

## Le manque à gagner de TVA en France

Cette étude propose une estimation de la taxe sur la valeur ajoutée (TVA) qui aurait dû être perçue mais ne l'a pas été du fait d'erreurs ou d'omissions déclaratives—qu'elles soient volontaires ou non. Cette estimation repose sur une extrapolation des taux d'irrégularités détectées par les contrôles de la DGFiP ; elle est validée par une expérience de contrôles aléatoires réalisée par les équipes de vérification de la DGFiP au cours de l'année 2022.

Le manque à gagner de TVA dû à la sous-déclaration des entreprises déclarant de la TVA serait ainsi compris dans une fourchette de 6 à 10 Md€, soit 4-5% du montant de TVA effectivement collecté. Cette estimation est un minorant du manque à gagner sur l'ensemble de la TVA, car elle n'inclut ni les entreprises ne déclarant pas de TVA (à tort ou du fait du régime de la franchise en base de TVA), ni les potentielles irrégularités qui ne seraient pas détectées par les contrôles fiscaux. Le secteur du commerce représente la plus grande part de ce manque à gagner (30 %), à hauteur de son poids dans la TVA perçue, tandis qu'en termes relatifs, ce sont dans les activités immobilières et dans l'hébergement et la restauration que les taux de sous-déclaration seraient les plus élevés, à 13 % de la TVA de chaque secteur. Pour 44 % de ce manque à gagner, les irrégularités portent sur la dissimulation ou l'omission d'opérations qui auraient dû être assujetties à la TVA.

*Auteurs : Nicolas Charnacé, Acher Elbaz et Pierre Barnouin*

### Introduction

La taxe sur la valeur ajoutée (TVA) est la première recette fiscale de l'État et finance également d'autres administrations publiques. Cependant, une partie de l'impôt échappe à l'administration fiscale, du fait de déclarations erronées ou manquantes de la part des contribuables (volontairement ou non) et de l'impossibilité de collecter la TVA due dans certains cas. L'impôt non-perçu représente un manque à gagner pour les administrations publiques.

La DGFiP a parmi ses missions le contrôle fiscal, qui permet de veiller au respect du droit fiscal et donc du principe d'égalité devant l'impôt. Elle contribue ainsi à réduire la fraude fiscale, qui est l'évitement délibéré et illégal de l'imposition. Si le contrôle fiscal et les sommes qu'il permet de recouvrer font l'objet de rapports publics réguliers, il existe néanmoins, comme la section 2 le montre, peu de travaux permettant de quantifier de manière robuste l'ampleur économique de la fraude.

Cette étude propose une mesure du manque à gagner de TVA, c'est-à-dire des sommes non recouvrées par la DGFiP du fait d'irrégularités déclaratives, que celles-ci soient volontaires ou non. Elle s'appuie en cela sur une étude de l'Insee ([Quantin & Welter-Médée 2022](#)), qui consistait à extrapoler les montants de redressement notifiés aux entreprises contrôlées à l'ensemble des entreprises redevables de la TVA, en l'améliorant à

plusieurs titres : en ventilant les droits de TVA redressés par année d'affaire, en l'appliquant à une période plus contemporaine, en incluant un champ plus exhaustif de contrôles fiscaux, et en déclinant les résultats par secteur d'activité et par type d'irrégularité constatée.

Les chiffrages présentés ici consistent intuitivement à extrapoler à l'ensemble de la population des entreprises assujetties à la TVA et ayant déposé au moins une déclaration de TVA, un montant manquant de versement, à partir des informations provenant du sous-ensemble des entreprises contrôlées par l'administration fiscale. Le manque à gagner de TVA traduit donc l'ensemble des anomalies déclaratives des entreprises connues de l'administration fiscale.

Le manque à gagner de TVA parmi les entreprises déclarantes est ainsi estimé entre 6,9 et 9,1 Md€ en 2019, dernière année pour laquelle l'essentiel des contrôles ont été menés à terme. Une estimation provisoire pour 2020 donne un ordre de grandeur proche, mais doit être interprétée avec prudence du fait des spécificités de cette année de crise sanitaire. Ce montant est à voir comme un minorant du manque à gagner total de TVA, car l'extrapolation est d'une part limitée aux entreprises déclarant de la TVA et ne prend donc pas en compte les entreprises en défaillance déclarative complète ; et d'autre part, elle ne corrige pas un possible biais de détection. Sur dix ans, ce montant reste stable dans cet intervalle de 6-10 Md€.

L'hébergement-restauration et les activités immobilières se démarquent par le taux estimé d'irrégularités le plus élevé. Sur l'ensemble de l'économie, la dissimulation et l'omission d'opérations qui auraient dû être assujetties à la TVA sont les principales causes du manque à gagner ainsi mesuré.

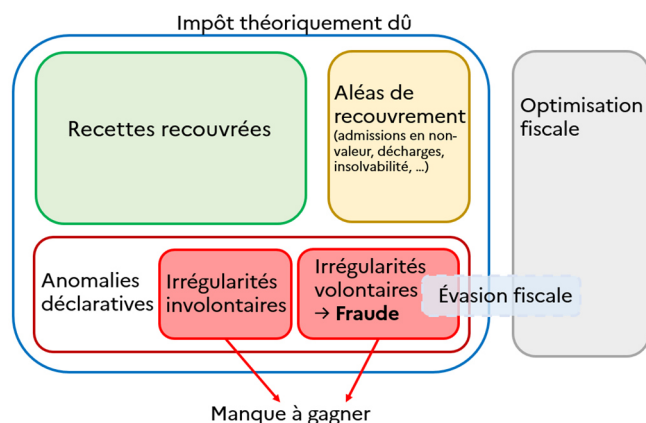
La méthodologie proposée ici permettra à la DGFIP de disposer d'estimations régulièrement actualisées des phénomènes de fraude à la TVA, ce qui permettra d'établir des points d'étape dans la lutte contre la fraude.

## 1. Peu de chiffrages pour appuyer le débat public sur la fraude fiscale

### 1.1. Manque à gagner : un concept plus large que la fraude

La fraude fiscale recouvre de multiples comportements et ne représente pas un concept univoque [graphique 1]. Beaucoup de pays le définissent dans leurs droits internes, comme la France qui caractérise la fraude par la violation délibérée de la législation fiscale (article 1741 du Code Général des Impôts). La frontière avec les comportements d'optimisation fiscale (avec en particulier les pratiques de transfert de bénéficiaires) est par ailleurs poreuse ; entre les deux, l'évasion fiscale se situe dans une zone grise entre légal et illégal.

Graphique 1 : Définitions de la fraude fiscale et du manque à gagner



Source : DGFIP

Il est impossible de mesurer la fraude directement, car c'est un phénomène qui non seulement est inobservable (par définition), mais qui contient un élément intentionnel qu'on ne peut approcher par des méthodes statistiques. À la place, l'écart fiscal, qui est un concept plus large ajoutant à la fraude fiscale les irrégularités involontaires et les aléas de recouvrement, permet de mesurer la différence entre les montants effectivement collectés par l'administration et ceux qui l'auraient été si toutes les obligations avaient été strictement respectées par les contribuables.

Deux principaux types de méthodes permettent d'estimer l'écart fiscal :

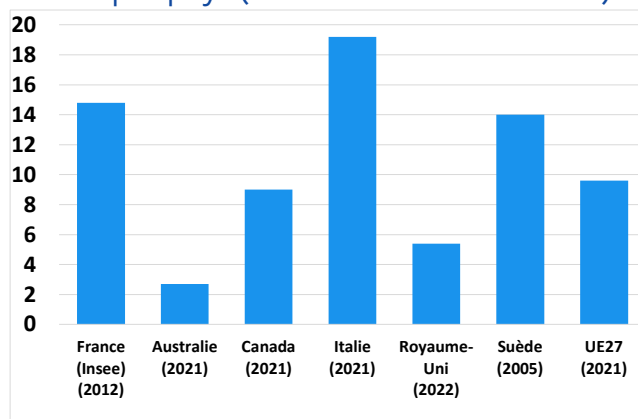
- (1) Les approches descendantes sont des méthodes macro-économiques qui simulent la taxation sur la base d'agrégats de comptes nationaux et en dérivent une estimation de l'impôt dû—une TVA théorique recalculée avec laquelle comparer la TVA collectée. Les approches descendantes nécessitent peu de ressources et utilisent des données publiquement disponibles et comparables entre pays ; elles peuvent être utilisées pour des comparaisons internationales (comme le rapport sur l'écart TVA de la Commission européenne). Elles ne permettent en revanche pas d'identifier directement les causes et les mécanismes de la fraude (cf. partie 1.2).
- (2) Les approches ascendantes sont, elles, des méthodes micro-économiques exploitant des données individuelles, telles que les résultats de contrôles fiscaux ou les déclarations fiscales. Ces méthodes permettent une plus grande granularité et d'appréhender les comportements spécifiques des contribuables, de désagréger les résultats (par industrie ou région), et d'établir des probabilités de fraude pour des profils précis. Elles sont en revanche plus coûteuses en ressources et dépendent de la qualité, souvent hétérogène, des données individuelles.

La disponibilité des données des contrôles fiscaux permet à la DGFIP d'utiliser une méthode ascendante et d'estimer un manque à gagner ventilé par secteur et par type d'irrégularité.

### 1.2. Les estimations des écarts fiscaux de TVA oscillent entre 3 % et 20 % selon les pays

À l'échelle internationale, ce sont en priorité les économies avancées qui s'efforcent d'estimer leurs écarts fiscaux, du moins pour les principaux impôts—à commencer par la TVA : 57% le font pour la TVA, 38% pour l'IR, et 36% pour l'IS<sup>1</sup>. Ces résultats issus d'enquête auprès des autorités fiscales étrangères confirment que, s'agissant de lutte contre la fraude, la majorité d'entre elles s'attellent en priorité à la TVA.

Graphique 2 : Estimations de l'écart fiscal sur la TVA par pays (en % de la TVA collectée)



Sources : publications des administrations fiscales ou des instituts statistiques nationaux sur des années différentes (indiquées entre parenthèses)

<sup>1</sup> Enquête ISORA sur les administrations fiscales, 2018

Plusieurs pays (Australie, Canada, Italie, Royaume-Uni, Suède) publient des estimations régulières de l'écart fiscal sur la TVA, généralement avec une combinaison de méthodes ascendantes et descendantes (ou uniquement des méthodes descendantes). La fourchette des estimations de manque à gagner est vaste, comprise entre 3 % du montant théorique de la TVA pour l'Australie et 19 % pour l'Italie [graphique 2].

La plupart des administrations fiscales estime que les contrôles aléatoires permettent de meilleures estimations de l'écart fiscal. Ainsi, l'Australie, le Canada, l'Irlande, le Danemark, la Suède, le Royaume-Uni, et les Etats-Unis effectuent des contrôles aléatoires, la plupart d'entre eux avec une actualisation chaque année.

### 1.3. Des estimations pour la France variant de 7 à 25 Md€

Beaucoup de débats ont lieu autour des estimations de la fraude fiscale alors que peu de chiffreages fiables et récents existent. Malgré tout, quelques travaux, se sont efforcés d'estimer l'écart fiscal sur la TVA en France :

- A l'aide d'une méthode descendante, la Commission européenne estime l'écart fiscal sur la TVA pour la France à 10-15 Md€ en 2020 et 2021. Cette méthode produit des résultats volatils, qu'il faut interpréter avec prudence (cf. partie 6). Elle englobe en outre dans l'écart fiscal les dispositions les plus granulaires du code fiscal (taux spécifiques sur certains produits, exemptions de certains contribuables), qui ne peuvent être simulées avec des agrégats macroéconomiques.
- Le Conseil des prélèvements obligatoires (CPO) a donné en 2007 une fourchette de 7 à 12 Md€ en extrapolant les redressements sans en contrôler le biais de sélection.
- L'Insee a publié en 2022 une estimation de la fraude à la TVA en corrigeant les données des contrôles fiscaux externes ciblés du biais de sélection ; le résultat, pour l'année d'activité 2012, est de l'ordre de 20-25 Md€.

Cette étude vise à répondre à ce besoin identifié de chiffreages plus robustes et récurrents au niveau français et à enrichir le débat public sur le sujet.

## 2. Les données de contrôle de la DGFIP

### 2.1. Le contrôle fiscal des entreprises

Le contrôle fiscal a pour objectif de détecter et de rectifier les manquements délibérés ou opérés de bonne foi par le contribuable, que celui-ci soit particulier ou professionnel.

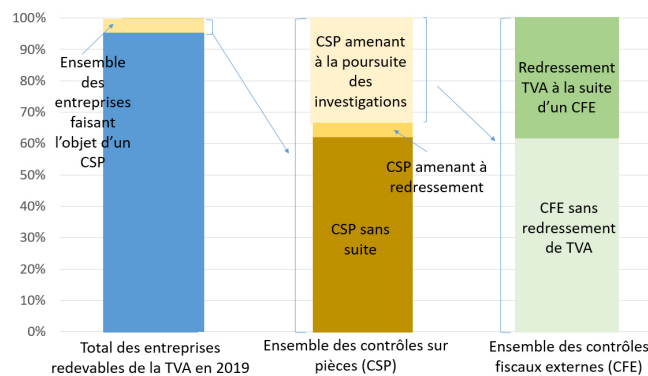
Pour les entreprises, trois types de structures réalisent des contrôles : la Direction des vérifications nationales et internationales (DVNI) pour les grandes entreprises, les Directions inter-régionales spécialisées de contrôle fiscal (Dircofi) pour les entreprises de taille moyenne et les directions régionales et départementales (brigades locales) pour le segment regroupant les plus petites entreprises. Si les entreprises dépendant de la DVNI

sont contrôlées à intervalles réguliers, les PME sont quant à elles contrôlées sur la base d'une analyse du risque de fraude.

Une fois une entreprise identifiée comme à risque, celle-ci peut faire l'objet d'un contrôle sur pièces (CSP). Il s'agit d'une analyse critique globale des déclarations déposées, à l'aide des renseignements et documents figurant au dossier de l'entreprise et d'outils informatiques permettant de recouper et d'analyser les données fiscales et comptables des entreprises.

A l'issue de cette première analyse, il existe trois possibilités : (1) le service de vérification décide de ne pas donner suite car il estime que l'entreprise n'est pas à rectifier ; (2) le service estime avoir assez d'éléments pour redresser l'entreprise sans intervention sur place ; ou : (3) le service estime que les éléments dont il dispose motivent la poursuite des investigations par l'engagement d'un contrôle fiscal externe (CFE) lors duquel les enquêteurs se rendent sur place [graphique 3].

Graphique 3 : Présentation du processus de contrôle fiscal au titre de 2019



**Lecture :** 4,5 % des entreprises redevables ont fait l'objet d'un contrôle fiscal sur pièces portant sur la TVA au titre de 2019

**Note :** Ce schéma est simplifié et ne prend pas en compte la part des contrôles fiscaux externes autoprogammés par les services de recherche et vérification

**Source :** DGFIP

Un contrôle fiscal externe peut prendre plusieurs formes, en fonction des cas :

- L'examen de comptabilité : un contrôle à partir de la transmission des comptes de l'entreprise.
- La vérification générale : un contrôle sur site, par le service, de la situation fiscale de l'entreprise par rapport à l'ensemble des impôts sur les trois années non encore couvertes par la prescription.
- La vérification ponctuelle : une vérification sur des années spécifiques ou pour un seul impôt.
- La vérification simple : une procédure simplifiée utilisée pour les petites entreprises et les bénéficiaires non commerciaux.

Cette étude fait l'hypothèse qu'une vérification générale conduit au contrôle systématique de la TVA.

Lorsque le contrôle détecte une irrégularité, l'entreprise se voit notifier un redressement correspondant au montant d'impôt manquant qui aurait dû être acquitté, ainsi que des intérêts de retard et d'éventuelles pénalités et sanctions. Autrement, le dossier de l'entreprise est classé sans suite.

Tableau 1 : Répartition des contrôles en fonction de leur année de clôture et d'affaires

Exercices contrôlés	Année de clôture du dossier												Nombre de contrôles, en milliers
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	
2009	60%	30%	8%	2%									116
2010	39%	37%	18%	5%	1%								197
2011	21%	31%	28%	15%	4%	1%							243
2012	5%	20%	29%	28%	14%	3%	1%						248
2013		4%	20%	29%	28%	14%	3%	1%					237
2014			4%	20%	30%	28%	14%	3%	1%				226
2015				4%	21%	29%	28%	13%	3%	1%			211
2016					4%	20%	30%	28%	10%	5%	2%		198
2017						4%	21%	31%	22%	16%	5%	1%	181
2018							4%	23%	25%	29%	15%	4%	161
2019								4%	17%	30%	33%	15%	148
2020									4%	21%	37%	37%	129

**Note :** Le nombre de contrôles au titre de l'exercice 2020 est provisoire, des dossiers portant sur cet exercice seront clos en 2024.

**Lecture :** 60 % des dossiers où la TVA a été contrôlée en 2009 ont été clôturés en 2012

**Source :** DGFIP

## 2.2. Des dossiers traités aux exercices contrôlés

En 2022, l'ensemble des dossiers de contrôles fiscaux clôturés ont permis de redresser 2,1 Md€ de droits nets de TVA [tableau 2].

Cependant, ce total recouvre plusieurs années d'affaires, c'est-à-dire les années au titre desquelles les entreprises sont redressées. Un dossier clôturé en 2022 peut ainsi, par exemple, redresser une entreprise au titre des années 2019, 2020 et 2021, années où le service de vérification aura constaté des montants d'impôt dus et non versés à l'administration.

Il existe donc un décalage entre année de clôture d'un dossier et les années au titre desquelles l'entreprise est contrôlée : par exemple, l'ensemble des dossiers contrôlant l'année 2013 ont pour la majorité été clôturés entre 2014 et 2016 [tableau 1]. Ce délai est borné dans le temps, du fait de la prescription fiscale en matière de TVA : lorsque l'administration fiscale engage un contrôle sur une entreprise, elle peut dans la grande majorité des cas exercer une éventuelle reprise de TVA sur l'année en cours ainsi que sur les trois années précédentes<sup>2</sup>. Toutefois, des crédits reportables générés sur des années antérieures peuvent être rectifiés lorsqu'ils ont été reportés sur les résultats des années de la période contrôlée.

Il est ainsi nécessaire de ventiler les redressements sur les différentes années contrôlées : les droits nets de TVA au titre de 2015 s'élèvent par exemple à 1,5 Md€ [tableau 3]. Les montants de TVA redressés par année d'exercice sont inférieurs à ceux par année de clôture du dossier diffusés dans les rapports d'activité de la DGFIP pour quatre raisons :

- (1) Le champ de cette étude se limite aux entreprises ayant déclaré de la TVA (même au titre d'une seule année). Par conséquent, il manque les entreprises redressées n'ayant pas déclaré de TVA et les

entreprises défailtantes. Les droits nets de TVA redressés des entreprises défailtantes s'élèvent autour de 0,4 Md€ par an.

- (2) Les motifs de redressement à la TVA conservés dans cette étude sont ceux retenus par la [Cour des comptes \(2019\)](#) et excluent certains motifs tels que la TVA à l'importation non déclarée non récupérable.
- (3) Les données du contrôle fiscal utilisées datent de janvier 2024, de sorte que les résultats des contrôles réalisés au titre de l'exercice 2020 sont pris en compte principalement dans les dossiers clos en 2021, 2022 et 2023. Cependant il reste des contrôles au titre de l'année 2020 dans les dossiers en cours de clôture en 2024 au moment où cette étude est rédigée. **Les résultats présentés pour l'année d'affaire 2020 sont donc provisoires.**
- (4) Les dossiers clos en 2020 ont redressé moins de TVA que ceux clos les autres années du fait d'une adaptation de l'activité des services de contrôle pendant la crise sanitaire. Moins de contrôles que les autres années ont donc été clôturés en 2020. Les années d'affaires 2017, 2018, et 2019 ont donc des redressements de droits de TVA plus faibles que les années précédentes.

<sup>2</sup> L'article L.176 du livre des procédures fiscales prévoit que le droit de reprise s'exerce jusqu'à la fin de la troisième année suivant celle au cours de laquelle la taxe est devenue exigible conformément aux dispositions de l'article 269 du code général des impôts.

Tableau 2 : Montants des droits nets de TVA redressés par année de clôture (en Md€)

Année de clôture	Droits nets de TVA redressés
2015	2,2
2016	2,3
2017	2,1
2018	2,0
2019	1,8
2020	0,9
2021	1,6
2022	2,1

Source : Rapports d'activité annuels de la DGFIP

Cette analyse des données de contrôle porte deux enseignements. Tout d'abord, pour estimer le manque à gagner d'une année, ce sont bien les montants de redressement par année d'affaires qu'il faut extrapoler. Raisonner par année d'affaires et non par année de clôture a l'avantage de déterminer le montant de TVA qui n'a pas été versée par les entreprises contrôlées au titre d'une année donnée et ainsi permettre une estimation d'un manque à gagner de TVA sur un exercice précis, qui peut être mis en regard de la TVA collectée au titre de cet exercice. En outre, il est nécessaire d'attendre deux à trois ans pour avoir suffisamment de données de contrôle, empêchant une estimation en temps réel de la fraude.

Tableau 3 : Montant des droits nets de TVA redressés par année d'affaires (en Md€)

Année d'affaires	Droits nets de TVA	Dont CFE	Dont CSP
2010	1,3	1,2	0,1
2011	1,6	1,4	0,2
2012	1,7	1,5	0,2
2013	1,6	1,4	0,2
2014	1,7	1,5	0,2
2015	1,6	1,4	0,2
2016	1,7	1,6	0,1
2017	1,2	1,1	0,1
2018*	0,9	0,8	0,1
2019*	0,6	0,5	0,1
2020*	0,6	0,5	0,1

Notes : Seules les entreprises redressées ayant déclaré de la TVA durant l'année d'affaires considérée sont comptabilisées. \* chiffres provisoires.

Source : DGFIP

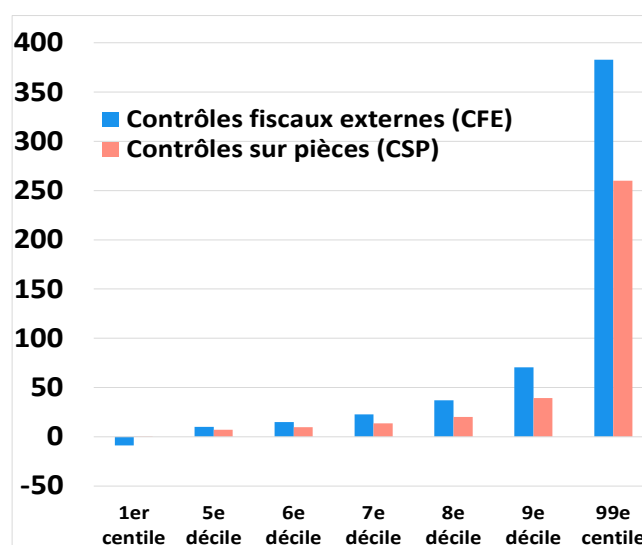
### 2.3. Une distribution des montants rectifiés concentrée sur de petits montants

Une part de la difficulté à extrapoler les données de redressement en un montant global de manque à

gagner de TVA consiste à bien calculer la probabilité de contrôle et le montant moyen de redressement en cas de contrôle, en fonction du type d'entreprise. Cette section étudie ainsi la typologie des entreprises contrôlées.

Si la plupart des secteurs d'activité ont une proportion équivalente d'entreprises contrôlées et d'entreprises redressées, le secteur de la construction semble plus faire l'objet de redressements que les autres : les entreprises de ce secteur représentaient en 2019 16 % des entreprises contrôlées, et 20 % des entreprises redressées à la TVA [tableau 4].

Graphique 4 : Distribution des redressements de TVA au titre de 2019 (en k€)



Note : Le premier centile laisse apparaître des redressements négatifs : en effet un contrôle fiscal peut se solder par une restitution si le service estime qu'il y a eu sur la période contrôlée trop d'impôt payé.

Lecture : en 2019, 99 % des redressements de TVA étaient inférieurs à 0,4 M€ pour le CFE et 0,3 M€ pour les CSP.

Source : DGFIP

Par ailleurs, les rectifications notifiées au titre de l'année 2019 laissent apparaître une moyenne tirée vers le haut par les plus gros redressements [graphique 4].

Ainsi, alors que les redressements médians au titre de la TVA pour 2019 sont de respectivement 7 k€ pour les contrôles sur pièces et de 10 k€ pour les CFE, les redressements moyens sont respectivement de 26 k€ et de 34 k€.

En moyenne, un CFE permet d'identifier un manque à gagner plus important qu'un contrôle sur pièces. Cet écart s'explique d'une part par la sélection plus stricte des dossiers passant en CFE, parce qu'ils sont *a priori* plus fraudogènes ou bien parce que le CSP a identifié un risque plus important. D'autre part, ces redressements plus importants s'expliquent par les investigations plus poussées que permet le contrôle fiscal externe, le service examinant sur place la comptabilité d'une entreprise en la confrontant à certaines données de fait ou matérielles afin de contrôler les déclarations souscrites.

Tableau 4 : Entreprises contrôlées et redressées par secteur au titre de 2019

Secteurs d'activité	Entreprises contrôlées		Entreprises redressées	
	Part dans le total	Nombre d'entreprises	Part dans le total	Nombre d'entreprises
Commerce de gros et de détail	22%	33 523	22%	4 496
Services aux entreprises	17%	24 805	18%	3 798
Construction	16%	23 253	20%	4 232
Activités immobilières	10%	15 046	9%	1 862
Industrie	9%	13 592	7%	1 351
Hébergement et restauration	5%	7 679	4%	839
Transports et entreposage	4%	5 365	6%	1 146
Activités financières et d'assurance	5%	7 646	3%	660
Information et communication	4%	6 113	4%	799
Agriculture, sylviculture et pêche	3%	3 877	2%	425
Autres activités de service	3%	3 817	3%	560
Enseignement, santé humaine et action sociale	2%	2 883	2%	456
<b>Total</b>	<b>100%</b>	<b>147 599</b>	<b>100%</b>	<b>20 624</b>

**Champ :** Ensemble des entreprises redevables de la TVA en 2019

**Source :** DGFIP

Tableau 5 : Redressements par motif pour l'année 2019

Motif de redressement	Redressement moyen (en k€)	Redressement total (en M€)
Dissimulations d'opérations assujetties à la TVA	24,8	281,5
TVA déduite à tort	18,7	158,8
Retard dans le versement de TVA	23,8	49,0
Irrégularités sur des dispositifs spécifiques (TVA immobilière, etc.)	25,4	40,4
Irrégularités relatives aux obligations de territorialité	77,9	29,4
Opérations exonérées à tort	58,0	11,4
Redevables non-éligibles à une exonération / taux réduit	25,6	10,2
Fraude intracommunautaire	22,7	7,8
Opérations exonérées à tort au titre d'exportations	17,1	7,5
Fraude carrousel	40,1	5,4
Autres	23,5	0,3

**Note :** Les motifs de redressement prévus par le code général des impôts sont regroupés ici par principaux types d'infraction. Les redressements négatifs ont été exclus.

**Source :** Données du contrôle fiscal, DGFIP.

Les données du contrôle fiscal permettent également d'identifier les motifs pour lesquels les entreprises sont redressées. La dissimulation (délibérée ou non) d'opérations assujetties à la TVA par les entreprises constituent ainsi le premier motif de redressements en 2019 [tableau 5]. Les redressements les plus importants en moyenne sont dus au non-respect des règles de territorialité édictées notamment aux articles 258 et 259 du Code général des impôts : ils représentent un redressement total de 30 M€ au titre de 2019, effectué sur moins de 400 entités.

### 3. Extrapoler les données de contrôle fiscal

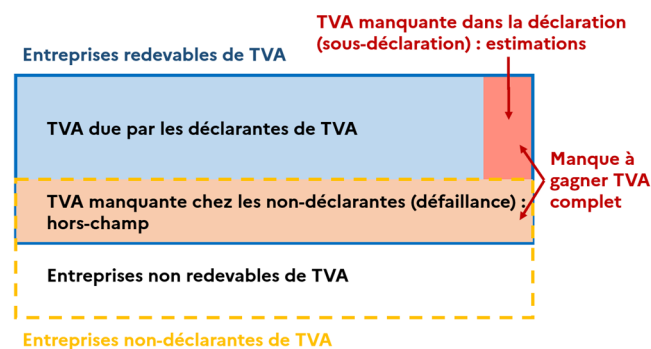
#### 3.1. Périmètre d'étude

Pour pouvoir extrapoler les contrôles fiscaux effectifs, il faut avoir de l'information auxiliaire sur les entreprises non contrôlées. Par conséquent, on se restreint ici aux entreprises ayant effectué au moins une déclaration de TVA pour l'année civile considérée. Ainsi, les entreprises sous le régime de la franchise de TVA ne sont pas retenues. De même, les entreprises défailtantes, c'est-à-dire celles ne déposant aucune déclaration de TVA, ne sont pas incluses. Le manque à gagner mesuré ici n'est donc pas exhaustif ; il se focalise sur la sous-déclaration des entreprises déposant une déclaration de TVA [graphique 5]. Une estimation complémentaire sur le manque à gagner des entreprises défailtantes qui devraient être redevables de TVA reste à mener.

Les montants pris en compte pour la présente étude ne concernent que les droits mis en recouvrement, tels que notifiés par la vérification, et excluent les pénalités, les sanctions, ou les procédures d'appel.

Le champ d'étude porte sur les CSP et CFE relatifs à la TVA. En revanche, les contrôles relatifs aux demandes de remboursement de crédits de TVA sont exclus de notre champ, car c'est une procédure automatique qui ne relève pas des mêmes mécanismes ciblés de détection des fraudes.

Graphique 5 : Périmètre de l'estimation du manque à gagner



**Note :** Représentation schématique, l'aire des différentes cellules n'est pas représentative des montants considérés.

### 3.2. Correction du biais de sélection

Plusieurs éléments rendent complexe l'exercice d'extrapolation des contrôles fiscaux pour estimer le manque à gagner.

- Il existe un **biais de sélection** : pour utiliser de manière optimale les moyens dont ils disposent, les services de la DGFIP ne sélectionnent pas les entreprises de façon aléatoire mais ciblent celles jugées les plus susceptibles d'être redressées. A cause de ce biais de sélection, l'extrapolation risque de conduire à une surestimation de la fraude (puisque l'on considérerait à tort que toutes les entreprises sont potentiellement aussi fraudeuses que celles ciblées par les contrôles).
- Il existe un **biais de détection** : la détection des fraudes dépend de la possibilité d'obtenir des informations externes et de la connaissance des mécanismes d'évitement de l'impôt. C'est notamment le cas pour les contrôles sur pièces, durant lesquels les services ont accès à moins d'information que lors d'un contrôle fiscal externe. A cause de ce biais, une extrapolation des contrôles fiscaux conduit à une sous-estimation de la fraude.
- Enfin, les **données sont plus ou moins exhaustives** en fonction du type de dossier ; en particulier, les services de contrôle peuvent avoir tendance à ne pas renseigner les CSP sans irrégularités, pour gagner du temps.

Cette étude s'efforce de corriger le biais de sélection, afin de fournir un minorant de la fraude fiscale. Comme [Quantin & Welter-Médée \(2022\)](#), elle s'inspire des techniques d'imputation du mécanisme de la non-réponse, en apparentant les entreprises contrôlées à des « répondantes » et les autres à des entités « enquêtées » (c'est-à-dire contrôlables) mais ne répondant pas. La probabilité d'inclusion dans l'enquête (égale à 1 car toutes les entreprises sont contrôlables) est pondérée par la probabilité de répondre, soit la probabilité d'être contrôlée. La pondération repose sur des groupes de contrôle homogène (GCH) construits à partir des quantiles de distribution des probabilités prédites d'être contrôlé. Après repondération, le manque à gagner est estimé par secteur d'activité, sur l'ensemble des entreprises contrôlées. Les secteurs d'activité correspondent à des

domaines qui se caractérisent par un haut niveau de corrélation TVA avec la variable à extrapoler, ici, le manque à gagner TVA.

L'hypothèse est faite que, dans chaque groupe, les entreprises ont un comportement homogène de fraude (cela se rapproche d'une méthode économétrique de stratification par score de propension) et qu'elles auraient un redressement moyen égal au redressement moyen des entreprises contrôlées de leur groupe.

Les estimations sont faites séparément pour chaque année, pour chaque secteur d'activité et par service de vérification (DVNI, Dircofi, et brigades locales).

#### 3.2.1. Probabilité de contrôle fiscal

Un modèle d'apprentissage supervisé (*boosting*) est utilisé pour associer à une entreprise une probabilité de contrôle, en fonction de ses caractéristiques. L'avantage par rapport à un modèle économétrique classique est que ce modèle permet de tenir compte de la complexité du processus de sélection, sans pour autant se restreindre à une liste prédéterminée de déterminants.

L'utilisation de tels algorithmes de *machine learning* vise ici, non à prédire avec justesse les contrôles fiscaux effectivement menés, mais à estimer au mieux des probabilités d'être contrôlé pour réduire, autant que possible, le biais de sélection, en construisant des groupes d'entreprises pour lesquelles l'éventualité d'un contrôle fiscal est similaire. Cela assure que les entreprises effectivement contrôlées dans un groupe de contrôle homogène soient représentatives en termes de comportement déclaratif de toutes les entreprises du même groupe.

Le modèle d'apprentissage choisi est le *gradient boosting* et les variables prédictives retenues sont les variables présentes dans les déclarations de TVA (CA3 et CA12), comme le montant de TVA déductible, le secteur d'activité, le code juridique ou les acquisitions intracommunautaires. Le *boosting* était le modèle de *machine learning* le plus adapté d'après [Quantin & Welter-Médée \(2022\)](#).

En pratique, un échantillon d'entreprises redevables est tiré au sort pour entraîner l'algorithme, c'est-à-dire ajuster les paramètres du modèle d'apprentissage. Le reste des entreprises forme l'échantillon de validation ; ce dernier sert à ajuster les hyper-paramètres qui permettent de contrôler le processus d'entraînement du modèle en ajustant sa vitesse d'apprentissage et son nombre d'itérations. La construction de l'échantillon de validation se fait par validation croisée.

Environ 4,5 % des entreprises étant contrôlées au titre de chaque année d'affaires, les qualités prédictives risquent d'être faibles. Une solution est de rééquilibrer l'échantillon d'entraînement, en sur-échantillonnant la classe minoritaire (les entreprises contrôlées) et en sous-échantillonnant la classe majoritaire. On crée ainsi des entreprises « synthétiques » à partir des entreprises contrôlées existantes, par combinaison linéaire des caractéristiques des plus proches voisins. Les individus

réels et synthétiques sont mieux répartis et permettent ainsi une meilleure généralisation.

Cette technique (le *Synthetic Minority Over-sampling Technique*) intervient seulement dans l'échantillon d'apprentissage. Pour calibrer le modèle, les probabilités de contrôle sont estimées seulement avec les entreprises réelles.

### 3.2.2. Montant de fraude qui aurait été détecté

Une fois que les probabilités d'être contrôlé sont estimées pour chacun des redevables, ces derniers sont regroupés en 50 quantiles de probabilité, et ce pour chaque service de vérification. Ces groupes de contrôle homogène (GCH) permettent d'homogénéiser les probabilités de contrôle prédites au sein de chaque groupe (en prenant la proportion d'entreprises contrôlées par GCH comme probabilité empirique moyenne), de regrouper des entreprises similaires en termes de probabilité et d'envergure de fraude, et d'extrapoler le montant de manque à gagner, en fonction du poids (égale à l'inverse de la probabilité empirique au sein du GCH) de l'entreprise dans l'ensemble des entreprises qui déclarent la TVA [encadré 1].

Pour éviter que les résultats ne soient tirés par les gros redressements décrits à la section 2.3, les redressements exceptionnels pour chaque service de vérification sont exclus de l'échantillon des entreprises contrôlées (en revanche, les montants redressés sont bien inclus dans le manque à gagner total, en considérant que ces redressements ne sont représentatifs que d'eux-mêmes). Plus précisément, pour chaque année, on identifie et exclut les 10 redressements qui s'écartent le plus de la médiane, en corrigeant du fait que la distribution des redressements dans certains secteurs d'activité est naturellement plus dispersée que dans d'autres secteurs.

L'extrapolation est complétée par un « calage » avec la TVA brute déclarée. Les estimations sont réalisées avec un estimateur par le ratio, qui est un estimateur asymptotiquement sans biais (supérieur donc, à l'extrapolation d'un montant de redressement). Le ratio de redressement dans un domaine, c'est-à-dire par GCH, par secteur d'activité et par service de vérification, est donc égal au quotient du montant total de redressement des entreprises contrôlées du domaine rapporté au total de TVA brute déclarée sur la période contrôlée. Puis, l'estimation du montant total est obtenue en multipliant le ratio estimé par la TVA brute totale déclarée par les entreprises du domaine considéré.

Finalement, le montant total estimé s'obtient simplement par somme des trois estimateurs correspondant à chaque service de contrôle, ces trois estimateurs étant parfaitement indépendants.

### 3.3. Tests de précision

Les probabilités de contrôle des entreprises, telles que prédites par le modèle de *boosting*, sont testées. Pour cela, on compare si les entreprises contrôlées sont bien celles ayant une probabilité prédite supérieure à 0,5.

	Probabilité d'être contrôlé ≥ 0,5	Probabilité d'être contrôlé ≤ 0,5
Entreprises contrôlées (=Positif)	VP (Vrais Positifs)	FN (Faux Négatifs)
Entreprises non contrôlées (=Négatif)	FP (Faux Positifs)	VN (Vrais Négatifs)

Le modèle fonctionne si la proportion d'entreprises effectivement contrôlées parmi celles qui se sont vues prédire un contrôle fiscal est élevée. La précision du modèle est ainsi définie comme le taux de vrais positifs parmi les positifs prédits, soit  $\frac{VP}{VP + FP}$  avec les notations du schéma ci-dessus. Les résultats sont de 0,19 pour les entreprises du champ de contrôle de la DVNI ; 0,19 pour celles du champ des Dircofi ; et 0,07 pour celles dépendant des directions locales. En d'autres termes, avec une règle de binaire à 0,5, seules 7% des prédictions de contrôle fiscal par les directions locales correspondent à des contrôles qui ont eu lieu.

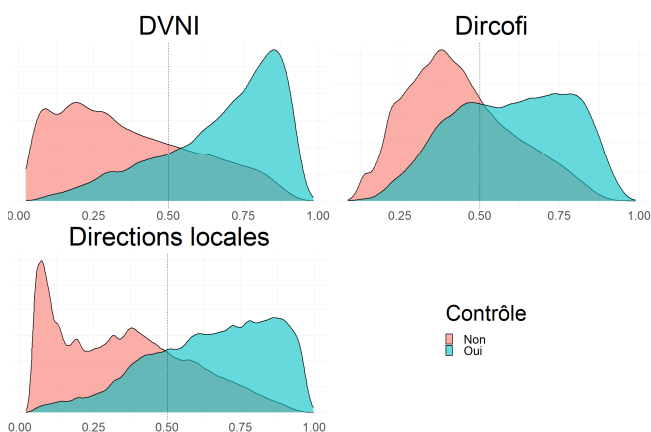
Par ailleurs, étant donné que l'échantillon est déséquilibré, une meilleure mesure est la précision de la prédiction, définie comme le taux de bonne prédiction, soit  $\frac{VP + VN}{VP + FP + VN + FN}$ . En l'espèce, la précision est de 0,68 pour les prédictions de contrôle de la DVNI ; 0,65 pour celles des Dircofi ; 0,73 pour celles des Directions locales. La prédiction est assez bonne pour les DVNI et les Directions locales mais moins pour les entreprises relevant des Dircofi.

Ces résultats montrent que cet algorithme ne reproduit que partiellement le ciblage des contrôles fiscaux effectués par la DGFIP. En effet, le *machine learning* est limité par le faible nombre d'entreprises contrôlées par rapport à l'ensemble de la population, l'impossibilité de modéliser l'expérience des vérificateurs ou les éventuelles informations complémentaires à leur disposition, et les ressources limitées des services de contrôle fiscal.

Cependant, les probabilités estimées servent seulement à construire des GCH par une méthode des scores, elles ne sont pas exploitées à des fins de prédiction car pour la correction du biais de sélection, aucune règle de décision n'est définie. L'important est donc que les probabilités prédites permettent de distinguer au mieux les entreprises avec des risques différents d'être contrôlées. D'après la distribution des probabilités prédites par service de vérification, le modèle est bien moins discriminant pour les entreprises dans le champ de contrôle des Dircofi que pour celles dépendant des autres services [graphique 6].



Graphique 6 : Densité des probabilités prédites d'être contrôlé (2019)



**Note :** Pour chaque structure de contrôle, le graphique propose pour les entreprises effectivement contrôlées (en bleu) et non contrôlées, la probabilité de contrôle prédite par la méthode d'apprentissage de *boosting*.  
**Source :** DGFIP

## 4. Résultats

### 4.1. Entre 6 et 10 Md€ de manque à gagner annuel sur la dernière décennie

Les données du contrôle fiscal donnent un redressement moyen de l'ordre de 5 000€ par entreprise contrôlée (en excluant les redressements exceptionnels et en incluant les CSP) [graphique 8]. Sous l'hypothèse erronée que le taux d'irrégularité est similaire pour les entreprises contrôlées et non contrôlées, qui ignorerait le biais de détection inhérent aux contrôles ciblés, une extrapolation « naïve » des contrôles fiscaux à toute la population donnerait un manque à gagner aux alentours de 12 Md€ par an.

En prenant en compte correctement le biais de sélection, cette étude parvient à réaliser une estimation du manque à gagner de TVA sur la période 2010-2020, en incluant tous les contrôles fiscaux (CFE et CSP). Les

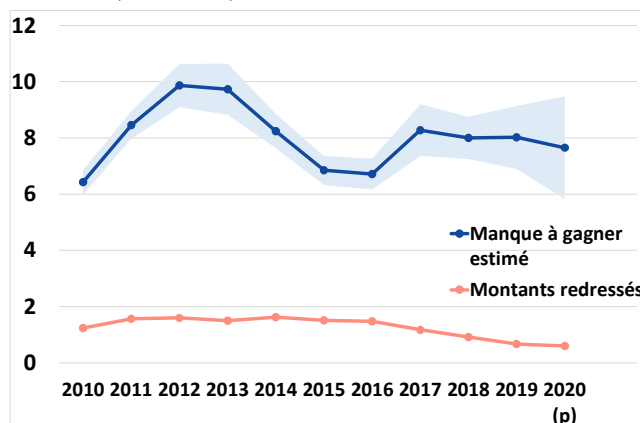
résultats se situent de manière relativement stable, entre 6 et 10 Md€, même si le manque à gagner semble avoir été plus élevé entre 2011 et 2014.

S'agissant de résultats d'inférence statistiques, ces estimations sont nécessairement entourées d'une part d'incertitude (la bande bleu clair du [graphique 7]), de sorte qu'il est préférable de commenter des ordres de grandeurs que les estimations centrales.

Ces résultats excluent les entreprises ne déclarant pas de TVA et ignorent les irrégularités non détectées par le contrôle fiscal, et doivent donc être interprétés comme des minorants du manque à gagner de TVA.

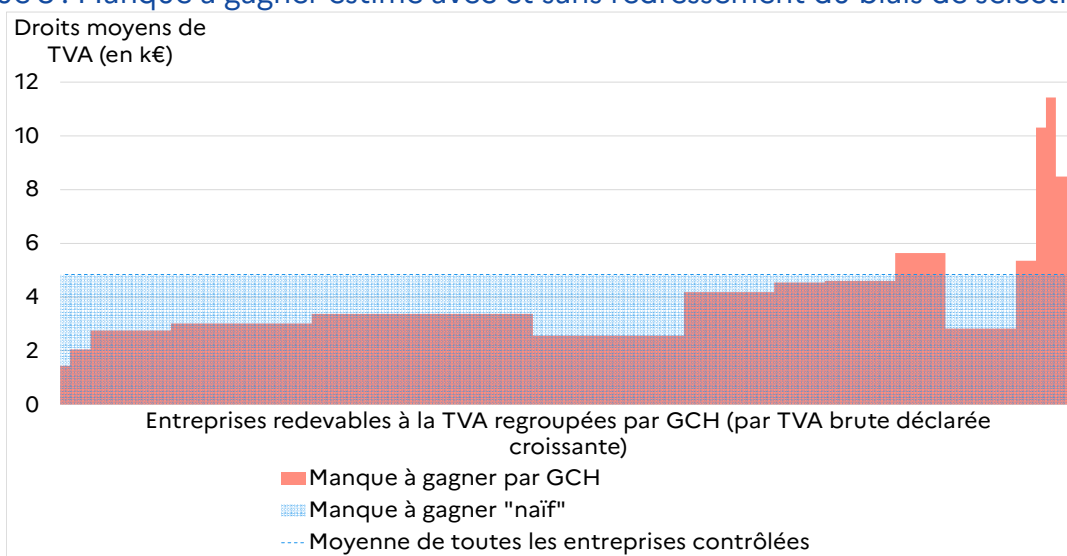
De ce manque à gagner « brut », on peut déduire les montants redressés lors des opérations de contrôle, soit entre 1,2 Md€ et 1,6 Md€ [tableau 3], pour obtenir un manque à gagner net de l'ordre de 5 à 9 Md€.

Graphique 7 : Manque à gagner sur la TVA en France (en Md€)



**Note :** (p) : provisoire. L'aire en bleu clair correspond à l'intervalle de confiance à 95% des estimations. Les montants redressés au titre des années 2017 à 2020 sont plus faibles du fait d'une part de moindres redressements dans les dossiers clos en 2020 suite à l'adaptation de l'activité des services de contrôle pendant la crise sanitaire et d'autre part du fait de dossiers non encore clôturés au titre de ces années à la date d'extraction des données.  
**Source :** DGFIP.

Graphique 8 : Manque à gagner estimé avec et sans redressement du biais de sélection



**Lecture :** L'aire en rose correspond aux redressements moyens restructurés par GCH selon le nombre d'entreprises dans l'échantillon tandis que l'aire bleu indique le redressement moyen naïf (qui donne un montant moyen de 5 k€).  
**Source :** DGFIP

Tableau 6 : Ventilation du manque à gagner par secteur (2019)

Secteurs d'activité	Part dans le manque à gagner (en %)	Montant (en Md€)	Part dans la TVA économique du secteur (en %)
Commerce de gros et de détail	30,1	2,4	5,0
Construction	14,8	1,2	9,0
Services aux entreprises	12,3	1,0	2,0
Activités immobilières	10,2	0,8	13,0
Enseignement, santé humaine et action sociale	6,2	0,5	0,0
Agriculture, sylviculture et pêche	4,7	0,4	0,0
Hébergement et restauration	4,7	0,4	13,0
Autres activités de service	4,7	0,4	8,0
Transports et entreposage	4,1	0,3	6,0
Industrie	3,9	0,3	2,0
Information et communication	2,6	0,2	2,0
Activités financières et d'assurance	1,7	0,1	3,0
<b>Total</b>	<b>100,0</b>	<b>8,0</b>	<b>5,0</b>

Source : DGFIP

#### 4.2. Des taux d'irrégularité plus forts dans l'hospitalité et dans l'immobilier.

Les secteurs du commerce (30 %), de la construction (15 %) et des services aux entreprises (10 %) sont les principaux contributeurs au manque à gagner sur la TVA en 2019 [tableau 6].

Les secteurs avec la plus grande part de TVA non-collectée par l'administration fiscale proportionnellement à la TVA économique du secteur, c'est-à-dire ceux qui ont le taux d'irrégularité le plus élevé, sont les activités immobilières (13 %), l'hébergement-restauration (13 %) et la construction (9 %).

#### 4.3. Principale infraction : la dissimulation d'opérations assujetties à la TVA

Lors d'un contrôle fiscal, les services renseignent les types d'irrégularités détectées pour justifier les redressements [tableau 7].

Ce sont les pratiques de dissimulation (délibérée ou non) d'opérations assujetties à la TVA qui contribuent au premier chef au manque à gagner de TVA en 2019 (44,5 % du montant) ; il s'agit par exemple d'entreprises ne déclarant pas une partie de leurs activités. Ce motif se rapporte notamment à l'article 256 du code général des impôts<sup>3</sup>. La seconde source de manque à gagner est la TVA déduite à tort (18 %), suivie par les irrégularités sur des régimes spécifiques comme celui relatif à la TVA immobilière (12,3 %) et les irrégularités relatives à la territorialité<sup>4</sup> (12,1 %).

Tableau 7 : Ventilation du manque à gagner par type d'irrégularité (2019)

	Part du montant total (en %)
Dissimulation d'opérations assujetties à la TVA	44,5
TVA déduite à tort	18,3
Irrégularité sur des dispositifs spécifiques (TVA immobilière...)	12,3
Irrégularités relatives aux obligations de territorialité	12,1
Retard dans le versement de TVA	8,3
Opérations exonérées à tort	1,1
Fraude intracommunautaire	1,0
Opérations exonérées à tort au titre d'exportations	1,0
Fraude carrousel	0,9
Redevables non-éligibles à une exonération ou à un taux réduit	0,5
Autres	0,0

Source : DGFIP

<sup>3</sup> L'article 256 du CGI pose le cadre de l'assujettissement des livraisons de biens corporels et des prestations de services à la TVA.

<sup>4</sup> Les irrégularités relatives à la territorialité correspondent à des situations où un acheteur de biens corporels ou de prestations de services situé en France n'a pas collecté la TVA sur ses importations comme il aurait dû (articles 258 et 259 du CGI).

## 5. Validation des estimations par des contrôles aléatoires

Les estimations présentées jusqu'à présent se basent sur les résultats des contrôles ciblés, qu'il faut corriger du biais de sélection selon des méthodologies complexes. Idéalement, pour pouvoir extrapoler les données de redressements fiscaux individuels sans se soucier du biais de sélection, il faudrait disposer d'un échantillon dont on connaît les probabilités de tirage. Le recours à des contrôles d'entreprises tirées au hasard, éventuellement après une stratification déterminée, permet en effet d'éliminer le biais de sélection (mais pas le biais de détection).

C'est ainsi qu'en 2022, à titre d'expérimentation, la DGFIP a soumis à des contrôles fiscaux aléatoires deux sous-secteurs particuliers, présentant entre eux des différences importantes.

### 5.1. Moins de redressements avec les contrôles aléatoires

Les entreprises sélectionnées pour constituer le champ des contrôles aléatoires étaient actives, créées avant le 1<sup>er</sup> juillet 2020, et avaient déposé au moins une déclaration de TVA. Cette démarche ne permet donc pas d'estimer le manque à gagner des entreprises ne déposant aucune déclaration.

Au total, sur 610 entreprises contrôlées aléatoirement, 539 ont fait l'objet d'un contrôle fiscal externe et 71 d'un contrôle sur pièces. L'expérience des contrôles aléatoires s'est limitée aux deux secteurs suivants :

- (1) Le secteur de la « fabrication d'instruments et de fournitures à usage médical et dentaire » (groupe 32.5 de la nomenclature d'activité) est composé de peu d'entreprises mais de taille relativement importante. Dans ce secteur, 2 542 entreprises entrent dans le champ de l'estimation et 329 (13 %) ont été contrôlées aléatoirement.
- (2) Le secteur de la « réparation de biens personnels et domestiques » (groupe 95.2) dont 281 (4 %) ont été contrôlées parmi les 7 982 entrant dans le champ de l'estimation.

Parmi les entreprises contrôlées aléatoirement, 22 % ont été redressées à la TVA pour un montant moyen de 16 k€, soit une probabilité et un montant de redressement moindres que les contrôles ciblés [tableau 8].

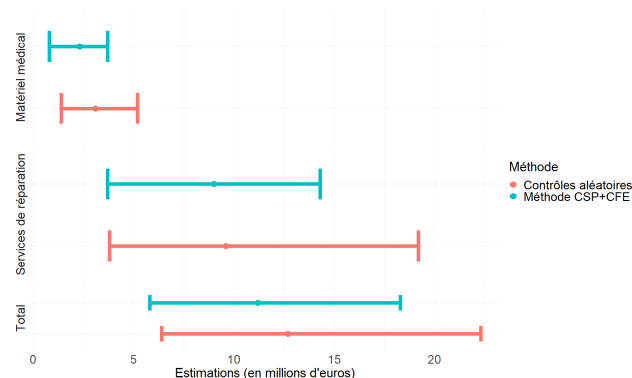
Ces résultats des contrôles aléatoires confirment donc l'existence du biais de sélection dans les contrôles ciblés à deux titres—la fiabilité du ciblage se manifeste dans la sélection d'entreprises suspectes de fraude et dans la sélection d'entreprises qui fraudent relativement plus que les autres.

### 5.2. Un manque à gagner extrapolé similaire

Lorsque les redressements des contrôles aléatoires sont extrapolés à l'ensemble des entreprises des deux sous-secteurs qui ont été l'objet de l'expérimentation, l'estimation du manque à gagner est similaire à la méthode par les contrôles ciblés [graphique 10].

Comme les contrôles aléatoires permettent en théorie de meilleures estimations du manque à gagner, ces résultats valident la robustesse de la correction du biais de sélection des estimations fondées sur les contrôles ciblés.

Graphique 9 : estimations par contrôles aléatoires



**Lecture :** Le manque à gagner pour l'ensemble des deux secteurs est estimé de 6 à 22 M€ avec les contrôles aléatoires contre 6 à 18 M€ avec la méthode (CSP+CFE) de redressement du biais de sélection des contrôles ciblés.

**Source :** DGFIP

Tableau 8 : Redressements à la TVA pour les contrôles aléatoires et ciblés

Secteur d'activité	Contrôles aléatoires			Contrôles programmés clôturés en 2022		
	Nombre de contrôles	Montant moyen rectifié (en k€)	Taux de redressement	Nombre de contrôles	Montant moyen rectifié (en k€)	Taux de redressement
Instruments médico-dentaires	329	17,3	19,8%	289	212,2	28,0%
Réparation	281	15,6	25,3%	268	70,1	39,6%
Total des deux secteurs	610	16,4	22,3%	557	131,7	33,6%

**Champ :** Contrôles CSP et CFE  
**Source :** DGFIP

## 6. Discussion des résultats

### 6.1. Les estimations présentées comportent deux limites

Les estimations présentées dans cette étude souffrent d'une limite de champ : elles ne montrent le manque à gagner que pour les entreprises déclarant de la TVA, pas pour l'ensemble de la population des entreprises. Cette limite incite ainsi à considérer les estimations présentées comme des minorants, pour chaque année, du manque à gagner de TVA. La deuxième limite provient du biais de détection, qu'il n'est pas possible de corriger. En effet, le montant et la nature des redressements opérés dépendant pour partie du jugement des vérificateurs, il est possible que seule une partie du manque à gagner ait été repérée. La fraude réelle serait donc plus importante que ce qu'indiquent ces estimations.

Ces deux limites expliquent en partie pourquoi les approches descendantes, telles que celle adoptée par la Commission européenne, aboutissent à des résultats plus importants. Ces estimations ont toutefois leurs propres limites, liées à une méthode appliquée à un niveau agrégé et ne prenant pas en compte toutes les spécificités de la réglementation de la TVA ni toutes les corrections appliquées dans le calcul des comptes nationaux, et ne sont pas nécessairement une meilleure estimation du manque à gagner, ni même un majorant de celui-ci.

Les travaux antérieurs de l'Insee aboutissaient à un manque à gagner de TVA plus important (pour l'année 2012). C'est parce qu'ils n'extrapolent que les CFE, là

où la présente étude utilise toutes les données de contrôle fiscal, y compris les CSP. Or, ces derniers conduisent en moyenne à des redressements moins élevés [graphique 4 pour 2019].

### 6.2. Une estimation pour le débat national plus que pour les comparaisons internationales

L'estimation de 6-10 Md€ de manque à gagner de TVA pour 2019 correspond à un ordre de grandeur compris entre 4 et 7 % du montant théorique de TVA si toutes les obligations fiscales relatives à cet impôt étaient respectées. Si cette estimation est supérieure à celle réalisée par l'Australie en 2021, qui estime son manque à gagner à 2,7 %, elle est inférieure à l'ensemble des estimations réalisées par les autres pays. Le Royaume-Uni estime ce manque à gagner à 5,4 % pour 2022, l'Union européenne à 9,6 % pour 2020 et l'Italie à 19,2 % pour 2020. Cependant, l'estimation présentée dans cette note pour la France ne porte que sur les entreprises déclarant de la TVA, et ne peut pas nécessairement être comparée aux estimations des autres pays, portant potentiellement sur des champs différents.

En revanche, cette méthodologie pourra facilement être reconduite chaque année, afin de jauger des progrès réalisés en matière de lutte contre la fraude. En particulier, au regard de la lourde responsabilité de la sous-déclaration de certaines entreprises sujettes à la TVA au manque à gagner national, la mise en place de la facturation électronique devrait permettre de recouper plus aisément ventes effectuées et déclarations de TVA, et pourra conduire à réduire le manque à gagner.

## Références

- [S. Quantin & C. Welter-Médée \(2022\) Estimation des montants manquants des versements de TVA : exploitation des données du contrôle fiscal](#)
- [C. Louvot-Runavot \(2011\) L'évaluation de l'activité dissimulée des entreprises sur la base des contrôles fiscaux et son insertion dans les comptes nationaux](#)
- [Inspection générale des finances \(2019\) Sécurisation du recouvrement de la TVA](#)
- [Cour des Comptes \(2019\) La fraude aux prélèvements obligatoires](#)
- [Ministère des finances \(2019\) Projet de loi de finances pour 2019, Evaluations des voies et moyens, Tome 1](#)
- [FMI \(2018\) Shadow Economies Around the World: What Did We Learn Over the Last 20 Years?](#)
- [CIAT, FMI, IOTA, et OCDE \(2018\) Enquête ISORA sur les administrations fiscales](#)
- [Commission européenne \(2023\) Rapport EU VAT gap](#)

### Encadré 1 : Éléments méthodologiques

#### Repondération par groupe de contrôle homogène (GCH)

Les groupes de contrôle homogène (GCH) sont construits à partir des probabilités d'être contrôlé estimées par *gradient boosting*. Les GCH permettent de corriger les poids de chaque entreprise pour prendre en compte leur poids de sélection. Pour chaque entreprise  $k$ , la probabilité d'inclusion  $\pi_k$  est repondérée par la probabilité d'inclusion de son groupe  $p_{GCH}$ , ce qui corrige le mécanisme de sélection dans la mesure où la probabilité de contrôle est constante dans chaque GCH. La probabilité d'inclusion d'un GCH est la probabilité empirique de contrôle observée dans le GCH, soit la proportion d'entreprises contrôlées par GCH :  $\frac{n_{GCH}}{N_{GCH}}$ , où  $N_{GCH(k)}$  est le nombre d'entreprise du GCH de l'entreprise  $k$  et  $n_{GCH(k)}$  le nombre d'entreprises contrôlées au sein de ce GCH. Le tirage est ici systématique pour toute entreprise  $k$ , donc la probabilité d'inclusion initiale  $\pi_k$  vaut 1. Le poids de l'entreprise  $k$ , qui mesure le nombre d'entreprises que l'entreprise contrôlé représente, s'écrit alors :

$$w_k = \frac{1}{\pi_k p_{GCH(k)}} = \frac{1}{p_{GCH(k)}} = \frac{N_{GCH(k)}}{n_{GCH(k)}}$$

#### Estimateur par le ratio

L'estimateur par le ratio est un cas particulier d'un estimateur par calage sur marge et s'appuie sur de l'information auxiliaire pour gagner en précision. La variable auxiliaire est la TVA brute déclarée qui est très corrélée aux montants de redressement de TVA (Quantin & Welter-Médée 2022). Cet estimateur est asymptotiquement sans biais sous l'hypothèse d'absence de sélection endogène.

L'ensemble  $U$  des entreprises déclarant de la TVA lors de l'année d'affaires  $N$  est partitionné en  $H$  domaines  $(U_h)_{h \in H}$ , définis comme un croisement du secteur d'activité, du GCH, et du type de service de contrôle fiscal. On note par la suite  $s$  l'échantillon (l'ensemble des entreprises contrôlées);  $s_h$  le sous-échantillon des entreprises contrôlées appartenant au domaine  $h$ ;  $Y_k$  le redressement de l'entreprise  $k$  en droit net de TVA au titre de l'exercice de l'année d'affaires  $N$  (c'est la variable d'intérêt);  $Y$  le manque à gagner de toute la population en  $N$ ;  $X_k$  la TVA brute de l'entreprise  $k$  déclarée au titre de l'année d'affaires  $N$  (la variable auxiliaire de calage).

Les estimateurs de Horvitz-Thompson repondérés des totaux sont :  $\widehat{Y}_h = \sum_{k \in s_h} w_k y_k$  et  $\widehat{X}_h = \sum_{k \in s_h} w_k x_k$ . En notant  $\widehat{R}_h$  le ratio entre ces deux estimateurs, l'estimateur par le ratio du manque à gagner sur un domaine  $h$  s'écrit alors :

$$\widehat{Y}_h = X_h \frac{\widehat{Y}_h}{\widehat{X}_h} = X_h \widehat{R}_h$$

L'estimateur du manque à gagner sur toute la population est la somme des estimateurs sur chaque domaine  $h$ ,  $\widehat{Y} = \sum_{h \in H} X_h \widehat{R}_h$

#### Construction des intervalles de confiance

La précision de l'estimation dépend de la variance de l'estimateur. Or, la variance de l'estimateur par le ratio n'a pas d'expression explicite connue, car  $\widehat{R}_h$  est une fonction non-linéaire des estimateurs  $\widehat{Y}_h$  et  $\widehat{X}_h$ . Une expression approximative de la variance est obtenue par des techniques de linéarisation de Taylor :

$$\widehat{V}(\widehat{R}_h) = \frac{N(1-f)}{N-1} \sum_{k \in U} (g_{h,k} - \overline{g}_h)^2$$

$$g_{h,k} = \frac{w_k I(k \in U_h) (y_k - x_k \widehat{R}_h)}{\sum_{k \in U} w_k I(k \in U_h) x_k}$$

où  $\overline{g}_h = \frac{1}{N} \sum_{k \in U} g_{h,k}$

L'estimateur de la variance du total estimé sur le domaine  $h$ , sachant que les montants  $X_k$  sont positifs, s'écrit alors comme :  $\widehat{V}(\widehat{Y}_h) = X_h^2 \widehat{V}(\widehat{R}_h)$ . On en déduit un estimateur de la variance du total estimé sur toute la population, ainsi que l'intervalle de confiance à 95 % associé, en supposant que les domaines sont indépendamment distribués et que le ratio suit une loi normale :

$$\widehat{V}(\widehat{Y}) = \sum_{h \in H} X_h^2 \widehat{V}(\widehat{R}_h) \quad ; \quad IC(\widehat{Y})_{95\%} = \left[ \widehat{Y} \pm 1,96 \sqrt{\sum_{h \in H} X_h^2 \widehat{V}(\widehat{R}_h)} \right]$$