

**L'INDUSTRIE DE DÉFENSE :
L'ÉCLAIRAGE DE LA PREMIÈRE ENQUÊTE STATISTIQUE EN FRANCE,
L'ENQUÊTE SUR LES ENTREPRISES DES INDUSTRIES DE DÉFENSE**

Matthieu WYCKAERT et Léa PRENÉ

Ministère des armées, Service statistique ministériel de la défense

matthieu1.wyckaert@intradef.gouv.fr ; lea.prene@intradef.gouv.fr

Mots-clés (6 maximum) : Échantillonnage, calage, imputation, appariement avec des données administratives, Base industrielle et technologique de défense

Domaine concerné : Théorie des sondages amont, Contrôle et redressement des données, data editing

Résumé

La France développe depuis des années son industrie de défense pour soutenir sa souveraineté nationale au travers de ce que l'on a pris l'habitude d'appeler la Base industrielle et technologique de défense (BITD). Bien que reconnu, le secteur de la défense n'est encore que trop peu connu et sa mesure reste parcellaire. En effet, la Nomenclature statistique des Activités économiques dans la Communauté Européenne ou sa déclinaison française ne permet que partiellement de mesurer le poids de l'industrie de défense. A cela s'ajoute le caractère dual des technologies employées : une même technologie peut à la fois servir à produire des matériels de guerre et des matériels civils, rendant difficile la séparation entre les usages civils ou militaires d'un produit à partir des seules données de la statistique publique collectées actuellement.

Une première approche pour délimiter et mesurer le poids de la BITD est d'utiliser des données administratives et de recourir à des dires d'experts. Outre les questions techniques de l'usage de données tirées de systèmes d'information de gestion, cette approche ne permet pas d'avoir un spectre suffisamment large pour avoir une vue d'ensemble de l'industrie de défense, notamment du fait de l'absence dans ces systèmes d'information de l'ensemble des relations inter-entreprises concourant à la production.

Ce problème de spectre peut être résolu par une approche statistique bien connue qu'est la réalisation d'une enquête. C'est pourquoi une première enquête de statistique publique sur le sujet de l'industrie de défense a été réalisée en 2018, l'enquête sur les Entreprises des industries de défense. L'enquête permet de mieux délimiter le périmètre des unités concernées par les activités militaires et de mieux mesurer leur sensibilité aux commandes publiques. Les données administratives utilisées jusqu'à présent conduisaient à minimiser le périmètre et le poids de l'industrie de défense.

De façon à limiter la charge de réponse, le questionnement a été optimisé en mettant à profit le rapprochement des données collectées avec d'autres dispositifs statistiques. Le champ de l'enquête a été délimitée à partir d'une sélection d'activité principale exercée tirée de l'information disponible

dans différentes sources administratives. La méthode d'échantillonnage utilisée est un sondage aléatoire simple stratifié, intégrant une coordination négative de la partie aléatoire de l'échantillon avec les échantillons des autres enquêtes statistiques. La correction de la non réponse utilisée a mobilisé des techniques d'imputation comme le cold-desk ou la méthode du plus proche voisin et des techniques de ré pondération des poids initiaux par la méthode du calage sur marge après modélisation de la non réponse.

Abstract

Although the importance of the defense sector is recognized, its measurement remains yet problematic. Indeed, the nomenclature of activity in the European Community or its French version (NAF) only partially allow us to measure the weight of the defense industry. Access to sensitive administrative data and recourse to experts constitute a first approach to demarcate the productive fabric contributing to the production of war materials. However, this approach remains fragmented due to inter-company relations not subject to administrative formalities. The alternative and second approach is to conduct a survey (survey on « defence industries businesses », EID in French). The sampling base is constituted from administrative data and the information already present in the official statistical bases. Pairing with other statistical surveys makes it possible to limit the size of the questionnaire and therefore to limit the response burden for companies.

Introduction :

La politique de défense des États nécessite la production de systèmes d'armement. En France, la production de ces systèmes est confiée à des entreprises qui coopèrent pour leurs conceptions avec la direction générale de l'armement (DGA) du ministère des armées. Ces entreprises peuvent être autorisées à exporter des armes vers d'autres pays en respect des traités internationaux. Ce tissu industriel est appelé la Base industrielle et technologique de défense (BITD), l'industrie de défense ou encore l'industrie de l'armement. La construction européenne a fait émerger des coopérations entre les firmes des différentes BITD en Europe, allant de relations inter-firmes à des rapprochements, au point que le sujet d'une BITD européenne se pose pour les économistes spécialistes de la question. Des entreprises transfrontalières européennes, souvent sous impulsion politique, ont aussi vu le jour. Le groupe Airbus en est un exemple emblématique. Devant l'émergence de la question de l'Europe de la défense, Eurostat investit le sujet des statistiques officielles sur le domaine de la défense dans le cadre du Plan d'action pluriannuel du Système statistique européen 2021-2027 (abrégé en MAP pour « Multi-annual Action Plan »).

L'article aborde la question de l'utilité d'une enquête pour mesurer le poids économique de l'industrie de défense et explicite le processus mis en œuvre en 2018 par le service statistique ministériel de la défense, l'Observatoire économique de la défense (OED). L'article commence par une définition de l'industrie de défense opérationnelle et traite des lacunes des données issues des autres processus que celui d'une enquête statistique dédiée. Il se poursuit en détaillant le plan de sondage adopté, les techniques de redressement des données mises en œuvre et les résultats obtenus.

1. État des lieux de la production de statistiques sur l'industrie de défense avant la mise en œuvre de l'enquête

La réalisation de statistiques cohérentes et exhaustives sur l'industrie de défense¹ nécessite de bien délimiter le sujet d'analyse pour discerner la meilleure méthode de constitution des données sur le

¹ Dans cette article, on assimile dans le même vocable « industrie de défense », « secteur de l'armement » et « Base industrielle et technologique de défense (BITD) par simplification, étant entendu que ces trois expressions ne sont, en fait, pas strictement équivalentes.

sujet. Dans le cas de l'industrie de défense, l'approche par les nomenclatures officielles actuelles ou par les données administratives disponibles s'avèrent déficientes.

1.1. Qu'est-ce que l'industrie de défense

1.1.1. Une définition tirée des travaux d'économie de défense

Au niveau de la définition de l'industrie de défense, C. Serfati propose une définition simple dès la première ligne de son ouvrage sur le sujet en ces termes : « L'industrie de défense [...] regroupe les entreprises et centres de recherche qui développent, produisent et assurent la maintenance des systèmes d'armes » (Serfati, 2014). On peut compléter cette définition avec celle de J. P. Dunne (que C. Serfati cite dans son ouvrage par la suite) qui propose une définition prenant en considération une approche technologique plus complète : on y trouve comme dans la définition précédente « les unités qui concourent à la production des systèmes d'armes et des équipements létaux », mais également les « unités qui fournissent des produits non létaux mais stratégiques » ainsi que « des unités qui fournissent des produits courants utilisés par les armées » (Dunne, d'après Moura, 2012).

L'industrie de défense a plusieurs particularités. Tout d'abord, le caractère dual de certaines productions de biens et de services rend souvent difficile le classement en civil ou en militaire. Une même technologie peut avoir aussi bien un débouché militaire qu'un débouché civil, sans que la distinction soit facile à opérer dans les faits. Ensuite, ce secteur a une forte présence de grands groupes d'entreprises. Leurs constructions juridiques et leurs degrés d'internationalisation influent sur la pertinence des statistiques usuelles (comme l'affectation de la valeur ajoutée entre les différents pays d'implantation). Enfin, la dynamique industrielle du secteur comporte un haut degré technologique impliquant une structuration des marchés et une recherche constante de l'innovation technologique, rendant le travail des nomenclaturistes plus délicats.

La définition retenue pour l'enquête sur les Entreprises des industries de défense (EID) est celle de S. Moura (Moura, 2012). L'industrie de défense est constituée des « entreprises qui permettent aux armées de conduire leurs opérations, spécifiquement les unités qui concourent à la production des systèmes d'armes et des équipements létaux, de la R&D jusqu'à l'entretien ». Elle correspond à la première partie de la définition de J. P. Dunne (Dunne, 1995). Cette définition approche la BITD par la constitution d'une liste d'entreprises qui concourent aux marchés liés à l'activité militaire par leur chiffre d'affaires ou par les travaux de R&D effectué pour leur élaboration, et non sur la technologie employée. Le chiffre d'affaires militaire est lui défini comme le chiffre d'affaires réalisé par la vente de biens ou de services dont la destination finale est les forces armées nationales ou étrangères, dans un cadre terrestre, naval, aérien ou spatial.

1.1.2. Les produits de l'industrie de défense

Pour définir les produits à inclure parmi les produits militaires d'intérêt, l'enquête EID s'est appuyée sur la liste des produits (ML, pour « *Military List* ») de la Liste commune des équipements militaires de l'Union européenne (JOUE n°2015/C129/01 du 21 février 2015, « Liste commune des équipements militaires de l'Union européenne » adoptée par le Conseil le 9 février 2015).

Il s'agit des véhicules terrestres et de leurs composants (chars d'assaut, véhicules militaires amphibies, véhicules blindés, véhicules de dépannage, véhicules servant à remorquer ou à transporter des systèmes d'armes ou des munitions), des navires de guerre et de leurs composants (navires de surface ou sous-marins, matériel naval spécialisé, moteurs, systèmes de propulsion et appareils de détection immergés), des aéronefs, des véhicules aériens avec ou sans équipage et de leurs composants (composants spécialement conçus, drones, lanceurs et matériel d'appui au sol, matériel pour leur commandement ou leur contrôle), des armes et des munitions et de leurs composants (armes à canon lisse, matériel pour le lancement de gaz dont pistolet, fusil, obusier, mortier, arme anti-char, lance-flammes, mitrailleuse, munitions et dispositifs de réglage de fusée dont cartouches), des bombes, des torpilles, des roquettes, des missiles et de leurs composants (grenades,

mines et systèmes de protection des aéronefs contre les missiles), des armes à énergie cinétique et de leurs composants (modèle d'essai et d'évaluation), des agents chimiques ou biologiques toxiques, des substances radioactives et de leurs composants (agents anti-émeutes et le matériel de protection et de décontamination), des matières énergétiques et de leurs composants (explosifs, combustibles, comburants), du matériel de conduite de tir et du matériel d'alerte et de leurs composants (viseurs d'armement, calculateur de bombardement, système d'acquisition, de désignation, de télémétrie et de poursuite de cible, et de leur matériel de contre-mesure), du matériel électronique, des « véhicules spatiaux » et de leurs composants (matériel de contre-mesure, de guidage et de navigation, matériel de brouillage des systèmes mondiaux de navigation par satellite), du matériel d'imagerie et de leurs composants (enregistreur, caméras, radars, sonars), de blindage et protection et de leurs composants (plaques de blindage, casques, vêtements de protection balistique), des appareils de plongée et de nage sous-marine et de leurs composants, du matériel d'entraînement et de simulateurs et de leurs composants (entraîneurs à la cible radar, aux tirs, à la guerre anti sous-marine, au lancement des missiles), des simulateurs de vols, des systèmes de génération d'images pour simulateurs, des systèmes d'armes à énergie dirigée et de leurs composants (systèmes lasers, systèmes à faisceau de particules, systèmes radiofréquence de destruction).

1.2. Le recours aux nomenclatures officielles ou aux données administratives ne permettent pas de produire des statistiques exhaustives

Les nomenclatures internationales et leur pendant national sont souvent de grand secours pour analyser un secteur ou une branche d'activité. Dans le cas de l'industrie de défense, elles se révèlent imparfaites. Les produits militaires sont répartis dans plusieurs classes de nomenclature.

On peut alors penser à utiliser des sources administratives propres au domaine concerné en scindant le chiffre d'affaires généré par l'activité militaire par ces deux destinations finales : les commandes passées par l'État français aux entreprises résidentes et les exportations de matériels de guerre vers d'autres États. Bien qu'étant une méthode intéressante, le recours aux données administratives disponibles ne permet pas d'estimer précisément le chiffre d'affaire total généré par les commandes militaires.

1.2.1. Les nomenclatures d'activité ou de produits ne permettent pas d'isoler l'industrie de défense

L'industrie de défense ne peut pas être analysée à partir de la nomenclature d'activité française (NAF), pendant de la nomenclature d'activité de la Communauté européenne (Nace). Elle n'y est pas présente comme une branche ou un secteur spécifique. Les produits à destination militaire sont inclus, sauf une exception notable, dans différentes activités au sens de la NAF. En effet, seul le code 30.40Z « Construction de véhicules militaires de combat » sépare clairement les véhicules de combat du reste de la production de véhicules. Les autres armements sont classés dans la classe qui se réfère au bien produit sans mention de l'usage final, civil ou militaire, du produit. C'est par exemple le cas pour les navires de guerre qui sont inclus comme les autres bateaux dans la classe 30.11Z « Construction de navires et de structures flottantes ». De même, la maintenance (« en condition opérationnelle » dans le cas des équipements militaires) n'est pas non plus séparable dans la nomenclature de la maintenance ou de la réparation de matériels civils. Par exemple, pour reprendre l'exemple des véhicules de combat, leurs activités d'entretien et de réparation font partie de la sous-classe 33.11Z « Réparations d'ouvrages en métaux ».

Au niveau des nomenclatures de produits, on retrouve ces limites. Elles n'ont pas nécessairement pour visée de distinguer les produits selon leur destination militaire ou civil. C'est le cas de la nomenclature Prodcod. Sa déclinaison française est dénommée ProdFra. La nomenclature ProdFra sépare certains équipements du reste. On peut citer le code 25401300H4 « Bombes, grenades, torpilles, mines » ou encore le code 25401300H5 « Missiles et similaires ». Cette nomenclature s'efforce aussi de séparer l'usage militaire de l'usage civil avec les codes 2540140000 « Parties et

accessoires pour armes autres que pour armes de guerre » et 25401450H0 « Parties et accessoires pour armes de guerre ». Cependant, cet effort de séparation ne couvre pas le large spectre des équipements militaires. Une autre nomenclature internationale est celle utilisée par les douanes pour établir les statistiques sur les échanges internationaux. Elle permet dans une certaine mesure de retracer les importations et les exportations de matériels de guerre. En effet, ces matériels sont distribués dans les sections et chapitres du tarif des douanes au niveau de la Nomenclature Combinée (NC8) de sorte qu'on retrouve, en croisant avec le degré de confidentialité ou avec le régime de restriction du produit, dans le chapitre 93 les « armes, munitions et leurs parties et accessoires », dans le chapitre 88 les avions militaires et dans le chapitre 89 les navires de guerre. Le recours à cette nomenclature éclaire la question des produits échangés à l'international mais trouve ses limites en ce qui concerne les biens à double usage (civil et militaire).

1.2.2. La BITD tirée de la base de données sur les Statistiques Annuelles sur la Défense, son Industrie et ses Entreprises

La base de données « Statistiques ANnuelles de la Défense, son Industrie et ses Entreprises » (Sandie) recensait les unités légales liées à la défense et opérant sur le territoire national, de façon à identifier *in fine* la BITD. Les premières versions remontent aux années 1990 mais elle est arrivée à sa forme ultime en 2011. Ce dispositif annuel a été entretenu de cette date jusqu'à l'année 2016. « Sandie répond à trois objectifs principaux. En premier lieu, il s'agit d'affiner la définition du périmètre des entreprises liées à la Défense dans le tissu des entreprises françaises. En second lieu, et grâce au recueil de données associées aux différents périmètres, la base fournit un socle à l'analyse statistique du suivi de l'activité de ces entreprises. Enfin, Sandie représente la source des études économiques, à la fois qualitatives et quantitatives, conduites par l'OED » (Médina, 2001). Sandie avait un champ plus étendu que la BITD, incluant notamment tous les fournisseurs du ministère de la défense hors armement.

Les statistiques sur la BITD étaient élaborées à partir d'une extraction de Sandie. L'approche consistait à retenir un périmètre à dire d'experts. D'abord, on ne gardait que les dépenses d'armement du ministère de la défense (tirées du système de pilotage des dépenses de l'État Chorus dès sa mise en route). Ensuite, on intégrait les sociétés exportant des équipements militaires (fournies par le Département des Statistiques et des Études Économiques de la direction générale des Douanes et Droits indirects) appartenant au secteur secondaire, à certains secteurs du commerce de gros, au secteur 62.02A « Conseils en systèmes et logiciels informatiques » ou au 71.12B « Ingénierie, études techniques ». On y ajoutait les fournisseurs des Maîtres d'ouvrage industriels (MOI) français travaillant avec la DGA ou de la Division des applications militaires du Commissariat à l'énergie atomique et aux énergies pour des montants supérieurs à 50 000 euros hors taxe pour les unités légales appartenant au secteur dont l'activité principale exercée (APE) soit commençait par « 20 » ou suivants jusqu'à « 33 » soit était les codes 71.12B ou 62.02A. Enfin, on y ajoutait les statistiques issus des programmes de l'Organisation conjointe de coopération en matière d'armement (Occar) et de l'Organisation du traité de l'Atlantique Nord (Otan). La base de données des entreprises retenues par cette sélection étaient ensuite enrichie de données de caractérisation par des sources entreprises issues des dispositifs de l'Insee, de façon à ne garder que les unités marchandes et exploitantes ou non exploitantes participant au système productif, hors sociétés en participation (SEP) et sociétés civiles immobilières (SCI) non employeuses.

Ce système quoique très abouti ne parvenait pas à fournir une bonne estimation du chiffre d'affaires généré dans les entreprises par les marchés militaires comme l'a montré l'expérience. D'abord, le traitement des informations sur les fournisseurs des MOI était délicate du fait de la dualité de certaines technologies. Ensuite, les relations inter-firmes sur le territoire national n'étaient qu'imparfaitement prises en compte par cette manière. Enfin, la délimitation en termes d'APE par dire d'expert a ses limites que l'enquête EID a montré. Certaines APE était conservées à tort. Par exemple, les entreprises des APE 26.80Z « Fabrication de supports magnétiques et optiques » ou

27.31Z « Fabrication de câbles de fibres optiques » ne se définissent pas comme faisant partie de l'industrie de défense.

1.3. La nécessité d'un dispositif d'enquête

Un dispositif d'enquête s'est avéré nécessaire du fait de toutes ces limites, d'autant plus qu'il permet de recueillir une information plus détaillée.

1.3.1. Une estimation plus réaliste avec l'enquête EID

Il est intéressant de mesurer les différences d'estimation du montant des ventes de produits militaires selon les sources citées auparavant pour se faire une idée du gain de l'enquête. Dans un souci de comparaison à champ constant, on se limite à la partie industrielle de la BITD en reprenant le champ retenu dans l'enquête EID 2018 et en prenant une année de référence commune : l'année 2016.

L'estimation à partir de l'EID (Cf. Tableau 1) semble plus raisonnable d'autant qu'elle est cohérente avec les montants inscrits aux programmes d'équipements des forces armées et aux montants des exportations facturées. Les différences s'expliquent selon les sources utilisées par : pour les données de l'Enquête annuelle de production, par le fait que les codes distinguant le « militaire » dans les nomenclatures sont peu nombreux et parcellaires ; pour les données des douanes, par son champ ; pour les données de la BITD version Sandie, par une approche trop restrictive au niveau des APE retenues et par une sous-estimation des ventes avec les données administratives disponibles.

Tableau 1 : Comparaison entre les sources disponibles pour estimer le chiffre d'affaires militaire généré en France en 2016

Optique de l'estimation	Montant total pour 2016 en M€
Estimation par l'APE 30.40Z « Construction de véhicules militaires de combat » (Source : Insee, Enquête annuelle de production)	336,3
Estimation par les codes ProdFra ayant explicitement une référence à un usage militaire (Source : Insee, Enquête annuelle de production)	305,9
Estimation avec les données des douanes (Méthode Douanes/OED)	8 296,0
Estimation avec le montant renseigné dans la BITD tirée de Sandie	15 588,1
Estimation avec l'enquête EID	24 864,2

NB : L'estimation avec les données des douanes est prise de l'article de M. Wyckaert (Wyckaert, 2019) en utilisant la méthodologie élaborée entre l'OED et le service statistique ministériel des douanes. Par construction, elle ne comprend que les exportations.

1.3.2. Les autres avantages d'un dispositif d'enquête

L'enquête EID 2018 permet de produire une meilleure estimation du chiffre d'affaires militaire. On peut ainsi mieux quantifier l'importance de l'activité militaire sur le territoire national, mieux prendre en compte la sensibilité du tissu productif français aux commandes du ministère des armées et de l'étranger, et de mieux cerner la chaîne de valeur.

L'enquête permet d'interroger les entreprises sur leur positionnement dans la chaîne de valeur de l'industrie de défense aux travers de leurs fonctions, du type d'armement vendu et de la participation ou non à l'effort de Recherche et de Développement (R&D). On recueille des informations sur la variété des savoir-faire valorisés : l'assemblage d'éléments, la production de ces éléments, l'ingénierie ou la participation au maintien en condition opérationnelle.

De façon à limiter le plus possible la charge de réponse, le dispositif a tenu compte des informations déjà présentes dans les autres enquêtes de statistique publique ou dans les registres administratifs. Ainsi, le questionnaire auto-administré est assez court ce qui a permis aux répondants d'y consacrer en moyenne moins d'un quart d'heure. Les informations demandées sont toutes présentes dans les comptes des entreprises soit dans leur comptabilité générale (les comptes de classe 4) soit dans leur comptabilité analytique.

2. Le plan de sondage

L'enquête EID 2018 est la première enquête en France sur le champ de l'industrie de défense. Elle s'est adressée à l'industrie manufacturière de France entière. La réussite de cette première expérience a permis de projeter une nouvelle version de l'enquête en prenant en compte les services industriels et technologiques avec la future enquête sur les Entreprises de défense des industries et des services (EDIS). Dans cet article, nous nous cantonnerons au dispositif EID.

2.1. La base de sondage

La base de données Sandie n'étant pas construite pour être un répertoire au sens statistique du terme, elle n'avait pas toutes les caractéristiques nécessaires. L'utiliser risquait de biaiser les résultats de l'enquête du fait de défauts de couverture. Il était donc plus prudent de s'appuyer sur un répertoire plus large et d'en sélectionner un champ.

2.1.1. Le répertoire Sirius pour constituer la base de sondage

Le répertoire Sirius (Système d'immatriculation au répertoire des unités statistiques) entretenu par l'Insee sert à une grande partie des enquêtes de statistique publique auprès des entreprises. Il est exhaustif (adossé au répertoire Sirene), contient des informations de caractérisation nécessaires à la stratification et est mis à jour régulièrement. La base de sondage de l'EID est extraite se limitant aux sociétés, non micro-entreprises de l'industrie. Le biais sur l'estimation du total de cette restriction est estimé comme marginal puisque d'après les informations de Sandie, les entrepreneurs individuels et les micro-entreprises de l'industrie représentent un montant très faible des commandes versées par le ministère des armées ou par un MOI (moins de 1 %).

Dès le départ, seules les NAF des divisions 20 à 33 (sauf 32 hormis groupe 325) ont été gardées. Ainsi, l'enquête ne concerne pas les activités textiles (la fabrication de tenues militaires), l'industrie du bois, l'agro-alimentaire (fournitures de rations de combat) et le raffinage.

Il peut exister un défaut de couverture dans le cas où la base de sondage contient des unités cessées ou dont les caractéristiques évoluent hors du champ de l'enquête. Pour corriger ce biais qui se traduit par des unités non répondantes mais hors champ, il a été procédé à un appariement *ex post* de la base de sondage et du référentiel Sirius le plus récent (Sautory, 2015).

2.1.2. La sélection sectorielle opérée

La délimitation sectorielle de l'enquête EID a été réalisée sur la base de la probabilité pour une entreprise d'une APE donnée d'appartenir à l'industrie de défense. Pour estimer cette probabilité, il a été utilisé les sources administratives disponibles (par le truchement de Sandie).

On note cette probabilité estimée comme suit : $P(\text{Def/Sect})$. Il s'agit de la probabilité d'appartenir à l'industrie de défense pour une société sachant qu'elle est classée dans une APE donnée.

Le seuil d'inclusion dans le champ de l'enquête EID a été fixé arbitrairement à 5 %. Un contrôle sur les unités concernées par cette exclusion a montré que l'on pouvait considérer comme négligeable l'approximation induite sur le chiffre d'affaires militaire total. Au final, la méthode a conduit à retenir 138 APE².

2.2. La méthode d'échantillonnage

De façon à gagner en précision des estimations, le choix s'est porté sur un sondage aléatoire simple stratifié. Les strates ont été construites avec les APE et un indicateur de présence dans Sandie (croisement APE*Sandie). Deux types de strates ont été constituées : une partie exhaustive et une partie aléatoire. Pour cette dernière, on sélectionne aléatoirement des unités à enquêter avec comme méthode de tirage une coordination négative des échantillons des enquêtes opérée par l'Insee (Guggemos & Sautory, 2012).

2.2.1. Constitution des strates exhaustives et formule de calcul du chiffre d'affaire militaire

Les strates exhaustives contiennent les unités considérées comme faisant partie du cœur de l'industrie de défense. Il s'agit donc, en premier lieu, des unités présentes dans la BITD version Sandie, puis, en second lieu, des unités légales appartenant à une APE jugée comme ayant une forte probabilité de contenir des entreprises de l'industrie de défense. Par exemple, on peut légitimement supposer que le code 25.40Z « Fabrication d'armes et de munitions » contient une forte proportion d'unités de l'industrie de défense.

Explicitement, une APE est attribuée à la partie exhaustive dès lors que la probabilité pour une entreprise de cette APE d'appartenir à l'industrie de défense, $P(\text{Def/Sect})$, dépasse les 40 %. Cette méthode conduit à retenir 32 APE industrielles dans la partie exhaustive soit 7 200 unités.

Les APE affectées en totalité à la partie exhaustive sont : 20.13A « Enrichissement et retraitement de matières nucléaires », 20.51Z « Fabrication de produits explosifs », 22.19Z « Fabrication d'autres articles en caoutchouc », 23.43Z « Fabrication d'isolateurs et pièces isolantes en céramique », 23.44Z « Fabrication d'autres produits céramiques à usage technique », 24.45Z « Métallurgie des autres métaux non ferreux », 24.52Z « Fonderie d'acier », 25.40Z « Fabrication d'armes et de munitions », 25.91Z « Fabrication de fûts et emballages métalliques similaires », 26.11Z « Fabrication de composants électroniques », 26.12Z « Fabrication de cartes électroniques assemblées », 26.51A « Fabrication d'équipements d'aide à la navigation », 26.51B « Fabrication d'instrumentation scientifique et technique », 26.60Z « Fabrication d'équipements d'irradiation médicale, d'équipements électromédicaux et électrothérapeutiques », 26.70Z « Fabrication de matériels optique et photographique », 26.80Z « Fabrication de supports magnétiques et optiques », 27.11Z « Fabrication de moteurs, génératrices et transformateurs électriques », 27.20Z « Fabrication de piles et d'accumulateurs électriques », 27.31Z « Fabrication de câbles de fibres optiques », 28.12Z « Fabrication d'équipements hydrauliques et pneumatiques », 28.13Z « Fabrication d'autres pompes et compresseurs », 28.14Z « Fabrication d'autres articles de robinetterie », 28.15Z « Fabrication d'engrenages et d'organes mécaniques de transmission », 28.21Z « Fabrication de fours et brûleurs », 28.24Z « Fabrication d'outillage portatif à moteur incorporé », 28.41Z « Fabrication de machines-outils pour le travail des métaux », 30.11Z « Construction de navires et de structures flottantes », 30.30Z « Construction aéronautique et spatiale », 30.40Z « Construction de véhicules militaires de combat », 33.14Z « Réparation d'équipements électriques », 33.16Z « Réparation et maintenance d'aéronefs et d'engins spatiaux », et 33.17Z « Réparation et maintenance d'autres équipements de transport ».

² Pour une description détaillée de l'enquête EID et des APE retenues, voire <https://www.insee.fr/fr/metadonnees/source/operation/s2056/presentation>

Par simplification, par la suite, sauf exception, nous allons expliciter les notations pour la seule variable d'intérêt qu'est le chiffre d'affaires militaire. Sans non réponse, le chiffre d'affaires militaire s'obtient en multipliant la part de chiffre d'affaires militaire, obtenue par l'enquête, par le chiffre d'affaires total, tiré des sources fiscales.

En notant CAM_i le chiffre d'affaires militaire de l'unité i , le chiffre d'affaire militaire total pour cette partie de l'échantillon s'obtient comme suit :

$$CAM_{total} = \sum_{i=1}^{7200} CAM_{total\ i}$$

2.2.2. Constitution des strates aléatoires et estimation du chiffre d'affaires militaire total

Les strates aléatoires contiennent les unités non présentes dans la partie exhaustive. Elles sont construites à partir du classement en APE, soit 106 strates aléatoires. Le nombre d'unités tirées dans l'ensemble des strates aléatoires est de 2 795 unités.

On note N_h la taille de la strate aléatoire U_h et n_h le nombre d'unités tirées dans cette strate.

Le chiffre d'affaires militaire total peut se décomposer sous la forme :

$$CAM_{total} = \sum_{h=1}^{106} CAM_{total\ h}$$

On peut estimer CAM_{total} avec l'estimateur sans biais que constitue l'estimateur de Horvitz-Thompson (les π_i étant strictement positifs sans sous-couverture), en affectant à chaque unité i une probabilité d'inclusion π_i .

Dans chaque strate U_h , on sélectionne un échantillon S_h de taille n_h .

On peut estimer sans biais CAM_{total} par $CAM_{total}^{\hat{}}$ avec :

$$CAM_{total}^{\hat{}} = \sum_{h=1}^{106} CAM_{total\ h}^{\hat{}}$$

Le total $CAM_{total\ h}^{\hat{}}$ est estimé sans biais par :

$$CAM_{total\ h}^{\hat{}} = \frac{\sum_{i=1}^{n_h} CAM_i}{\pi_i}$$

L'allocation retenue est une allocation proportionnelle. Les probabilités d'inclusion d'ordre 1 s'écrivent :

$$\pi_k = \frac{n_h}{N_h}$$

et le nombre d'unités tirées dans chaque strate est :

$$n_h = n \frac{N_h}{N}$$

3. Les redressements opérés des données collectées

Une fois les données collectées, un processus de contrôle et d'apurement a précédé l'étape de redressement de façon à augmenter la qualité des estimations. Deux principes ont été mis en œuvre : corriger les erreurs qui ont un impact sur les résultats (Granquist, 1995) et ne changer que le minimum de données lorsqu'il faut corriger une erreur (Fellegi & Holt, 1976).

3.1. L'étape d'apurement des données collectées

3.1.1. Les micro-vérifications

Les vérifications des questionnaires ont été menées de façon séquentielle, en commençant par la première question pour finir par la dernière de façon à corriger les erreurs repérées.

Il s'agit, d'abord, de vérifier que les règles de validité des réponses sont respectées. L'une d'entre elles est de vérifier que les sommes des pourcentages font bien 100 % : La répartition du chiffre d'affaires militaire total pour 2016 et 2017, et si la part de chiffre d'affaires militaire est positive, la répartition du chiffre d'affaires militaire par fonction et par type de produit. La première concerne les sommes de pourcentages proches de 100 %. Dans ce cas, les pourcentages ont été simplement dilatés et arrondis de façon à rétablir l'égalité de la somme à 100 %.

Ensuite, il s'agit de vérifier les règles de cohérence. L'une d'entre elles consiste à comparer la part de chiffre d'affaires militaire déclarée avec les répartitions par fonction ou par type de produit hors modalité « Autres ». Si le pourcentage de chiffre d'affaires militaires d'une année donnée égale la somme des pourcentages respectivement de l'ensemble des fonctions hors modalité « autres fonctions » et l'ensemble des types de produits hors « autres produits », alors la modalité « autres » a été retirée et les pourcentages de la question incriminée ont été dilatés pour que leurs sommes fassent 100 % (le contact n'avait vraisemblablement pas bien compris les consignes explicitées et avait mis dans la modalité « autres » des questions sur les fonctions et les types de produits des réponses sur le chiffre d'affaires civil).

3.1.2. Traitements des hors champs

Suite à l'appariement avec le référentiel Sirius le plus récent, certaines sociétés non répondantes se sont avérées cessées ou mises en sommeil et ont donc été mises hors champ. L'appariement avec le référentiel Sirius a également permis de détecter 393 micro-entreprises. Celles-ci ont été exclues des traitements ultérieurs.

Au travers des précisions sur la nature du chiffre d'affaires mentionnées dans le questionnaire par les répondants, certaines APE sélectionnées initialement à partir de la base de données Sandie ont finalement été classées hors champ au regard de la définition retenue pour l'enquête. Il s'agit entre autres de sociétés pharmaceutiques, de sociétés fabricantes de matériels de cuisine et de sociétés du bâtiment. 19 APE sur les 106 de départ dans la partie aléatoire de l'échantillon ont été ainsi exclues. Il s'agit par exemple du code 23.11Z « Fabrication de verre plat » ou encore du code 30.99Z « Fabrication d'autres équipements de transport n.c.a. ». En prenant en compte ces exclusions d'APE du champ, le taux de réponse de l'enquête EID 2018 s'établit à 85,0 %.

3.2. La correction de la non réponse

Il existe deux types de non réponse : la non réponse partielle qui implique que la question sans réponse ne soit pas une question considérée comme majeure (il s'agit de la question sur l'appartenance à l'industrie de défense) et la non réponse totale qui provient d'une absence de retour de l'entreprise (ou un retour insuffisant).

3.2.1. La correction de la non réponse partielle

La non réponse partielle a été très limitée dans l'enquête EID 2018. Seules 2 % des entreprises répondantes n'ont pas rempli soit la répartition de leur chiffre d'affaires par fonction soit celle par type de produit. La non réponse partielle pour la question sur la R&D a été plus importante avec 10 % des entreprises répondantes ayant omis de remplir cette question.

Lorsque la non réponse partielle concerne les répartitions, on a deux cas de figure. Le premier cas consiste en un oubli du remplissage d'une des deux années demandées. Il a été alors considéré que la répartition de l'année remplie vaut pour les deux années dès lors que la part de chiffre d'affaires militaire de 2016 et 2017 était bien renseignée.

Lorsque il n'y avait aucune des deux années de rempli, il a été procédé à une imputation par le plus proche voisin pour la question sur les fonctions à l'aide d'une macro Sas fournie par la Division Sondage de l'Insee (Macro PPV_TRI). Les variables auxiliaires utilisées ont été par ordre d'importance : le secteur d'activité (APE), le(s) type(s) de produit déclaré(s) dans l'Enquête annuelle de production de l'Insee et la part de chiffre d'affaires militaire. La valeur manquante est ainsi remplacée par la valeur de son plus proche répondant.

On note s_r l'ensemble des répondants et $FONC_i$ le vecteur des valeurs remplies par l'unité i pour la question sur les fonctions, $Z=(z_1, \dots, z_3)$ le vecteur de variables auxiliaires disponibles pour toutes les unités concernées.

On a : $\tilde{FONC}_i = FONC_j$ pour $j \in s_r$ tel que $dist(Z_i, Z_j)$ est minimale.

Pour la question sur les travaux engagés de R&D sur des produits militaires, l'imputation a été faite par cold-deck en utilisant l'information disponible de l'Enquête annuelle sur les moyens consacrés à la R&D expérimentale dans les entreprises du MESRI-SIES de 2015 et 2016 (2017 n'étant pas disponible à l'époque). Lorsque le montant de ressources externes consacrées à la R&D dans l'enquête contenait des financements du ministère des armées, la réponse manquante a été codée à « oui » et dans le cas contraire à « non ».

3.2.2. La correction de la non réponse dans la partie exhaustive

Les unités non répondantes de la partie exhaustive de l'échantillon ont été traitées par imputation par cold-deck. Cette méthode a été préférée à la repondération. Il a été considéré que toutes les unités de la partie exhaustive sont des unités non substituables (Neiter & Buisson, 2010). Pour ce cœur de l'industrie de défense, une liste d'entreprises complète permet de faciliter les travaux futurs des chercheurs en économie de défense. De plus, la correction de la non réponse par imputation dans cette partie d'échantillon peut facilement se faire avec l'information auxiliaire existante.

On pose :

d_i le poids de départ de l'unité i tel que :

$$d_i = \frac{1}{\pi_i}$$

r_i une variable indicatrice de réponse pour l'unité i tel que :

$$r_i = \begin{cases} 1 & \text{si l'unité } i \text{ est répondante} \\ 0 & \text{sinon} \end{cases}$$

On définit un estimateur imputé de la part de chiffre d'affaires militaire avec prise en compte de la non réponse tel que :

$$CAM_{total}^{\hat{}} = \sum_{i=1}^m d_i r_i CAM_i + \sum_{i=1}^m d_i (1-r_i) \tilde{CAM}_i$$

où \tilde{CAM}_i désigne la valeur imputée utilisée pour remplacer la valeur manquante CAM_i .

Le modèle d'imputation est :

$$\tilde{CAM}_i = f(Z_i, \beta) + \varepsilon_i$$

Avec $\varepsilon_i = 0$; $Z = (z_1, \dots, z_5)$ un vecteur de variables auxiliaires disponible pour toutes les unités concernées ; β un vecteur de paramètres inconnus.

Pour une année considérée, l'imputation de la part du chiffre d'affaires tient compte des montants facturés par l'unité légale enregistrés dans le système Chorus, des exportations enregistrées par les douanes, et des montants payés par l'Otan et l'Occar.

Pour la question sur la R&D, l'imputation a tenu compte de l'information disponible dans l'Enquête annuelle sur les moyens consacrés à la R&D dans les entreprises du MESRI-SIES dès lors que les unités considérées déclarent avoir un financement du ministère des armées dans l'enquête. La question sur les fonctions a été imputée par la méthode du plus proche voisin selon le même modèle que pour la non réponse partielle. La question sur les types de produit a été imputée selon la même méthode que l'imputation du chiffre d'affaires militaire en faisant correspondre les différentes informations présentes dans les différentes sources au produit le plus probable.

Au total, parmi les unités concernées, la non réponse totale corrigée par imputation concerne 10 % des unités de l'industrie de défense telles que définies dans l'EID 2018.

3.2.3. La correction de la non réponse par repondération

La partie aléatoire de l'échantillon a été corrigée de la non réponse totale par repondération. Le cadre général consiste à modifier le système de pondération de départ en deux étapes : une étape d'étude du mécanisme de non réponse, pour déterminer des poids ajustés et une seconde étape de calage sur marge, et obtenir les pondérations finales.

On note : \tilde{w}_i le poids ajusté, w_i le poids final de l'unité i .

L'objet de la première étape est de passer du système de pondération de départ au système de pondération ajusté en dilatant les pondérations de départ par la méthode des groupes de réponses homogènes (GRH).

On pose l'estimateur ajusté de CAM_{total} :

$$CAM_{total}^{\hat{}} = \sum_{i=1}^m \frac{d_i}{\hat{r}_i} CAM_i$$

Avec \hat{r}_i la probabilité de réponse estimée de l'unité i .

On suppose le modèle :

$$\hat{r}_i = f(z_i, \hat{Y})$$

z_i est un vecteur de variables auxiliaires disponibles pour toutes les unités échantillonnées et \hat{Y} un vecteur de paramètres inconnus.

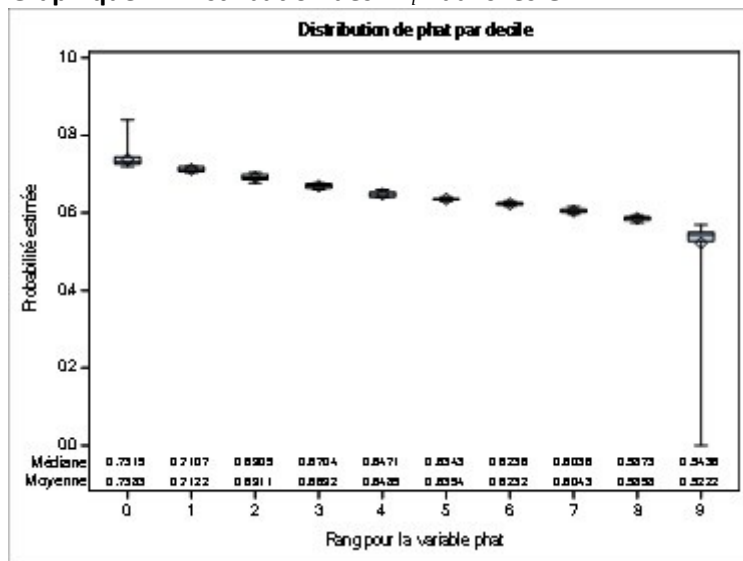
On obtient un système de pondération ajusté pour la non réponse \tilde{w}_i pour toute unité i de l'échantillon où $\hat{w}_i = \frac{d_i}{\hat{r}_i}$.

On définit un modèle non paramétrique basé sur les probabilités de réponse estimées par la méthode des GRH. On forme des classes homogènes par rapport aux probabilités de réponse en passant par leur modélisation. Pour former ces classes, nous avons utilisé la méthode des scores. Dans un premier temps, on cherche un modèle pour estimer la probabilité de réponse à partir de variables auxiliaires disponibles pour toutes les unités de l'échantillon. On obtient un score \hat{r}_i pour l'unité i . On prend soin pour estimer ce score d'enlever les strates sans non réponse. Pour se prémunir d'une augmentation de la variance, on a ensuite sélectionné les variables auxiliaires de façon à ce qu'elles soient liées au chiffre d'affaires militaire (via un modèle linéaire généralisée), corrélées à l'appartenance à l'industrie de défense (via un test du V de Cramer) et explicatives de la non réponse observée (par un modèle logistique). Cela a conduit à retenir un modèle logistique avec deux co-variables en procédant par *stepwise* : un regroupement de strates et la région du siège social de l'unité légale. Le test de Hosmer-Lemeshow valide l'adéquation du modèle.

Le modèle d'estimation est de la forme : $\hat{r}_i = f(\text{Groupe de strates}_i, \text{Région}_i, \hat{Y})$

Ensuite, on construit des GRH par rapport à \hat{r}_i . La séquence de regroupement des unités i se fait avec la méthode des quantiles égaux. On regroupe en classe de tailles approximativement égales les unités i en ordonnant les valeurs de \hat{r}_i en ordre croissant. Le nombre de classes ainsi défini s'établit à dix, en respectant la contrainte d'un nombre suffisant de répondants dans chaque classe. Les dix GRH sont construits sur les déciles des probabilités estimées de réponse par le modèle déterminé ci-dessus. On obtient une dilatation moyenne des poids initiaux sur l'ensemble de l'échantillon de 1,56 (rapport entre le poids ajusté et le poids initial) avec un coefficient de variation de 9,3 %. La variance de la pondération passe de 0,898 à 2,041, soit une augmentation assez contenue.

Graphique 1 : Distribution des \hat{r}_i dans les GRH



La seconde étape consiste à une étape de calage sur marges pour passer du système de pondération ajusté au système de pondération final. Pour assurer une cohérence et une comparabilité entre les résultats de l'enquête et les résultats des statistiques structurelles sur les entreprises calculés par l'Insee, le calage se fait sur les totaux connus pour les secteurs concernés du chiffre d'affaires total et de la valeur ajoutée hors taxe.

L'estimateur de calage est de la forme :
$$CAM_{total}^{\hat{}} = \sum_{i=1}^m w_i CAM_i$$

Avec les contraintes suivantes sur w_i :

- Minimiser $\sum_{i=10}^m \frac{d_{(i)} G(w_i/d_i)}{q_i}$ où G est une fonction de pseudo-distance et q_i un coefficient de pondération indiquant l'importance de chaque unité dans le calcul de la distance ;

$$\left\{ \begin{array}{l} \sum_{i=1}^m w_i CA_i^{2016} = CA_{total}^{2016} \\ \sum_{i=1}^m w_i CA_i^{2017} = CA_{total}^{2017} \\ \sum_{i=1}^m w_i VA_i^{2016} = VA_{total}^{2016} \\ \sum_{i=1}^m w_i VA_i^{2017} = VA_{total}^{2017} \end{array} \right. \text{ avec } \left\{ \begin{array}{l} CA_{total}^t = \sum_{i \in APEduc champ} CA_i^t \\ VA_{total}^t = \sum_{i \in APEduc champ} VA_i^t \end{array} \right.$$

On peut écrire l'estimateur de calage comme : $CAM_{total}^{\hat{}} = \sum_{i=1}^m d_i F(d_i \hat{Y}' x_i) CAM_i$

Où F est la fonction de calage (fonction inverse de G') et x_i le vecteur des totaux connus.

La fonction de distance prise est celle de la méthode Logit de façon à limiter la dispersion des poids et de ne pas avoir de poids finaux négatifs (que l'on sait difficilement compréhensible pour les non initiés).

La fonction de pseudo-distance est donnée par :

$$G(w_i/d_i) = \begin{cases} \left(a_i \log\left(\frac{a_i}{1-L}\right) + b_i \log\left(\frac{b_i}{M-1}\right) \right) A^{-1}, & L < \frac{w_i}{d_i} < M \\ \infty & \text{sinon} \end{cases}$$

où :

$$a_i = \frac{w_i}{d_i} - L$$

$$b_i = M - \frac{w_i}{d_i}$$

$$A = \frac{M-L}{(1-L)(M-1)}$$

Par tâtonnement, on retient les bornes L=0,2 et M=3 qui permettent de minimiser l'écart entre les poids ajustés et les poids finaux et de faire converger l'algorithme de calage. Le coefficient de variation des poids finaux passe à 25,4 %.

4. Résultats de l'enquête

Le succès du processus d'enquête a permis de faire un retour sur l'exploitation des fichiers administratifs et de publier pour la première fois une estimation solide du chiffre d'affaires militaires dans l'industrie manufacturière.

4.1. De Sandie au Répertoire des entreprises fournisseurs de la défense

14^e édition des Journées de méthodologie statistique de l'Insee (JMS 2022)

A partir de 2017, l'OED n'a plus produit la base de données Sandie. En revanche, le dispositif de suivi des entreprises fournisseurs de la défense s'est enrichi pour devenir un répertoire statistique le Répertoire des entreprises fournisseurs de la défense (REFD)³ qui est référencé dans le référentiel de métadonnées statistiques de l'Insee.

Le REFD est un répertoire statistique recensant annuellement les montants des achats faits par le ministère des armées à ses entreprises fournisseurs. Il est constitué de données sur l'ensemble des dépenses effectuées par le ministère des armées, de données sur les fournisseurs, de bases statistiques sur les entreprises et de fichiers de nomenclatures officielles.

4.2. Estimation et calcul de précision

4.2.1. Estimation sur le chiffre d'affaires total

Le chiffre d'affaires militaire estimé avec l'enquête EID 2018 est de 29,6 milliards (Wyckaert, 2019). « Le segment aéronautique est le premier marché avec 36 % des ventes 2017 [...]. Le segment des navires de guerre arrive en seconde position avec 16,5 % des ventes. Le troisième marché est celui du matériel électronique et d'imagerie à usage militaire [...], avec 9,6 % du chiffre d'affaires militaire total »⁴.

Les données d'enquête ont aussi permis de commencer à chiffrer comment l'industrie se structure autour des systèmes d'armement. C. Serfati (Serfati, 2014) mentionnait l'existence d'un noyau. « En schématisant, la filière des industries de défense comporte deux niveaux. Une « périphérie » composée de sociétés présentes sur plusieurs marchés et marginalement sur les marchés liés à l'activité militaire, soit 40 % des sociétés du secteur pour 12 % du chiffre d'affaires total. Un « noyau » qui a une part significative de son chiffre d'affaires dans le militaire. Il comprend les fabricants d'engins explosifs à propulsion, de matériels électroniques militaires, d'équipements aéronautiques, terrestres et navales, les sous-traitants spécialisés, l'ingénierie en matières énergétiques à vocation militaire, la maintenance en condition opérationnelle et les sociétés de démantèlement ».

4.2.2. Calcul de précision

La non réponse à l'enquête EID 2018 a été assez limitée. Comme nous l'avons vu, elle diffère selon les secteurs (Cf. Graphique 2).

En présence de non réponse (et en supposant que notre estimation est non biaisée) et lorsque celle-ci est corrigée à l'aide d'un modèle d'imputation, la variance d'un total est composée de trois termes : un terme de variance d'échantillonnage, un terme de variance liées à la non réponse et un terme de covariance entre l'erreur qui est due à l'échantillonnage et l'erreur qui est due à l'imputation (Vallée, 2014).

$$\text{On a : } V_{total} = V_{sam} + V_{NR} + 2V_{mix}$$

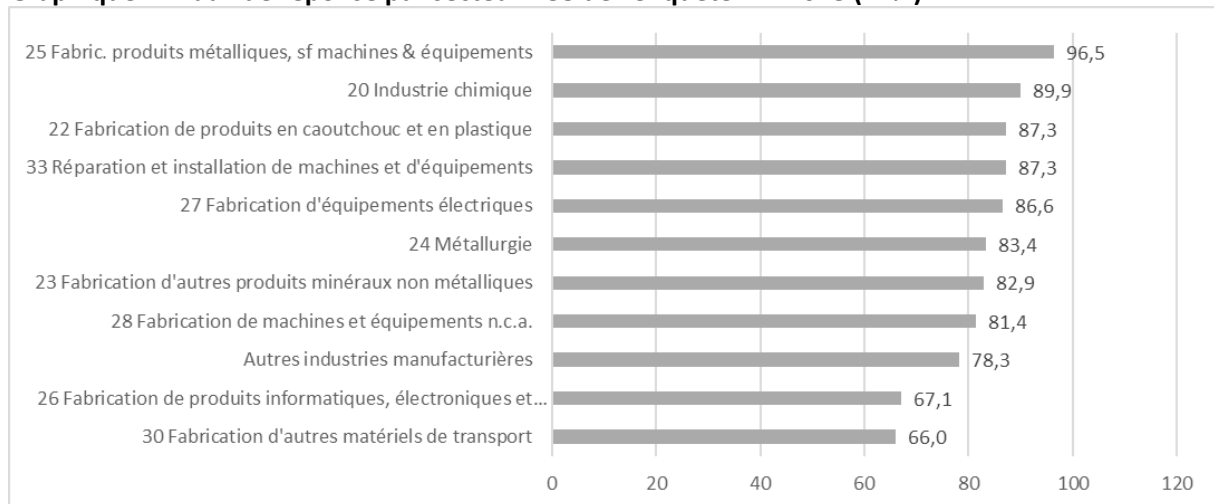
Lorsque l'erreur de non réponse est corrigée par un modèle de non réponse par repondération, la variance d'un total est obtenue avec les deux premiers termes.

$$\text{On a : } V_{total} = V_{sam} + V_{NR}$$

³ Pour des compléments, voire <https://www.insee.fr/fr/metadonnees/source/serie/s1043> et <https://www.defense.gouv.fr/sga/le-sga-en-action/economie-et-statistiques/qualite-statistique/fiches-methodologiques>

⁴ On peut trouver une vidéo dessinée produite par l'OED sur l'industrie de défense sur la chaîne YouTube de l'OED : <https://www.youtube.com/watch?v=5hsP0p9EPq8>

Graphique 2 : Taux de réponse par secteur A88 de l'enquête EID 2018 (En%)



Source : OED, EID 2018.

La forme générale de variance du chiffre d'affaires militaire total est un mélange entre les deux décompositions, sachant qu'en l'occurrence le terme de covariance entre l'erreur due à l'échantillonnage et l'erreur due à l'imputation est nul par construction. Il est de la forme :

$$V_{total} = V_{sam} + V_{NR}$$

avec :

- V_{sam} la variance d'échantillonnage pour la partie aléatoire de l'échantillon ;
- V_{NR} la variance liée à la non réponse qui se décompose en deux termes : un terme de variance liée à la repondération et un terme lié à l'imputation, telle que : $V_{NR} = V_{NR1} + V_{NR2}$

A partir de cette variance, on peut définir un indicateur de précision de notre estimation du total avec le coefficient de variation de la forme :

$$CV_p(\widehat{CAM}_{total}) = \frac{\sqrt{V_p(\widehat{CAM}_{total})}}{\widehat{CAM}_{total}}$$

Plusieurs méthodes de calcul de la variance seraient possibles : approximation de type Jackknife, de type bootstrap ou approximation analytique. On peut procéder à une estimation en séparant la partie aléatoire de la partie exhaustive, puisque par construction les deux parties sont indépendantes l'une de l'autre.

On peut estimer la variance totale du chiffre d'affaires militaire total V_{total} par :

$$\widehat{V}_{total} = \widehat{V}_{sam} + \widehat{V}_{NR1} + \widehat{V}_{NR2}$$

On peut alors estimer le coefficient de variation de l'estimation du chiffre d'affaires militaire total par :

$$\widehat{CV}_p(\widehat{CAM}_{total}) = \frac{\sqrt{\widehat{v}(\widehat{CAM}_{total})}}{\widehat{CAM}_{total}}$$

Sans non réponse, la variance d'échantillonnage est de la forme :

$$V_p(\widehat{CAM}_{total}) = \sum_{h=1}^{106} N_h^2 \frac{1 - \frac{n_h}{N_h}}{n_h} S_{CAM_h}^2 \quad \text{avec} \quad S_{CAM_h}^2 = \frac{1}{N_h - 1} \sum_{k \in U_h} (\widehat{CAM}_k - \overline{\widehat{CAM}_k})^2$$

Que l'on peut estimer par :

$$V_p(\widehat{CAM}_{total}) = \sum_{h=1}^{106} N_h^2 \frac{1 - \frac{n_h}{N_h}}{n_h} s_{CAM_h}^2 \quad \text{avec } s_{CAM_h}^2 = \frac{1}{n_h - 1} \sum_{k \in U_h} (\widehat{CAM}_k - \overline{\widehat{CAM}_k})^2$$

On a procédé à une première estimation de la variance d'échantillonnage de l'enquête EID. La variance d'échantillonnage de l'enquête \widehat{V}_{sam} peut être estimée numériquement avec la fonction R Qvar issue du package Gustave mis à disposition par l'Insee⁵ (Paliod & Chevalier, 2018).

De façon à avoir un premier indicateur de précision, même partiel, sur la seule partie aléatoire de l'échantillon, on définit le coefficient de variation suivant :

$$CV_{aléa}(\widehat{CAM}_{Total\ Strates\ aléatoires}) = \frac{\sqrt{V_{aléa}(\widehat{CAM}_{Total\ Strates\ aléatoires})}}{\widehat{CAM}_{Total\ Strates\ aléatoires}}$$

estimé par :

$$\widehat{CV}_{aléa}(\widehat{CAM}_{Total\ Strates\ aléatoires}) = \frac{\sqrt{v(\widehat{CAM}_{Total\ Strates\ aléatoires})}}{\widehat{CAM}_{Total\ Strates\ aléatoires}}$$

La résolution numérique obtenu avec le package Gustave est :

$$\widehat{CV}_{aléa}(\widehat{CAM}_{Total\ Strates\ aléatoires}) = 20,3 \%$$

Par rapport à l'estimation du chiffre d'affaires total en France en 2017, la contribution de l'estimation du total provenant de la partie aléatoire est marginale (moins de 1 % du total). Des travaux sont en cours pour réaliser une estimation du coefficient de variation de l'estimation du chiffre d'affaires militaire total, soit en suivant la méthode de J. Shao et K. J. Thompson (Shao et Thompson, 2009) par une approximation de type Jackknife ou en utilisant la macro Sas « Sevani » de Statistique Canada (Beaumont et Mitchell, 2002).

Bibliographie

- [1] Beaumont J.-F. et Mitchell C., « Système pour l'estimation de la variance due à la non-réponse et à l'imputation (SEVANI) », Modélisation des données d'enquête pour la recherche économique et sociale, Recueil du Symposium 2002 de Statistique Canada, 2002.
- [2] Demoly E., Fizzala A. et Gros E., « Méthodes et pratiques des enquêtes entreprises à l'Insee », Journal de la Société Française de Statistique, Vol. 155 N° 4, 2014.
- [3] Dunne J. P., « The Defense Industrial base », Handbook of Defense Economics, vol. 1, p. 1139-1176, Repec, 1995.
- [4] Fellegi I. P. et Holt D., « A systematic approach to automatic edit and imputation », Journal of the American Statistical Association, n°71, p. 17-35, 1976.
- [5] Granquist L., « Improving the traditional editing process », In *Business Survey Methods*, (Eds., B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge and P.S. Kott), John Wiley & Sons, Inc., p. 385-401, 1995.
- [6] Guggemos F. et Sautory O., « Sampling Coordination of Business Surveys Conducted by INSEE », Proceedings of the Fourth International Conference of Establishment Surveys, Montréal, Canada, 11-14 juin 2012..

⁵ Ce package ne prend pas en compte le traitement par imputation de la non réponse partielle, assimilant ces données imputées à des données d'enquête par simplification. Pour des approfondissements sur le package
14^e édition des Journées de méthodologie statistique de l'Insee (JMS 2022)

- [7] Haziza D., « Imputation and inference in the presence of missing data, Handbook of statistics », Vol. 29, *Sample Surveys : Theory Methods and Inference*, Editors : C.R. Rao and D. Pfeffermann, p. 215-246, 2009.
- [8] Médina V., « La base de données Sandie : un outil d'analyse économique », *Ecodef* n° 14, mars 2001.
- [9] Moura S., « La base industrielle et technologique de défense : identification et caractéristique », *Ecodef Études*, n° 58, janvier 2012.
- [10] Moura S., « La dualité dans la base industrielle et technologique de défense », *Ecodef Études*, n° 66, janvier 2014.
- [11] Neiter B. et Buisson B. « Comment redresser une enquête thématique », Document de travail, N° 2010/01, Direction des statistiques d'entreprises, Insee, 2010.
- [12] Paliot N. et Chevalier M., « Estimation de variance dans les enquêtes de l'Insee : le package R gustave », 10ème Colloque francophone sur les sondages, 26 octobre 2018 .
- [13] Sautory O., « Les enjeux méthodologiques liés à l'usage de bases de sondage imparfaites », Acte des 12èmes Journées de Méthodologie Statistique, 31 mars au 2 avril 2015.
- [14] Shao J. et Thompson K. J., « Estimation de la variance en présence de non-répondants et de strates à tirage complet », *Techniques d'enquête* Vol. 35, N°2, Statistique Canada, p. 235-246, décembre 2019.
- [15] Serfati C., « L'industrie française de défense », La documentation française, 2014.
- [16] Vallée A. A., « Estimation de la variance en présence de données imputées pour des plans de sondage à grande entropie », Département de mathématiques et de statistique, Faculté des arts et des sciences, Université de Montréal, juillet 2014.
- [17] Wyckaert M., « 6,4 milliards d'euros d'excédent commercial lié aux matériels de guerre en 2018 », *Ecodef Statistiques*, n° 126, mai 2019.
- [18] Wyckaert M., « Près de 30 milliards de chiffre d'affaires militaire pour les entreprises industrielles de la BITD en 2017 », *Ecodef Statistiques*, n° 133, septembre 2019.