de la comparaison du contenu des produits génériques obtenus par les deux procédures. Les différences observées concernent principalement les produits génériques mono-NC8 et ceux regroupant un très grand nombre de codes NC8.

Ces différences entre procédures trouvent leur explication dans l'information utilisée pour définir les produits génériques. Comme présenté dans la section 3, nous considérons le triplet (code NC8/ddeb/dfin), et non uniquement le couple code NC8 et date de création. Cette précision est importante, car il arrive que certains triplets composés d'un même code NC8, ne soient pas liés dans la table pré-décès-successeur. Plus précisément, deux triplets qui se suivent d'une année sur l'autre peuvent être composés d'un même code NC8, mais comporter des dates de début et dates de fin de validité différentes. Cela implique que ces triplets n'ont pas de lien entre eux. Dans ce cas, deux produits génériques distincts sont créés avec notre procédure. À l'inverse, la procédure de Pierce et Schott attribue le même index à ces deux triplets et conduit à la création d'un seul produit générique. C'est par exemple la raison pour laquelle nous obtenons un produit générique à 53 codes NC8 à la suite de l'utilisation de la procédure de Pierce et Schott, alors que, avec notre procédure, le produit générique le plus agrégé comprend 31 codes NC8. La prise en compte du triplet (code NC8/ddeb/dfin) nous permet d'assurer une plus grande cohérence dans la définition du produit générique.

Intérêt de la procédure de correction

Le cas des produits agroalimentaires

Dans la suite de cet article, nous appliquons notre procédure au fichier des douanes françaises (dénommé par la suite fichier des « douanes ») pour les produits agroalimentaires échangés sur la période 1995-2010. La comparaison des évolutions obtenues, avec ou sans l'utilisation de la procédure, permet de souligner l'intérêt de la méthode, car nous montrons qu'ignorer le problème de la concordance dans le temps de la NC peut conduire à des conclusions erronées et de sens opposés.

Les données des douanes françaises sont mises à disposition par le département des statistiques de la direction générale des douanes et droits indirects. Ce service a en charge le suivi des échanges extérieurs de la France avec les 26 autres États membres de l'Union européenne et les autres pays du reste du monde (nommés « pays tiers »). Il publie les résultats du commerce extérieur (importations et exportations en valeurs et en quantités) désagrégés par firme, pays partenaire et produit au niveau NC8.

1. La création de produits multi-NC8 génère-t-il un biais d'agrégation ?

Préalablement à l'étude de l'évolution des échanges commerciaux, il convient de s'interroger sur d'éventuels biais que pourrait introduire l'utilisation de notre procédure. En particulier, le regroupement de plusieurs codes NC8 sous un même index pourrait artificiellement augmenter les volumes des échanges pour cet index générant un biais d'agrégation (pour rappel, un produit générique est identifié par un index). Afin de quantifier l'importance des produits multi-NC8 dans notre application, nous déterminons dans un premier temps la part en valeur de ces produits dans les échanges internationaux de la France. Plus précisément, après

Tableau 3. Répartition des échanges par catégorie de produit générique

Partie A : Exportations										
	Nb de codes NC8 par produit générique									
	1	2	3	4	5	7	8	11	15	17
Exportation (%)	79.1	5.9	2.7	1.0	0.7	0.1	0.6	2.0	0.5	7.4
Nb Firmes	14 218	3 922	2 119	1 608	1 087	111	1 320	1 955	626	7 102
Nb Pays de destination	197	159	128	127	89	37	77	109	71	151
Partie B : Importations										
	Nb de codes NC8 par produit générique									
	1	2	3	4	5	7	8	11	15	17
Importation (%)	84.5	6.4	4.7	1.0	1.6	0.1	0.5	0.3	0.6	0.3
Nb Firmes	17 176	5 195	3 027	1 653	954	283	237	490	217	591
Nb Pays d'origine	193	113	126	70	81	32	22	46	28	26

Notes : Les données sont reportées pour l'année 2000. Les produits sont détaillés selon le nombre de codes NC8 qui les compose. Les exportations (importations) sont exprimées en pourcentage de la valeur totale des exportations (importations). Au total, 22 119 entreprises exportent et 18 872 entreprises importent en 2000. On dénombre 210 pays de destination des exportations et 196 pays d'origine des importations.

Source : les auteurs.

avoir appliqué notre procédure à la NC afin de générer les produits génériques, et ainsi assurer la concordance dans le temps de la nomenclature, nous regardons la répartition des firmes exportatrices et importatrices, ainsi que celle des pays de destination, en fonction du nombre de codes NC8 par produit générique.

Nous reportons dans la partie A du tableau 3 la part des exportations pour les produits composés de n codes NC8 ($n = 1, \dots, 17$), ainsi que le nombre de firmes et pays destinataires associés. On observe que les exportations sont principalement concentrées sur les produits mono-NC8. Autrement dit, notre procédure de reconstitution de produits concerne 21 % des exportations en valeur. À l'inverse, les produits multi-NC8 concernent un grand nombre de pays de destination et de firmes¹⁷.

Concernant les importations (partie B du tableau 3), nous pouvons observer que notre procédure de suivi dans le temps de

produits impacte seulement 15,5 % des importations en valeur confirmant la concentration des importations sur les produits mono-NC8. Le nombre de pays d'origine des produits multi-NC8 et des firmes important ces produits est très sensiblement inférieur à celui issu des exportations.

Dans un deuxième temps, nous vérifions que la création de produits génériques composés de plusieurs codes NC8 n'induit pas de biais dans les volumes exportés et importés au niveau de la firme. Nous reportons dans le tableau 4 quelques statistiques descriptives résumant les distributions des valeurs moyennes à l'exportation et à l'importation par firme et par produit. Une distinction est faite entre les montants échangés selon le degré d'agrégation du produit générique (mono-NC8 vs multi-NC8). À titre d'exemple, nous observons que les 90^{es} percentiles des montants exportés par couples firme-produit diffèrent sensiblement entre les produits mono et multi-NC8 (33.17 vs 27.52 milliers d'euros). Dans l'ensemble, cette différence entre les types de produits est visible sur l'ensemble des supports des distributions en raison

17. Les produits composés de 11 et 17 codes NC8 concernent essentiellement le secteur du vin.

Tableau 4. Distribution des exportations et importations par firme/produit générique

	Firmes exportatrices		Firmes importatrices	
	Produit mono-NC8	Produit multi-NC8	Produit mono-NC8	Produit multi-NC8
P10	0.07	0.13	0.08	0.11
P25	0.61	0.88	0.45	0.58
P50	2.28	2.64	2.18	2.38
P75	8.92	8.72	10.56	10.82
P90	33.17	27.52	41.03	40.58

Notes : Les résultats sont détaillés par percentiles définis sur la distribution des montants exportés et importés par firme et par type de produit générique pour l'année 2000. Les valeurs reportées sont exprimées en milliers d'euros.

Source : les auteurs.

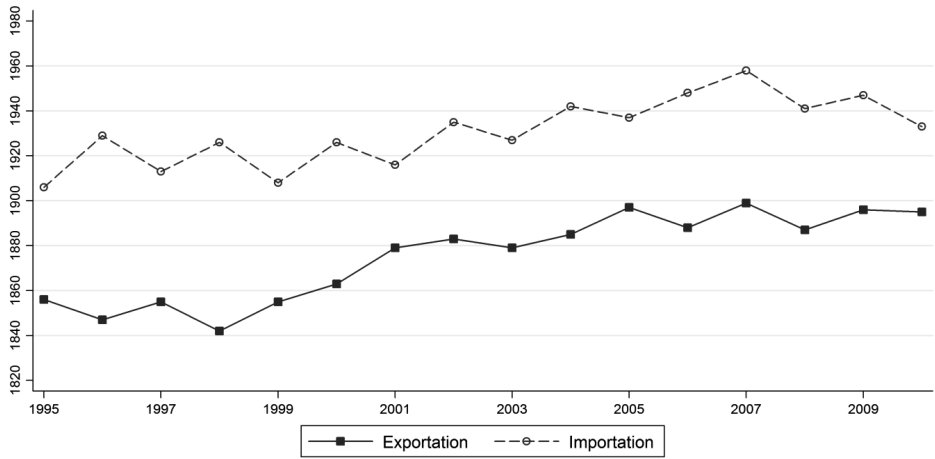
d'une distribution moins étalée pour les produits multi-NC8. L'emploi d'un test de Komogorov-Smirnov (p-value = 0.0001, dans les deux cas) permet de confirmer statistiquement ces différences entre distributions. Néanmoins, la comparaison des distributions des montants échangés entre produits mono et multi-NC8 ne permet pas d'affirmer l'existence d'un biais d'agrégation, car nous ne contrôlons pas pour les déterminants spécifiques à la firme et au pays partenaire dans les échanges. À cet effet, nous conduisons une analyse économétrique où nous régressons les volumes exportés (importés) par produit au niveau des firmes sur une variable binaire égale à 1 pour les produits multi-NC8. Les différentes composantes des échanges au niveau de la firme et du pays de destination (d'origine) sont contrôlées à l'aide d'un ensemble d'effets-fixes définis au niveau en question. L'échantillon utilisé est limité aux firmes qui exportent (ou importent) à la fois des produits mono et multi-NC8, ce qui permet de contrôler la capacité propre à la firme d'exporter (importer) des volumes élevés par exemple. Les coefficients estimés sont reportés dans le *tableau 6* en annexes. Le coefficient associé à la variable « produit multi-NC8 » n'étant pas positif et significatif, nous pouvons conclure à une absence de biais d'agrégation.

2. Exemple de correction du biais de « concordance »

La concordance dans le temps de la nomenclature étant assurée, nous allons à présent montrer l'intérêt de la procédure à partir d'un cas d'étude. La *figure 5* représente l'évolution du nombre de produits génériques exportés et importés chaque année par la France¹⁸. La représentation graphique permet d'observer une légère augmentation dans le temps du nombre de produits exportés, surtout en fin de période, passant de 1 856 produits en 1995 à 1 895 produits en 2010 (soit une hausse d'environ 2,1 %). L'accroissement de la concurrence internationale ne semble donc pas induire une réduction du nombre de produits disponibles sur les marchés internationaux en provenance de France. De même, le nombre de produits importés a légèrement augmenté sur cette même période, environ 2 % (de 1906 à 1933 produits). Ces constats demeurent si on utilise notre procédure de nomenclature pour les codes valides dès 1988 ou valides à partir de 2000.

18. Le *tableau 7* en annexes détaille le nombre de produits (définis soit au niveau du code NC8, soit au niveau du code NC6, ou soit au niveau du produit générique) existant dans la nomenclature officielle. Le nombre de produits réellement échangés par la France (importés ou exportés) est logiquement inférieur.

Figure 5. Évolution du nombre de produits génériques échangés par la France



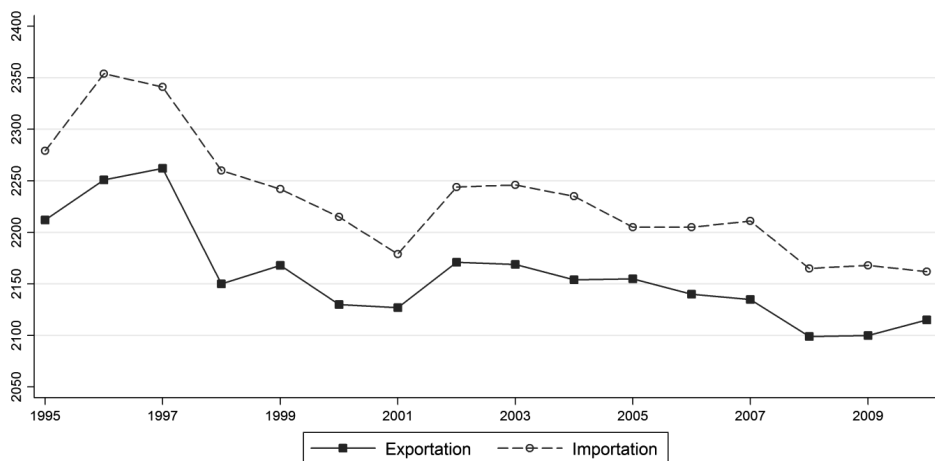
Source : Statistiques issues de la procédure appliquée aux douanes françaises 1995-2010.

A contrario, si l'on reproduit les mêmes calculs à partir du fichier des douanes sans corriger la nomenclature à l'aide de notre procédure, on constate une baisse du nombre de produits NC8 exportés et importés dans le temps (figure 6). Ceci révèle bien tout l'intérêt de notre procédure. En effet, dans le cas non corrigé, la baisse du nombre de produits échangés est due à une

hausse de la fréquence des agrégations de codes NC8 lors des révisions de la NC.

À la lumière de ces résultats, il apparaît intéressant d'utiliser notre procédure pour revisiter la question des gains aux échanges liés à l'introduction de nouvelles variétés. La littérature en économie internationale s'est longuement penchée sur cette question, sans toutefois prendre en

Figure 6. Évolution du nombre de codes NC8 échangés par la France



Source : Statistiques issues des Douanes françaises 1995-2010.

considération le problème soulevé par les révisions de nomenclatures. Dans notre exemple, nous souhaitons savoir si les nouveaux produits exportés ou importés en 2010 ont contribué fortement ou faiblement à l'accroissement de la valeur des exportations (ou importations). Pour évaluer la contribution de ces nouveaux produits, nous avons adopté la démarche développée par Broda et Weinstein (2006).

Pour les exportations, nous calculons le ratio correspondant à la valeur des exportations de produits génériques présents à la fois en 1995 et en 2010 (au prix de 2010) divisée par la valeur de l'ensemble des produits génériques exportés en 2010 ($\lambda_{x,générique,2010}$), soit :

$$\lambda_{x,générique,2010} = \frac{\sum_{g \in I_g} Val_{x,générique,2010}}{\sum_{gt \in I_{gt}} Val_{x,générique,2010}}$$

où I_g est l'ensemble des produits génériques communs présents en 1995 et 2010 et I_{gt} est l'ensemble de produits génériques présents en 2010, $Val_{x,générique,2010}$ est la valeur exportée par la France en 2010 pour les produits génériques concernés.

La valeur obtenue nous renseigne ainsi sur le poids des produits nouvellement introduits en 2010 en comparaison de ceux déjà exportés en 1995 : plus ce ratio est faible, plus le poids des nouvelles variétés est important. Une fois notre procédure appliquée, nous observons que 1 856 produits génériques sont exportés par la France en 1995 et ce chiffre s'établit à 1 895 pour 2010. Au total, 1 789 produits génériques sont communs aux deux années. Précisément, 106 nouveaux produits sont exportés en 2010 tandis que 67 produits présents en 1995 ne sont plus exportés en 2010. Nous obtenons un ratio très proche de 1 ($\lambda_{x,générique,2010} = 0,999$), traduisant une très faible contribution en valeur des nouveaux produits à l'évolution des exportations. À l'inverse, si nous reproduisons ce

calcul sans appliquer la correction de notre procédure, autrement dit si les produits sont identifiés par leur code NC8¹⁹, nous obtenons un ratio sensiblement différent ($\lambda_{x,NC8,2010} = 0,817$). Nous montrons ainsi que sans l'application de notre procédure, nous surestimons de 18,2 % les gains aux échanges liés à l'introduction de nouvelles variétés.

Le biais de concordance de la NC est encore plus sévère si l'on reproduit notre calcul sur les importations. Sans l'application de notre procédure nous obtenons un ratio $\lambda_{m,NC8,2010} = 0,700$, alors qu'une fois assurée la concordance dans le temps de la NC à l'aide de notre procédure, nous déterminons un ratio proche de 1 ($\lambda_{m,générique,2010} = 0,999$), soit un biais de près de 30 %.

Cet exemple soulève l'intérêt central de notre procédure dès lors que l'on souhaite analyser l'évolution des échanges d'un produit défini à un niveau très fin de la nomenclature, et ce quelle que soit l'entité à l'origine de la transaction (pays ou firme).

*
**

Dans cet article, nous présentons un algorithme permettant de suivre dans le temps les codes associés aux catégories d'une nomenclature à un niveau d'information très détaillé. Après avoir exposé les différentes étapes de l'algorithme, nous appliquons notre procédure aux données des douanes françaises pour les produits agroalimentaires. Nous obtenons ainsi une nouvelle classification des produits, construite à partir de la NC à huit chiffres, qui a pour avantage d'être insensible aux révisions de la nomenclature. À travers l'exemple de l'analyse de l'évolution des

19. Si l'on considère les exportations françaises pour les mêmes années, on compte 2 212 NC8 exportés par la France en 1995, 2 115 NC8 en 2010 et 1 379 NC8 communs aux deux années, soient 833 codes NC8 qui ne sont plus exportés et 736 nouveaux exportés.

produits échangés par la France entre 1995 et 2010, nous montrons que l'absence de concordance dans le temps de la NC peut conduire à des résultats contraires à ceux obtenus une fois ce problème solutionné à l'aide de notre procédure.

Plus généralement, les biais mesurés lors d'un simple calcul des gains à l'échange induits par les variétés nouvellement

introduites incitent à appliquer systématiquement ce type de procédure dès lors que l'on souhaite étudier la dynamique des échanges à un niveau de catégorisation très détaillé. ■

Les auteurs remercient les deux rapporteurs pour leurs commentaires et suggestions qui ont permis d'améliorer la qualité de leur article.

ANNEXES

Tableau 5. Comparaison de la procédure avec Pierce et Schott (2011)

Nb de codes NC8 par produit générique	Nb de produits génériques		Nb de produits génériques identiques
	Procédure	Pierce et Schott (2011)	
1	1 463	1 473	1 463
2	274	266	264
3	147	145	145
4	45	46	45
5	40	38	38
6	16	16	16
7	11	11	11
8	8	9	8
9	4	4	4
10	1	1	1
11	1	0	0
13	1	1	1
14	1	1	1
15	2	2	2
16	1	1	1
18	2	0	0
20	1	0	0
26	1	1	1
31	1	1	1
53	0	1	0
Total	2020	2017	

Note : La dernière colonne du tableau recense le nombre de produits génériques regroupant les mêmes codes NC8 entre la procédure décrite dans cet article et celle de Pierce et Schott (2011).

Source : les auteurs.

FAITS ET CHIFFRES

Tableau 6. Régressions des volumes exportés et importés par firme et par produit

Variable	Firmes exportatrices		Firmes importatrices	
	(1)	(2)	(3)	(4)
Produit multi-NC8	-10.0291*** (2.4531)	-5.7292 (6.4610)	-5.8696 (4.4054)	-1.9367 (7.5778)
Constante	65.7093*** (0.9693)	64.0102*** (2.5531)	94.9238*** (1.7450)	93.3660*** (3.0016)
Effets-Fixes Firmes	Oui	Non	Oui	Non
Effets-Fixes Firme-Pays	Non	Oui	Non	Oui
Clusters	Firme	Pays	Firme	Pays
R ²	0.2306	0.5094	0.1888	0.5670
Obs	75 706	75 706	122 968	122 968

Notes : *, **, *** indique une significativité à 10 %, 5 %, 1 % (respectivement). Les données correspondent à l'année 2000. Les erreurs standards robustes sont données entre parenthèses et clusterisées soit au niveau des firmes, soit au niveau du pays de destination ou d'origine. Les volumes exportés et importés sont exprimés en milliers d'euros.

Source : les auteurs.

Tableau 7. Nombre de produits dans la NC selon le niveau d'agrégation

Année	Nb de NC8	Nb de NC6	Nb de produits génériques
1995	2 435	697	2 019
1996	2 475	704	2 020
1997	2 475	704	2 020
1998	2 379	704	2 020
1999	2 378	704	2 020
2000	2 333	704	2 020
2001	2 302	704	2 020
2002	2 354	729	2 020
2003	2 360	729	2 020
2004	2 333	729	2 020
2005	2 313	729	2 020
2006	2 301	729	2 020
2007	2 292	722	2 020
2008	2 262	722	2 020
2009	2 252	722	2 020
2010	2 251	722	2 020

Source : les auteurs.