

Les 50 ans de la revue Statéco : analyse rétrospective du profil thématique et de ses auteurs

A. S. Berete, P. Brion, M. Kouki et C. Torelli*

En s'appuyant sur un corpus d'articles publiés au cours des cinquante années d'existence de la revue Statéco, l'article propose une analyse rétrospective tant du statut des auteurs que des thématiques abordées par la revue. D'une part, elle met en évidence la richesse de la production en méthodologie et analyse statistique et économique ; d'autre part, la montée progressive de la part des auteurs issus des pays en développement, en particulier d'Afrique subsaharienne et du Maghreb. Au regard du projet éditorial initial, si l'objectif semble atteint, il n'en demeure pas moins que, dans un contexte de demande accrue de statistiques au plan international et d'expansion de l'utilisation des données massives ouvrant de nouvelles perspectives à la statistique, il est plus qu'important de continuer à capitaliser les réflexions et actions méthodologiques entreprises dans les pays en développement, et de les diffuser au plus grand nombre d'acteurs de la coopération et plus largement de la communauté du développement en France et à l'étranger.

Introduction

Statéco a 50 ans : plus de 600 articles ont été publiés dans 116 numéros, rédigés ou co-rédigés par environ 550 auteurs. Créée en 1972 par l'Insee, Statéco est une revue méthodologique et analytique à destination des statisticiens et des économistes travaillant sur les enjeux socio-économiques et démographiques des pays en développement et en transition. Depuis 2004, l'Insee publie la revue en partenariat avec l'Institut de recherche pour le développement (IRD) et Afristat, renforçant ainsi son caractère scientifique et de terrain. La revue assure une double mission : celle d'informer sur les méthodes et outils statistiques, qui de ce fait donne aux cadres statisticiens et enseignants-chercheurs une tribune pour la valorisation de leurs travaux.

Dans l'éditorial du premier numéro, en 1972, Albert Ficatier insistait sur le côté « Bulletin de liaison¹ » que devait avoir la revue, afin de répondre à la demande

formulée par les statisticiens des « Etats africains et malgache » de sortir d'un certain isolement. L'idée était de prendre du recul par rapport à l'état des connaissances et des pratiques en matière de statistiques, qui s'appuyait alors principalement sur celles des pays du Nord, et de favoriser les échanges d'expériences entre les statisticiens du développement, du Nord et du Sud.

Cinquante ans après, on peut dire que la revue a largement réalisé cet objectif. Ces cinquante années ayant connu beaucoup de bouleversements, tant au niveau des objets d'étude des statisticiens que des méthodes qu'ils ont utilisées, on constate qu'au-delà du legs méthodologique unique que la somme d'articles publiés constitue, celle-ci raconte l'histoire de la coopération statistique française pendant cette période.

Entre 1972 et 2022, des évolutions techniques majeures ont eu lieu, avec des impacts forts pour les statisticiens. En particulier, l'utilisation de l'informatique a beaucoup évolué, et l'apparition d'Internet a modifié la

¹ *Bulletin de liaison des statisticiens et économistes exerçant leurs activités en Afrique, premier numéro paru en février 1972.*

*Aboubacar Sidiky Berete est ingénieur statisticien aberete327@gmail.com, Mokhtar Kouki est professeur à l'Université de Carthage, au sein de l'École supérieure de la statistique et de l'analyse de l'information de Tunis mokhtar.kouki@gmail.com, Philippe Brion a été rédacteur en chef de Statéco (1992-1997) philippe.brion55@gmail.com et Constance Torelli était chargée des programmes de coopération à l'Insee au moment de l'écriture de cet article constance.torelli@gmail.com

donne en matière d'accès à l'information, et aussi pour la mise à disposition des résultats produits. Les outils de collecte ont également évolué, intégrant des nouvelles technologies comme la saisie portable, ou la télédétection.

Dans le même temps, les pays concernés par la coopération statistique ont connu des histoires variées, avec pour certains des crises économiques et politiques majeures, avec également l'influence grandissante des organisations internationales (en particulier au travers des plans d'ajustements structurels² dans les années 1980 et 1990), l'impact de décisions politiques supranationales comme la dévaluation du franc CFA de 1994, ou encore les bouleversements internationaux comme la chute du mur de Berlin en novembre 1989.

La situation a également évolué dans les services statistiques des pays ciblés par la coopération en matière de statistiques. Les organisations internationales, avec la mise en place de grands programmes d'enquêtes, ont beaucoup pesé sur leur organisation, au risque de la déséquilibrer parfois. Parallèlement, la production des services a dû s'adapter à des problématiques nouvelles, comme les objectifs du Millénaire pour le développement en 2000, ou plus récemment les objectifs de développement durable sous l'égide des Nations unies (UN DESA, 2015).

Durant la même période, la coopération française en matière de statistiques a connu de gros bouleversements. Alors que l'on comptait, dans les services statistiques des pays en développement – très majoritairement en Afrique subsaharienne - environ 80 assistants techniques français (dont une trentaine d'agents de l'Insee) à la fin des années 1970 (Winter, 2010) et encore une dizaine jusqu'à la fin des années 2000, il n'existe quasiment plus d'assistants techniques français en poste dans ces services. Par ailleurs, les modalités de l'appui technique en matière de statistiques ont été revues dans le sens d'une plus grande mutualisation, avec en particulier la création d'Afristat, institution internationale résultant de la signature d'un traité inter-Etats en 1993 (Diop, 1995).

La revue Statéco a non seulement « survécu » à tous ces changements, mais elle les a aussi accompagnés. Plusieurs articles publiés ont déjà apporté des éclairages sur l'évolution de son contenu depuis ses débuts. Le numéro 94, publié en 1999, fournit un index des articles disponibles à l'époque selon les auteurs et une grille de classement prédéfinie, qui constitue un outil très précieux. Ensuite, à l'occasion du numéro 100³, publié en 2006, a été opérée une mise en

perspective historique de la revue selon différents prismes, au travers de plusieurs articles et de l'éditorial (Cling et Leenhardt, 2006). Doligez (2016), puis Rousset et Sessego (2020) se sont, quant à eux, intéressés à la revue au travers des articles qui y ont été publiés par des membres du groupe AMIRA⁴ (sur lequel on revient plus loin dans ce papier). Enfin, d'autres articles du présent numéro de Statéco apportent des éclairages spécifiques sur des domaines ciblés.

Cet article propose de poursuivre et d'approfondir cette analyse rétrospective tant du statut des auteurs que des thématiques abordées par la revue. L'article est organisé comme suit. La première section présente brièvement la méthodologie utilisée pour compiler et analyser les articles publiés au cours des cinquante années d'existence de la revue Statéco. La deuxième section livre une lecture de la production, telle que le nombre d'articles publiés suivant les périodes, et s'attachera à caractériser la population des auteurs, en soulignant la montée en puissance des contributeurs issus des pays du Sud. La troisième section utilise l'analyse textuelle pour établir une cartographie des thématiques les plus fréquentes dans la revue suivant les périodes de publication et l'affiliation institutionnelle des auteurs. Enfin, face à l'héritage de la revue, la dernière section conclut et trace quelques perspectives, mettant en évidence les enjeux de la statistique publique dans un contexte de *data revolution* et de valorisation des expériences appliquées, en particulier en Afrique sub-saharienne, tout en respectant le format académique de la revue.

Méthodologie

L'analyse de données textuelles (Lebart et Salem, 1994) s'est développée grâce aux outils informatiques et statistiques qui permettent d'extraire et d'analyser l'information contenue dans des textes.

Pour notre démarche, le corpus de recherche a été composé en plusieurs étapes (cf annexe 1 pour la méthodologie détaillée). La première phase a consisté à collecter l'ensemble des articles publiés dans la revue depuis sa création. Cette compilation a été facilitée par l'utilisation de l'index du numéro 94 paru en 1999, du Cd-rom inclus dans le numéro 100 publié en 2006, ainsi que la plateforme numérique disponible sur le site de l'Insee⁵. Nous avons développé un algorithme de *data scraping* qui a permis d'extraire et d'harmoniser les informations nécessaires pour la constitution de la base de données (titre et résumé de l'article, nom de(s)

² Programmes d'aide assortis de conditionnalités mis en place par les institutions de Bretton Woods (FMI et Banque mondiale) en direction des pays pauvres endettés.

³ Le CD-ROM inclus dans le numéro 100 contient l'ensemble des articles de tous les numéros parus depuis la création de la revue (1 à 100), dont l'indexation est basée sur celle établie

par le numéro 94 de 1999 qui classait tous les numéros de 1 à 93.

⁴ AMIRA : Amélioration des méthodes d'investigation en milieu rural africain.

⁵ BNSP : bibliothèque numérique de la statistique publique disponible sur www.insee.fr

auteur(s) et leur(s) affiliation(s) institutionnelle(s), numéro et date de publication de l'article).

La deuxième étape consacrée au nettoyage des données, a consisté à supprimer d'abord les éléments non informatifs (ponctuation et chiffres), et non significatifs - *stopwords* – (pronoms, adverbes...). Les éléments informatifs ont ensuite été standardisés afin de les rendre comparables (correction d'orthographe, harmonisation, mise en format neutre des mots – *lemmatization*).

Pour la troisième étape, les techniques d'extraction de mots-clés ou d'ensembles de mots ont été utilisées sur les titres et les résumés afin de constituer un lexique thématique. Cette étape exigeait un bon niveau de connaissance de la revue pour choisir les mots ou combinaisons de mots les plus pertinents à inclure dans l'algorithme de recherche.

Au total 635 articles écrits pour la majorité en français (les articles en anglais bien qu'acceptés sont peu nombreux), ont été retenus dans la base (les documents présentés sous forme de notes de lecture ont été exclus du champ d'étude (une quarantaine)). Plus de 550 auteurs ont été identifiés, quelques reprises manuelles ont permis dans une dernière étape de compléter les organismes d'appartenance des auteurs lorsque l'information était manquante. A la fin du processus itératif, nous avons établi une liste de thèmes regroupés en grandes thématiques. Au total, 54 thématiques ont été retenues (annexe 2).

L'ensemble des traitements réalisés pour constituer la base de données et l'analyse des résultats s'est appuyé sur le logiciel statistique R et sur le langage Python.

Une production inégale sur la période 1972-2022

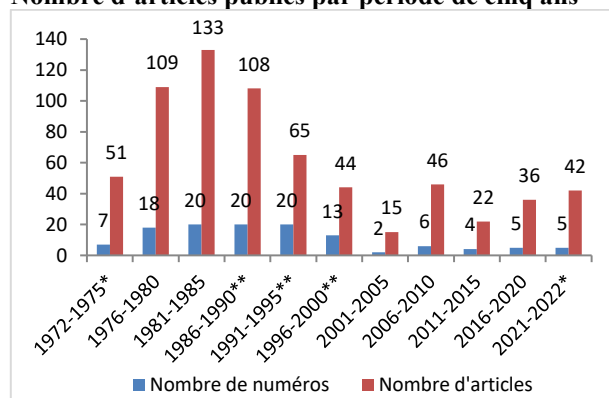
Le rythme de production variable au cours du temps, permet de distinguer plusieurs périodes (graphique 1). Avec une périodicité trimestrielle, le début de la période (années 1976-1990) est caractérisé par une production plus intense. Plusieurs facteurs peuvent l'expliquer et en premier lieu, le vivier d'auteurs constitué par les assistants techniques en poste dans les services statistiques des pays en développement (essentiellement les pays d'Afrique subsaharienne francophone) dans les années 1970, lequel va progressivement se tarir.

L'existence du groupe AMIRA, de 1975 à 1988, constitue un autre facteur explicatif de cette production intense. Ce groupe, créé conjointement par l'Orstom (aujourd'hui IRD), l'Insee et le Ministère français de la

coopération, rassemblait des chercheurs et statisticiens du développement dans une logique pluridisciplinaire, essayant de combiner approches qualitative et quantitative (Couty, 1983,1996 ; Charmes, 2006 ; Winter, 2010 ; Doligez, 2016 ; Rousset et Sessego, 2020). Statéco a été un de ses vecteurs de diffusion. Un des objectifs du groupe était de tenir compte au mieux des réalités de terrain dans les approches utilisées, ce qui l'a conduit à aborder des sujets originaux, comme la question des unités statistiques à considérer dans les enquêtes auprès des ménages, ou encore les problèmes liés à la diversité linguistique d'un pays. Il en est résulté une grande richesse au niveau des thèmes des articles publiés à cette époque.

Graphique 1

Nombre d'articles publiés par période de cinq ans



Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

* Statistiques « ajustées » par rapport au nombre d'années de la période considérée (pour tenir compte de leur durée plus faible).

** 1986-1990 : dont deux numéros doubles ; 1991-1995 : dont cinq numéros doubles ; 1996-2000 : dont trois numéros doubles et un numéro triple.

Les années 90 ont ensuite été marquées par une baisse du nombre d'articles publiés. L'élargissement des zones géographiques couvertes par les activités de coopération de l'Insee, suite à la chute du mur de Berlin, aurait pu entraîner une réorientation forte de la revue. Cela n'a pas été le cas, peut-être faut-il voir là le fait que, les actions de coopération menées vis-à-vis des pays de l'Est ayant essentiellement pour objectif de transformer les systèmes statistiques concernés afin qu'ils se conforment à certains standards de l'Union européenne, il y avait moins d'incitations pour des propositions d'articles abordant des sujets originaux. Ainsi, la revue Statéco ne publie plus que deux numéros par an à partir du milieu des années 1990.

Après trois années d'interruption entre 2001 et 2003, la revue paraît à nouveau en 2004 à l'initiative de Dial⁶,

⁶ Equipe réunissant à l'époque des chercheurs de l'Institut de recherche pour le développement (IRD), des statisticiens de l'Insee et des experts de l'Agence française de développement

(AFD), devenue une équipe de recherche du Laboratoire d'Economie de Dauphine (LEDA) depuis 2010 et qui réunit des enseignants-chercheurs de l'université de Dauphine, de l'IRD et du CNRS.

dans le cadre d'une coédition entre l'Insee, Afristat et Dial. En élargissant le champ couvert à des travaux d'analyse économique quantitative sur les pays en développement, Statéco se donne alors pour objectif, avec cette nouvelle formule, de souscrire aux standards d'une revue scientifique (comité de rédaction, système de relecture anonyme, etc.). Le partenariat mis en place permet de suivre sa double mission : continuer à traiter des évolutions méthodologiques de la statistique publique de manière appliquée en incluant une dimension recherche, et accompagner les auteurs du Sud (statisticiens, économistes, démographes, chercheurs) à la rédaction d'articles scientifiques (encadré 1).

Encadré 1

Extrait de l'introduction générale du numéro spécial Statéco sur les statistiques sur la Gouvernance, Paix et Sécurité (Calvo, Razafindrakoto, Roubaud, Sougané, n°116, 2022).

Parmi les 25 auteurs des articles, 18 proviennent de pays du Sud, dont 14 d'Afrique sub-saharienne. Un tel résultat n'a pas été atteint spontanément. Ayant cet objectif en tête, il a fallu orienter notre rôle d'éditeurs scientifiques aux différentes étapes du processus, notamment à deux stades : d'abord, au moment de la diffusion de l'appel à contributions, en veillant à ce que ce dernier atterrisse bien dans les institutions susceptibles de contribuer mais pas forcément attentives ou connectées sur les réseaux internationaux ; ensuite, au moment de la rédaction et de la finalisation des articles sélectionnés par un accompagnement continu et personnalisé de tous les auteurs en ayant accepté le principe. C'est notamment le cas des jeunes chercheurs, ou encore des auteurs issus du monde de la statistique, pour qui les codes standards de l'écriture d'articles académiques ne sont pas toujours bien maîtrisés ou même connus, voire hors du champ de leur pratique quotidienne, en dépit du potentiel de leur savoir pour l'avancée de la connaissance scientifique. Une telle stratégie a clairement un coût (en temps surtout), mais elle est enrichissante pour tous et nous espérons montrer par l'exemple que cette voie est possible (et souhaitable).

La revue a maintenant une périodicité annuelle. Plusieurs numéros ont repris la formule du numéro spécial dédié à une problématique particulière : le numéro 95-96-97 (2000) portant sur le projet Madio⁷ à Madagascar, le numéro 99 (2005) sur la mise en œuvre et l'analyse des résultats des enquêtes 1-2-3 dans la zone UEMOA, le numéro 100 (2006) sur le bilan et les nouvelles formes de la coopération économique et statistique avec l'Afrique, le numéro 104 (2009) sur le traitement de l'emploi et du secteur informel, le numéro 109 (2015) sur les migrations, marché du travail et dynamiques démographiques en Afrique sub-saharienne. Sur la période récente, le numéro 111

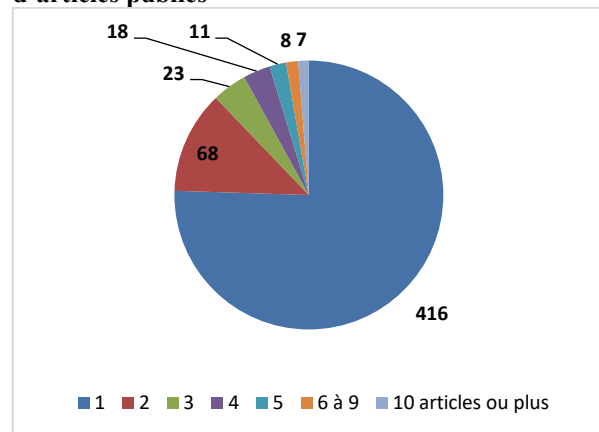
(2017) a été consacré aux comptes nationaux en Afrique sub-saharienne, le numéro 113 (2019) consacré au jumelage réalisé sous l'égide de la Commission européenne avec la Tunisie, le numéro 114 (2020) consacré aux objectifs de développement durable, et enfin le numéro 116 (2022) au thème « gouvernance, paix et sécurité » relevant de l'objectif de développement durable, ODD 16. A cette occasion, il faut souligner l'intérêt suscité par ces numéros et dossiers spéciaux, dont témoigne le grand nombre de lecteurs à travers quelques statistiques sur le nombre de vues et de téléchargements des articles à partir des sites web des trois partenaires coéditeurs et où sont mis en ligne les articles⁸.

Des auteur(e)s fidèles et d'origines variées

La population des auteur(e)s regroupe des chercheurs, des statisticiens, des économistes, des démographes, etc. L'analyse par auteur montre plusieurs évolutions de la revue. Au total environ 550 auteur(e)s différents ont publié (ou co-publié) dans Statéco depuis sa création. La majorité d'entre eux n'ont rédigé qu'un seul article, plus d'une centaine ont écrit au moins deux articles, et sept au moins 10 articles (graphique 2). Les auteur(e)s signent seul(e)s les deux tiers des articles. Cependant, avec une pratique d'écriture collective qui s'est développée au fil du temps, il est logique que le nombre d'auteurs augmente par numéro. Nous passons ainsi de 6 à 15 auteurs en moyenne par numéro entre les 45 premiers numéros (1972-1985) et les 20 derniers (2001-2022).

Graphique 2

Distribution des auteur(e)s selon le nombre d'articles publiés



Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

⁷ Ce projet mis en place à Madagascar de 1994 à 1999, visait la réhabilitation de l'appareil statistique national.

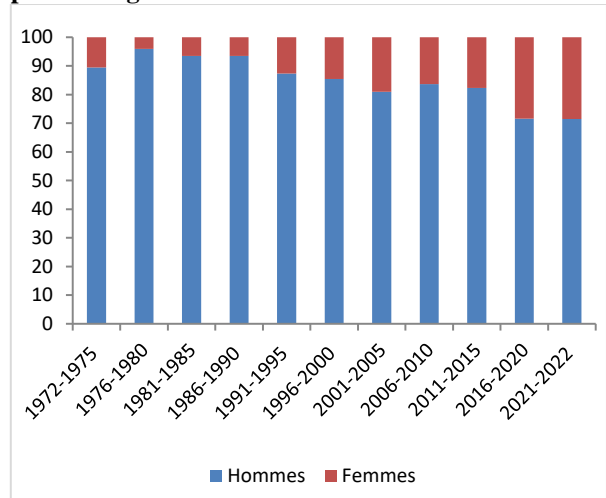
⁸ Quelques statistiques sur les quatre derniers numéros Statéco : N°116 (1098 vues, 2923 téléchargements et 2938 visites) ; N°115 (943 vues, 1797 téléchargements et 1683

visites) ; N°114 (905 vues, 3519 téléchargements et 3420 visites) ; N°113 (1407 vues, 4132 téléchargements et 3756 visites). Consultation du site www.insee.fr le 12 avril 2023. A cela il faudrait ajouter les mêmes données observées à partir des sites web de Dial et d'Afristat.

La part des auteures femmes a également progressé sur la période. En effet, si elles représentent seulement 14 % de la population des auteur(e)s, elles sont plus significativement présentes à partir des années 2000 (graphique 3). Cette tendance pourrait être liée à la combinaison entre la féminisation croissante de la population diplômée en statistique au sein des INS et celle concernant les sciences économiques et sociales (IRD par exemple).

Si l'on s'intéresse aux auteur(e)s les plus prolifiques (graphique 4, où les lecteurs fidèles reconnaîtront un certain nombre d'auteurs « historiques » ...), on peut constater que ceux-ci manifestent une grande fidélité à la revue dans la durée : sur les quinze auteur(e)s ayant publié au moins 6 articles, cinq ont publié sur une période de plus de vingt ans (tableau 1).

Graphique 3
Évolution de la catégorie genre des auteur(e)s en pourcentage



Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

Note de lecture : 28,4 % des auteur(e)s ayant contribué aux numéros de Statéco entre 2016 et 2020 sont des femmes. Parmi elles, certaines ont signé ou co-signé plusieurs articles durant la période.

Tableau 1
Longueur de la période de publication pour les auteur(e)s de plus de 6 articles

Longueur de la période	< 10 ans	10 à 14 ans	15 à 20 ans	20 à 24 ans	25 à 30 ans	> 30 ans
Nombre d'auteur(e)s	2	2	6	1	1	3

Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

Graphique 4
Auteur(e)s ayant publié dans la revue (la taille du nom dépend du nombre d'articles produits)



Remarque : le positionnement des noms sur le graphique n'est pas signifiant, et n'indique donc pas de proximité entre les auteurs (par exemple sur les thématiques qu'ils ont abordées).

Il est également intéressant de définir une typologie des auteurs, selon leurs organismes d'appartenance. Pour cela, une grille a été proposée (tableau 2), qui a été appliquée aux articles : plus précisément, en cas

d'auteurs multiples pour un article donné, on affecte des coefficients à chaque catégorie concernée dont la somme est égale à 1 : par exemple, si un article est co-écrit par un statisticien d'un pays en développement et

un représentant d'une organisation internationale, il se voit affecter une valeur 0.5 pour la catégorie 5 et une valeur 0.5 pour la catégorie 6.

Tableau 2
Nomenclature pour l'appartenance des auteurs

Catégorie	Descriptif
1. Insee - coopération ⁹	Service de coopération, puis Département des relations internationales et de la coopération, puis Unité de la coopération internationale, puis Département Europe et relations internationales, puis Département de la coordination statistique et internationale depuis 2012
2. Insee - autres	Insee hors coopération, services statistiques ministériels français, retraités Insee
3. Ministère coopération	Ministère français de la coopération, Agence française de développement, Caisse française de développement
4. Assistants techniques	Statisticiens français en poste dans les services statistiques des pays en développement
5. Statisticiens pays	Statisticiens appartenant aux services statistiques des pays en développement
6. Organisations internationales	Nations unies, Afristat, Eurostat, Banque mondiale, FMI, etc.
7. IRD	IRD, Orstom ¹⁰ , Dial
8. Autres	Universités (pays développés et pays en développement), organismes de recherche, consultants indépendants, autres

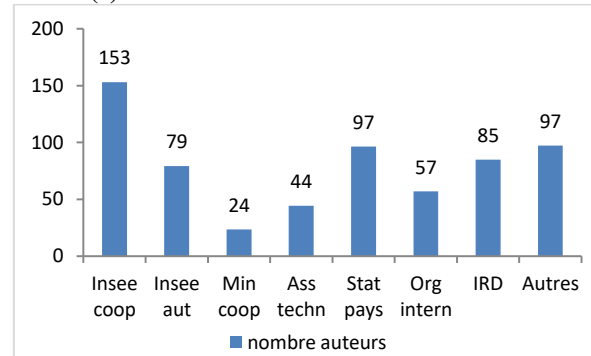
Remarque : la codification s'intéresse à l'organisme d'appartenance de l'auteur (ou des auteurs) au moment de la publication de l'article : certains auteurs auront pu voir leur codification changer au cours de leur carrière : par exemple, un agent publiant un article alors qu'il travaille au service de la coopération de l'Insee sera codé 1, et il sera codé 2 s'il publie un article après son passage en retraite. Cette codification est donc dédiée aux articles.

Si l'on somme les valeurs ainsi codées, on obtient des résultats qui sont résumés par le graphique 5. Le service/unité/département en charge de la coopération à l'Insee a été, comme on pouvait s'y attendre, le premier fournisseur d'articles pour la revue, suivi d'un ensemble de quatre catégories (Insee autres, statisticiens pays, IRD y compris Orstom et Dial, autres) qui ont fourni de l'ordre de 80 à 100 articles chacune. Les autres catégories sont moins représentées.

Cependant, ce premier résultat doit être relativisé, car le nombre d'articles publiés a considérablement varié au cours du temps (graphique 1) : il était beaucoup plus

important pour les années 1970 et 1980 grâce notamment à une parution trimestrielle. Il est donc intéressant de regarder la part de chaque provenance d'auteurs selon les périodes pour voir apparaître un autre constat (graphique 6).

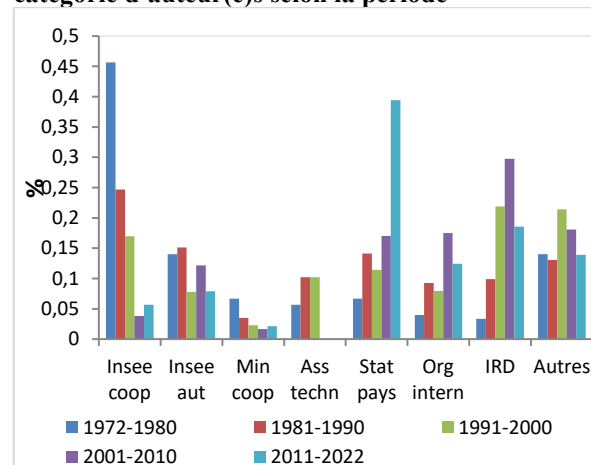
Graphique 5
Répartition des articles selon l'appartenance des auteur(e)s



Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

D'une part, le service de la coopération de l'Insee, qui était très prolifique dans les années 1970 et 1980, voit son importance considérablement réduite ensuite ; même constat pour le Ministère français de la coopération. L'évolution est encore plus marquée pour la catégorie des assistants techniques, qui, comme on l'a mentionné précédemment, a maintenant quasiment disparu.

Graphique 6
Pourcentage d'articles provenant de chaque catégorie d'auteur(e)s selon la période



Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

Note de lecture : sur la période 1972-1980, le service coopération de l'Insee a produit environ 45% des articles de Statéco.

⁹ Compte-tenu des différentes appellations des entités en charge de la coopération à l'Insee au fil du temps, le terme générique « service de coopération de l'Insee » sera la plupart du temps employé dans cet article.

¹⁰ Orstom : Office de la recherche scientifique et technique en outre-mer devenu d'abord en 1984 Institut français de recherche scientifique pour le développement en Coopération (tout en gardant le sigle Orstom), puis Institut de recherche pour le développement (IRD) en 1998.

A l'inverse, des catégories ont beaucoup progressé pendant la période, à savoir les organisations internationales, l'IRD/Orstom/Dial (rappelons que l'IRD-Dial assure la coédition de la revue depuis 2004), et surtout, sur la dernière période, la catégorie des « statisticiens pays », ce qui répond à un des souhaits exprimés lors du lancement de Statéco.

Par ailleurs, principalement francophone, la revue bénéficie d'une large audience, en particulier en Afrique sub-saharienne et dans les pays du Maghreb, et, de fait, encourage prioritairement les auteurs originaires des pays avec lesquels souvent une collaboration technique est mise en place. Durant la première moitié d'existence de la revue, la part des auteurs du Sud est restée modeste (15 %), et ce n'est qu'à partir des années 2000 que leur contribution s'est progressivement renforcée jusqu'à atteindre plus de 50 % au cours de la dernière décennie (graphique 7). Plusieurs raisons peuvent expliquer cette augmentation. En premier lieu, c'est dans le cadre de numéros ou dossiers thématiques qu'une large partie des articles sont rédigés ou co-rédigés par des auteurs du Sud. En effet, des projets menés dans les pays ou la région associant généralement des équipes pluridisciplinaires (chercheurs, statisticiens, économistes, démographes) débouchent régulièrement sur un numéro spécial consacré à la présentation des méthodologies et des résultats des travaux issus des projets. Ces numéros sont déjà cités dans la deuxième section de cet article, néanmoins on peut mentionner ici par exemple le numéro spécial 95-96-97 (2000), consacré au projet Madio à Madagascar ou plus récemment le numéro 116 consacré aux ODD (2022). Ou encore, bien que le présent numéro ne fasse pas partie du corpus de recherche, signalons que parmi les quatorze auteurs des articles de ce numéro, neuf proviennent d'Afrique sub-saharienne et de Tunisie.

En deuxième lieu, on peut retenir la fidélité des auteurs et du lectorat de la revue. Il n'est pas rare qu'un auteur ayant publié une première fois dans Statéco, envoie à nouveau d'autres propositions d'articles au comité de rédaction. Plus généralement la revue reçoit spontanément des contributions issues des auteurs du Nord et du Sud, même si toutes ne sont pas retenues ou n'aboutissent pas à la finale. Souscrivant aux standards d'une revue scientifique, le système d'évaluation par des référés en double aveugle mis en place lors de la reprise en 2004, permet par ailleurs d'accompagner les auteurs dans la finalisation de leurs articles (cf. encadré 1).

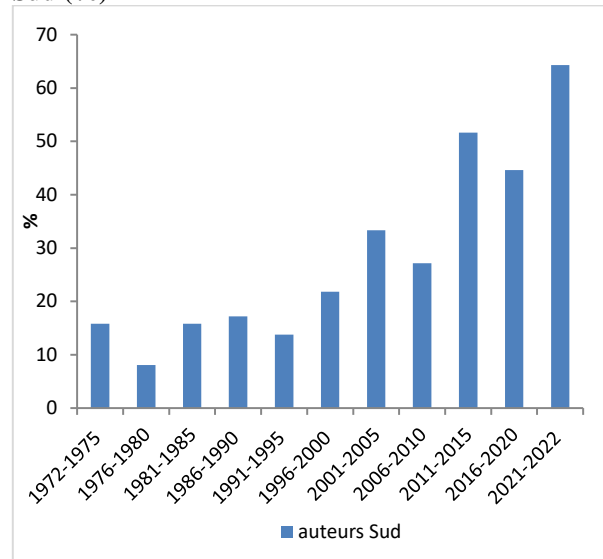
En troisième lieu, une plus grande visibilité est donnée à la revue, à travers sa liste d'abonnés, les appels à contribution pour des dossiers thématiques, les réseaux sociaux et les webinaires organisés lors de la publication d'un numéro, etc.

Enfin, l'attractivité de la revue a permis d'inscrire cette dynamique dans l'élargissement du comité de rédaction

de Statéco aux enseignants-chercheurs des écoles de statistique africaines en 2021.

Graphique 7

Évolution de la part des auteur(e)s originaires du Sud (%)



Source : Base de données constituée sur la revue Statéco, 1972-2022, calcul des auteurs.

Note de lecture : 45 % des auteur(e)s ayant contribué aux numéros de Statéco entre 2016 et 2020 sont originaires du Sud. Parmi eux, certains ont signé ou co-signé plusieurs articles durant la période.

Des thématiques couvrant un large spectre

Les évolutions qui viennent d'être présentées ont eu des conséquences au niveau des thématiques qui ont été abordées par la revue. Plutôt que de procéder à une simple énumération de celles-ci, il est préférable d'observer leur évolution au fil du temps, ainsi que la caractérisation de leurs auteurs.

Pour ce faire, l'analyse textuelle a été appliquée à la base de données constituée selon les principes exposés plus haut.

Même si cette technique présente des limites, en particulier dues au fait que les résumés des articles sont plus longs et complets dans les années récentes qu'ils ne l'étaient au début de la publication de Statéco, elle permet d'avoir une approche plus ouverte que le simple classement d'un article dans une thématique et une seule. En effet, un article peut cocher une ou plusieurs thématiques. La liste des thématiques, fournie dans l'annexe 2, comprend des thématiques générales, comme « développement » ou « coopération institutionnelle », et des thématiques plus spécifiques, comme « séries temporelles ».

Graphique 8

Nuage de thématiques abordées dans la revue (la taille est proportionnelle au nombre de fois où la thématique a été détectée dans le corpus de titres et résumés des articles)



Source : Base de données constituée sur la revue Statéco, 1972-2022, calculs des auteurs.

Note de lecture : méthodologie est la thématique la plus fréquente dans l'ensemble du corpus de titres et de résumés de la revue Statéco.

Remarque : le positionnement des thématiques sur le graphique n'est pas signifiant, et n'indique donc pas de proximité entre les différentes thématiques.

Le graphique 8 montre que ce sont les thématiques générales qui sont le plus fréquemment détectées : méthodologie, économie, institution, développement, ... thématiques qui répondent aux objectifs fixés a priori par la revue. Mais d'autres thématiques sont plus de l'ordre de « domaines métiers » pour les statisticiens-économistes : enquête, comptabilité nationale, démographie, sondage, etc.

Nous allons nous intéresser plus spécifiquement, parmi ces dernières, à celles qui sont le plus souvent citées, au regard des auteurs concernés et de leurs périodes de publication, en mentionnant parfois certains articles à titre d'exemple.

Les enquêtes constituent évidemment un domaine majeur pour les statisticiens, présent pendant la cinquantaine d'années de publication (avec une intensité plus marquée avant les années 2000). La catégorie des « assistants techniques », des « statisticiens pays », des agents du service de coopération de l'Insee et de la sphère IRD/ Orstom/ Dial sont surreprésentés parmi les auteurs. Cette

thématique est fréquemment associée à la démographie, à l'emploi et au secteur informel, thèmes sur lesquels nous reviendrons, mais aussi à la technique des sondages, ainsi qu'à la thématique « budget / consommation ».

Les articles abordant le thème des enquêtes ont permis non seulement de présenter, à des fins pédagogiques, différentes opérations menées dans les pays en développement, mais également de proposer une réflexion approfondie et originale sur les méthodes spécifiques à mettre en place dans ce contexte (Blaizeau, 1992) ; avec par exemple des papiers innovants sur les unités statistiques (Gastellu, 1979). A noter que, si le sujet des enquêtes est présent tout au long de la vie de la revue, certaines périodes, comme les années 1980, ou les années 1995 – 2000¹¹, ont vu un nombre d'articles consacrés à la thématique « budget consommation » plus important. Enfin, d'autres thématiques sont également associées aux enquêtes : agriculture, santé, éducation, inégalités, pauvreté, entreprises.

¹¹ En particulier, suite à la dévaluation du franc CFA de 1994, il y a eu à cette époque une demande de révision des

structures de consommation des ménages des pays concernés pour les indices de prix.

La comptabilité nationale constitue un deuxième thème largement présent dans la revue, tout au long de la période, mais avec plus d'intensité dans les années 1980 et 1990. Elle constitue évidemment une matrice fédérant les différentes productions statistiques. Bien que cette thématique concerne l'ensemble des catégories d'auteurs, les Insee « hors service coopération » y sont sur-représentés (montrant ainsi l'intérêt des comptables nationaux français pour l'adaptation des méthodes à des contextes différents de celui sur lequel ils travaillent habituellement). Il en est de même pour la catégorie « autres auteurs » qui rassemble des universitaires, des membres d'organismes de recherche ou des consultants. Là encore, les articles publiés ont permis, d'une part, de présenter des exemples concrets d'élaboration des comptes nationaux dans les pays en développement, mais également de proposer des pistes pour répondre à des problématiques complexes posées dans ce contexte (traitement des aides, inflation, etc.). Par ailleurs, l'outil ERE-TES, porté conjointement depuis une trentaine d'années par l'Insee et la Commission européenne (Brion et Sérurier, 1995 ; Aziz, 2017), a permis de fournir un support aux services statistiques des pays en développement, ainsi qu'un cadre d'harmonisation des méthodes pour l'élaboration des comptes nationaux (Cling et Perrault, 2021).

La démographie, troisième grand thème, a été abordée dans beaucoup d'articles des numéros anciens de Statéco (en particulier dans les années 1976-1985), avant d'être moins présente dans les années qui ont suivi¹². Les auteurs de la période faste faisaient essentiellement partie du service coopération de l'Insee et de la catégorie « autres auteurs ». Malgré sa moindre place dans les numéros récents, il faut noter que la thématique continue à être présente. Elle est souvent associée à la méthodologie d'enquête, ainsi qu'à des thématiques comme la santé ou l'étude des migrations.

La prise en compte du secteur informel pour comprendre le mode de fonctionnement des pays en développement a de longue date constitué un sujet d'intérêt pour les économistes spécialistes de ces pays, que ce soit pour l'appréhension des problématiques d'emploi ou pour l'élaboration des comptes nationaux. Les thèmes « emploi » et « secteur informel », souvent associés dans les articles publiés dans la revue, ont de ce fait été présents pendant l'ensemble de la période, mais avec une intensité plus marquée au début des années 2000. Deux catégories d'auteurs se distinguent : les chercheurs de la sphère Orstom / IRD / Dial, et les statisticiens pays qui ont été très impliqués sur ces sujets. Les réflexions menées initialement sur le thème de l'informel ont conduit à proposer des instruments pour le mesurer, en particulier les enquêtes 1-2-3 qui

ont fait l'objet de nombreux articles dans la revue (Razafindrakoto, Roubaud et Torelli, 2009).

La modélisation a également été un sujet fréquemment abordé dans Statéco, pour une bonne partie dans les années 1980 avec la présentation de différents modèles s'appuyant sur les comptes nationaux, mais avec un regain d'intérêt dans les années 2010. Les auteurs appartiennent souvent à des universités ou des organismes de recherche, mais les statisticiens pays sont également bien représentés.

Suite au besoin exprimé au milieu des années 1970 d'utilisation de méthodes de prévision annuelle dans les pays en développement (Winter, 2010), la thématique « prévision-conjoncture » a fait l'objet de différents articles dans les années 1980, avec une sur-représentation des auteurs du service coopération de l'Insee, de la sphère « Ministère de la coopération », ainsi que des statisticiens pays. Il a été beaucoup moins présent par la suite. Le même constat peut être appliqué au thème « finance », très présent dans les années 1970 et 1980, et souvent associé à d'autres thématiques comme l'économie, la comptabilité nationale, et au thème agriculture, présent dans les années 1970 – 1980 – 1990.

Les articles abordant la thématique « entreprises » ont été publiés sur l'ensemble de la période, avec un peu plus d'auteurs assistants techniques ou « Insee hors coopération » que la moyenne. Mais cette thématique apparaît finalement assez peu, si l'on compare sa place à celle qu'elle peut avoir dans les pays développés : peut-être faut-il voir là une conséquence de l'angle mort concernant la catégorie des petites et moyennes entreprises, coincée entre les entreprises dites « modernes », bien suivies pour l'élaboration des comptes nationaux, et le secteur informel qui a fait l'objet de nombreuses investigations (Savoie, 2016).

On peut maintenant passer en revue quelques thématiques liées aux activités des instituts statistiques : organisation statistique, informatique, formation. Ces thématiques ont été abordées essentiellement au début de la période (avec cependant un regain d'articles sur la formation dans les années 2006 – 2010), essentiellement par les agents du service de coopération de l'Insee, de l'Insee hors coopération, et les assistants techniques. La problématique de l'organisation des systèmes statistiques nationaux (SSN), reste au cœur des préoccupations des statisticiens. Dans ce numéro, l'article de Diop et Soulé (2023) dresse à cette occasion un état des lieux des réformes entreprises sur les SSN en Afrique subsaharienne. Le sujet de l'informatique constitue un cas à part, puisque les années 1970 correspondaient à une

¹² Sans doute faut-il y voir un effet de la création du Ceped (Centre Population et Développement) en 1988 associant 5 tutelles (EHESS, l'Ined, Insee, Orstom, et Université de Paris

VI), qui a développé sa propre gamme de publications. En 2008, le Ceped s'est transformé en une Unité mixte de recherche (UMR) associant l'IRD, l'Université Paris Descartes et l'Ined.

époque de montée en puissance dans les services statistiques, suivie par une banalisation de la place de l'informatique au sein de ces services.

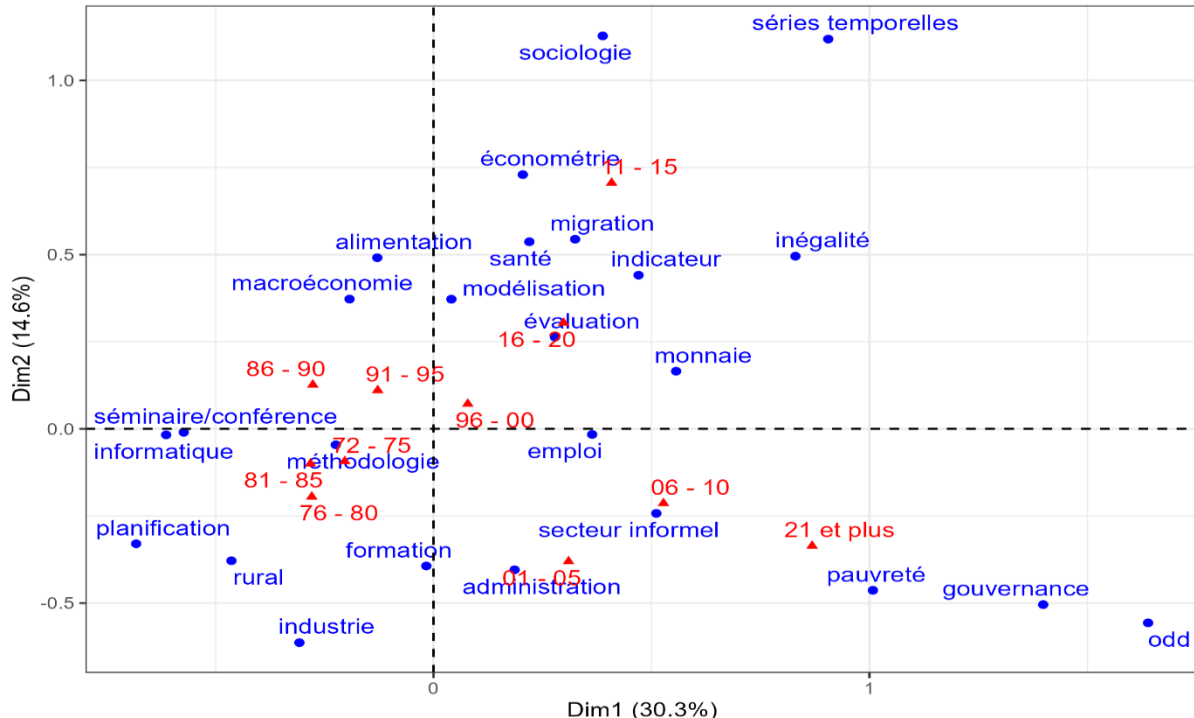
Enfin, on peut observer l'apparition de thématiques nouvelles dans les années récentes : séries temporelles, gouvernance, ODD. Ces thématiques montrent que les demandes adressées aux statisticiens du développement ont évolué, soit pour affiner certaines méthodes, soit pour aller sur des terrains non couverts jusqu'alors. Sur ces trois thématiques, il faut noter que les statisticiens pays sont très présents parmi les auteurs, ce qui, il faut le répéter, répond à un des objectifs fixés au départ à la revue. Les chercheurs de l'Orstom/IRD/Dial sont également présents sur les deux sujets gouvernance et ODD, et les représentants des organisations internationales sur les ODD.

Afin de compléter ces premiers éléments, on peut affiner l'analyse en mettant en regard thématiques et périodes, d'une part, et thématiques et catégorisation des auteurs, d'autre part. Les graphiques 9 et 10 sont issus d'analyses factorielles des correspondances appliquées aux deux tableaux croisant thématiques et périodes de parution d'une part, thématiques et organismes d'appartenance des auteurs d'autre part. Plus précisément, ces graphiques représentent les

Graphique 9

Résultat d'une analyse factorielle passée sur le tableau croisant thématiques (en bleu) et périodes (en rouge ; la période 1972-1975 est indiquée comme 72-75, par exemple)

Représentation des profils lignes et colonne selon les axes 1 et 2



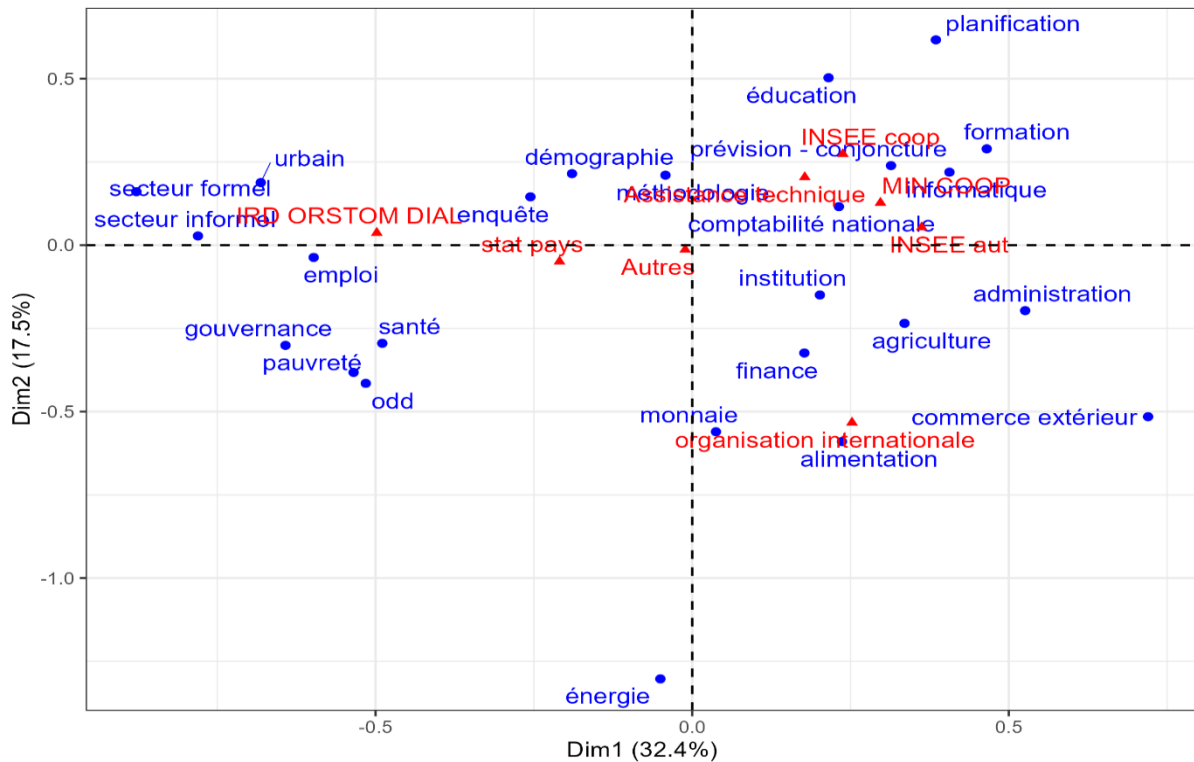
Note : seules les thématiques les plus contributrices sur les deux premiers axes sont indiquées sur le graphique.

¹³ Même si cette visualisation est incomplète, puisque de l'information est également apportée par les autres axes factoriels...

Graphique 10

Résultat d'une analyse factorielle passée sur le tableau croisant thématiques (en bleu) et organismes d'appartenance des auteurs (en rouge)

Représentation des profils lignes et colonne selon les axes 1 et 2



Note : seules les thématiques les plus contributrices sur les deux premiers axes sont indiquées sur le graphique.

Le graphique 10 permet de voir les spécificités de chaque catégorie d'auteurs en matière de thématiques : thèmes anciens (planification, prévision – conjoncture, formation, etc.) pour le service coopération de l'Insee et la sphère « Ministère de la coopération » ; thèmes plus spécifiques pour les organisations internationales (finances, commerce extérieur, etc.) ; secteur informel/emploi et thématiques nouvelles (gouvernance, ODD) pour la sphère Orstom/IRD/Dial ; ces thèmes sont souvent associés aux statisticiens pays, présents également sur le secteur des enquêtes ; ce dernier thème est partagé avec d'autres catégories, comme les assistants techniques qui ont également écrit sur des sujets liés à l'organisation et au mode de fonctionnement des services statistiques ; les « Insee autres » sont présents sur des thématiques comme la comptabilité nationale ou des sujets liés au mode de fonctionnement des services statistiques ; enfin les « autres auteurs », situés au centre du graphique¹⁴

interviennent sur des sujets comme la comptabilité nationale ou la modélisation.

Proposition d'une typologie des articles

Pour prolonger l'analyse précédente qui porte sur les thématiques, il est intéressant de voir si l'on peut dégager une typologie des articles, ceux-ci abordant souvent une ou plusieurs thématiques. A cet effet, un algorithme de classification a été appliqué à la base des articles¹⁵, et, à partir d'un nombre de classes établi à 25, on peut détailler leurs compositions (l'annexe 4 fournit des éléments sur la méthode de classification utilisée).

Ces classes peuvent elles-mêmes être regroupées en catégories, que l'on peut détailler¹⁶ selon un axe chronologique en essayant de les caractériser sur les thématiques qu'elles abordent, même s'il existe une certaine hétérogénéité au sein des classes. Leur

¹⁴ Mais celui-ci ne contient que l'information des deux premiers axes factoriels ...

¹⁵ Plus précisément constituée de 592 articles, les notes de lecture ayant été enlevées ainsi que quelques articles qui n'ont coché aucune thématique de la liste.

¹⁶ Trois classes très peu fournies (moins de 4 articles) n'ont pas été détaillées ici.

importance en termes de nombre d'articles est également indiquée.

La première catégorie comprend six classes d'articles publiés en grande partie avant 1990 :

- une première classe (19 articles), axée sur la planification, avec des auteurs du service coopération de l'Insee ou de la catégorie « autres » ;
- 26 articles s'intéressant aux questions de méthodologie d'enquête en milieu rural, avec des auteurs de l'Insee coopération, de l'Orstom/IRD/Dial, de la catégorie « autres auteurs », ainsi que des statisticiens pays ;
- 18 articles sur la question des politiques alimentaires (avec une présence des thèmes alimentation, gouvernance, agriculture), avec une surreprésentation des auteurs « Insee autres », des membres des organisations internationales et des « autres auteurs » ;
- une classe fournie (49 articles) dédiée à la démographie et majoritairement publiée dans les années 1976 - 1985, avec des auteurs du service coopération de l'Insee et de la sphère Orstom/IRD/Dial ; les thématiques migration, santé ou sociologie y sont souvent présentes ;
- 10 articles sur la thématique du commerce extérieur, avec une surreprésentation des auteurs des organisations internationales ; cette catégorie est cependant à nouveau présente dans des numéros récents de la revue ;
- enfin, un ensemble de 9 articles axés sur l'industrie, avec des auteurs « Insee autres » ou statisticiens pays, où les thématiques nomenclatures, commerce, entreprises apparaissent également.

La deuxième catégorie comprend cinq classes qui rassemblent des articles publiés essentiellement avant l'an 2000 :

- 22 articles publiés essentiellement dans les années 1980 par les assistants techniques et les statisticiens pays sur les enquêtes budget-consommation, ces articles cochant souvent les thématiques sondages ou pauvreté ;
- une classe de 20 articles sur les questions de modélisation (cochant souvent également les thématiques macroéconomie et comptabilité nationale), rédigés en particulier par des auteurs « Ministère de la coopération » ou des « autres auteurs » ;
- un ensemble de 8 articles dédiés à la problématique de l'ajustement structurel (essentiellement dans les années 1986-1995), avec une surreprésentation des assistants techniques et des organisations internationales ;
- 17 articles relatifs à la prise en compte de la géographie dans la production de statistiques et

pour lesquels des thématiques variées peuvent être cochées, par exemple migration, rural, urbain, agriculture), avec une légère surreprésentation des auteurs Orstom/IRD/Dial et assistants techniques ; si ces articles ont été publiés en grande partie avant 1995, cette classe continue cependant à exister dans des numéros récents ;

- enfin, un ensemble de 34 articles consacrés aux questions d'organisation des systèmes statistiques, rédigés souvent par des statisticiens pays, des assistants techniques, ainsi que des agents de l'Insee hors service coopération, et pour lesquels des thématiques comme coopération institutionnelle ou « séminaire / conférence » sont souvent présentes.

Une troisième catégorie regroupe des classes présentes tout au long de la période :

- un ensemble de 45 articles portant également sur la méthodologie d'enquête, avec des auteurs d'appartenances diversifiées ;
- une classe de 20 articles centrée sur les questions d'utilisation de données administratives, en particulier pour l'appréhension du secteur formel (données de comptabilité d'entreprise), avec une surreprésentation des auteurs « autres », des organisations internationales et des Insee hors coopération ;
- 13 articles consacrés à l'utilisation de méthodes économétriques, en particulier pour la mise au point d'indicateurs ou de modèles, avec des auteurs variés ;
- 19 articles relatifs aux questions de formation, souvent rédigés par des membres de l'Insee coopération et du ministère de la Coopération ;
- 12 articles relatifs à la prise en compte des questions monétaires (inflation, prix), dans les études économiques, rédigés par des statisticiens pays, des représentants d'organisations internationales et des auteurs « autres » ;
- 17 articles centrés majoritairement sur les indices de prix, avec une bonne part de statisticiens pays parmi les auteurs ; la thématique séries temporelles y apparaît fréquemment ;
- 21 articles sur les problématiques liées à la diffusion des données, avec une légère surreprésentation des statisticiens pays, assistants techniques et membres des organisations internationales ;
- 36 articles consacrés au secteur informel, en lien avec les thématiques emploi, pauvreté, enquêtes, comptabilité nationale ; les auteurs sont majoritairement de la sphère Orstom/IRD/Dial et des statisticiens pays ;
- un groupe fourni de 58 articles centrés sur la comptabilité nationale (différentes thématiques étant également détectées comme entreprises,

méthodologie, nomenclatures, secteur informel, etc.), avec une origine des auteurs variée (Insee coopération, Insee autres, statisticiens pays, autres auteurs) ;

- et pour finir une catégorie « fourre-tout », composée de 89 articles, qui « cochent » un ensemble de thématiques générales : économie, développement, institutions, coopération institutionnelle : il s'agit d'articles de réflexion générale sur la coopération, rédigés par des auteurs de toutes appartenances.

Enfin, une nouvelle classe de 20 articles, apparue depuis les années 2000 sur les questions de gouvernance et les objectifs de développement durable,

Graphique 11

Représentation de la classification des articles publiés dans Statéco, selon quatre catégories relatives à la période où les articles ont été majoritairement publiés



Remarques :

- les tailles des bulles représentant les classes sont proportionnelles aux nombres d'articles des classes.
- le positionnement des bulles sur le graphique n'indique pas de proximité entre les classes, seules les couleurs signifient une caractérisation de celles-ci sur la période où les articles ont été majoritairement publiés.

L'héritage de Statéco et perspectives

Les considérations précédentes mettent en évidence le fait que Statéco ne doit pas être vue comme une revue théorique sur les problématiques de développement, mais comme une revue de méthodologie appliquée, animée entre autres, dans ses objectifs initiaux, par la volonté de la coopération française de ne pas appliquer de manière trop directe les méthodes du Nord aux pays en développement (Savoie, 2016). Les articles qui y ont été publiés ont permis de partager les réflexions menées par des auteurs de tous horizons sur les problèmes posés par la production ou l'utilisation de données statistiques, et de présenter des cas concrets d'application. Cinquante années de publication de la revue ont ainsi conduit à constituer une mémoire unique sur ce domaine.

Cette mémoire est accessible en ligne¹⁷, et l'index publié dans le n°94 peut aider à chercher des articles dans les numéros anciens. Il faut par ailleurs rappeler ici le rôle fondamental joué, à une époque où Internet n'existait pas, par la diffusion gratuite de la revue en format papier¹⁸ au sein des services statistiques des pays en développement : cette diffusion lui conférait alors un statut relativement unique, en matière de documentation méthodologique francophone disponible dans ces services.

Par ailleurs, il faut souligner que, même si la revue est restée aux « marges » de la statistique publique en France, elle a pu avoir un impact sur les opérations qui y sont menées, par les réflexions et le recul nécessités par la mise en place de dispositifs dans le contexte des pays en développement. Ainsi, les travaux menés par le groupe AMIRA sur les unités statistiques à considérer ont pu soulever des questions qui seront présentes quelques années plus tard dans les pays développés, sur les ménages recomposés par exemple. Autre exemple, l'utilisation de la télédétection pour la démographie, abordée dans la revue dès les années 1980 (Dureau, Guillaume, 1984). Ce constat rejoint celui de Desrosières (2013), selon lequel la mesure du développement constitue un terrain propice aux innovations méthodologiques.

Enfin, si la revue publie quelques articles en anglais, davantage au cours des dernières années, on peut néanmoins regretter que la diffusion de Statéco soit restée, pour une grande part, cantonnée au monde francophone. Une revue « jumelle » en langue anglaise,

Inter-Stat, a été co-éditée par Eurostat, le DFID britannique (*Department for International Development*) et l'Insee entre 1990 et 2004, mais a cessé ensuite de paraître : elle reprenait un certain nombre d'articles de Statéco dans une version traduite, leur donnant alors une visibilité dans le monde anglo-saxon. Plus récemment, la revue de l'Association internationale de statistique officielle (*Statistical Journal of the IAOS (International Association for Official Statistics)*) a fait paraître un numéro reprenant une sélection d'articles du numéro 116 de Statéco dans une version traduite en anglais (Calvo, Cling, Razafindrakoto, Roubaud et Sougané, 2023).

Les instituts de statistique du monde entier sont désormais confrontés à de nouveaux défis : nouvelles demandes de statistiques – en particulier avec les objectifs de développement durable –, nouvelles technologies, nouvelles sources de données, etc. Confrontés à ces questions complexes, les statisticiens du développement devront mettre au point des méthodes dédiées à leur résolution. La récente révision de l'offre pédagogique dispensée dans les écoles de statistique africaine permet aux futurs statisticiens de relever les nouveaux défis dans le cadre de la révolution des données liée aux enjeux de la statistique publique (Bédécarrats, Cling et Roubaud, 2016 ; Coquet, Diouf et Laffiteau, 2020)).

Dans le même temps, les opérations lourdes (comptabilité nationale, enquêtes ménages de référence, etc.) continueront à constituer le socle de leur activité, tout comme les infrastructures nécessaires à leur bon fonctionnement (bases de sondage, répertoires, nomenclatures). Deux articles de ce numéro (Muller, 2023 ; Bonkougou, Gbossa et Métreau, 2023) donnent des pistes sur les évolutions possibles autour de l'élaboration des comptes nationaux : amélioration des répertoires d'entreprises, prise en compte de l'environnement, de la mondialisation, de la numérisation de l'économie.

Par ailleurs, on observe, dès à présent, dans certains pays en développement, des avancées sur des aspects méthodologiques spécifiques nouveaux, comme les estimations sur petits domaines (Molina, 2022). Celles-ci créent certainement des pistes pour des articles à venir dans la revue. D'autres sujets donnent lieu, à l'heure actuelle, à des développements dans les pays les plus avancés, comme les questions liées à la collecte multimode (Beck, Castell, Legleye et Schreiber, 2022), les techniques d'appariement (Mecatti, Singh, 2014), l'utilisation des données de sources non probabilistes

¹⁷ Par exemple sur le site de l'Insee, à l'adresse <https://www.insee.fr/fr/information/2116798>

¹⁸ Le tirage papier est passé de 1 000 exemplaires en 1980 à 1 600 exemplaires en 1998, atteignant 2 000 exemplaires jusqu'au numéro 100. Ensuite, pour tenir compte de la

version électronique proposée aux lecteurs, l'impression papier a diminué au fur et à mesure, avec moins de 250 exemplaires aujourd'hui (les tirages sont plus importants pour les numéros spéciaux). Le format papier est généralement destiné aux INS dans les pays africains.

(Beaumont, 2020) ou la problématique des unités statistiques à prendre en compte pour les statistiques d'entreprises (Lorenc, Smith et Bavdaz, 2018 ; Haag, 2019) : cette liste n'est sans doute pas exhaustive mais les sujets mentionnés, comme d'autres, conduiront très certainement à des expérimentations suivies de mises

en place de dispositifs pérennes dans les pays en développement, et à des propositions d'articles. Et les articles plus « généraux » sur l'organisation et la gouvernance des services statistiques auront bien entendu toute leur place dans les numéros de Statéco à venir.

Références bibliographiques

Aziz, J.-S. (2017), « ERE-TES, un outil en constante évolution », Statéco n°111.

Beaumont, J.-F. (2020), « Les enquêtes probabilistes sont-elles vouées à disparaître pour la production des statistiques officielles ? », Techniques d'enquêtes vol. 46, n°1.

Bédécarrats, F., Cling, J.-P., et Roubaud, F. (2016), « Révolution des données et enjeux de la statistique en Afrique. Introduction thématique », Afrique contemporaine, vol. 258, no. 2.

Beck, F., Castell, L., Legleye, S. et Schreiber, A. (2022), «Le multimode dans les enquêtes auprès des ménages – Une collecte modernisée, un processus complexifié », Courrier des Statistiques n°7, Insee.

Bengfort, B., Bilbro, R. et Ojeda, T. (2018), “Applied Text Analysis with Python”, O'REILLY Media.

Blaizeau, D. (1992), « Mieux répondre aux besoins d'information sur les ménages : quelques réflexions pour une approche cohérente », Statéco n°72.

Bonkougou, M., Gbossa, H.V.H. et Métreau, E. (2023), « Les comptes nationaux en Afrique : progrès depuis 40 ans et perspectives », Statéco n°117.

Brion, Ph. et Sérurier, M. (1995), « Le projet ERE-TES », Statéco n°81-82.

Calvo, T., Cling, J.-P., Razafindrakoto, M., Roubaud, F. et Sougané, A. (2023), *Governance Statistics and the UN Praia City Group*, dossier spécial du Statistical Journal de l'IAOS, vol. 39(2), juin.

Charmes, J. (2006), « L'héritage d'AMIRA », Statéco n°100.

Cling, J.-P. et Leenhardt, B. (2006), éditorial du numéro spécial « Bilan et nouvelles formes de la coopération économique et statistique avec l'Afrique », Statéco n°100.

Cling, J.-P. et Perrault, F. (2021), « L'assistance technique internationale à l'Insee : une transformation en profondeur au cours des années 2010 », Statéco n°115.

Coquet, F., Diouf, A. et Laffiteau, E. (2020), « Enjeux de la formation statistique en Afrique francophone : l'expérience des écoles de statistique africaines », Statéco n°114.

Couty, Ph. (1983), « Qualitatif et quantitatif », Statéco n°34.

Couty, Ph. (1996), « Les apparences intelligibles – une expérience d'économiste en Afrique », éditions Arguments, Paris.

Desrosières, A. (2013), « La mesure du développement : un domaine propice à l'innovation méthodologique », Revue Tiers Monde n°213.

Doligez, F. (2016), « Retour sur l'expérience d'Amira, 1975-1988 », Afrique contemporaine, vol. 258, n° 2.

Diop, L. (1995), « Afristat, un outil d'intégration régionale au service du renforcement des capacités statistiques en Afrique subsaharienne », Statéco n°81-82.

Diop, L. et Soulé, O.M. (2023), « La réforme des systèmes statistiques nationaux en Afrique Sub-saharienne : une symphonie inachevée », Statéco n°117.

Djirigui, J. et Sémécurbe, F. (2019), « Cinquante ans de résumés d'Economie et Statistique », Economie et Statistique n°510-511-512.

Dureau, F. et Guillaume, A. (1984), « La démographie depuis l'espace : un nouveau système d'observation ? », Statéco n°38.

Gastellu, J.M. (1979), « Mais où sont donc ces unités économiques que nos amis cherchent tant en Afrique ? », Statéco n°19.

- George, A. (2022)**, “Python Text Mining: Perform Text Processing, Word Embedding, Text Classification and Machine Translation”, BPB Publications.
- Haag, O. (2019)**, « Le profilage à l’Insee : une identification plus pertinente des acteurs économiques », *Courrier des Statistiques* n°2.
- Lebart, L. et Salem, A. (1994)** « Statistique textuelle », Dunod.
- Lorenc, B., Smith, P.A. et Bavdaz, M. (2018)**, “The unit problem and other current topics in business survey methodology”, Cambridge Scholars Publishing.
- Mecatti, F. et Singh, A.C. (2014)**, “Estimation in multiple frame surveys: a simplified and unified review using the multiplicity approach”, *Journal de la Société Française de Statistique*, Vol. 155 n°4.
- Molina, I. (2022)**, “Disaggregating data in household surveys using small area estimation methodologies”, *Statistics series n°97*, Economic Commission for Latin America and the Caribbean (ECLAC).
- Muller, P. (2023)**, « Entretien », *Statéco* n°117.
- Razafindrakoto, M., Roubaud, F. et Torelli, C. (2009)**, « La mesure de l’emploi et du secteur informels : leçons tirées des enquêtes 1-2-3 en Afrique », *Statéco* n°104.
- Rousset, C. et Sessego, A. (2020)**, « Ouvrir la boîte noire des statistiques du développement : le groupe AMIRA (Amélioration des méthodes d’investigation en milieu rural africain) dans le revue *Statéco* (Insee) », *Revue Statistique et Société*, vol.8, n°3.
- Savoie, B. (2016)**, « Approches françaises de la statistique en Afrique – Quels enseignements tirer de quatre décennies de coopération intense », *Revue Afrique contemporaine*, N°258, n°2.
- Statéco (2000)**, numéro spécial « Projet Madio à Madagascar » n°95-96-97.
- Statéco (2005)**, numéro spécial « Résultats des enquêtes 1-2-3 dans la zone UEMOA », n°99.
- Statéco (2009)**, numéro spécial sur le traitement de l’emploi et du secteur informel, n°104.
- Statéco (2015)**, dossier spécial MIMADEM, Migrations marché du travail et dynamiques démographiques, n°109.
- Statéco (2017)**, dossier spécial sur les comptes nationaux en Afrique sub-saharienne, n°111.
- Statéco (2019)**, numéro spécial sur le Jumelage avec la Tunisie, n°113.
- Statéco (2020)**, numéro spécial ODD2, n°114.
- Statéco (2022)**, numéro spécial sur les statistiques de Gouvernance, Paix et Sécurité, n°116.
- UN DESA (2015)**, “Transforming our world: the 2015 agenda for sustainable development”, *Rapport UN Sustainable Development Knowledge Platform*, New York.
- Wavresky, P. (2020)**, « Croisement de classifications issues de différents corpus », communication aux 15^{es} Journées internationales d’Analyse statistique des Données Textuelles.
- Winter, G. (2010)**, « A la recherche du développement – Un fonctionnaire au service d’une passion », Karthala.

Annexe 1 : Constitution de la base de données et utilisation de l’analyse textuelle

Cette annexe décrit la manière dont la base de données d’articles a été construite, en particulier au regard des thématiques qui y sont abordées. Le point de départ est une base de données issue des titres et des articles de la revue au format PDF. Différentes étapes permettent alors de constituer la base finale.

Etape 1 : les titres et résumés des articles sont collectés en format PDF et transformés en format texte, en utilisant la bibliothèque pdftools du logiciel RStudio.

La figure 1 donne un aperçu du document texte récupéré. Les chaînes de caractère «\n » correspondent à un retour à la ligne.

Etape 2 : On remarque que, dans la figure 1, les articles sont « balisés ». Pour chaque article, on retrouve l’intitulé de l’article, l’auteur, le résumé. Le premier traitement consiste à localiser dans le texte les trois chaînes de caractères précédemment citées (i.e. intitulé de l’article) et ensuite de les récupérer dans l’ordre dans une table. Cette opération est effectuée en utilisant les trois bibliothèques gsubfn, stringi et stringr du logiciel RStudio.

Pour les numéros de 1 à 93 de la revue Statéco, il a été possible de capitaliser sur un travail précédent de synthèse des articles publiés selon la structure : année, titre, auteur, résumé et numéro de l'article.

Pour les numéros allant de 94 à 115, nous avons utilisé d'autres méthodes d'extraction : nous avons défini un algorithme qui en utilisant le lien (url) de l'article renvoie une liste composée de l'année de publication, l'auteur (ou les auteurs), l'intitulé, le numéro, le résumé et l'affiliation de l'auteur (ou des auteurs) produite à titre indicatif pour quelques articles.

Figure 1 : document texte récupéré

```
[1] " Statéco n° 1 - Février 1972\n\nIntitulé de l'article : Orientations actuelles des études démographiques en Afrique tropicale\nAuteur : Blanc R.\nRésumé : R. Blanc fait le point des enquêtes et recensements démographiques réalisés en Afrique tropicale jusqu'à 1971, et il présente\nle groupe de travail constitué par l'Orstom (Office de la recherche scientifique et technique d'outre-mer), l'Insee, l'Ined (Institut national d'études démographiques) et le secrétariat d'État aux Affaires étrangères pour étudier en commun les problèmes de la\ndémographie africaine.\n\nIntitulé de l'article : Les limites de la comptabilité nationale dans la mesure du bien-être social et dans les comparaisons\ninternationales\nAuteur : Stoupy J.\nRésumé : L'article de J. Stoupy pose le problème de l'évaluation du bien-être social, dans les pays industrialisés et dans ceux en voie\n'd'industrialisation. Il met l'accent sur les difficultés propres à chacune des deux situations.\n\nIntitulé de l'article : Technique statistique et analyse du développement rural. Opération de diagnostic\nAuteur : Raymond H.\nRésumé : H. Raymond explique les raisons de la nécessité d'une enquête socio-économique préalablement à toute opération de\ndéveloppement rural. Il détaille ensuite la méthodologie de cette enquête (plan de sondage, questionnaires, traitement des données).\n\nIntitulé de l'article : Les statistiques de l'éducation\nAuteur : Deblé I.\nRésumé : L'art
```

Etape 3 : La base de données consolidée de tous les articles de Statéco ayant 635 lignes et 6 colonnes est ensuite exploitée en utilisant la bibliothèque Spacy du logiciel Python, qui permet d'appliquer des méthodes d'analyse textuelle (voir par exemple Bengfort, B., Bilbro, R., Ojeda, T. (2018), George, A. (2022)). L'analyse textuelle consiste en :

- La suppression des caractères spéciaux (stopwords) tels que virgule, point-virgule, deux points, articles, ... Sachant que certains articles sont rédigés en anglais, nous avons fusionné les caractères spéciaux français et anglais.
- La suppression des préfixes et suffixes (lemmatizer).
- L'application des techniques d'extraction des mots clés : celles-ci consistent à extraire les mots clés d'un texte (par exemple: démographie, population, entreprise, enquête, sondage, comptabilité nationale, comptes nationaux) en utilisant différents algorithmes d'apprentissage non supervisé sous Python tels que rake_nltk, yake (car préalablement on ne savait pas quels mots clés sont associés à un titre ou un résumé).
- La définition d'un dictionnaire de thématiques. Ainsi, à chaque mot clé ou groupe de mots clés on associe une thématique. Par exemple tous les mots clés suivants « démographie », « démographies », « démographes », « population », « recensement de la population », « natalité », « population indigène », « population urbaine », « structure familiale », « statistiques démographiques », ..., renvoient à la thématique « démographie ».
- L'extraction automatique de l'information en termes de thématiques n'est pas sans surprise. En effet, et à titre d'exemple, lorsque la chaîne de caractère intervient en début de phrase elle n'est pas détectée. En conséquence, il faut rajouter un espace (' '), avant et après la chaîne de caractères, ce qui permet d'identifier la chaîne de caractère et par conséquent la thématique. En effet, si par exemple on cherche le mot « formation » et si on n'ajoute pas un espace avant et après le mot « formation », on risque de faire correspondre le mot « information » à la thématique car « formation » en est une composante. Le même problème est valable pour le mot « ins » qui devrait correspondre à la thématique « institution » alors que « ins » est une composante des mots « instrument », « instabilité », « médecins », « soins », « vaccins », ...
- Sur un autre plan, si la chaîne de caractères « recensement » renvoie à la thématique « population », elle peut poser problème dans le cas d'une chaîne de caractères comme « recensement des entreprises », qui renvoie à la thématique entreprise. Ainsi, si l'extraction des chaînes de caractères sous forme de mots et de groupes de mots permet d'affiner l'extraction, elle doit être complétée par un apport « manuel » de validation.

Etape 4 : La base de données a ensuite été complétée à la main par une codification des organismes d'appartenance des auteurs, selon la nomenclature présentée dans le tableau 2.

On trouvera dans Wavresky (2020) et Djiriguan et Sémécurbe (2019) des exemples de travaux similaires menés sur d'autres publications (publications de l'Inra, Economie et Statistique).

Annexe 2 : Liste des 54 thématiques retenues

administration	indice
agriculture	industrie
ajustement structurel	inégalité
alimentation	informatique
artisanat	institution
budget-consommation	macroéconomie
commerce	methodologie
commerce extérieur	migration
comptabilité d'entreprise	modélisation
comptabilité nationale	monnaie
coopération institutionnelle	nomenclature
démographie	odd
développement	organisation statistique
diffusion	pauvreté
économétrie	planification
économie	prévision - conjoncture
éducation	prix
emploi	rural
énergie	santé
enquête	secteur formel
entreprise	secteur informel
évaluation	séminaire/conférence
finance	séries temporelles
finance publique	sig
formation	sociologie
gouvernance	sondage
indicateur	urbain

Annexe 3 : Nuage de thématiques abordées dans la revue sur l'ensemble de la période et par décennie



Note : La taille est proportionnelle au nombre de fois où la thématique a été détectée dans le corpus de titres et résumés des articles.

Annexe 4 : Méthode utilisée pour la classification des articles

Pour déterminer la classification des articles, la méthode k-means (ou k-moyennes) a été utilisée. Elle a été appliquée aux variables indicatrices, disponibles pour chaque article, relatives aux 54 thématiques.

1. Algorithme utilisé dans la méthode k-means.

Etape 1 : Définir un nombre de classes/clusters (k) à construire

Etape 2 : On tire au hasard k points ; clusters formés par un point chacun. On note ce vecteur :

$$C_0 = (C_{01}, C_{02}, \dots, C_{0k})$$

Etape 3 : On calcule les distances euclidiennes de tous les points i aux k points choisis initialement,

$$C_{0j}, j = 1, \dots, k.$$

$$d_{0j}^i = \|X_i - C_{0j}\|, i = 1, \dots, n, j = 1, 2, \dots, k$$

... et on associe chaque point au cluster le plus proche.

On calcule les nouveaux centres des clusters :

$$C_{0j} = \frac{1}{n_j} \sum_{l \in j} X_j^l$$

Avec X_j^l , les coordonnées du point l appartenant au cluster j .

Etape 4 : On calcule la somme des carrés des écarts intra-clusters (en anglais *Within Cluster Sum of Squares*) :

$$WCSS(k) = \sum_{j=1}^k \sum_{l \in j} \|X_j^l - C_{0j}\|^2$$

Etape 5 : répéter les étapes 3 et 4 jusqu'à ce que la somme des carrés des écarts intra-clusters se stabilise.

2. Critère de nombre optimal de clusters : le critère du coude

On calcule la somme des carrés des écarts intra-cluster (WCSS) pour différentes valeurs de k . On représente le graphe WCSS en fonction du nombre de clusters. Le nombre optimal de cluster correspond au point d'apparition du coude. Dans la figure ci-dessous, un nombre « optimal » correspond à 25.

