

**DIRECTION DES AFFAIRES FINANCIÈRES ET DES ENTREPRISES  
COMITÉ DE LA CONCURRENCE**

**Groupe de travail n° 3 sur la coopération et l'application de la loi**

**Les outils de filtrage des données dans les enquêtes de concurrence – Note de référence du Secrétariat**

28 novembre 2022

Ce document a été établi par le Secrétariat de l'OCDE et servira de note de référence à l'appui du point 3 de l'ordre du jour de la 136e réunion du Groupe de travail le 28 novembre 2022.

Les opinions qui y sont exprimées et les arguments qui y sont avancés ne reflètent pas nécessairement le point de vue officiel de l'Organisation ou des gouvernements de ses pays membres.

D'autres documents relatifs à cette discussion sont disponibles à l'adresse suivante :  
[www.oecd.org/competition/data-screening-tools-for-competition-investigations.htm](http://www.oecd.org/competition/data-screening-tools-for-competition-investigations.htm)

Pour toute question relative au présent document veuillez contacter :

Mme Despina Pachnou [Despina.pachnou@oecd.org]

M. Daniel Westrik [Daniel.westrik@oecd.org]

**JT03505422**

# Les outils de filtrage des données dans les enquêtes de concurrence\*

---

Les outils de filtrage des données dans les enquêtes de concurrence sont des méthodes empiriques qui utilisent des ensembles de données numériques afin d'évaluer des marchés et le comportement des entreprises sur ceux-ci, d'identifier des modèles et de tirer des conclusions sur la base de paramètres testés spécifiques. Cette note se concentre sur les filtres visant à détecter des ententes, puisqu'ils sont, et de loin, les plus fréquemment utilisés.

Il n'est pas facile d'estimer la véritable étendue et le succès du filtrage. Il est fréquent que les autorités de la concurrence n'annoncent pas publiquement leurs initiatives de filtrage, de crainte que cela ne conduise des entreprises se livrant à des comportements anticoncurrentiels à devenir plus créatives et à échapper à la détection. Les autorités de la concurrence, même si elles déclarent publiquement utiliser des filtres, détaillent rarement leur étendue, leur fréquence, leur méthodologie et leur taux de succès. Il n'en demeure pas moins que plusieurs affaires de répression des ententes, couronnées de succès, se sont appuyées sur le filtrage de données relatives à la passation des marchés publics. La doctrine et les annonces récentes des autorités de la concurrence révèlent un intérêt croissant pour les filtres. Les filtres se fondent sur des hypothèses à propos des différentes manières dont un marché fonctionne et des types de comportement illégal qui peuvent survenir sur ce marché. En conséquence, il n'existe aucun filtre parfait qui puisse, à lui seul, identifier toutes les violations : différents filtres sont donc conçus pour s'adapter à différents types potentiels de violations sur différents marchés. Étant donné que les volumes de données disponibles ne cessent de croître, des techniques d'apprentissage automatique permettent de combiner différents filtres et d'améliorer la fiabilité des résultats du filtrage.

---

\* Cette note a été rédigée par Despina Pachnou et Daniel Westrik, de la Division de la concurrence de l'OCDE, et a bénéficié de l'assistance de recherche de Eduardo Mangada Real de Asúa, et des commentaires de Ori Schwartz, Antonio Capobianco et Sabine Zigelski, tous membres de la Division de la concurrence de l'OCDE.

La disponibilité et la qualité des données déterminent l'analyse que les autorités de la concurrence peuvent effectuer et sont l'un des principaux obstacles empêchant l'utilisation systématique des filtres. En effet, les filtres, si bien conçus soient-ils, échouent s'ils sont appliqués à des données défectueuses ou incomplètes.

Les filtres ne fournissent généralement pas la preuve suffisante d'une violation, excepté dans de rares cas exceptionnels. Les autorités de la concurrence se fondent plus souvent sur des filtres pour détecter un comportement suspect afin d'ouvrir une enquête ou de prioriser des affaires.

---

# Table des matières

Les outils de filtrage des données dans les enquêtes de concurrence	2
1. Contexte	6
2. Les développements intervenus dans le domaine des filtres numériques	8
2.1. La classification des filtres en filtres structurels et comportementaux est toujours valable	8
2.2. Les filtres d'ententes suscitent un intérêt croissant	9
2.3. Méthodes de filtrage numérique présentées dans la littérature académique récente	13
2.4. Le risque de faux négatifs ou de faux positifs	16
2.5. L'apprentissage automatique améliore le filtrage numérique	18
2.6. Filtrage numérique par des entreprises privées	19
3. Les conditions requises pour un filtrage efficace, en termes de données et de ressources humaines	22
3.1. De bons filtres exigent de disposer de données	22
3.2. Focus sur les filtres de détection des soumissions concertées : l'importance des données relatives à la passation des marchés publics	23
3.3. Le filtrage exige des connaissances et compétences spécialisées	28
3.4. La coopération internationale en matière de partage des données et de développement de filtres	30
4. La valeur du filtrage	33
4.1. Étayer la décision d'ouvrir ou de clôturer une procédure	33
4.2. Fournir des preuves à l'appui d'une décision constatant l'infraction	36
5. Conclusions	39
Annexe A. Littérature académique relative aux filtres d'ententes de 2015 à 2022	40
Méthodes basées sur la distribution des offres/prix	40
Arrondi des offres	43
Méthodes basées sur la cointégration	43
Indépendance conditionnelle entre les offres	43
Doubles différences	44
Fractions d'offre historiques	45
Méthodes de réseau	45
Méthodes probabilistes	45
Méthodes de discontinuité de régression	46
Méthodes de rupture structurelle	46

Regroupements suspects dans les données spatiales géographiques	47
Combinaison de plusieurs filtres	47
<b>Annexe B. Méthodes d'apprentissage automatique</b>	<b>49</b>
Régression linéaire	49
Régression logistique	50
Méthode de régression ridge	50
Méthode de l'opérateur de sélection et de retrait le moins absolu (LASSO)	50
Réseaux neuronaux	50
Arbres de régression	51
Machines à vecteurs de support	51
Forêts aléatoires	51
Ensemble	52
<b>Notes de fin de document</b>	<b>53</b>
<b>Bibliographie</b>	<b>58</b>
<b>Tableaux</b>	
Tableau 1. Indicateurs de filtres d'ententes correspondant au type de collusion dans les marchés publics	16
Tableau 2. Liste complète idéale des variables pour le filtrage des ententes en matière de passation des marchés publics	26
Tableau A 0.1. Indicateurs testés	47
<b>Graphiques</b>	
Graphique 1. Nombre total de demandes de clémence, par région, 2015-2020	10
<b>Encadrés</b>	
Encadré 1. Exemples de filtres numériques d'ententes : Brésil, Colombie, Singapour, Espagne	11
Encadré 2. Le système coréen d'analyse des indicateurs de soumissions concertées (BRIAS)	17
Encadré 3. La Deutsche Bahn utilise le filtrage pour identifier des ententes dans sa chaîne d'approvisionnement	20
Encadré 4. Des données décentralisées ont altéré le fonctionnement de l'outil britannique de filtrage des ententes	24
Encadré 5. Conditions requises pour que les bases de données sur les marchés publics soient utiles dans l'optique du droit de la concurrence	27
Encadré 6. Exemples d'équipes chargées des données au sein des autorités de la concurrence	29
Encadré 7. L'outil Bid Viewer de l'Autorité danoise de la concurrence et de la consommation	31
Encadré 8. Exemples de cas dans lesquels des filtres d'ententes ont conduit à l'ouverture d'enquêtes	33
Encadré 9. Prioriser les affaires au moyen de filtres	35
Encadré 10. Marché public d'achat de médicaments au Mexique	37
Encadré 11. Les méthodologies et hypothèses économétriques doivent être divulguées aux parties	38

# 1. Contexte

1. Les outils de filtrage des données dans les enquêtes de concurrence sont des méthodes empiriques qui utilisent des ensembles de données numériques afin d'évaluer des marchés et le comportement des entreprises sur ceux-ci, d'identifier des modèles et de tirer des conclusions sur la base de paramètres testés spécifiques. Dans cette note, les termes « filtres de données » et « filtres numériques » sont utilisés de manière interchangeable. Les filtres visant à détecter des ententes sont de loin les plus répandus, et cette note se concentre donc sur eux.

2. Les autorités de la concurrence peuvent utiliser des filtres numériques afin de rechercher des comportements suspects qui sont signalés si certains critères sont remplis et justifient l'ouverture d'une enquête. Le filtrage des données peut également faire suite à une plainte ou au signalement d'un lanceur d'alerte, afin de vérifier la validité potentielle des allégations et de trouver des indices révélant qu'un comportement spécifique peut effectivement être problématique. Les filtres peuvent également aider à prioriser la répression, en sélectionnant les affaires qui présentent de forts indices d'activité illégale et un plus grand potentiel de preuve de cette activité, et en clôturant d'autres affaires, si les indices d'illégalité sont moins nombreux ou moins concluants.

3. Il n'est pas facile d'estimer l'étendue et le succès du filtrage. Il est fréquent que les autorités de la concurrence n'annoncent pas publiquement leurs initiatives de filtrage, de crainte que cela ne conduise des entreprises se livrant à des comportements anticoncurrentiels à échafauder des systèmes plus sophistiqués et, ce faisant, à rendre le filtrage inefficace. Il n'en demeure pas moins que plusieurs affaires de répression des ententes, couronnées de succès, se sont appuyées sur le filtrage de données relatives à la passation des marchés publics.

4. L'OCDE a précédemment traité des filtres d'ententes dans l'OECD Business and Finance Outlook 2021 (OCDE, 2021, pp. 95-148<sup>[1]</sup>), un atelier sur la gestion des affaires d'ententes complexes (OCDE, 2018<sup>[2]</sup>), un atelier sur le filtrage des ententes à l'ère du numérique (OCDE, 2018<sup>[3]</sup>), une table ronde sur les algorithmes et la collusion (OCDE, 2017<sup>[4]</sup>), une table ronde sur les enquêtes d'office relatives aux ententes et à l'utilisation de filtres pour détecter des ententes (OCDE, 2013<sup>[5]</sup>), et des travaux nationaux consacrés à la [lutte contre les soumissions concertées dans les marchés publics](#).

5. À l'époque où l'OCDE a discuté des enquêtes d'office relatives aux ententes en 2013, peu de juridictions utilisaient les filtres numériques d'ententes. Depuis lors, certaines autorités de la concurrence ont cessé d'utiliser les filtres qu'elles utilisaient auparavant, tandis que d'autres ont développé de nouveaux filtres. L'expansion des technologies numériques a suscité un intérêt généralisé croissant pour l'utilisation de la technologie et de l'intelligence artificielle afin de soutenir l'application du droit de la concurrence<sup>1</sup>. Plusieurs autorités de la concurrence s'appliquent actuellement à développer ou ont déjà développé et utilisent de nouveaux filtres. Ces développements sont intervenus à la fois dans l'apprentissage automatique et dans la recherche universitaire récente. La présente note vise à explorer ces développements en matière de répression et de recherche. Elle passe également en revue l'état de la littérature sur ce sujet en 0 et explique brièvement les techniques d'apprentissage automatique en 0.

6. Les exigences fondamentales afin d'assurer le succès des filtres, et les plus grands défis pour les autorités de la concurrence, semblent être principalement de deux ordres : l'accès à des données fiables, d'une part, et les compétences et connaissances requises de la part du personnel de l'autorité de concurrence afin d'organiser la collecte, le nettoyage et la manipulation des données, de telle sorte qu'elles

produisent des résultats utiles. Un grand nombre d'autorités de la concurrence se sont lancées dans de vastes projets de collecte et de traitement des données, et ont engagé des experts en mégadonnées et des spécialistes en informatique, afin de compléter les compétences des juristes et économistes. À un moment où les autorités œuvrent pour développer des filtres numériques, la coopération internationale peut fournir des avantages substantiels, particulièrement en termes de partage d'expérience, de ressources, de code et de données potentiellement uniformes.

7. Cette note de référence ne couvre pas la conduite des investigations numériques menées dans le cadre de l'informatique légale (afin d'analyser les éléments découverts dans des appareils numériques) ni l'eDiscovery (utilisé afin de traiter des preuves électroniques dans le cadre d'affaires). Ces sujets ont été discutés au cours de la session 2020 du Forum de la concurrence pour l'Amérique latine et les Caraïbes, consacrée au recueil des preuves dans les enquêtes pour ententes (OCDE, 2020<sup>[6]</sup>). Cette note ne couvre pas non plus l'utilisation de filtres numériques dans d'autres domaines que la lutte contre les ententes<sup>2</sup>.

8. La présente note est organisée comme suit :

- La **Section 2.** examine les développements intervenus dans le domaine des filtres numériques.
- La **Section 3.** détaille les exigences auxquelles les filtres numériques doivent satisfaire pour être efficaces, et les bénéfices de la coopération internationale.
- La **Section 4.** discute de la valeur probante des résultats du filtrage.
- La **Section 5.** est consacrée à la conclusion.

## 2. Les développements intervenus dans le domaine des filtres numériques

9. La présente note de référence se concentre sur les filtres visant à identifier des ententes horizontales anticoncurrentielles, car elles sont, et de loin, les plus courantes.

10. Il existe deux grands types de filtres : les filtres structurels et les filtres comportementaux. Les deux approches, leurs caractéristiques, leur utilisation, leurs avantages et leurs inconvénients ont été analysés par l'OCDE en 2013 (OCDE, 2013<sup>[5]</sup>). Cette classification des types de filtres et les théories économiques sur lesquels ils se fondent n'ont pas changé au fil des années et demeurent largement valables aujourd'hui. Toutefois, des évolutions sont intervenues, à la fois dans la littérature universitaire et dans la pratique des autorités de la concurrence en matière de filtres numériques comportementaux, qui visent dans leur grande majorité à identifier des ententes.

### 2.1. La classification des filtres en filtres structurels et comportementaux est toujours valable

11. Les filtres structurels visent à identifier des marchés présentant des traits propices à la collusion, sur la base des caractéristiques structurelles du marché et des produits en cause, notamment la concentration du marché et l'homogénéité des produits (Harrington, 2006<sup>[7]</sup>). Ces filtres permettent aux autorités de la concurrence de filtrer des marchés ou des industries et d'attirer l'attention sur ceux qui sont les plus exposés aux ententes, c'est-à-dire d'établir une liste initiale des marchés qui méritent un examen plus attentif (OCDE, 2013<sup>[5]</sup>).

12. Les marqueurs des filtres structurels, c'est-à-dire les facteurs qui peuvent influencer les gains et coûts potentiels, - et, partant, la rationalité et la stabilité des ententes et de la collusion sur un marché – peuvent être à la fois des facteurs liés à l'offre et des facteurs liés à la demande. Ils se fondent sur les conditions qui déterminent le succès de l'entente<sup>3</sup>. Font partie des facteurs structurels qui facilitent la collusion le faible nombre de concurrents, les obstacles importants à l'entrée (y compris des indicateurs d'entrée comme le taux d'attrition), la communication fréquente entre les entreprises (par exemple dans le cadre d'appels d'offres répétés) et la transparence du marché. Les facteurs liés à la demande comprennent la stabilité des conditions de la demande et la faible élasticité de la demande. Enfin, les facteurs liés à l'offre sont la maturité du secteur, l'homogénéité des produits, la lenteur du rythme de l'innovation, la symétrie des coûts et la possibilité de mise en commun des coûts, la symétrie des capacités, l'excédent de capacité, les contacts multi-marchés, les liens structurels (par exemple, des liens structurels entre concurrents, que ce soit en termes de partage de la propriété des entreprises concernées ou de membres des conseils d'administration), et des antécédents de comportements anticoncurrentiels sur ce marché (OCDE, 2013<sup>[5]</sup>), (Zlatcu and Suci, 2017<sup>[8]</sup>).

13. Les filtres comportementaux recherchent les activités d'une entreprise qui pourraient indiquer une collusion sur un marché donné, en raison de comportements inhabituels et inexplicables susceptibles de montrer qu'un accord d'entente est en œuvre. La conception de filtres comportementaux se déroule pour l'essentiel en deux étapes. La première étape consiste à définir des marqueurs de collusion qui permettent



de faire la part entre un comportement conforme au jeu de la concurrence et un comportement qui peut s'expliquer par la collusion. La deuxième étape porte sur la recherche de ruptures structurelles ou de chocs exogènes (par exemple, la modification des coûts des facteurs de production) qui permettent d'expliquer une modification du comportement des entreprises (OCDE, 2013<sup>[5]</sup>).

14. Les filtres comportementaux n'ont pas changé beaucoup au fil des années. Les principaux marqueurs (indices) que les filtres comportementaux visent à détecter se fondent sur ce que la théorie économique et l'analyse des ententes découvertes nous disent sur les facteurs qui marquent la création, la vie et la rupture d'une entente. Les marqueurs incluent généralement les prix (ou la valeur des offres dans le cas des soumissions concertées), et peuvent prendre en considération les coûts, les marges, les quantités ou les parts de marché. Les marqueurs habituels des filtres d'ententes sont respectivement les suivants (Zlatcu and Suci, 2017<sup>[8]</sup>) :

1. Marqueurs basés sur les prix, en partant du postulat qu'une entente réussie aboutira normalement à des prix plus élevés. Les marqueurs les plus courants sont les méthodes de comparateur des prix et la corrélation des prix. Les méthodes de comparateur, telle la méthode des « doubles différences », comparent l'évolution des prix entre le groupe de traitement et un groupe de contrôle. Les méthodes de corrélation des prix examinent les hausses et baisses de prix coordonnées : si elles ne s'expliquent pas par les coûts ou la demande, elles pourraient indiquer une collusion entre des entreprises.
2. Marqueurs basés sur les variations des prix. De faibles variations des prix et un manque de sensibilité aux coûts pourraient indiquer l'existence d'une collusion. Les modèles de séries chronologiques qui analysent la volatilité des prix peuvent servir à évaluer l'impact de la collusion sur les prix et la variation des prix avant, pendant et après les périodes d'activité de l'entente.
3. Marqueurs basés sur la part de marché. Des parts de marché stables au fil du temps pourraient être un indice de collusion.
4. Marqueurs pour les soumissions concertées. Il existe plusieurs indices de soumissions concertées, notamment : (i) la même entreprise est souvent le soumissionnaire le moins disant ; (ii) le fait que certains fournisseurs se retirent de façon inattendue de la procédure d'appel d'offres ; (iii) le soumissionnaire gagnant sous-traite de façon répétée des travaux à des soumissionnaires perdants ; (iv) le soumissionnaire gagnant n'accepte pas le marché, mais s'avère ensuite en être le sous-traitant ; (v) la soumission d'offres identiques ; (vi) l'élimination soudaine de remises prévues ; (vii) une augmentation de prix ne pouvant pas s'expliquer par les coûts ; (viii) de grandes différences entre l'offre gagnante et les offres perdantes ; (ix) de grandes différences entre les offres faites par le même fournisseur dans deux procédures d'appel d'offres similaires (OCDE, 2009<sup>[9]</sup>).

15. On peut décrire la distinction entre les filtres d'ententes structurels et comportementaux de la manière suivante : les filtres structurels cherchent à identifier des marchés sur lesquels il est très probable qu'une entente *se formera*, tandis que les filtres comportementaux cherchent à identifier les marchés sur lesquels une entente *s'est formée* (Harrington and Imhof, 2022<sup>[10]</sup>). Les filtres structurels et comportementaux peuvent être utilisés ensemble, les filtres structurels identifiant des marchés à risque justifiant une enquête, et les filtres comportementaux étant ensuite appliqués sur ces marchés.

## 2.2. Les filtres d'ententes suscitent un intérêt croissant

16. Il n'existe pas beaucoup d'informations publiques sur l'utilisation et le succès des filtres. De nombreuses autorités de la concurrence ne souhaitent pas annoncer publiquement leurs initiatives de filtrage, de crainte que cela ne conduise des entreprises se livrant à des comportements anticoncurrentiels à échafauder des systèmes plus sophistiqués et, ce faisant, à rendre le filtrage inefficace. C'est pourquoi la véritable étendue du filtrage n'est pas connue et ne peut pas être estimée facilement.

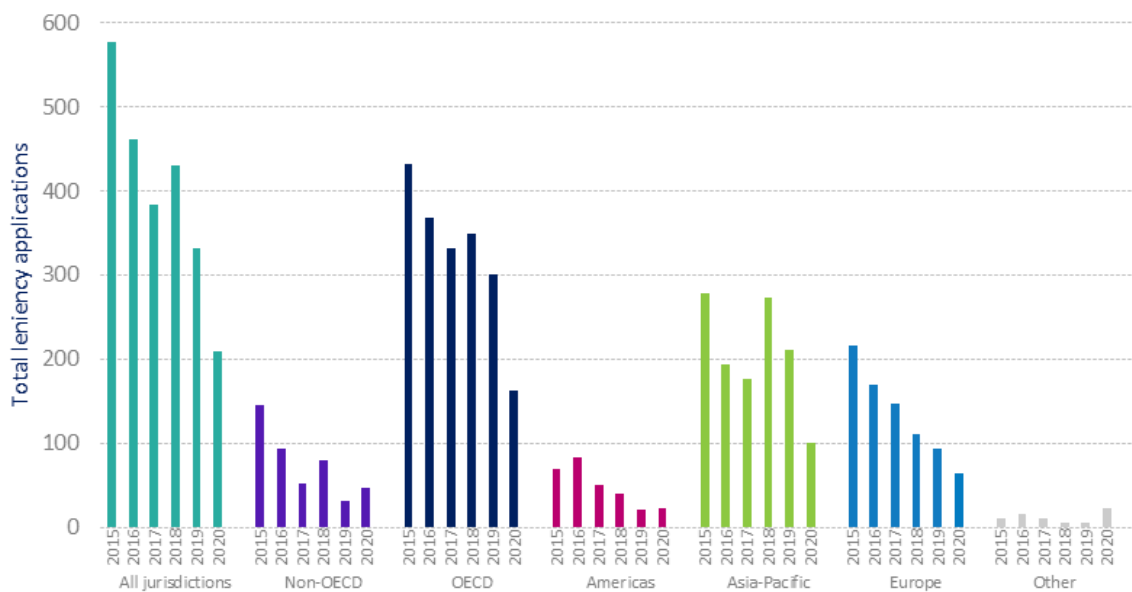
17. À l'époque où l'OCDE a organisé une table ronde pour débattre de l'utilisation de filtres afin de détecter des ententes en 2013, un petit nombre de juridictions utilisaient des filtres d'ententes de manière systématique comme outils de détection ex ante (OCDE, 2013, pp. 9,82-223<sup>[5]</sup>). Plusieurs juridictions procédaient certes à une certaine forme de filtrage des ententes, mais c'était plus sur une base ad hoc et souvent pour corroborer des soupçons d'entente déjà existants (Beth and Gannon, 2022, p. 81<sup>[11]</sup>). Par ailleurs, l'accent était mis davantage sur des filtres structurels, permettant aux autorités de la concurrence d'identifier des marchés potentiellement à risque justifiant des investigations plus poussées, tandis que l'utilisation des filtres comportementaux était plus limitée.

18. Depuis 2013, les filtres suscitent un intérêt croissant. Déjà en 2016, 15 des 27 autorités de la concurrence auprès desquelles le Réseau international de la concurrence a effectué une enquête avaient déclaré procéder à un certain filtrage (Harrington and Imhof, 2022<sup>[10]</sup>). L'Autorité danoise de la concurrence et de la consommation a annoncé au cours de la Conférence CMA de 2022 sur le thème Data, Technology and Analytics, qu'elle développait l'outil de filtrage des ententes Bid Viewer (Encadré 7) en collaboration avec plusieurs autorités et échangeait des vues sur les filtres avec plus de 15 autorités<sup>4</sup>. Beth et Gannon présentent une liste de juridictions qui ont récemment annoncé publiquement qu'elles utilisent déjà des filtres d'ententes ou projettent de développer et d'utiliser ces filtres (Beth and Gannon, 2022<sup>[11]</sup>).

19. La disponibilité croissante de grands volumes de données numériques sur les prix et les quantités, ainsi que l'émergence de nouvelles technologies qui permettent d'extraire et d'analyser des données d'une manière de plus en plus automatisée, ont permis d'élaborer de nouvelles méthodes de filtrage (notamment l'apprentissage automatique, détaillé dans l'Encadré 7), qui peuvent améliorer la fiabilité des résultats du filtrage. Plusieurs autorités de la concurrence envisagent, ont récemment commencé de développer, ou utilisent déjà, des filtres numériques nouveaux ou améliorés afin de détecter des comportements anticoncurrentiels. La plupart de ces filtres se concentrent sur des accords horizontaux.

20. Les progrès techniques de l'environnement numérique ne sont peut-être pas la seule raison qui explique l'intérêt croissant pour les filtres. Ces dernières années, le nombre de demandes de clémence, - soit la méthode de détection des ententes la plus importante -, a baissé<sup>5</sup>. En effet, entre 2015 et 2020, les demandes de clémence ont baissé de 64 % dans le monde, et, en Europe, les demandes de clémence en 2020 ont baissé de 71 % par rapport à 2015 (OCDE, 2022, p. 46<sup>[12]</sup>).

**Graphique 1. Nombre total de demandes de clémence, par région, 2015-2020**



Note : Ces chiffres incluent les 48 juridictions qui ont fourni des données complètes sur les demandes de clémence pour les six années complètes et qui ont un programme de clémence en vigueur.

Source : Graphique 5.4 (OCDE, 2022, p. 46<sup>[12]</sup>). Base de données CompStats OCDE.

21. Simultanément, l'activité des ententes ne montre aucun signe de fléchissement. Pendant la seule période de six ans comprise entre 2010 et 2016, un nombre record de 75 nouvelles ententes internationales ont été découvertes chaque année (Connor, 2016<sup>[13]</sup>)<sup>6</sup>, et la poursuite des ententes injustifiables demeure une priorité de l'action répressive des autorités de la concurrence des Membres de l'OCDE.

22. L'activité des ententes, combinée au déclin du nombre de demandes de clémence, signifie qu'il faut intensifier les actions de répression de manière plus volontariste, au moyen d'enquêtes d'office, afin de découvrir de nouveaux cas d'ententes et de booster les programmes de clémence en créant une menace crédible de découverte des ententes. Les filtres d'ententes sont un complément important des programmes de clémence (Abrantes-Metz and Metz, 2019<sup>[14]</sup>).

23. Les autorités de la concurrence améliorent leurs filtres existants et en développent de nouveaux. L'Encadré 1 donne des exemples de filtres qui ont conduit à découvrir des ententes (au Brésil) et de filtres développés par les autorités de la concurrence (en Colombie, à Singapour et en Catalogne).

### Encadré 1. Exemples de filtres numériques d'ententes : Brésil, Colombie, Singapour, Espagne

#### Brésil

L'autorité brésilienne de la concurrence, CADE, a développé Cérebro (le « Cerveau »), outil qui se fonde sur l'exploration de données et des tests statistiques pour détecter des systèmes de soumissions suspects dans la passation des marchés publics. Cérebro inclut :

1. un entrepôt de données qui combine des bases de données publiques et privées dans une unique base de données consultable ;
2. une exploration des données sur i) les modèles et similitudes de comportement des concurrents, ii) des faits suspects, iii) des signes de simulation de la concurrence ;
3. des tests (modèles) statistiques basés sur i) la littérature universitaire consacrée aux filtres d'ententes statistiques, ii) de précédentes affaires d'ententes, iii) la théorie microéconomique.

Les outils d'exploration des données permettent d'automatiser des analyses qui exigeaient auparavant des travaux de la part des enquêteurs et des fonctionnaires chargés des affaires. Les objectifs de Cérebro sont doubles : trouver des indices d'ententes dans les procédures de passation des marchés publics, comme des faits non plausibles ou des modèles comportementaux, et fournir des preuves pour la réalisation de perquisitions, d'une part, et, d'autre part, soutenir et améliorer les investigations.

Plus précisément, cet outil recherche les modèles de comportement suivants dans les données sur la passation des marchés : suppression d'offres, offres de couverture, rotation des offres, soumissionnaires perdants superflus, part de marché stable, modèles de fixation des prix, similitudes des textes et métadonnées des dossiers soumis.

En 2018, la CADE a lancé sa première enquête fondée sur les constatations livrées par l'outil, lequel est devenu la procédure quotidienne de détection des soumissions concertées. Le processus d'exploration des données et les marqueurs de Cérebro continuent d'être améliorés grâce à des tests et à l'usage, et une petite équipe est entièrement dédiée à ces tâches.

Source : OCDE (2021) Fighting bid rigging in Brazil: a review of federal public procurement <https://www.oecd.org/daf/competition/fighting-bid-rigging-in-brazil-a-review-of-federal-public-procurement.htm>; OCDE (2018) "Workshop on cartel screening in the digital era", <https://www.oecd.org/competition/workshop-on-cartel-screening-in-the-digital-era.htm>.

### **Colombie**

La Superintendencia de Industria y Comercio (SIC) a développé un outil, dénommé Sherlock, qui analyse les données sur la passation des marchés publics afin d'aider les enquêteurs à identifier des signes ou schémas qui suggèrent un comportement collusif.

Au cours de la première phase du projet, la SIC a créé un outil permettant aux enquêteurs d'accéder à des données publiquement accessibles récupérées sur internet dans un format lisible. Le but est d'aider les responsables chargés du traitement des affaires à identifier, par exemple, les procédures de passation des marchés présentant des caractéristiques similaires à celles des procédures examinées ; les soumissionnaires ; les marchés adjugés par la même entité aux mêmes soumissionnaires ; les relations entre les représentants légaux et les participants concernés, etc. Le logiciel donne accès à une base de données de la SIC qui présente des données nettoyées et structurées et fournit des statistiques descriptives simples. L'outil permet en outre la visualisation grâce à un tableau de bord qui peut appliquer des filtres et générer des graphiques et des comparaisons.

Dans la seconde phase du projet, qui est toujours en cours, la SIC va automatiser l'application de filtres pour des marqueurs identifiés par des organisations internationales comme l'OCDE. Cette automatisation sera réalisée en utilisant des techniques d'apprentissage automatique et d'apprentissage profond.

### **Singapour**

La Commission de la concurrence et des consommateurs de Singapour (Competition and Consumer Commission of Singapore (« CCCS »)) dispose d'un outil de détection des soumissions concertées (Bid Rigging Detection Tool (« BRDT »)) qu'elle a développé en interne. Le BRDT analyse les prix et les schémas de soumission sur la base d'indicateurs quantitatifs qui signalent un comportement suspect de la part de soumissionnaires.

Le CCCS, en collaboration avec l'Agence gouvernementale de technologie, développe également un logiciel de comparaison de documents, qui permet une exploration en profondeur des dossiers d'offres soumis dans le cadre de soumissions « suspectes ». Le logiciel utilise des techniques d'analyse de textes afin de générer des scores de similitude permettant d'effectuer des comparaisons au niveau des phrases et d'un document entier. Les enquêteurs peuvent ainsi se concentrer sur des phrases et des documents similaires, plutôt que de devoir passer au peigne fin de grands volumes de documents, ce qui réduit considérablement le temps et les efforts nécessaires pour analyser les preuves, et minimise le risque de négliger des documents importants en raison d'une erreur humaine.

### **Catalogne (ACCO)**

En Espagne, l'autorité de la concurrence catalane (ACCO) a développé un outil de filtrage des ententes dénommé « Smart Administrative Procurement Collusion Research Tool » (ERICCA). Cet outil informatique donne un accès complet aux données sous-jacentes, et fournit des statistiques descriptives de base sur la procédure d'appel à la concurrence et les soumissionnaires. Il signale en outre les groupes suspects d'entreprises, en utilisant un modèle d'apprentissage automatique non supervisé.

L'ACCO a rencontré des problèmes liés à la disponibilité de données de base, l'identification des soumissionnaires et l'accès aux prix de toutes les offres (et non pas uniquement au prix de l'offre gagnante). Il élabore actuellement des filtres supplémentaires pour les groupements d'entreprises suspectes, notamment afin d'identifier les soumissionnaires gagnants en fonction de l'autorité chargée

de la passation du marché et de la période de soumission des offres gagnantes. L'ACCO ajoutera également des données et informations issues d'affaires précédentes de l'ACCO, au fur et à mesure de leur disponibilité.

Source (Colombie, Singapour et Espagne) : Schrepel, Thibault and Groza, Teodora, The Adoption of Computational Antitrust by Agencies: 2021 Report (21 juin 2022), 2 Stanford Computational Antitrust, 78 (2022), SSRN : <https://ssrn.com/abstract=4142225>

24. Les développements détaillés dans l'Encadré 1 reprennent des informations publiques sur les initiatives de filtrage des autorités de la concurrence concernées. D'autres autorités de la concurrence peuvent également élaborer et utiliser des filtres, sans toutefois l'annoncer. Les autorités de la concurrence qui préfèrent garder ainsi le secret le font notamment de crainte qu'en dévoilant l'existence, et en particulier les caractéristiques de leurs filtres, les entreprises se livrant à des activités illégales utilisent ces informations pour adapter leur comportement, tenir le filtre en échec et, partant, éviter la détection.

25. D'autres autorités voient la publicité autour de l'utilisation de filtres d'ententes comme une forme de dissuasion, incitant les entreprises à s'inquiéter du risque de détection, et, avec un peu de chance, à cesser leurs activités illégales ou à solliciter des mesures de clémence (OCDE, 2013<sup>[5]</sup>). Un filtre d'entente efficace inciterait donc les entreprises à adapter leur comportement afin d'éviter la détection, ce qui réduirait probablement l'efficacité de l'entente et accroîtrait les coûts de coordination, ce qui réduirait finalement les incitations à conclure une entente illégale (Kawai, 2022<sup>[15]</sup>). Une protection complète contre la détection exigerait de ne procéder à aucune augmentation des prix<sup>7</sup>. Étant donné qu'il existe déjà beaucoup d'informations publiques sur les filtres d'ententes, et que la plupart des méthodes de filtrage des ententes sont issues de la littérature universitaire, les ententes peuvent déjà adapter leur conduite pour échapper à des filtres potentiels. Cette situation exigerait de concevoir des outils de filtrage de plus en plus sophistiqués, dans une sorte de jeu du chat et de la souris<sup>8</sup>.

### 2.3. Méthodes de filtrage numérique présentées dans la littérature académique récente

26. La littérature académique relative aux filtres d'ententes existe depuis des décennies. Les premières contributions importantes portaient sur les fonctions d'offres<sup>9</sup> (notamment celles de (Porter and Zona, 1993<sup>[16]</sup>), (Porter and Zona, 1997<sup>[17]</sup>), (Porter and Zona, 1999<sup>[18]</sup>), (Bajari and Ye, 2003<sup>[19]</sup>)), tandis que d'autres présentaient des filtres comportementaux statistiques (dont (Harrington, 2006<sup>[7]</sup>), (Abrantes-Metz et al., 2006<sup>[20]</sup>)).

27. Cette littérature a été considérablement étoffée ces dernières années. L'0 présente un aperçu de quelques articles importants depuis 2015, classés par approche de filtrage. Les approches comprennent les méthodes basées sur la distribution des offres/prix, l'arrondi des offres, les méthodes basées sur la cointégration, l'indépendance conditionnelle entre les offres, les doubles différences, les fractions d'offre historiques, les méthodes de réseau, les méthodes probabilistes, les méthodes de discontinuité de la régression, les méthodes de rupture structurelle, les regroupements suspects dans les données géographiques spatiales et la combinaison de plusieurs ententes<sup>10</sup>. La plupart des articles concernent les données relatives aux marchés publics.

28. La littérature académique présentée en 0 examine principalement les filtres comportementaux, considérés comme les plus efficaces. Ces filtres exigent de disposer de données et de savoir quelles informations rechercher dans ces données. Harrington et Imhof (2022<sup>[10]</sup>) identifient trois grandes approches aux filtres comportementaux utilisant les données sur les offres et les prix : les marqueurs de collusion, les ruptures structurelles et les anomalies<sup>11</sup>. Un marqueur de collusion est une tendance dans les données qui est plus cohérente avec la collusion qu'avec la concurrence. Une rupture structurelle est

un changement brusque dans le processus de génération des données. Une anomalie est une tendance dans les données qui est inexplicable ou incompatible avec la concurrence, mais qui peut se révéler compatible avec la collusion<sup>12</sup>. Une combinaison de marqueurs de collusion et de ruptures structurelles appliquée aux marchés publics constituerait un bon point de départ pour une autorité de la concurrence désireuse de lancer un programme de filtrage des ententes avec des ressources limitées<sup>13</sup>.

29. **Les méthodes basées sur la distribution des offres/prix** (parfois appelées « filtres basés sur la variance ») sont des filtres statistiques comportementaux qui prennent principalement en compte les moments des offres/prix, tels que la distance relative normalisée, la différence en pourcentage, l'écart type, le coefficient de variance, la dispersion, l'asymétrie et l'aplatissement. Ces filtres présument que l'activité de l'entente réduit la variance des offres ou des prix. Plusieurs articles récents utilisent des techniques d'apprentissage automatique pour combiner plusieurs filtres basés sur la distribution des offres.

30. **L'approche de l'arrondi des offres** prend en compte le nombre de zéros consécutifs à la fin du prix de l'offre. Par exemple, le « niveau d'arrondi » pour un prix d'offre de 12 300 000 JPY serait de cinq. On part de l'hypothèse que les soumissionnaires de l'offre gagnante choisiront des chiffres ronds pour leurs offres afin d'éviter les erreurs de communication. Le « niveau de l'arrondi » sera plus élevé en cas de collusion.

31. **Les filtres d'entente basés sur la cointégration** sont une méthodologie économétrique de séries chronologiques qui permet de vérifier s'il existe un co-mouvement statistiquement significatif des prix entre les entreprises, ce qui suggère une collusion potentielle. Deux séries temporelles sont dites cointégrées s'il existe entre elles une relation d'équilibre à long terme qui peut être représentée comme une fonction linéaire de ces deux séries temporelles.

32. **L'indépendance conditionnelle à l'aide des fonctions d'offre** est une méthode populaire qui suppose que l'offre concurrentielle d'une entreprise ne devrait pas être corrélée aux offres concurrentielles des autres entreprises, une fois que l'on a contrôlé les informations susceptibles d'affecter la valeur des offres. Cette approche estime une régression de la fonction d'offre qui teste l'indépendance conditionnelle entre les offres. La dépendance conditionnelle entre les offres démontre qu'il s'agit d'offres complémentaires potentielles résultant d'une collusion.

33. **La méthode des doubles différences** est une approche économétrique standard qui utilise des données de panel pour déterminer l'effet moyen du traitement (par exemple, l'augmentation des offres/prix due à la collusion) en calculant la différence entre le changement moyen dans le temps pour le groupe de traitement (différence un) et le changement moyen dans le temps pour le groupe de contrôle (différence deux). Dans le contexte du filtrage des ententes, la variable dépendante pourrait être l'offre/le prix. Le groupe de traitement pourrait être constitué des offres/prix influencés par l'entente présumée, et le groupe de contrôle serait constitué de concurrents opérant sur un marché géographique distinct, non affecté par l'entente présumée.

34. **L'approche de la fraction de l'offre historique** prend en compte la distribution des fractions des offres, qui peut être calculée comme la propre offre du soumissionnaire moins l'offre la plus basse (c'est-à-dire l'offre la plus proche). Si les parties à une entente évitent de présenter des offres proches de celle du soumissionnaire retenu, peu d'offres se situeront autour de zéro, ce qui traduit un comportement suspect.

35. **Les méthodes de réseau** permettent de déterminer s'il existe des points chauds où les entreprises interagissent plus fréquemment et où la collusion peut donc être plus probable et plus facile à maintenir. Grâce aux techniques d'apprentissage non supervisé (voir la section 2.5), les enquêteurs peuvent déterminer les groupes d'entreprises qui interagissent fréquemment dans un ensemble donné d'appels d'offres.

36. **Les méthodes probabilistes** utilisent la distribution des offres « honnêtes » non collusoires pour déterminer la probabilité pour qu'une offre donnée potentiellement collusoire soit le fruit du hasard. Les

offres à prix relativement élevé (celles qui se situent dans un percentile élevé) sont réputées plus susceptibles d'être collusoires. On peut établir une distribution distincte pour les offres « honnêtes » non collusoires en fonction du nombre de soumissionnaires dans la procédure d'appel d'offres (par exemple, une distribution distincte pour un soumissionnaire, deux soumissionnaires, trois soumissionnaires, etc.), afin de prendre en compte le fait qu'une offre non collusoire peut avoir un prix plus élevé lorsqu'il y a moins de concurrence pour une enchère donnée.

37. **Le plan de discontinuité de la régression (RDD)** identifie un effet moyen du traitement en prenant en compte les différences dans une variable de résultat pour les observations qui sont marginalement au-dessus ou au-dessous d'un seuil de traitement. Cette méthode sert à évaluer l'avantage de l'entreprise en place ou la rotation des offres. En effet, la méthode RDD peut déterminer si les soumissionnaires qui ont remporté un appel d'offres de façon marginale étaient plus susceptibles d'être des entreprises en place ou non. Elle suppose que les soumissionnaires qui ont remporté ou perdu des appels d'offres sont considérés comme ayant reçu de manière aléatoire une variable de résultat (comme le fait d'être ou non une entreprise en place). Si les gagnants de l'offre marginale sont plus souvent des entreprises en place, cela sous-entendrait que les parties à l'entente attribuent les appels d'offres aux entreprises en place (c'est-à-dire que les autres parties à l'entente promettent de ne pas soumissionner à ces appels d'offres).

38. **Les filtres d'entente à rupture structurelle** permettent d'estimer s'il existe des ruptures structurelles dans la fonction prix, c'est-à-dire des variations de prix inexplicables au cours de la période qui seraient très probablement dues à une entente. Une rupture structurelle est un changement soudain dans le processus de génération des données sur les prix qui peut être le résultat du début, de la fin ou de l'interruption d'une entente. La fonction prix peut être estimée à l'aide d'une régression de forme réduite qui inclut des variables qui déterminent les prix du marché, notamment les variables de l'offre et de la demande, ainsi que des contrôles pour les ruptures structurelles qui sont induites par l'activité potentielle des ententes.

39. **Les regroupements suspects dans les données géographiques spatiales** permettent de déterminer s'il existe des regroupements géographiques d'entreprises pratiquant des prix suspects. Par exemple, le coefficient de variation (voir la section « Méthodes basées sur la distribution des offres/prix » ci-dessus) peut être utilisé pour identifier les prix suspects, puis un algorithme sera appliqué pour vérifier si ceux-ci sont regroupés sur le plan géographique. On pourrait alors y discerner un accord local visant à augmenter les prix, comme ce fut le cas pour les prix de l'essence dans les stations-service.

40. Certains articles combinent plusieurs filtres. (Fazekas et al., 2022<sup>[21]</sup>) et associent des filtres tels que des méthodes basées sur la distribution des offres/prix, des modalités de soumission d'offres et des filtres de concentration du marché, en utilisant une approche d'apprentissage automatique supervisé par l'intermédiaire de forêts aléatoires (voir Annexe B. Méthodes d'apprentissage automatique). Les auteurs élaborent une classification théorique des types de collusion dans les marchés publics qui devrait couvrir « *la plupart, sinon la totalité, des comportements collusoires* » et identifier les indicateurs de filtrage des ententes qui peuvent permettre de détecter ces types d'ententes (Tableau 1). Ils estiment que certains filtres fonctionneront mieux pour certains types d'ententes et qu'une combinaison de filtres est nécessaire pour identifier des cas de collusion. Leurs conclusions empiriques vont dans ce sens et soulignent que les filtres uniques ne sont généralement pas adaptés et que l'approche multi-filtres est la plus efficace.

**Tableau 1. Indicateurs de filtres d'ententes correspondant au type de collusion dans les marchés publics**

Type de collusion				Nom de l'indicateur									
Type de collusion	Structure du marché	Technique de collusion	Répartition de la rente	Prix		Modalités de soumission des offres						Structure du marché	
				Prix relatif	Loi de Renford	Offre unique	Soumissionnaires	Sous-traitance	Consortium	Position de rupture	Probabilité d'être	Structure de marché	Structure de marché
				1	2	3	4	5	6	7	8	9	10
A	Concentrée	Offre retenue	Sous-traitants	Oui	Oui	Oui	Oui	Oui	Non	Non	Oui	Oui	Non
B	Concentrée	Offre perdante	Sous-traitants	Oui	Oui	Non	Non	Oui	Non	Oui	Non	Oui	Non
C	Concentrée	Offre conjointe	Consortium	Oui	Oui	Oui	Oui	Non	Oui	Non	Oui	Oui	Non
D	Stable	Offre retenue	Sous-traitants	Oui	Oui	Oui	Oui	Oui	Non	Non	Oui	Non	Oui
E	Stable	Offre perdante	Sous-traitants	Oui	Oui	Non	Non	Oui	Non	Oui	Non	Non	Oui
F	Stable	Offre retenue	Appel d'offres coordonné	Oui	Oui	Oui	Oui	Non	Non	Non	Oui	Non	Oui
G	Stable	Offre perdante	Appel d'offres coordonné	Oui	Oui	Non	Non	Non	Non	Oui	Non	Non	Oui

Note : « Oui » signifie que l'indicateur peut détecter le type de collusion, tandis que « Non » signifie qu'il ne le peut pas. La plupart, sinon la totalité, des comportements collusoires sur les marchés publics peuvent être classés en trois catégories : (1) technique de collusion (offres retenues, offres perdantes/offres non concurrentielles, offres conjointes) ; (2) forme de partage de la rente (sous-traitant, consortiums/propriété conjointe, offres coordonnées, paiements parallèles informels) ; (3) structure de marché résultante (structure de marché concentrée où les entreprises attribuent des marchés géographiques ou de produits spécifiques à une entreprise donnée, ou structure de marché stable où les entreprises imitent un marché concurrentiel avec rotation du gagnant, avec pour corollaire des parts de marché stables dans le temps). (Fazekas et al., 2022, p. 14<sup>[21]</sup>) présentent une description détaillée de cette répartition de l'entente.

Source : Secrétariat de l'OCDE, sur la base du Tableau 8 dans (Fazekas and Tóth, 2016<sup>[22]</sup>) et des Tableaux 1.1 et 1.3 dans (Fazekas et al., 2022<sup>[21]</sup>).

## 2.4. Le risque de faux négatifs ou de faux positifs

41. Les filtres (numériques ou autres) comportent un risque inhérent de faux positifs ou de faux négatifs : ils fournissent uniquement une preuve économique qui est souvent ambiguë, car elle peut correspondre soit à une action concertée, soit à une action indépendante. Les autorités de la concurrence qui ont participé à la table ronde de l'OCDE de 2013 consacrée aux enquêtes d'office relatives aux ententes et à l'utilisation de filtres pour détecter les ententes (OCDE, 2013<sup>[5]</sup>) ont particulièrement souligné le risque de faux positifs qui peuvent les inciter à se saisir d'une affaire alors qu'aucune activité illégale n'a eu lieu en réalité, ce qui entraîne une perte de temps et de ressources.

42. Les faux positifs posent particulièrement problème en ce qui concerne les filtres structurels, car ils sont plus larges et moins sophistiqués. Ils pourraient conduire à signaler des industries concentrées, comportant peu d'acteurs et des barrières importantes à l'entrée, alors même qu'elles ne sont pas nécessairement collusives<sup>14</sup>. Les variables omises sont au cœur de ce problème (Harrington, 2006<sup>[7]</sup>).

43. Les filtres comportementaux emportent le même risque, lorsqu'ils ne saisissent pas des circonstances qui peuvent expliquer le comportement suspect. À titre d'exemple, des corrélations de prix peuvent être la conséquence d'une collusion (auquel cas elles sont illégales), d'une collusion tacite (auquel cas elles ne constituent pas une violation dans la plupart des juridictions) ou d'une coïncidence. Les résultats du filtrage doivent donc être analysés prudemment afin d'éviter de sauter à des conclusions hâtives, et afin de distinguer, par exemple, des cas de collusion tacite ou de chocs exogènes pouvant expliquer les variations de prix.



44. Il est apparu, à un certain moment, que le système d'analyse des indicateurs de soumissions concertées (*Bid Rigging Indicator Analysis System* (BRIAS)) de l'autorité coréenne de la concurrence avait produit trop de résultats positifs. Le système BRIAS a fait l'objet de plusieurs améliorations depuis 2018, et son utilisation est actuellement en progression.

### Encadré 2. Le système coréen d'analyse des indicateurs de soumissions concertées (BRIAS)

En 2006, l'autorité coréenne de la concurrence (Korean Fair Trade Commission (KFTC)) a lancé un système automatisé de détection des ententes, basé sur l'affectation de coefficients de pondération à différents indicateurs. Cet outil, connu sous le nom de système d'analyse des indicateurs de soumissions concertées (*Bid Rigging Indicator Analysis System* (BRIAS)), a pour but de détecter des soumissions concertées en utilisant des filtres appliqués aux données électroniques sur les appels d'offres publics, émanant de KONEPS, qui est la plateforme centrale de dématérialisation des marchés publics.

Le système BRIAS collecte des données sur la passation des marchés publics des autorités gouvernementales et des entreprises publiques au-delà d'un certain montant. Il sélectionne des marchés (portant sur des valeurs égales ou supérieures à 100 millions KRW pour l'achat de biens et de services, à 5 milliards KRW pour des projets généraux de construction et à 500 millions KRW pour des constructions spécialisées), et les filtre sur la base de critères spécifiques en fonction des types d'offres et de collusion. Le système BRIAS produit ensuite un score reflétant la probabilité de soumission concertée, basé sur l'affectation de valeurs pondérées à des facteurs spécifiques au secteur considéré.

La KFTC s'est servie de ses expériences passées en matière de répression des ententes illégales et des signaux d'alerte pouvant révéler l'existence de soumissions concertées afin de concevoir les indicateurs de cet outil et leurs poids respectifs. Les scores de probabilité de collusion les plus élevés sont basés sur : (i) un taux élevé d'adjudication des marchés au profit d'une entreprise ; (ii) un petit nombre de soumissionnaires répondant à l'appel d'offres ; (iii) un grand nombre d'offres supérieures au prix estimé ; (iv) l'utilisation de procédés de soumission non concurrentiels ; et (v) de grands écarts entre l'offre gagnante et les offres perdantes. L'efficacité de l'outil dépend de l'équilibrage des poids attribués à chacun des indicateurs.

Il s'est avéré que le système BRIAS avait généré un nombre important de résultats présentant des scores élevés, de telle sorte qu'il est devenu très difficile de sélectionner les affaires justifiant une enquête.

En 2015, la KFTC a décidé d'améliorer certaines fonctions du système BRIAS et ces améliorations sont devenues opérationnelles en 2018. Elles comprennent une nouvelle fonction de recherche et un moyen plus efficace de connecter les données sur les soumissions au système BRIAS, de telle sorte que l'utilisation de ce système est actuellement en progression. La KFTC a utilisé le système BRIAS pour enquêter sur 16 affaires et les sanctionner entre 2007 et 2021, dont 13 à partir de 2018.

Source : Kim, Daein. (2021). Korean Public Procurement Law, K-Law Academy, Korea Legislation Research Institute, [https://www.researchgate.net/publication/348618992\\_Korean\\_Public\\_Procurement\\_Law](https://www.researchgate.net/publication/348618992_Korean_Public_Procurement_Law) ; OCDE (2013) Ex officio cartel investigations and the use of screens to detect cartels, [www.oecd.org/daf/competition/exofficio-cartel-investigations.htm](http://www.oecd.org/daf/competition/exofficio-cartel-investigations.htm); Informations fournies au Secrétariat par la Korean Fair Trade Commission

45. Les filtres peuvent également produire des faux négatifs. Les filtres structurels sont moins susceptibles de produire des faux négatifs, étant donné que la collusion est difficile à soutenir sur des marchés qui ne présentent pas des caractéristiques « propices à la collusion ». En revanche, les filtres comportementaux risquent davantage de produire des faux négatifs lorsqu'ils sont appliqués dans un

contexte différent de celui pour lequel ils ont été conçus, étant donné que le modèle peut échouer à reconnaître des indices d'activité illégale qui n'ont pas été intégrés dans sa conception<sup>15</sup>.

46. L'un des moyens de minimiser les risques d'erreur pourrait consister à réaliser une série de tests, plutôt qu'un seul test, en utilisant un filtre composite. L'approche multi-filtres permise par l'apprentissage automatique et la disponibilité d'un plus grand nombre de données, devrait être utile à cet égard.

## 2.5. L'apprentissage automatique améliore le filtrage numérique

47. Les filtres sont conçus à partir d'hypothèses sur la manière dont un marché fonctionne et sur les types de comportement illégal qui peuvent se produire. Autrement dit, les différents filtres sont personnalisés en fonction des différents types potentiels de violations et des différents marchés, et il n'existe aucun filtre unique parfait qui permette d'identifier toutes les violations sur tous les marchés. Toutefois, des techniques d'apprentissage automatique peuvent permettre de mieux prédire si un comportement correspond à une collusion. Au cours des 5 dernières années, la littérature académique et la pratique des autorités de la concurrence dans le contexte des filtres d'ententes se sont largement focalisées sur l'apprentissage automatique.

48. L'apprentissage automatique est un sous-domaine de l'intelligence artificielle (« *la science et l'ingénierie permettant de créer des machines intelligentes* »), qui donne aux « *ordinateurs la capacité d'apprendre sans être explicitement programmés* » (OCDE, 2017, p. 9<sup>[4]</sup>). Il s'agit « *d'une application d'algorithmes de structure minimale et de filtrage par motif afin (i) d'inférer une règle de classification à partir d'un ensemble de données d'entraînement et (ii) de faire des prédictions utiles sur des données nouvelles* » (Abrantes-Metz and Metz, 2018<sup>[23]</sup>). Pour exprimer les choses simplement, l'apprentissage automatique poursuit le même objectif principal que d'autres filtres empiriques « classiques » (classifier et prédire), mais approche cet objectif d'une manière moins structurée et plus pilotée par les données, qui peut s'affranchir (davantage) d'hypothèses initiales et gagner en sophistication et en précision à l'usage.

49. En particulier, plusieurs auteurs soutiennent que s'il existe des données suffisantes (y compris en fusionnant des sources de données), l'apprentissage automatique permet de combiner plusieurs filtres (en utilisant des méthodes comme l'apprentissage d'ensemble - voir 0) d'une manière qui peut conduire à une identification plus précise de la collusion. En effet, les résultats du filtrage peuvent être regroupés dans un indice composite, chacun se voyant affecter un coefficient de pondération (Fazekas et al., 2022<sup>[21]</sup>). En d'autres termes, sachant qu'il n'existe aucun filtre unique qui puisse correctement signaler une activité collusive dans tous les cas, l'apprentissage automatique peut combiner plusieurs filtres d'une manière susceptible d'identifier au mieux des cas de collusion potentielle et, partant, de conduire à l'ouverture d'un plus grand nombre d'enquêtes sur la base d'indicateurs de collusion (Huber and et Imhof, 2021<sup>[24]</sup>; Imhof and Wallimann, 2021<sup>[25]</sup>; Huber and Imhof, 2019<sup>[26]</sup>; Imhof, Karagök and Rutz, 2018<sup>[27]</sup>).

50. Il existe trois grands types d'approches de l'apprentissage automatique : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage de renforcement. La plupart des études universitaires consacrées aux filtres d'ententes utilisent des approches d'apprentissage supervisé. Les autres utilisent essentiellement des approches d'apprentissage non supervisé (généralement afin d'identifier des groupes d'entreprises qui interagissent fréquemment). Très peu, si tant est qu'il en existe, utilisent des approches d'apprentissage de renforcement.

51. **L'apprentissage supervisé** (Hastie et al., 2009, pp. 9-42<sup>[28]</sup>) utilise des données d'entrée (également appelées variables indépendantes) qui servent à estimer un résultat (également désigné sous le terme de réponse ou de variable dépendante). Il se fonde habituellement sur un ensemble de données d'entraînement relatives à des affaires résolues, connues sous le nom de données « signalées » ou « étiquetées », qui fournit un échantillonnage de données d'entrée afin de produire un résultat. L'apprentissage supervisé peut être pensé comme un apprentissage par l'exemple ou un apprentissage

avec un enseignant. L'ensemble des données d'entraînement fournit des exemples des valeurs des filtres d'ententes pour des soumissions concertées et des soumissions de pleine concurrence. Le modèle est formé sur cet ensemble de données, en estimant les poids à accorder à chacun des filtres afin d'optimiser l'identification de comportements illégaux. Le modèle d'entraînement est ensuite appliqué à des données de test contenant les valeurs des filtres d'ententes (de préférence sur le même marché ou un marché comparable), afin de prédire si un comportement donné peut effectivement constituer une soumission concertée ou une soumission respectant les règles de concurrence. L'ensemble des données d'entraînement doit avoir une taille suffisante pour minimiser le risque d'erreur et contenir des classifications correctes des résultats, afin de les considérer comme révélant une « collusion » ou non (Abrantes-Metz and Metz, 2018<sup>[23]</sup>). L'apprentissage supervisé est souvent le mieux adapté au filtrage des ententes, car il peut être utilisé pour prédire si une offre/un prix est ou non collusif. Toutefois, il exige de disposer de données « étiquetées » sur des exemples existants d'offres/de prix collusifs et non collusifs.

52. **L'apprentissage non supervisé** (Hastie et al., 2009, pp. 485-586<sup>[28]</sup>) poursuit le même objectif que l'apprentissage supervisé (c'est-à-dire utiliser des données d'entrée pour estimer un résultat). La distinction clé tient au fait que l'apprentissage non supervisé utilise des données « non signalées » ou « non étiquetées », c'est-à-dire des données qui contiennent uniquement des valeurs d'entrée et non une valeur de sortie associée. Ainsi, l'apprentissage non supervisé peut être pensé comme un apprentissage sans enseignant. Il s'agit plutôt d'un ensemble de données d'entrée avec une distribution de probabilité sous-jacente, et l'objectif est ici de déterminer cette distribution de probabilité sans l'aide d'un superviseur ou d'un enseignant indiquant quand une allocation est correcte. En d'autres termes, le modèle essaie d'apprendre une structure à partir de données « non signalées » ou « non étiquetées ». L'apprentissage non supervisé peut être une alternative utile à l'apprentissage supervisé car il n'exige pas de disposer de données étiquetées, et identifie plutôt des valeurs aberrantes suspectes qui sont les plus éloignées de la « norme » (Deng, 2017<sup>[29]</sup>).

53. **L'apprentissage de renforcement** (Barto and Dietterich, 2004<sup>[30]</sup>), comme l'apprentissage non supervisé, utilise des données « non signalées » ou « non étiquetées », c'est-à-dire de données qui contiennent uniquement des valeurs d'entrée et non une valeur de sortie associée. Toutefois, à la différence de l'apprentissage non supervisé, l'apprentissage de renforcement utilise un critère de performance qui récompense un résultat positif et pénalise un résultat négatif (c'est-à-dire l'apprentissage selon les méthodes des « essais et des erreurs »). L'algorithme de l'apprentissage de renforcement sélectionne des données d'entrée et observe le résultat à partir du critère de performance, en choisissant finalement la donnée d'entrée qui fournit la valeur maximale selon ce critère de performance.

54. Les techniques spécifiques d'apprentissage supervisé et non supervisé utilisées dans les études universitaires consacrées aux filtres d'ententes sont décrites en 0.

## 2.6. Filtrage numérique par des entreprises privées

55. L'utilisation de filtres d'ententes n'est pas limitée aux autorités de la concurrence. Certaines entreprises ont également commencé à utiliser des filtres afin d'enquêter sur le risque d'ententes au sein de leurs chaînes d'approvisionnement<sup>16</sup> ou dans le cadre d'un programme de conformité<sup>17</sup>. L'Encadré 3 illustre la manière dont la Deutsche Bahn a élaboré des filtres d'ententes.

56. Le filtrage des ententes comporte plusieurs avantages pour les entreprises. Il peut dissuader de conclure des ententes dans la chaîne d'approvisionnement de l'entreprise en raison du risque de détection, fournir des preuves à l'appui de demandes d'indemnisation, aider les entreprises à identifier des marchés exposés au risque de collusion et leur permettre de concevoir leurs stratégies d'approvisionnement de manière à contrer ces risques (Beth and Gannon, 2022, pp. 85-86<sup>[11]</sup>). Les filtres d'ententes peuvent être particulièrement bénéfiques pour les entreprises qui achètent des matières premières ou lancent des

appels d'offres uniquement fondés sur le prix, étant donné que les ententes sont extrêmement courantes sur ces marchés (Harrington, 2021<sup>[31]</sup>; 2015<sup>[32]</sup>).

57. Les filtres d'ententes utilisés par des entreprises du secteur privé peuvent s'avérer encore plus efficaces que les filtres publics, étant donné que les entreprises disposent de données sur les fournisseurs que les autorités de la concurrence n'ont pas, ce qui permet au filtre de s'appuyer sur de meilleurs ensembles de données. Par exemple, les entreprises qui recourent à des appels d'offres pour se fournir en produits ont accès à des données en temps réel sur les soumissions au niveau de l'entreprise, et disposent d'une connaissance du marché d'approvisionnement qui leur permet de déterminer si les conditions régnant sur ce marché sont concurrentielles (ou au contraire collusives) (Beth and Gannon, 2022, p. 85<sup>[11]</sup>).

58. Les entreprises peuvent également utiliser des filtres d'ententes dans le cadre de leur programme de conformité (Abrantes-Metz, 2012<sup>[33]</sup>; 2020<sup>[34]</sup>), (Deng, 2017<sup>[29]</sup>), (Johnson and Sokol, 2020<sup>[35]</sup>), (OCDE, 2021, p. 40<sup>[36]</sup>). Une entreprise peut bénéficier de la découverte d'un comportement illégal, en étant la première à faire une demande de clémence, et en bénéficiant, ce faisant, d'une immunité de sanctions ou d'une réduction des sanctions (Abrantes-Metz and Metz, 2020<sup>[34]</sup>).

### Encadré 3. La Deutsche Bahn utilise le filtrage pour identifier des ententes dans sa chaîne d'approvisionnement

La Deutsche Bahn (qui est la compagnie ferroviaire nationale en Allemagne) recourt à un filtrage structurel pour identifier des marchés exposés au risque d'entente, et oblige ses fournisseurs opérant sur des marchés à haut risque à introduire ou maintenir des programmes de conformité antitrust efficaces. La Deutsche Bahn projette également d'introduire des filtres comportementaux afin d'identifier des ententes dans sa chaîne d'approvisionnement.

Les filtres visent à la fois à détecter les ententes et à dissuader de les conclure. En effet, les fournisseurs savent dès lors que leurs schémas de soumission sont contrôlés non seulement par l'autorité de la concurrence compétente mais également par leur client. Si la dissuasion fonctionne, cela peut contribuer à réduire la population globale d'ententes. Si la détection fonctionne, les résultats du filtrage privé devront être communiqués aux autorités de la concurrence. Il est important que les autorités de la concurrence aient les outils nécessaires pour évaluer les preuves quantitatives du filtrage des ententes qui leur sont fournies par des entreprises dans le cadre de plaintes formelles. Par ailleurs, les résultats du filtrage peuvent être utilisés par l'entreprise victime afin de réclamer une indemnisation aux fournisseurs membres de l'entente.

La Deutsche Bahn a pour objectif de mettre en place une gamme différente d'algorithmes capables d'identifier la plupart des formes d'accords d'entente, et de tirer profit de l'apprentissage automatique afin d'automatiser le filtrage dans la plus large mesure possible.

Source : Beth, H. et O. Gannon (2022), "Cartel screening—can competition authorities and corporations afford not to use big data to detect cartels?", *Competition Law & Policy Debate* 2022, Vol. 7, No. 2, pp. 1-12, <https://doi.org/10.4337/clpd.2022.0001> ; Beth, H. et T. Reimers (2019), "Screening Methods for the Detection of Antitrust Infringements", [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3501700](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3501700) ; Clemens, G. (2017), "Raising Rivals' Costs Through Cartel Detection—Why Downstream Buyers Rather Face an Upstream Cartel than Downstream Competition"

59. La coopération entre les autorités de la concurrence et les entreprises privées est importante pour aider les entreprises à décider dans quels cas un comportement suspect doit être signalé aux autorités de la concurrence. Les filtres d'ententes peuvent également aider à recueillir des preuves lorsqu'il existe d'autres indications de comportement suspect, par exemple des plaintes de consommateurs ou de

concurrents ou le signalement d'un lanceur d'alerte. Si un comportement suspect lui est signalé, l'autorité de la concurrence devra comprendre les résultats du filtrage et pourquoi l'entreprise concernée a découvert une activité suspecte. C'est pourquoi il peut être nécessaire que les résultats du filtrage opéré par l'entreprise, les données sous-jacentes, les formules utilisées dans le cadre du filtrage et les algorithmes de filtrage soient divulgués aux autorités de la concurrence à titre confidentiel. Cela permettra aux autorités de la concurrence d'évaluer et même de répliquer l'analyse, afin de déterminer s'il y a lieu ou non de prendre des mesures supplémentaires (Beth and Gannon, 2022, pp. 87-88<sup>[11]</sup>).

# 3. Les conditions requises pour un filtrage efficace, en termes de données et de ressources humaines

60. Le filtrage n'est pas qu'une question de choix de modèle. Il dépend également de la qualité des données auxquelles il s'applique, et de l'expertise de l'équipe qui le réalise. La coopération entre les autorités de la concurrence peut aider à partager des connaissances, voire même des ensembles de données et des logiciels et permettre, ce faisant, d'améliorer la conception et les résultats du filtrage.

## 3.1. De bons filtres exigent de disposer de données

61. La disponibilité et la qualité des données déterminent le type d'analyse empirique que les autorités de la concurrence sont en mesure de réaliser, et sont essentielles pour appliquer des filtres numériques, même simples. Les données doivent être accessibles, solides et utilisables ; même s'ils sont bien conçus en théorie, des filtres échoueront en pratique s'ils sont testés sur des données défectueuses ou incomplètes.

62. En 2013, l'OCDE a noté que « *des informations et des données suffisantes, pertinentes et exactes sont nécessaires à tous les stades de la mise en œuvre des filtres, depuis leur conception jusqu'à leur application et à l'interprétation des résultats* » (OCDE, 2013<sup>[61]</sup>). Les autorités de la concurrence se heurtent à plusieurs difficultés, à commencer par l'existence même des données, puis l'accès à ces données (et en particulier des données désagrégées et brutes, et des données qui ne sont pas publiquement disponibles), le format, l'intégrité de la qualité des données, et pour terminer la recherche, le nettoyage et l'utilisation des données. L'expérience des autorités de la concurrence et la littérature sur la question suggèrent toutes deux qu'il est important de commencer par régler ces problèmes de données, avant de pouvoir déployer des filtres (voir 0).

63. Lorsqu'elle souhaite procéder au filtrage d'un marché, l'autorité de la concurrence doit commencer par consulter les sources de données disponibles, ou potentiellement disponibles. Ces sources peuvent être :

- Des informations publiquement disponibles, y compris celles qui proviennent des registres des sociétés, des chambres de commerce et des plateformes de dématérialisation des marchés publics (l'accès aux informations sur la passation des marchés est détaillé à la section 3.2). Plusieurs filtres d'ententes identifient un comportement collusif en comparant les résultats (par exemple, les offres ou les prix) obtenus par le jeu d'ententes avec ceux obtenus dans des conditions de pleine concurrence. En conséquence, les autorités de la concurrence ont besoin de données sur des offres/prix qui se sont avérés être le fruit d'une entente et ceux dont il a été établi qu'ils étaient concurrentiels, telles que ces données résultent d'affaires précédentes, en combinant éventuellement plusieurs jeux de données. Le filtrage des ententes peut être efficace et rentable s'il est possible d'utiliser des données aisément disponibles (Harrington, 2021<sup>[31]</sup>).

- Des informations détenues par des autorités du secteur public, y compris des autorités de régulation sectorielle, des organes gouvernementaux et des entités responsables de la passation des marchés publics. Si les informations ne sont pas publiques, il peut se poser des problèmes de protection des données et de confidentialité, en particulier selon que le partage d'informations est autorisé ou non, y compris entre des entités du secteur public. En Australie, par exemple, la loi de 2018 qui a modifié le code pénal (Crimes Legislation Amendment (Powers, Offences and Other Measures) Act 2018) (« la Loi CLAPOOM ») permet le partage de données entre les autorités chargées de l'application de la loi dans les cas de comportement frauduleux<sup>18</sup>, bien que le comportement constitutif d'une entente ne soit pas explicitement couvert par la définition de la fraude.
- Le Web-scraping, c'est-à-dire une méthode « *d'exploration automatique des sites Web, afin d'en extraire automatiquement des données structurées qu'ils contiennent. L'utilisation d'algorithmes peut grandement faciliter le processus de collecte de données et l'analyse des données. Ces outils ont déjà été utilisés dans des enquêtes menées en vertu du droit de la concurrence* » (Lianos, 2021<sup>[37]</sup>). Le web-scraping est une activité chronophage et peut ne pas être une solution viable à long terme ; il peut toutefois fournir des indications initiales suffisantes pour justifier l'achat d'informations commerciales auprès de fournisseurs de données tiers.
- Données achetées à des fournisseurs de données commerciales.

64. Plusieurs filtres d'ententes identifient un comportement collusif en comparant les résultats (par exemple, les offres ou les prix) obtenus par le jeu d'ententes avec ceux obtenus dans des conditions de pleine concurrence. En conséquence, les autorités de la concurrence auront besoin de données sur des offres/prix qui se sont avérés être le fruit d'une entente et ceux dont il a été établi qu'ils étaient concurrentiels, telles que ces données résultent d'affaires précédentes, et pourront devoir combiner plusieurs jeux de données.

65. En outre, si les données sont utilisées régulièrement, l'autorité de la concurrence peut créer des pipelines de données (« *un ensemble de phases de traitement des données depuis une source de données jusqu'à un ensemble de données de destination, les données de sortie générées par une phase constituant les données d'entrée de la phase suivante* » (Hunt, 2022<sup>[38]</sup>)). Les pipelines de données peuvent être très utiles si des sources de données sont utilisées de manière répétée, afin de fournir un jeu de données qui peut être utilisé pour le filtrage.

66. La disponibilité de données ne signifie pas que les données présentent le niveau requis de granularité ou de qualité, sont exemptes d'omissions ou d'erreurs, ou ont été convenablement traitées de manière à maintenir leur intégrité. Le niveau de détail et la qualité des données ont fait l'objet d'une analyse approfondie dans le contexte des données sur les marchés publics (qui sont examinées à la section 3.2). Étant donné qu'il existe des données numériques sur la passation des marchés publics, et que les autorités de la concurrence choisissent fréquemment de commencer leurs initiatives de détection des ententes dans ce secteur, les caractéristiques des données requises et les pénuries éventuelles de données ont reçu une attention particulière.

### 3.2. Focus sur les filtres de détection des soumissions concertées : l'importance des données relatives à la passation des marchés publics

67. Un grand nombre de filtres numériques se concentrent sur la passation des marchés publics, où des données sont disponibles dans les systèmes nationaux et sous-nationaux de dématérialisation des marchés publics. Par exemple, BRIAS (Encadré 2), Cérebro (Encadré 1) et Bid Viewer (Encadré 7) ont tous analysé des données numériques sur la passation des marchés publics afin d'y détecter des indices de soumissions concertées. Le type de données numériques nécessaires afin de filtrer des soumissions

concertées, ainsi que les défauts ou omissions possibles des données disponibles ont donc tous fait l'objet d'une analyse fouillée et ont été testés de manière empirique.

68. L'obtention de données sur la passation des marchés publics peut se heurter à un premier obstacle, tenant à l'absence de centralisation et à la fragmentation des données. L'absence de données centralisées (ce qui peut par exemple être souvent le cas dans des pays fédéraux, où les données sur les appels d'offres peuvent être tenues au niveau national, régional/de l'État et parfois municipal) signifie, pour commencer, qu'il n'existe aucun jeu complet de données à filtrer. L'une des raisons de l'abandon par la CMA de son outil de filtrage des ententes en 2020 tenait précisément à l'absence d'un ensemble centralisé de données.

#### Encadré 4. Des données décentralisées ont altéré le fonctionnement de l'outil britannique de filtrage des ententes

En décembre 2017, la Competition and Markets Authority britannique (CMA) a lancé son outil de filtrage des ententes (Screening for Cartels (SfC)) afin d'aider les responsables de la passation des marchés publics à filtrer les données de leurs dossiers d'appels d'offres pour détecter des signes de comportement constitutif d'une entente, en utilisant divers algorithmes qui repéraient un comportement inhabituel de la part de soumissionnaires et des schémas de fixation des prix pouvant être révélateurs d'une soumission concertée.

SfC était une application permettant aux responsables de la passation des marchés publics d'entrer des données sur les procédures d'appel d'offres, d'effectuer une analyse et de produire un rapport contenant une série d'indicateurs potentiels de soumission concertée. L'outil se basait en effet sur douze tests ciblant trois domaines clés : (i) le nombre et le type d'appels d'offres ; (ii) les systèmes de fixation des prix ; et (iii) l'analyse technique des métadonnées afin de retracer l'origine des documents et de détecter des soumissions révélant une faible motivation de leurs auteurs. Des algorithmes séparés étaient utilisés pour évaluer chacun des tests, conduisant à des contrôles de réussite/d'échec qui étaient ensuite combinés dans un seul score pondéré de suspicion. Ce score était présenté dans un format facile à lire qui permettait aux responsables de la passation des marchés publics d'évaluer le risque de collusion dans leurs procédures d'appel d'offres. La CMA a insisté sur le fait que des scores élevés ne devaient pas être interprétés comme la preuve de l'existence d'ententes, mais plutôt « *inciter à revenir aux documents de soumission, à les réexaminer et à poser des questions* ».

Plusieurs problèmes juridiques liés à la propriété et à la disponibilité des données sur la passation des marchés publics ont contraint la CMA à opter pour un « modèle distribué ». En d'autres termes, les responsables de la passation des marchés publics pouvaient désormais télécharger l'application sur le site internet de la CMA, puis l'utiliser dans le cadre de leurs propres systèmes afin d'analyser leurs appels d'offres respectifs. La CMA a souligné la flexibilité du modèle distribué, en prétendant qu'il permettrait à SfC de fonctionner indépendamment de la CMA et élargirait le champ des possibilités futures de personnalisation et de développement de ses paramètres.

En pratique, toutefois, le déploiement décentralisé de SfC signifiait que les données nouvelles entrées dans l'algorithme par un responsable de la passation des marchés publics restaient inconnues des versions parallèles de l'algorithme utilisé par des responsables de la passation des marchés publics différents. En conséquence, les algorithmes de l'outil demeuraient rigides : étant donné que les autorités concernées entreraient uniquement des données sur leurs propres appels d'offres, il n'existerait pas suffisamment d'informations pour améliorer les paramètres et calibrer les algorithmes. Cette situation était due à l'absence d'une base de données fiable et centralisée sur la passation des marchés publics, pouvant servir à former les algorithmes. Les biais potentiels dans les résultats du score de suspicion ont alors commencé à se multiplier.



Parallèlement, les douze paramètres ont été critiqués en raison du caractère arbitraire de leur désignation et de leur pondération : plutôt que d'être élaborés grâce à l'entraînement des algorithmes, ils se fondaient sur la théorie. En outre, les trois indicateurs ciblés par les paramètres ne semblaient pas reposer sur des informations relatives à la structure ou la dynamique du marché, ce qui biaisait les résultats générés par l'outil.

Les vices de conception, combinés à l'absence de données afin de développer une version plus perfectionnée de SfC, ont conduit à remettre en cause la fiabilité de l'outil et sa capacité à détecter des soumissions concertées. La CMA a donc retiré SfC qui n'est plus utilisé depuis le 20 janvier 2020.

Source : Competition and Markets Authority <https://www.gov.uk/government/publications/screening-for-cartels-tool-for-procurers/about-the-cartel-screening-tool#:~:text=cartel%2Dscreening%2Dtool-.About%20the%20tool,competing%20for%20customers%20or%20contracts>  
 Sanchez-Graells, A. (2019). 'Screening for Cartels' in Public Procurement: Cheating at Solitaire to Sell Fool's Gold? *Journal of European Competition Law & Practice*, 10(4), 199-211. <https://academic.oup.com/jeclap/article/10/4/199/5537135?login=true>;

69. Des outils centralisés à l'intention des responsables de la passation des marchés publics continuent d'être développés. En juin 2022, le Bureau de la concurrence du Canada a lancé un outil d'évaluation du risque de collusion (Collusion Risk Assessment Tool) afin de lutter contre les soumissions concertées dans la passation de marchés publics (Bureau de la concurrence Canada, 2022<sup>[39]</sup>). Il s'agit d'une ressource en ligne disponible pour les responsables de la passation des marchés à la fois dans le secteur public et privé et pour les agents chargés des achats. Il est destiné à être utilisé dans le cadre des activités quotidiennes de vigilance afin de protéger des procédures d'appel d'offres respectant les règles de la concurrence. Il comporte un questionnaire de 10 minutes, que les responsables de la passation des marchés peuvent remplir en donnant des détails sur tout appel d'offres qu'ils projettent. L'outil produit ensuite un score de risque de collusion basé sur les caractéristiques spécifiques du projet, et propose des stratégies d'atténuation pouvant être adoptées afin de minimiser ces risques. Il peut donc alerter précocement les responsables de la passation des marchés sur les risques potentiels de soumissions concertées, et leur fournir des suggestions d'atténuation de ces risques. Toutefois, l'outil n'est pas un outil de détection en soi : il peut être utile pour améliorer la procédure d'appel d'offres et prévenir le risque d'ententes, mais n'est pas susceptible de fournir des pistes d'enquête au Bureau de la concurrence.

70. Une autre difficulté se pose fréquemment, qui tient au fait que les données sur les soumissions (dans le cadre des appels d'offres) peuvent ne pas être lisibles par machine et ne peuvent donc pas être entrées directement dans des outils numériques de filtrage sans traitement supplémentaire<sup>19</sup>. En outre, les données sur les soumissions ne sont parfois pas enregistrées dans un format cohérent ou contiennent des erreurs<sup>20</sup>. À titre d'exemple, des données comme les dénominations ou raisons sociales des entreprises peuvent ne pas être standardisées d'un ensemble de données à l'autre, ce qui peut rendre difficile de rapprocher ces ensembles de données. Le nettoyage manuel de ces variables (par exemple, afin d'identifier quand deux noms d'entreprises légèrement différents se rapportent à la même entreprise) pourrait être effectué en extrayant, structurant et nettoyant manuellement ces données pour former un ensemble de données cohérent, mais cela peut être une activité à forte intensité de ressources. Par exemple, l'autorité espagnole de la concurrence, la CNMC, construit sa propre base de données sur la passation des marchés publics, sur la base de données téléchargées à partir de la plateforme nationale de passation de ces marchés, qu'elle filtre et dont elle nettoie les erreurs (Campuzano, 2021<sup>[40]</sup>)<sup>21</sup>. Des algorithmes d'appariement peuvent automatiser utilement un grand nombre de ces travaux, ce qui permet d'économiser du temps et des ressources (Fazekas et al., 2022<sup>[21]</sup>). À titre d'exemple, des algorithmes d'appariement peuvent créer un score de similitude basé sur le nombre de caractères que deux noms d'entreprises différents peuvent avoir en commun, en limitant la supervision humaine aux seules entreprises qui ont un score de similitude élevé (Bharadwaj et al., 2022<sup>[41]</sup>).

71. En outre, il est nécessaire de disposer d'une base de données suffisamment large pour que le filtrage soit efficace. Les systèmes nationaux de dématérialisation des marchés publics ont parfois une

couverture vaste mais contiennent peu d'informations, c'est-à-dire peu de variables sur lesquelles tester le filtre. Par exemple, certaines plateformes numériques nationales n'enregistrent que les soumissionnaires gagnants. La plupart des bases de données européennes sur les marchés publics ne saisissent que les phases de soumission et d'évaluation des risques, mais ne contiennent aucune information sur l'exécution du marché ou ses modifications (Fazekas et al., 2022<sup>[21]</sup>).

72. Si des données cruciales pour un filtre font défaut dans un ensemble de données spécifique, pour l'une des raisons précitées, les résultats du filtrage risquent de ne pas être fiables ou utiles, ou, en tout état de cause, de ne pas être aussi utiles qu'elles pourraient l'être si les bonnes variables y étaient incluses. Fazekas et Tóth ont proposé une « liste complète idéale » des variables pour le filtrage des ententes (Tableau 2)<sup>22</sup>.

**Tableau 2. Liste complète idéale des variables pour le filtrage des ententes en matière de passation des marchés publics**

Groupe de variable	Variable	Inclus dans l'annonce		
		CFT	CA	CC
<b>Acheteur</b>	Nom de l'acheteur	●	●	●
	Département/service de l'acheteur	●	●	●
	ID unique de l'acheteur	●	●	●
	Adresse de l'acheteur	●	●	●
	Type d'acheteur	●	●	●
<b>Soumissionnaire/Offres</b>	Nom du soumissionnaire		●	●
	ID/ID fiscal du soumissionnaire		●	●
	Adresse du soumissionnaire		●	●
	Nombre d'offres soumises		●	
	Nombre d'offres exclues		●	
	Prix de l'offre (détails sur le prix total et les prix unitaires)		●	●
	Date exacte de la soumission de l'offre		●	
	Type de l'offre (offre gagnant/perdant)		●	
<b>Appel d'offres/Marché</b>	Bénéficiaires effectifs		●	●
	Type de procédure	●	●	
	Contrat-cadre (1ère/2ème étape)	●	●	
	Critères d'adjudication	●	●	
	Seuil (inférieur/supérieur aux seuils de l'UE ?)	●	●	
	Prix estimé (détails sur le prix total ou les prix unitaires)	●	●	
	Type de marché (prestation de services, fournitures, travaux)	●	●	●
	Codes CPV (% de la valeur du marché par produit)	●	●	●
	Code(s) NUTS de l'exécution du marché	●	●	●
Statut (annulé, en cours, etc.)	●	●	●	
<b>Dates</b>	Date de publication de l'avis d'appel d'offres	●	●	●
	Date limite de soumission des offres	●		
	Dates de début et de fin du marché	●	●	●
	Date de publication de l'adjudication du marché		●	
	Date de signature du marché			
	Date de publication de l'achèvement du marché			●
<b>Sous-traitance</b>	Nom et ID unique (ID fiscal) du sous-traitant		●	●
	Part du sous-traitant		●	●
<b>Groupement</b>	Nom et ID unique (ID fiscal) des membres du groupement		●	●
	Part des membres du groupement		●	●
<b>Exécution du marché</b>	Date de fin de l'exécution du marché			●
	L'exécution a-t-elle été conforme au marché ?			●

Explication si l'exécution n'a pas été conforme au marché			●
Informations sur la modification du marché			●
Informations sur la qualité de l'exécution			●

Note : CFT= appel d'offres, CA= adjudication du marché, CC= achèvement du marché  
Source : (Fazekas and Tóth, 2016<sup>[22]</sup>)

73. Ces listes de variables peuvent être utiles afin de promouvoir la concurrence, en faisant mieux connaître les variables et le format des données que les autorités chargées de la passation des marchés publics ont besoin de collecter pour faire appliquer le droit de la concurrence. Outre cette sensibilisation, il serait plus sûr que les gouvernements, ou des autorités déléguées à la passation des marchés publics ou à la surveillance des marchés publics, publient des orientations sur la soumission et le stockage de données sur les marchés publics, par exemple en imposant l'utilisation de plateformes numériques respectant des normes de qualité prédéfinies<sup>23</sup>, et en garantissant la collecte de données pertinentes d'une manière uniforme dans des champs de données prédéterminés. Contrôler la soumission, le contenu et la qualité des données au point d'entrée des données aiderait à recueillir des informations fiables, à jour et, élément important, comparables (Fazekas and Tóth, 2016<sup>[22]</sup>)<sup>24</sup>.

74. De la même manière, il serait utile que les gouvernements prennent des mesures (y compris en adoptant une législation à cet effet) afin de garantir que les autorités de régulation qui détiennent des données sectorielles, et les autorités publiques chargées de la passation des marchés publics, qui détiennent des données sur les appels à la concurrence, partagent ces données avec les autorités de la concurrence. Ce partage des données peut être autorisé sous réserve de protections appropriées en matière de confidentialité, afin de garantir la sécurité du stockage, de l'accès et de l'utilisation des données, et à condition qu'aucune autre divulgation ne soit faite autrement qu'à des fins licites définies.

75. Dans le cadre des projets de lutte contre les soumissions concertées dans les marchés publics menés par l'OCDE dans différents pays, le Secrétariat a analysé différents systèmes nationaux de passation des marchés publics et fourni des recommandations sur les moyens de constituer des bases de données sur les marchés publics qui soient à la fois complètes et pertinentes pour les autorités chargées de la passation de ces marchés et les autorités de la concurrence. Le Secrétariat a recommandé de bonnes pratiques pour la collecte, la qualité et le stockage des données et l'accès à celles-ci, et a suggéré des variables permettant d'analyser les systèmes de soumissions concertées (Encadré 5).

### Encadré 5. Conditions requises pour que les bases de données sur les marchés publics soient utiles dans l'optique du droit de la concurrence

#### L'expérience de l'OCDE dans les projets de lutte contre les soumissions concertées dans les marchés publics

Les orientations suivantes visent à garantir que les bases de données sur les marchés publics soient utiles afin d'appliquer la loi réprimant les soumissions concertées :

**Ciblage des données.** Les données doivent être adaptées à l'objet de l'analyse proposée. L'identification préalable de cet objet informe sur le type et le format des données collectées. Les autorités responsables de la passation des marchés publics devraient se concerter avec les autorités de la concurrence afin d'élaborer des bases de données sur les marchés publics qui puissent être utilisées afin de détecter des cas potentiels de collusion. À titre indicatif, toutes les soumissions contenant des données au niveau de l'entreprise et des métadonnées doivent être enregistrées.

**Qualité des données.** Il est de la plus haute importance de disposer de données de bonne qualité afin de produire des résultats utiles pouvant être interprétés correctement. L'entrée des données et les

méthodes de validation doivent garantir l'enregistrement des données d'une manière standardisée, cohérente et exempte d'erreurs. À titre d'exemple, les champs relatifs à la tarification et aux prix unitaires doivent être uniformes ; les champs de texte et les conventions de dénomination doivent être définis ; et des contrôles des anomalies de codage doivent être intégrés dès le stade d'entrée des données.

**Exploitabilité des données.** Les informations doivent être stockées dans un format consultable qui permette de les traiter et de les utiliser aisément (par exemple, dans des tableurs ou des bases de données, plutôt que des images scannées des marchés), de telle sorte que les filtres nécessaires et les techniques analytiques puissent être aisément appliqués. Les bases de données tenues par les différentes autorités publiques devraient être interoperables, en termes de formatage et de référencement croisé, afin de permettre de les combiner et de les filtrer à la recherche d'indices de soumissions concertées.

**Accès aux données.** Les droits d'accès aux bases de données devraient être clairement définis, à la fois en termes de données d'entrée (bases de données centralisées ou décentralisées) et de données de sortie. Les autorités de la concurrence devraient avoir accès aux bases de données pour pouvoir mener leurs actions d'application de la loi.

Note : Pour plus d'informations sur les travaux par pays que l'OCDE a consacrés à la lutte contre les soumissions concertées dans les marchés publics, voir [www.oecd.org/daf/competition/oecdrecommendationonfightingbidrigginginpublicprocurement.htm](http://www.oecd.org/daf/competition/oecdrecommendationonfightingbidrigginginpublicprocurement.htm)  
Source : OCDE (2018), Fighting Bid Rigging in IMSS Procurement: Impact of OECD Recommendations, pp. 167-170, [www.oecd.org/daf/competition/fighting-bid-rigging-in-imss-procurement-impact-of-oecd-recommendations.htm](http://www.oecd.org/daf/competition/fighting-bid-rigging-in-imss-procurement-impact-of-oecd-recommendations.htm)

76. La révision prochaine de la Recommandation sur la lutte contre les soumissions concertées dans les marchés publics (OCDE, 2012<sup>[42]</sup>) prévoit de conseiller d'utiliser des systèmes électroniques de passation des marchés publics, depuis la publication de l'ouverture de l'appel d'offres jusqu'à la clôture du contrat. Cela permettrait aux autorités de la concurrence de retracer et d'évaluer toutes les actions liées à la passation des marchés, en tant que de besoin. La Recommandation révisée est également susceptible de donner des orientations sur le niveau d'informations que doivent contenir les bases de données afin de permettre la détection des schémas de soumissions concertées (notamment des données sur les offres gagnantes et perdantes et la sous-traitance), et de recommander de permettre aux autorités de la concurrence d'avoir accès à ces bases de données.

### 3.3. Le filtrage exige des connaissances et compétences spécialisées

77. Les autorités de la concurrence ont commencé à investir dans les compétences et les connaissances numériques, y compris pour le développement de filtres numériques d'ensembles de données. Alors que le développement de ces filtres reposait initialement sur des économistes, la plupart des autorités de la concurrence emploient désormais des spécialistes de la technologie, tels des experts en mégadonnées et des informaticiens, qui travaillent avec des économistes pour analyser des données et développer des filtres. L'engagement croissant d'experts en mégadonnées et d'informaticiens par les autorités de la concurrence est comparable à l'engagement d'économistes dans les années 80 (Lianos, 2021, p. 17<sup>[37]</sup>).

78. Un grand nombre d'autorités de la concurrence se sont dotées d'unités dédiées aux données, qui ont pour tâche de collecter et de nettoyer les données et d'assister les équipes travaillant sur des affaires concernant les données et marchés numériques. Dans certains cas, ces unités participent également à l'élaboration d'outils de filtrage. D'autres autorités n'ont pas d'unités dédiées mais ont engagé des directeurs des nouvelles technologies, et/ou ont intégré un personnel chargé des données massives, de l'intelligence artificielle et de l'apprentissage automatique dans d'autres unités ou équipes (Hunt, 2022<sup>[38]</sup>).

79. En 2021, la Commission hellénique de la concurrence a publié un rapport sur le droit de la concurrence et l'économie à l'ère de l'informatique, qui énumère les évolutions intervenues au sein des autorités de la concurrence dans les pays membres de l'OCDE et les pays BRICS (Brésil, Russie, Inde, Chine et Afrique du Sud), et contient un tableau détaillé, donnant des informations sur les équipes ou unités spécialisées dans les technologies numériques et judiciaires, et sur l'emploi de directeurs des nouvelles technologies et d'experts en mégadonnées (Lianos, 2021, pp. 119-148<sup>[37]</sup>).

## Encadré 6. Exemples d'équipes chargées des données au sein des autorités de la concurrence

### Royaume-Uni : la DaTA team de la Competition and Markets Authority

En 2019, la Competition & Markets Authority britannique (CMA) a créé une unité spécialisée dénommée DaTA (Data, Technology and Analytics) et engagé un Chief Data & Technology Insight Officer pour la diriger. La DaTA emploie actuellement près de 50 personnes couvrant les disciplines suivantes : science des données, ingénierie, expertise technologique, science comportementale, eDiscovery et criminalistique numérique. Elle fournit des conseils d'expert en matière de données et de technologie, et pilote l'acquisition, le traitement et l'utilisation de données. Elle procède également à l'extraction de données et développe des pipelines de données.

DaTA développe actuellement un pipeline de données extraites des dossiers de toutes les sociétés anonymes immatriculées auprès du Companies House, c'est-à-dire le registre des sociétés du Royaume-Uni. La CMA a fréquemment besoin de ces données pour de nombreuses raisons, y compris i) obtenir des informations sur des sociétés soupçonnées de participer à une entente, ii) comprendre les structures de propriété sur un marché et iii) comprendre l'état d'un marché, en particulier les niveaux de concentration et de rentabilité. Habituellement, le personnel de la CMA utilise l'outil de recherche publiquement disponible et télécharge ensuite manuellement les données collectées grâce à cette navigation, ce qui prend beaucoup de temps et peut conduire à des erreurs. Le pipeline intégrera régulièrement toutes les données, les dédupliquera et les rendra disponibles dans un outil conçu pour les besoins de la CMA.

Source : Hunt, S. (2022), The technology-led transformation of competition and consumer agencies: the Competition and Markets Authority's experience, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1082753/Stefan\\_discussion\\_paper.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082753/Stefan_discussion_paper.pdf).

### Grèce : l'unité criminalistique informatique de la Commission hellénique de la concurrence

La Commission hellénique de la concurrence (CHC) a créé une unité criminalistique informatique en octobre 2020, qui est dirigée par un économiste et coopère avec plusieurs experts en mégadonnées, qui interviennent en tant qu'experts externes pour la Commission.

La CHC est également sur le point de mettre en place une plateforme d'infrastructure de gestion des données massives, conçue sur mesure par un contractant externe, qui permettra de télécharger automatiquement des données publiques en temps réel émanant de sources différentes (observatoire des prix des supermarchés, prix des carburants, prix des fruits et légumes, données sur les marchés publics, etc.). La CHC a nommé des experts pour concevoir un programme d'extraction de données brutes à partir d'informations non structurées disponibles sur Internet au format pdf et dans d'autres formats, et de les extraire sous forme de fichiers éditables au format csv (valeurs séparées par des virgules). Les données doivent principalement servir à la détection des ententes, mais fournissent également un environnement intégré d'analyse des données, grâce à plusieurs outils sur mesure ou logiciels standard qui permettent de visualiser et d'analyser des données. La CHC a également engagé des prestataires de services afin de développer un modèle de données et des tableaux de bord intégrés, ainsi que des progiciels sur mesure.

Source : Lianos, I (2021), Computational law and economics: an inception report, <https://www.epant.gr/en/enimerosi/computational-competition-law-and-economics.html>

80. Il est important de noter que bien que la science des données soit nécessaire pour le filtrage des données, la décision à propos des variables à inclure et de la forme qu'elles doivent prendre doit également impliquer la participation d'économistes et des responsables chargés des affaires en cause (Abrantes-Metz and Metz, 2018<sup>[23]</sup>). Il est indispensable d'avoir des connaissances et de l'expérience dans la reconnaissance et le traitement des comportements illégaux pour pouvoir choisir la bonne méthode de filtrage, définir les paramètres des outils de filtrage et évaluer les résultats du filtrage<sup>25</sup>.

### 3.4. La coopération internationale en matière de partage des données et de développement de filtres

81. Le développement de filtres numériques exige des autorités de la concurrence qu'elles investissent aujourd'hui, sachant qu'elles récolteront les fruits de cet investissement plus tard. Pour avoir du sens, cet investissement doit donc être fait à long terme, afin de développer, de maintenir et d'améliorer des outils et des méthodes.

82. Cet investissement à long terme bénéficierait de la coopération entre les autorités de la concurrence (Hunt, 2022<sup>[38]</sup>). En effet, il est fréquent que les autorités acquièrent des compétences, construisent des ensembles de données et développent des filtres de données en parallèle. La coopération pourrait permettre d'économiser du temps et des ressources, et elle est essentielle étant donné que les ententes deviennent plus sophistiquées. Les autorités peuvent se partager leur expertise technique et leur expérience, et, s'il existe de bonnes relations de travail et un climat de confiance, le code<sup>26</sup> et, sous certaines conditions, des données pour entraîner des modèles de filtrage. Lianos (2021<sup>[37]</sup>) estime que les autorités de la concurrence devraient se partager les données sur les ententes qu'elles ont acquises grâce aux filtres existants, afin de créer un ensemble de données contenant des exemples de collusion, pour les filtres qui utilisent des méthodes d'apprentissage automatique. Tout nouveau comportement suspect identifié grâce à cette analyse de l'ensemble de données pourrait être partagé afin d'aider les outils de filtrage à devenir plus perfectionnés et plus fiables. Des recherches sont en cours afin de déterminer l'utilité d'entraîner des algorithmes de filtrage des ententes en utilisant des données sur les ententes provenant d'autres juridictions (Huber, Imhof and Ishii, 2020<sup>[43]</sup>).

83. Étant donné que le partage des données peut poser des problèmes de sensibilité et de confidentialité des données et est donc susceptible de se heurter à des contraintes légales, l'anonymisation des données ou la création d'ensembles de données synthétiques (« *une approche de la confidentialité, selon laquelle, plutôt que de diffuser des données réelles, seules des données synthétiques générées par un ou plusieurs modèles de population sont diffusées* ») (OCDE, n.d.<sup>[44]</sup>) pourrait rendre le partage plus facile<sup>27</sup>. Les données synthétiques gagnent du terrain dans le domaine de l'apprentissage automatique (Contrôleur européen de la protection des données, n.d.<sup>[45]</sup>). Le partage d'algorithmes entraînés serait plus facile, étant donné qu'il ne révélerait pas, en principe, des données sous-jacentes sensibles<sup>28</sup>.

84. La coopération internationale peut soutenir le développement conjoint de logiciels de filtrage. La DCCA a développé un nouvel outil de filtrage, baptisé Bid Viewer, basé sur l'apprentissage automatique. La DCCA a coopéré à cet effet avec les autorités de la concurrence espagnole et suédoise (entre autres).

### Encadré 7. L'outil Bid Viewer de l'Autorité danoise de la concurrence et de la consommation

L'Autorité danoise de la concurrence et de la consommation (ADCC) a travaillé sur les méthodes algorithmiques et développé un logiciel de filtrage pour la détection des soumissions concertées, dénommé « Bid Viewer ». Cet outil utilise des indicateurs statistiques, l'apprentissage automatique, l'analyse des schémas de soumission des entreprises et des modèles de réseaux neuronaux créés par l'intelligence artificielle.

Les indicateurs statistiques comprennent : (i) la distance relative normalisée ; (ii) la différence en pourcentage ; (iii) le coefficient de variation ; (iv) le coefficient d'asymétrie ; et (v) le coefficient d'aplatissement. Les indicateurs sont testés sur les prix des offres soumises. Bid Viewer combine les indicateurs statistiques en utilisant l'apprentissage automatique.

Le modèle d'apprentissage automatique est formé sur un ensemble de données relatives à des ententes avérées. Le modèle est ensuite testé sur un second ensemble de données relatives à des ententes avérées. Si l'analyse du second ensemble de données est fiable, le modèle est alors appliqué aux données réelles sur la passation des marchés et signale les offres présentant un risque élevé de collusion.

Bid Viewer teste à la fois le partage d'un marché géographique et d'un marché non géographique (en identifiant des entreprises qui ne soumissionnent pas dans le cadre des mêmes appels d'offres). Bid Viewer teste également l'existence de fausses soumissions, c'est-à-dire des offres qui sont intentionnellement trop élevées et sont donc destinées à perdre. Il est possible de calculer le nombre prévisionnel de soumissions gagnantes d'une entreprise donnée en utilisant le nombre d'appels d'offres auxquels elle participe et le nombre de concurrents participant à ces appels d'offres. Bid Viewer compare le nombre de soumissions gagnantes effectivement remportées par l'entreprise par rapport au nombre prévisionnel qu'elle était censée remporter. Les entreprises sont signalées si elles se situent trop au-delà ou en-deçà de ces valeurs.

Dans une récente présentation au cours d'une réunion de l'OCDE, l'ADCC a souligné l'importance de disposer de données au niveau de l'offre. Les analyses de filtrage ne seraient pas possibles sans des informations sur les offres gagnantes et perdantes : les offres gagnantes ne pourraient pas, à elles seules, montrer si deux entreprises n'ont jamais participé au même appel d'offres. Idéalement, les données devraient inclure les offres retirées et les offres non valides.

Bid Viewer a été mis à la disposition d'autres autorités nationales de la concurrence, y compris les autorités espagnole et suédoise. La Competition and Markets Authority britannique est sur le point de commencer à l'utiliser.

Source : Compte rendu succinct de la 134<sup>e</sup> réunion du Groupe de travail n° 3 sur la coopération et l'application de la loi (30 novembre 2021), DAF/COMP/WP3/M(2021)2 (accès réservé aux Membres et Partenaires de l'OCDE) ;

Kultima, J. R. (2022) "Collusion detection in public procurement using computational methods", [https://www.en.kfst.dk/media/cnldn11q/bid-viewer\\_56\\_seneste.pdf](https://www.en.kfst.dk/media/cnldn11q/bid-viewer_56_seneste.pdf),

Hunt, S. (2022), The technology-led transformation of competition and consumer agencies: the Competition and Markets Authority's experience,

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1082753/Stefan\\_discussion\\_paper.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082753/Stefan_discussion_paper.pdf)

85. Les autorités de la concurrence peuvent s'échanger leurs connaissances et leur expertise dans le cadre de forums internationaux, comme l'OCDE et le Réseau international de la concurrence (RIC), et à l'occasion de conférences consacrées à l'utilisation de la science des données et de l'intelligence artificielle, particulièrement dans le cadre de la lutte contre les ententes. La conférence organisée par la

CMA sur le thème « Data, Technology and Analytics » (2022) a été l'une des premières conférences à permettre un forum d'échange entre les experts en mégadonnées et les experts en technologie.

86. Il convient de noter qu'un grand nombre de types de coopération (notamment le partage d'expérience, de connaissances et d'expertise) ne se heurteraient pas à des obstacles juridiques, étant donné qu'ils seraient normalement considérés comme le partage d'informations confidentielles des autorités concernées<sup>29</sup>. Plusieurs répondants à une enquête réalisée en 2019 auprès des membres de l'OCDE et du RIC sur la coopération internationale en matière d'application de la loi ont déclaré que leur législation n'imposait aucune restriction en matière de partage d'informations confidentielles des autorités et que ce partage pourrait être très utile. Certains répondants ont noté, dans le cas où les informations partagées seraient sensibles (auquel cas le partage de ces données serait réputé sensible), qu'il faudrait un haut degré de confiance entre les autorités, ainsi que la connaissance des procédures et processus respectifs de traitement des informations sensibles et confidentielles, sans oublier une garantie de l'autorité receveuse de n'utiliser les informations qu'aux fins auxquelles elles sont destinées (OCDE/RIC, 2021<sup>[46]</sup>).



# 4. La valeur du filtrage

87. En règle générale, le résultat d'un filtrage qui identifie un comportement suspect ne suffit pas, en soi, à prouver une violation du droit de la concurrence<sup>30</sup>. Réciproquement, une analyse de filtrage qui ne repère pas un problème ne signifie pas qu'il n'existe aucun comportement illégal.

88. Les résultats du filtrage peuvent étayer la décision d'ouvrir des enquêtes et de prioriser certains dossiers plus souvent et plus facilement, plutôt que d'aider à statuer sur une affaire sur la seule base des preuves économiques indirectes qu'ils fournissent. La raison principale en est que la norme conduisant à ouvrir une enquête, ou à prendre des mesures d'enquête comme des perquisitions (même si elles supposent de convaincre un juge de délivrer un mandat de perquisition - Encadré 8), est généralement moins élevée que la norme en matière de preuve afin d'obtenir une décision finale de constatation de l'infraction, puisqu'il s'agit alors de rapporter une preuve suffisante des faits sur lesquels cette décision se fonde et de pouvoir convaincre le tribunal saisi de l'affaire.

## 4.1. Étayer la décision d'ouvrir ou de clôturer une procédure

89. Les autorités peuvent s'appuyer sur les résultats du filtrage pour décider d'ouvrir une enquête, et utiliser leurs outils d'enquête habituels, - perquisitions et demandes d'information (Heijnen, Haan and Soetevent, 2015<sub>[47]</sub>) – afin de trouver une meilleure preuve, et, idéalement, une preuve directe, du comportement suspecté (OCDE, 2013<sub>[5]</sub>). Il est néanmoins toujours nécessaire d'analyser soigneusement les résultats du filtrage avant d'ouvrir une enquête, afin d'éviter des biais qui conduiraient à engager des procédures avec une faible probabilité de découvrir un comportement illégal.

90. L'Encadré 8 décrit une procédure couronnée de succès qui a été ouverte par l'autorité suisse de la concurrence, COMCO, sur la base de résultats de filtrage, ainsi qu'une affaire brésilienne dans laquelle le juge compétent a délivré un mandat de perquisition sur la foi des conclusions du filtrage (alors que les juges brésiliens n'accordent généralement des mandats de cette nature que dans les cas où une demande de clémence a été présentée).

### Encadré 8. Exemples de cas dans lesquels des filtres d'ententes ont conduit à l'ouverture d'enquêtes

#### L'entente pour la construction de routes dans la région de See Gaster en Suisse

L'autorité suisse de la concurrence (COMCO) a développé un outil de filtrage des ententes, dans le cadre d'un vaste projet à long terme de lutte contre les soumissions concertées. La COMCO a utilisé cet outil pour mettre au jour une entente de soumissions concertées dans le secteur de la construction de routes dans la région de See Gaster en Suisse, en vertu de laquelle les participants s'étaient entendus sur les prix de leurs offres et déterminé qui devrait remporter les marchés entre 2002 et 2009.

La COMCO a initialement utilisé deux filtres simples : le coefficient de variation (déviation standard des offres, divisée par la moyenne des offres pour un appel d'offres donné) et la distance relative (différence entre les deux offres les plus basses, divisée par la déviation standard des offres perdantes pour un

appel d'offres donné). Toutefois, ces deux filtres simples n'ont pas fourni de preuves concluantes. La COMCO a réalisé que cela pouvait être dû à une collusion partielle, c'est-à-dire une collusion qui n'impliquait pas toutes les entreprises ou tous les appels d'offres, et elle a donc développé un filtre pour identifier une collusion partielle.

La COMCO a appliqué une procédure en quatre étapes pour détecter une collusion partielle. Lors de la première étape, elle a identifié des offres suspectes qui présentaient un faible coefficient de variance et une distance relative élevée. Dans la seconde étape, elle a identifié s'il existait un groupe d'entreprises qui soumettaient fréquemment des offres pour ces marchés suspects. Au cours de la troisième étape, elle a examiné la couverture de ces offres suspectes afin de déterminer s'il y avait d'autres entreprises actives dans la région et si la collusion était susceptible d'être stable. La quatrième étape a consisté à appliquer un filtre pour identifier la rotation des offres (c'est-à-dire les cas dans lesquels le soumissionnaire gagnant soumet une offre basse, tandis que les autres soumissionnaires soumettent des offres élevées destinées à perdre), en examinant la distribution des offres normalisées entre des paires d'entreprises suspectes.

La COMCO a ouvert une enquête sur la base des résultats du filtrage. C'est la première fois qu'une analyse statistique de données ex ante a conduit à ouvrir une procédure. L'enquête a duré entre avril 2013 et juillet 2016 et a abouti à condamner huit entreprises à des amendes d'un montant total d'environ 5 millions CHF.

Source : Imhof, D., Y. Karagök et S. Rutz (2018), "Screening for Bid Rigging—Does It Work?", *Journal of Competition Law & Economics* 14(2), pp. 235-261, <https://academic.oup.com/jcle/article-pdf/14/2/235/25640157/nhy006.pdf>. Wallimann, H., D. Imhof et M. Huber (2020), "A machine learning approach for flagging incomplete bid-rigging cartels", arXiv preprint arXiv:2004.05629, <https://arxiv.org/pdf/2004.05629.pdf>. COMCO (2018), OECD-BWB Workshop on Complex Cartel Case Management, [www.slideshare.net/OECD-DAF/swiss-competition-commission-on-cartel-detection-and-screening-in-public-procurement](http://www.slideshare.net/OECD-DAF/swiss-competition-commission-on-cartel-detection-and-screening-in-public-procurement)

### **Le marché « protection incendie » au Brésil**

En septembre 2021, la Direction générale de la CADE a publié une note technique concernant un cas de soumissions concertées dans le cadre du marché « lutte et prévention contre les incendies/brigade de pompiers » (Procédure administrative de la CADE n° 08700.004914/2021-05). Le Tribunal administratif de la CADE, qui doit statuer sur cette affaire, n'a pas encore rendu sa décision.

Dans leur réponse à l'appel d'offres portant sur des équipements de protection incendie pour 33 entités du secteur public, deux soumissionnaires ont fourni la même adresse de messagerie électronique de contact (qui était l'adresse de l'un d'eux). L'autorité responsable de la passation du marché a noté ce fait, puis identifié une personne qui travaillait pour les deux entreprises et avait également des liens avec un plus grand nombre d'entreprises sur le même marché. La CADE a ensuite procédé à une analyse empirique afin d'identifier des comportements de soumissions suspects et découvert des indices de collusion. Plus précisément, la CADE a filtré les appels d'offres dans lesquels les entreprises suspectes liées par cette personne commune avaient soumissionné, identifié tous les concurrents participant à ces appels d'offres « suspects » et filtré tous les appels d'offres dans lesquels ces concurrents « suspects » avaient soumissionné. La CADE a progressivement élargi la couverture de son ensemble de données, et le nombre de participants possibles à l'entente.

L'analyse empirique a consisté en filtres comportementaux, y compris l'offre moyenne, la variance moyenne des offres, le coefficient de variation des offres, le nombre moyen de participants à chaque appel d'offres et le nombre de fois où une entreprise a soumis une offre aberrante. L'analyse a également comporté des Tests de Benford et d'Entropie, et l'analyse des groupes d'entreprises afin d'identifier les groupes qui ont interagi plus fréquemment et de comparer les résultats des filtres comportementaux pour ces groupes.

Le Test de Benford identifie d'éventuelles soumissions fictives ou de couverture. Il s'appuie sur la loi de Benford qui est une « *formule mathématique [qui] décrit la distribution régulière des nombres* » (OCDE, 2013, p. 55<sup>[5]</sup>). Les valeurs des soumissions qui s'écartent de la distribution décrite par la loi de Benford peuvent être un signe de soumissions fictives ou de couverture. Le Test d'Entropie est une analyse de classement visant à identifier des cas extrêmes de rotation des soumissionnaires. Ils comprennent les « éternels gagnants » qui remportent toujours les appels d'offres, ou les « éternels perdants », qui perdent toujours. Ces résultats des soumissions peuvent être des signes de collusion étant donné que, dans des contextes concurrentiels, le classement des soumissions varie normalement.

Les filtres appliqués dans cette affaire ne le sont pas systématiquement par la CADE. Ils ont été appliqués à la suite d'un signal d'alerte initial (même adresse électronique, une personne en commun) et ont fourni une preuve utile pour la détection de l'entente et l'ouverture de l'enquête. La CADE a sollicité et obtenu un mandat judiciaire afin de réaliser des perquisitions sur la base des résultats du filtrage. Ce cas est exceptionnel, puisque les juges brésiliens n'accordent généralement des mandats de perquisition qu'à la suite de demandes de clémence, étant donné que ces perquisitions sont jugées intrusives. En l'occurrence, les résultats du filtrage et le forage des données ont été convaincants. La CADE a réalisé des perquisitions dans les locaux de 14 entreprises, qui ont fourni la preuve que des entreprises avaient effectivement échangé des informations concurrentielles sensibles.

Source : [Technical note N° 7/2021/SG-TRIAGEM CONDUTAS/SGA2/SG/CAD](#) ; Schrepel, Thibault et Groza, Teodora, The Adoption of Computational Antitrust by Agencies: 2021 Report (21 juin 2022). 2 Stanford Computational Antitrust, 78 (2022), SSRN: <https://ssrn.com/abstract=4142225>

91. Les filtres aident à prioriser les affaires afin de se concentrer sur celles qui paraissent les plus fondées ou présentent une meilleure chance de succès, et sur les affaires closes qui risquent d'être définitivement classées. À titre d'exemple, l'autorité brésilienne de la concurrence (CADE), ayant reçu un grand nombre de plaintes pour comportement anticoncurrentiel sur le marché de la vente au détail de fuel, a développé un filtre afin de séparer les affaires qui semblent mériter une enquête plus poussée de celles auxquelles il semble moins justifié de donner suite (Ragazzo, 2012<sup>[48]</sup>).

## Encadré 9. Prioriser les affaires au moyen de filtres

### Secteur de la vente au détail de fuel au Brésil

L'autorité brésilienne de la concurrence (CADE) a développé un filtre afin d'identifier celles des nombreuses plaintes reçues dans le secteur de la vente de fuel au détail au Brésil auxquelles il serait le plus utile de donner suite.

La méthodologie de filtrage incluait trois tests statistiques : (i) l'évolution des marges bénéficiaires des détaillants de la ville dans laquelle l'entente opérait prétendument ; (ii) la corrélation entre les marges des détaillants et le coefficient de variation (niveau de dispersion des prix) pour cette ville ; et (iii) la corrélation entre la marge bénéficiaire des détaillants de la ville, comparée à celle de l'État (de la région) concerné.

La méthodologie partait du postulat que les trois tests seraient effectués et que l'analyse pourrait donner l'un des deux résultats suivants : (i) aucune probabilité d'entente (c'est-à-dire, réduction des marges bénéficiaires des détaillants de temps à autre, association positive entre les hausses de marges et la variabilité des prix, et marges des détaillants évoluant d'une manière similaire à la moyenne observée dans l'État concerné) ; ou (ii) probabilité d'entente (c'est-à-dire, hausse des marges bénéficiaires des

détaillants au fil du temps ; association négative entre les marges des détaillants et la variabilité des prix, et marges bénéficiaires traduisant une évolution disparate ou une évolution non similaire à la moyenne observée dans l'État concerné).

Le filtre a été appliqué sur un très grand volume de données existantes pour le marché de la vente au détail de fuel, collectées par l'Agence nationale des pétroles, et a permis de rejeter des plaintes sans fondement. Dans un petit nombre de cas, les résultats du filtrage ont détecté un possible comportement d'entente.

Source : Ragazzo, C. (2012), Screens in the gas retail market : the Brazilian experience, <https://www.competitionpolicyinternational.com/screens-in-the-gas-retail-market-the-brazilian-experience/>.

## 4.2. Fournir des preuves à l'appui d'une décision constatant l'infraction

92. Les résultats du filtrage et les indications de violations qu'ils peuvent détecter ne constituent pas la preuve directe d'une activité illégale. Ce sont uniquement des indications, qui doivent être interprétées pour parvenir à une conclusion, et elles peuvent servir de preuve indirecte. La Section 0 mentionne des circonstances dans lesquelles des filtres échouent à distinguer entre des faux positifs et une conduite illégale, par exemple lorsqu'un filtre identifie une corrélation de prix mais ne parvient pas à opérer la distinction entre une collusion explicite (illégale) et une collusion tacite (légale) ou une coïncidence.

93. Dans la plupart des juridictions et la plupart des cas, à tout le moins en matière d'ententes, la preuve directe d'une coordination explicite est requise pour conclure définitivement à l'existence d'une infraction. Il est difficile de se fonder sur des preuves économiques obtenues grâce au seul filtrage, étant donné, en particulier, que la responsabilité potentielle encourue pour avoir violé les dispositions anti-ententes du droit de la concurrence est généralement lourde. Dans les pays où les ententes donnent lieu à des poursuites pénales, l'infraction au droit de la concurrence doit être prouvée au-delà de tout doute raisonnable, ou selon une norme de preuve aussi élevée. En conséquence, il est encore plus important, dans ces cas, de rapporter la preuve directe d'un accord d'entente pour justifier une action pénale, afin de respecter la norme de preuve requise et de convaincre le tribunal. Toutefois, la preuve indirecte est admissible et utile, y compris dans le cas d'une procédure pénale (OCDE, 2006<sup>[49]</sup>)<sup>31</sup>.

94. L'analyse économique peut donc jouer un rôle de soutien, mais n'est généralement pas la preuve qui va déterminer la décision dans une affaire (Harrington, 2006<sup>[7]</sup>). Néanmoins, pratiquement tous les pays qui ont fait une contribution écrite ou orale lors de la table ronde organisée par l'OCDE en 2006 sur les poursuites contre les ententes sans preuve directe d'un accord ont décrit au moins une affaire dans laquelle une preuve circonstancielle a été utilisée et retenue ; la preuve économique est l'un des deux types de preuves circonstanciels (l'autre étant la preuve de communications entre les membres de l'entente suspectée) (OCDE, 2006<sup>[49]</sup>).

95. Peu d'affaires ont été entièrement fondées sur les preuves économiques issues des filtres. L'Encadré 10 présente une affaire dans laquelle divers facteurs (modèles) de preuves économiques basées sur le filtrage ont été considérés dans leur ensemble, contribuant à brosser un tableau du comportement en cause, qui correspondait de manière convaincante à une collusion. Les preuves livrées par le filtrage ont été jugées suffisamment nombreuses, claires et décisives pour que les magistrats statuant en appel considèrent que les faits sur lesquels se fondait la décision de première instance étaient prouvés.

### Encadré 10. Marché public d'achat de médicaments au Mexique

L'autorité mexicaine de la concurrence, la CFC (désormais COFECE) a enquêté sur une entente dans le cadre de la passation d'un marché par l'Institut de la sécurité sociale mexicain (IMSS). L'enquête a démarré à la suite d'une plainte de l'IMSS, à propos de différents appels d'offres lancés entre 2003 et 2006, portant sur la fourniture d'insuline humaine et de solutions électrolytiques et intraveineuses.

La CFC a utilisé des filtres de prix sur des données fournies par l'IMSS. Elle a identifié des appels d'offres caractérisés par des prix d'adjudication identiques et une rotation des soumissionnaires gagnants, ainsi que des filtres portant sur les parts de marché et a découvert que les soumissionnaires détenaient des parts de marché similaires qui convergeaient au fil du temps. Les prix des offres ne semblaient pas correspondre aux coûts, et, alors que les membres de l'entente offraient en moyenne les mêmes prix avec un taux de variation mineur, la situation a changé lors de l'entrée d'un nouveau concurrent sur le marché, après quoi les prix ont diminué et leur dispersion a augmenté.

La CFC a conclu à l'existence d'une collusion, et a infligé des amendes aux entreprises pharmaceutiques Baxter, Fresenius, Eli Lilly et Pisa. La décision prise dans cette affaire s'est fondée sur les preuves économiques fournies par le filtrage. La CFC a également obtenu des preuves des communications intervenues entre les entreprises concernées, établissant qu'elles avaient eu des occasions d'interagir, et qu'elles se connaissaient, voyageaient ensemble et communiquaient.

La Cour suprême du Mexique a confirmé la décision de la CFC en 2015, confirmant ainsi la légitimité de l'analyse économique en tant que preuve d'une pratique anticoncurrentielle.

Source : Mena-Labarthe, C. (2012), "Mexican experience in screens for bid-rigging", *Competition Policy International Antitrust Chronicle*, March, <https://www.competitionpolicyinternational.com/assets/Cartel-Column-June-New-Format-Full.pdf> COFECE (2015), The Mexican Supreme Court of Justice upholds the decision of the Competition Authority concerning a bid-rigging in the pharmaceutical sector (Baxter, Fresenius, Eli Lilly and Pisa), [www.competitionpolicyinternational.com/assets/COFECE-009-2015-English.pdf](http://www.competitionpolicyinternational.com/assets/COFECE-009-2015-English.pdf)

96. Les techniques d'apprentissage automatique, qui permettent une classification des comportements pilotée par les données, libérée (ou plus libérée) de la formulation d'hypothèses de base et gagnant en précision au fil de son utilisation (Section 2.5), pourraient accroître la fiabilité des données fournies par le filtrage, et, partant, leur valeur probante. Cela dépendra en partie de la mesure dans laquelle les données fournies par le filtrage, et la manière dont elles ont conduit à un résultat donné, pourront être présentées d'une manière claire et convaincante à un tribunal. Il est probable que l'utilisation accrue des filtres numériques et des techniques d'analyse algorithmique conduira à élaborer une jurisprudence spécifique sur les normes de preuve correspondantes et, en particulier, sur la valeur des preuves obtenues grâce à l'analyse des données (Lianos, 2021<sup>[37]</sup>).

97. Dans la mesure où les résultats du filtrage concourent à une décision finale défavorable, on peut soutenir que les parties affectées ont un droit d'accès à la méthodologie de filtrage et aux méthodes de traitement des données (von Bonin Andreas, 2020<sup>[50]</sup>). Tout dépend des voies de recours de chaque juridiction, mais, d'une manière générale, l'utilisation de modèles économiques, sans explication de leur méthodologie ou hypothèses de base, peut exposer la décision constatant l'infraction à une annulation, si l'entreprise prouve qu'elle aurait pu se défendre contre ces informations, si elle y avait eu accès (OCDE, 2019<sup>[51]</sup>).

98. En conséquence, les autorités de la concurrence peuvent utiliser des filtres, pour autant que les hypothèses, la méthodologie et les résultats du filtrage soient suffisamment expliqués aux parties faisant l'objet de l'enquête, en particulier si les résultats de ce filtrage sont utilisés contre elles. Les parties devraient également avoir des opportunités appropriées de présenter leurs propres arguments sur le processus et l'interprétation du filtrage, avant qu'une décision finale ne soit prise à leur encontre. Cela

exigerait probablement un accès au code et aux données sous-jacentes du filtrage, moyennant des protections appropriées en termes de confidentialité, notamment une divulgation limitée à un cercle restreint de personnes tenues d'obligations spécifiques en matière de traitement des informations divulguées (OCDE, 2019<sup>[52]</sup>).

99. Cette divulgation serait conforme aux dispositions de la Recommandation de l'OCDE sur la transparence et l'équité procédurale dans la mise en œuvre du droit de la concurrence (OCDE, 2021<sup>[53]</sup>) qui exige « [d'offrir] *aux parties l'opportunité de présenter une défense adéquate avant le prononcé d'une décision finale* », y compris « [en leur donnant] *accès aux preuves pertinentes collectées par, ou soumises à l'autorité de la concurrence ou au tribunal* » et « [en donnant] *aux parties une opportunité véritable de présenter une réponse complète aux allégations et de soumettre des preuves à l'appui de leurs arguments devant les décideurs clés* ».

100. Dans une affaire de contrôle des concentrations (Encadré 11), la Cour de Justice de l'Union européenne a jugé que la décision de la Commission européenne de bloquer l'opération de concentration, sans divulguer la méthodologie et les hypothèses du modèle économique qu'elle avait utilisé pour analyser les effets de cette opération, avait violé les droits de la défense des parties à l'opération. Bien que cet arrêt ait concerné une opération de concentration, l'argumentation de la Cour se fondait sur le respect des droits de la défense et s'appliquerait d'une manière similaire à une affaire d'entente.

### Encadré 11. Les méthodologies et hypothèses économétriques doivent être divulguées aux parties

#### Utilisation de modèles économétriques dans le contrôle des concentrations d'entreprises

En 2012, United Parcel Service, Inc. (« UPS ») a notifié à la Commission son projet d'acquisition de son concurrent TNT Express NV (« TNT »). Les deux entreprises opéraient sur les marchés des services internationaux de distribution expresse de petits colis. La Commission a bloqué cette opération de concentration en 2013.

En appel devant le Tribunal de l'Union européenne, UPS a allégué que la Commission avait violé ses droits de la défense, en se fondant sur un modèle économétrique de concentration des prix différent de celui qui avait été divulgué aux parties à l'opération pendant la procédure administrative. Le Tribunal a accueilli ce moyen et annulé la décision de la Commission. La Commission a formé un pourvoi contre cet arrêt devant la Cour de Justice de l'Union européenne.

La Cour a confirmé l'arrêt du Tribunal de l'Union européenne et rejeté le pourvoi de la Commission. La Cour a jugé « *que la requérante a suffisamment démontré [...] qu'elle aurait pu avoir une chance, même réduite, de mieux assurer sa défense* » si elle avait connu les détails du modèle économétrique sur lequel la décision de la Commission était fondée. La Cour a donc considéré que les droits de la défense de la requérante avaient été violés et qu'il convenait donc d'annuler la décision de la Commission.

Source : Affaire C-265/17 P United Parcel Service, Inc. c./ Commission

# 5. Conclusions

101. Les filtres d'ententes suscitent un intérêt croissant, largement dû à l'émergence d'un environnement numérique permettant leur application. De très grands volumes de données sont désormais accessibles et peuvent être traités. De nouvelles technologies permettent d'analyser ces données d'une manière de plus en plus automatisée, et ont permis d'élaborer de nouvelles méthodes de filtrage, en utilisant l'apprentissage automatique en particulier. De récentes publications universitaires, annonces des autorités de la concurrence et conférences indiquent que le filtrage numérique gagne en popularité.

102. Il n'existe pas beaucoup d'informations publiques sur l'utilisation et le succès des filtres numériques. Nous connaissons plusieurs affaires qui ont commencé sur la base des résultats d'un filtrage, par exemple celles qui se sont appuyées sur Cérebro (Encadré 1), l'outil de la COMCO (Encadré 8) et le système BRIAS (Encadré 2). Toutefois, étant donné que plusieurs autorités de la concurrence n'annoncent pas publiquement leurs initiatives de filtrage, le succès et l'impact réels du filtrage sont inconnus et ne peuvent pas être estimés facilement.

103. Deux choses sont certaines : le filtrage des données sur la passation de marchés publics a peu de chances d'être efficace s'il ne s'appuie pas sur un grand nombre de données de bonne qualité, d'une part, et le succès des filtres, ainsi que le temps et les efforts exigés pour leur déploiement, dépendent de la disponibilité de données, d'autre part. La CMA a abandonné son outil de filtrage des ententes en partie en raison de difficultés d'importation et d'enrichissement des données (Encadré 4), Cérebro s'appuie sur une extraction massive de données (Encadré 1), tandis que plusieurs autorités de la concurrence construisent leurs propres bases de données afin de pouvoir les utiliser à l'avenir pour la détection des ententes.

104. L'avenir du filtrage aura tout à gagner de la coopération entre les autorités de la concurrence et dépendra, dans une certaine mesure, de cette coopération. Il est fréquent que les autorités de la concurrence acquièrent des compétences, construisent des ensembles de données et développent des filtres de données en parallèle. La coopération pourrait donc permettre d'économiser du temps et des ressources. Les autorités peuvent partager leur expertise technique, leurs succès et leurs échecs, le code et, sous certaines conditions, des données. Dans la mesure où les ententes peuvent gagner en sophistication, la coopération internationale peut être la réponse.

## Annexe A. Littérature académique relative aux filtres d'ententes de 2015 à 2022

La littérature académique relative aux filtres d'ententes existe depuis plusieurs décennies. Les premières contributions importantes à ce sujet sont celles de (Porter and Zona, 1993<sup>[16]</sup>), (Porter and Zona, 1997<sup>[17]</sup>), (Porter and Zona, 1999<sup>[18]</sup>), (Bajari and Ye, 2003<sup>[19]</sup>), (Harrington, 2006<sup>[7]</sup>), (Abrantes-Metz et al., 2006<sup>[20]</sup>), (Bolotova, Connor and Miller, 2008<sup>[54]</sup>), (Abrantes-Metz and Metz, 2012<sup>[55]</sup>), et (Jiménez and Perdiguero, 2012<sup>[56]</sup>).

Cette littérature a été considérablement étoffée ces dernières années. Dans cette annexe figurent quelques-unes des contributions les plus importantes depuis 2015, présentées par approche. Les approches comprennent les méthodes basées sur la distribution des offres/prix, l'arrondi des offres, les méthodes basées sur la cointégration, l'indépendance conditionnelle en utilisant les fonctions d'offre, les doubles différences, les fractions d'offre historiques, les méthodes de réseau, les méthodes probabilistes, les méthodes de discontinuité de la régression, les méthodes de rupture structurelle, les regroupements suspects dans les données géographiques spatiales, et les articles qui traitent de plusieurs filtres. Cette classification est en partie fondée sur une présentation récente de Jens Roat Kultima, de l'Autorité danoise de la concurrence et de la consommation, lors de la conférence CMA Data, Technology and Analytics 2022 sur « la détection de la collusion dans les marchés publics à l'aide de méthodes informatiques »<sup>32</sup>.

### Méthodes basées sur la distribution des offres/prix

Au rang des premières contributions notables relatives aux méthodes de filtrage de la distribution des offres et des prix (parfois aussi appelées « méthodes basées sur la variance ») se trouvent celles de (Abrantes-Metz et al., 2006<sup>[20]</sup>) et (Bolotova, Connor and Miller, 2008<sup>[54]</sup>). (Jiménez and Perdiguero, 2012<sup>[56]</sup>) fournissent une revue utile des études empiriques qui utilisent des méthodes de filtrage basées sur la variance.

Imhof (2017<sup>[57]</sup>) applique des filtres statistiques simples en utilisant des données sur la construction de routes et l'acquisition de revêtements dans le canton du Tessin en Suisse (« Entente du Tessin »). Aucun paiement annexe n'a été effectué dans le cadre de cette entente. Pour cette entente, le soumissionnaire retenu était l'entreprise présentant le coût le plus bas, afin de maximiser les profits ex ante de l'entente. Par la suite, le prix de l'offre gagnante et les prix de l'offre de couverture perdante ont été fixés de manière concertée. Toutes les entreprises ont participé à l'entente et toutes ont truqué les offres pendant cinq ans. Par conséquent, la grande portée et la longue durée de cette entente ne sont peut-être pas représentatives de la plupart des ententes où les offres sont truquées. L'auteur utilise des filtres qui déterminent la distribution des prix des offres notamment : (i) le coefficient de variation (calculé comme l'écart type des offres soumises pour un contrat donné divisé par la moyenne arithmétique des offres soumises pour un contrat donné) ; (ii) la distance relative (calculée comme la distance entre les deux offres les plus basses divisée par l'écart type des offres perdantes) (iii) l'aplatissement ; et (iv) l'asymétrie. L'auteur calcule ces statistiques pour la période de l'entente et les compare à la période précédant et suivant l'entente. Une variation plus faible et une distribution en dents de scie des offres pendant la période de l'entente indiquent un truquage des offres avec des offres de couverture perdantes artificielles. L'auteur applique cette



analyse ex post en connaissant les dates présumées de l'entente. En cas d'application ex ante, les auteurs recommandent de rechercher des changements structurels inexplicables dans ces statistiques de test.

Imhof et al. (2018<sub>[27]</sub>) appliquent des filtres statistiques simples en utilisant des données sur les marchés publics de la construction routière en Suisse dans le district de See-Gaster du canton de Saint-Gall (« Entente de See-Gaster »). La COMCO (autorité suisse de la concurrence) s'est servie de cette analyse pour effectuer des perquisitions, au cours desquelles elle a trouvé des preuves qui ont donné lieu à une enquête et à des sanctions. Les auteurs adoptent une « approche boîte à outils » qui intègre plusieurs méthodes basées sur la distribution des offres/prix, notamment : (1) le coefficient de variation (calculé comme l'écart type des offres soumises pour un contrat donné divisé par la moyenne arithmétique des offres soumises pour un contrat donné) ; (2) le filtre des offres de couverture utilisant la distance relative (calculé comme la distance entre les deux offres les plus basses divisée par l'écart type des offres perdantes) ; (3) le filtre de la collusion partielle ; et (4) le filtre de la rotation des offres.

Huber et al. (2019<sub>[26]</sub>) combinent des filtres statistiques avec des techniques d'apprentissage automatique en utilisant des données relatives à des ententes avérées visant à truquer des offres dans le secteur suisse de la construction routière (qui comprend « l'Entente du Tessin » et « l'Entente de See Gaster » (voir ci-dessus) ainsi que deux autres ententes dont les auteurs gardent les détails confidentiels). Les auteurs utilisent sept filtres statistiques comportementaux qui tiennent compte de la distribution des valeurs des offres dans chaque appel d'offres (écart-type, coefficient de variation, aplatissement, différence en pourcentage, asymétrie, rapport de différence relatif et distance normalisée). Les auteurs utilisent deux techniques d'apprentissage automatique pour combiner ces filtres statistiques : (1) régression logit de l'opérateur de sélection et de retrait le moins absolu ; (2) méthode d'ensemble, qui consiste en une moyenne pondérée de prédictions basées sur des arbres de régression ensachés, des forêts aléatoires et des réseaux neuronaux. Ces approches d'apprentissage automatique comportent deux étapes générales : (1) répartir les données en données d'entraînement et données de test, et entraîner le modèle sur les données d'entraînement, ce qui revient à pondérer chaque filtre ; et (2) appliquer ce modèle entraîné aux données de test pour prédire si les appels d'offres sont collusoires ou concurrentiels. Selon les auteurs, les deux techniques d'apprentissage automatique ont permis de déterminer que 80 % des appels d'offres sont soit des ententes, soit concurrentiels. Deux filtres (distance normalisée et coefficient de variation) présentent, et de loin, le meilleur pouvoir prédictif, c'est-à-dire que les modèles d'apprentissage automatique leur ont attribué le poids le plus élevé. Le test de robustesse a montré que même lorsque ces deux filtres étaient exclus, les deux méthodes d'apprentissage automatique affichaient toujours une précision supérieure à 80 %, car les autres filtres « entrent en jeu », ce qui signifie que même avec un ensemble plus limité de prédicteurs, l'approche peut être très précise.

Wallimann et al. (2020<sub>[58]</sub>) ont élaboré une méthode de détection du truquage incomplet des offres pour les ententes avérées visant à faire des soumissions concertées dans des appels d'offres suisses portant sur des projets de construction routière et de génie civil (« Entente du Tessin », « Entente de See-Gaster » (voir ci-dessus) et une entente relative à l'asphalte dans les Grisons). Un truquage est considéré complet lorsque toutes les entreprises participant à un appel d'offres sont impliquées dans l'entente, et incomplet lorsque quelques entreprises seulement font partie de l'entente. Dans le cas d'un truquage incomplet, les offres concurrentielles incluses dans un appel d'offres peuvent fausser les filtres comportementaux traditionnels appliqués aux appels d'offres, dont la précision s'en trouve réduite. Pour résoudre ce problème, les auteurs appliquent des filtres sur tous les sous-groupes possibles de trois ou quatre offres dans un appel d'offres. Les auteurs utilisent une approche d'apprentissage automatique de type « forêt aléatoire ». Les auteurs ont noté que la méthode d'apprentissage automatique est plus performante que les filtres comportementaux individuels traditionnels, car elle se sert d'un ensemble plus large de filtres comportementaux comme données d'entrée et les pondère en fonction des données, en choisissant les meilleurs prédicteurs dans chaque cas.

Huber et al. (2020<sub>[43]</sub>) combinent des filtres statistiques avec des techniques d'apprentissage automatique, comme suggéré par (Imhof, Karagök and Rutz, 2018<sub>[27]</sub>), (Huber and Imhof, 2019<sub>[26]</sub>) et (Wallimann, Imhof

and Huber, 2020<sup>[58]</sup>) qui ont été développés dans le contexte des données sur les marchés publics de la construction routière en Suisse. Les auteurs appliquent ces filtres de truquage des offres à une entente avérée et sanctionnée sur les marchés publics japonais dans la préfecture d'Okinawa en matière de génie civil et de construction de bâtiments (« Données japonaises »). Ils utilisent également les données des marchés publics de (Huber and Imhof, 2019<sup>[26]</sup>) (« Données suisses »). Ils emploient trois types de filtres différents : (i) des filtres pour la variance des offres (coefficient de variance, écart, statistique d'aplatissement) ; (ii) des filtres pour l'asymétrie des offres (différence en pourcentage, distance, distance relative, distance normalisée, différence normalisée alternative, statistique d'asymétrie) ; et (iii) un filtre pour l'uniformité des offres (statistique de Kolmogorov-Smirnov). Les auteurs se servent de l'apprentissage automatique comme suit : (i) les forêts aléatoires uniquement ; et (ii) la méthode d'ensemble, qui est une moyenne pondérée de trois algorithmes : arbres de régression ensachés, forêts aléatoires et réseaux neuronaux. Les auteurs ont constaté qu'en utilisant les données japonaises, le taux de classification correcte hors échantillon varie entre 88 % et 93 %. Toutefois, lorsque ce taux est entraîné sur une combinaison de données suisses et japonaises, le pourcentage le plus élevé est de 85 %. Cette précision a été davantage réduite lorsque le taux a été entraîné pour un pays et testé sur l'autre pays, en partie parce que les valeurs de filtres pour les soumissions concurrentielles dans un pays sont similaires aux soumissions anticoncurrentielles dans l'autre pays. Par exemple, le coefficient de variance des soumissions concertées suisses était similaire à celui des soumissions concurrentielles japonaises, de telle sorte que lorsque le modèle a été entraîné sur les données japonaises puis appliqué aux données suisses, le coefficient de variance n'a pas été utile pour détecter les soumissions concertées suisses.

Silveira et al. (2021<sup>[59]</sup>) utilisent une approche économétrique de séries chronologiques pour évaluer la variance des prix à la consommation pour les ententes avérées sur les marchés locaux de l'essence au Brésil. Les auteurs évaluent un (i) modèle MS-GARCH (Markov-Switching Generalized Auto Regressive Conditional Heteroskedasticity) ; et (ii) un modèle LGC (Local Gaussian Correlation). Le filtre MS-GARCH entend identifier le comportement du marché marqué par des ententes en examinant la volatilité et la dispersion des prix (faible volatilité des prix pendant la période de l'entente). Le modèle LGC utilise la relation dynamique entre la marge bénéficiaire moyenne et la variation du prix de détail moyen de l'essence (avec une augmentation de la marge bénéficiaire et un prix plus stable attendu pendant la période de l'entente, ce qui signifie une relation intertemporelle négative entre la marge bénéficiaire et la variation du prix indiquant peut-être un accord collusif). Les auteurs constatent que le modèle MS-GARCH est plus performant que le modèle LGC.

Silveira et al. (2022<sup>[60]</sup>) combinent des filtres statistiques et des techniques d'apprentissage automatique afin de déterminer les prix à la consommation pour les ententes avérées sur les marchés locaux de l'essence au Brésil. Les auteurs utilisent cinq filtres statistiques (écart type, coefficient de variance, dispersion, asymétrie et aplatissement). Ils calculent ces filtres statistiques au niveau de la ville (il y a quatre villes dans les données) au cours d'une semaine donnée (entre 200 et 300 semaines environ dans chacune des périodes d'entente ou hors entente) à l'aide des données hebdomadaires sur les stations-service. Les auteurs combinent ces filtres statistiques à l'aide de cinq méthodes supervisées d'apprentissage automatique (logit, logit de l'opérateur de sélection et de retrait le moins absolu, logit ridge, forêt aléatoire et réseau neuronal). Ils estiment que le coefficient de variance et l'écart type sont les prédicteurs les plus performants, bien que cela dépende du contexte. Ils concluent donc que tous les filtres statistiques sont pertinents. Les données sous-jacentes et le code (Python et Stata) de l'analyse sont décrits dans l'annexe A du document.

Rodriguez et al. (2022<sup>[61]</sup>) combinent des filtres statistiques aux techniques d'apprentissage automatique en utilisant des données sur les ententes avérées dans différentes industries, extraites des données sur les marchés publics du Brésil, de l'Italie, du Japon, de la Suisse et des États-Unis. Les auteurs utilisent sept filtres statistiques, notamment : (i) le coefficient de variation ; (ii) l'écart ; (iii) la différence entre les deux offres d'enchères les plus basses ; (iv) la différence relative ; (v) l'asymétrie ; (vi) l'aplatissement ; (vii) le test de Kolmogorov-Smirnov. Les auteurs se servent de 11 algorithmes d'apprentissage

automatique (apprentissage supervisé), notamment : (i) la descente de gradient stochastique (modèle linéaire) ; (ii) les arbres supplémentaires (méthode d'ensemble) ; (iii) la forêt aléatoire (méthode d'ensemble) ; (iv) Ada Boost (méthode d'ensemble) ; (v) Gradient Boosting (méthode d'ensemble) ; (vi) les machines à vecteurs de support ; (vii) K Neighbours ; (viii) le perceptron multicouche (modèle de réseau neuronal) ; (ix) Naïve Bayes de Bernoulli (Naïve Bayes) ; (x) Naïve Bayes gaussienne (Naïve Bayes) ; et (xi) le processus gaussien. Les auteurs effectuent 500 itérations (de sous-ensembles de données d'entraînement et de test) pour chaque algorithme. Les trois algorithmes les plus performants ont été les arbres supplémentaires, les forêts aléatoires et Ada Boost (toutes des méthodes d'ensemble). Dans le scénario où toutes les informations sur les enchères étaient disponibles, ces algorithmes ont eu une précision (taux de détection) comprise entre 81 % et 95 %. Toutes les données sous-jacentes et le code (Python) sont fournis dans un fichier supplémentaire.

## Arrondi des offres

Ishii (2014<sub>[62]</sub>)<sup>33</sup> adopte « l'arrondi des offres » en utilisant les données sur les offres faites par les membres d'une entente avérée, ayant fourni des soumissions concertées dans le domaine du bâtiment et des travaux publics au Japon. Le « niveau de l'arrondi » d'une offre est la longueur de la série de zéros consécutifs à la fin du prix de l'offre, par exemple, le « niveau de l'arrondi » pour l'offre 12 300 000 JPY est de cinq. On part de l'hypothèse que les soumissionnaires de l'offre gagnante choisiront des chiffres ronds pour leurs offres afin d'éviter des erreurs de communication. L'auteur utilise plusieurs méthodes statistiques et économétriques pour tester de multiples hypothèses. L'auteur constate que le « niveau de l'arrondi » est plus élevé en cas de collusion, entre autres.

## Méthodes basées sur la cointégration

Kurdoglu et Yucel (2022<sub>[63]</sub>) adoptent un filtre d'entente basé sur la cointégration en utilisant les données d'une entente avérée relative au ciment turc. L'approche basée sur la cointégration est une méthodologie économétrique de séries chronologiques qui teste s'il existe un co-mouvement statistiquement significatif des prix entre les entreprises, ce qui suggérerait une collusion potentielle. Deux séries temporelles (dans ce cas, les séries temporelles des prix chez deux fabricants de ciment) sont dites cointégrées s'il existe une relation d'équilibre à long terme liant les séries temporelles respectives, qui peut être représentée comme une fonction linéaire de ces deux séries temporelles (Hamilton, 1994, pp. 571-572<sub>[64]</sub>). Kurdoglu et Yucel notent, dans plusieurs régions géographiques de Turquie, que la cointégration des séries temporelles des prix des cimentiers a été plus importante pendant la période de l'entente que hors entente.

## Indépendance conditionnelle entre les offres

(Bajari and Ye, 2003<sub>[19]</sub>) est une étude de référence qui expose l'idée d'indépendance conditionnelle, c'est-à-dire que les offres scellées sont censées être indépendantes les unes des autres (après contrôle des facteurs observables). Des articles plus récents analysent cette approche en se fondant sur la littérature relative à l'économétrie spatiale. L'élément spatial n'est plus géographique mais basé sur l'environnement des offres. Pour ce faire, ces articles utilisent des fonctions d'offres.

Bergman et al. (2020<sub>[65]</sub>) adoptent une approche économétrique spatiale en utilisant les données d'une entente avérée visant à faire des soumissions concertées dans l'industrie suédoise du revêtement d'asphalte. Selon cette approche, l'offre concurrentielle d'une entreprise ne saurait être corrélée aux offres concurrentielles des autres entreprises, une fois que les informations susceptibles d'influer sur la valeur des offres (telles que le degré de concurrence et le coût du projet) ont été vérifiées. Les auteurs testent l'indépendance conditionnelle entre les offres. La dépendance conditionnelle entre les offres est un signe

qu'il s'agit d'offres complémentaires potentielles résultant d'une collusion. Les auteurs testent l'indépendance conditionnelle à l'aide d'une fonction d'offre qui réduit plusieurs variables indépendantes, notamment une matrice de pondérations spatiales (le coefficient d'intérêt) et des variables qui peuvent avoir un impact sur les valeurs des offres (contrôles), par rapport aux valeurs des offres (variable dépendante). Ils notent une corrélation positive pendant la période de l'entente (c'est-à-dire un coefficient positif et significatif sur la matrice des pondérations spatiales), et une corrélation non significative pendant la période post-entente, ce qui suggère que les parties à l'entente ont adopté des offres complémentaires pendant la période de l'entente. Cet article vérifie l'existence des ententes plutôt que d'appliquer un filtre ex ante, car la méthodologie des auteurs repose sur la connaissance des ententes. (Bergman et al., 2015<sup>[66]</sup>) semble être une version de travail antérieure de ce document.

Lundberg (2017<sup>[67]</sup>) applique une approche économétrique spatiale en utilisant les données relatives à une entente avérée de soumissions concertées dans l'industrie suédoise du revêtement d'asphalte (liée au travail de (Bergman et al., 2015<sup>[66]</sup>)). L'auteur emploie le I de Moran, un test de corrélation spatiale, afin de détecter des offres complémentaires potentielles. La statistique du test I de Moran est calculée directement à l'aide des valeurs des offres, ainsi que des données résiduelles d'une fonction d'offre.

De Andrade Lima et Resende (2021<sup>[68]</sup>) appliquent une approche économétrique spatiale en utilisant les données d'une entente présumée visant à faire des soumissions concertées sur le marché brésilien des dispositifs cardiaques implantables. Les auteurs emploient le I de Moran, un test de corrélation spatiale, afin de détecter des offres complémentaires potentielles. Les auteurs calculent la statistique du test I de Moran à l'aide des données résiduelles d'une fonction d'offre.

## Doubles différences

La méthode des doubles différences est une approche économétrique standard qui utilise des données de panel pour déterminer l'effet moyen du traitement en calculant la différence entre le changement moyen dans le temps pour le groupe de traitement (différence un) et le changement moyen dans le temps pour le groupe de contrôle (différence deux).

Clark et al. (2020<sup>[69]</sup>) utilisent une approche des doubles différences pour comparer l'isolement des offres gagnantes et le regroupement des offres dans une entente avérée visant à faire des soumissions concertées dans l'industrie canadienne de l'asphalte. La variable dépendante est une variable muette basée sur la fraction historique de l'offre (la propre offre du soumissionnaire moins le minimum des autres offres, divisé par le prix de réserve) ou la différence des offres (lorsqu'aucun prix de réserve n'est disponible). Elle est égale à un lorsque la fraction historique des offres ou la différence des offres se situe dans un intervalle donné. Le groupe de traitement est l'industrie de l'asphalte de Montréal. Le groupe de contrôle est l'industrie de l'asphalte de la ville de Québec. La période concernée correspond à celle de l'entente (avant l'enquête) et celle d'après l'entente (après l'enquête). La régression contrôle les caractéristiques de l'adjudication (telles que le prix moyen décalé du pétrole brut, la quantité d'asphalte indiquée dans l'appel d'offres et l'indice Herfindahl-Hirschman (spécifique à la ville)). Les auteurs ont trouvé des preuves de l'existence d'offres gagnantes isolées (le gagnant souhaitant une petite marge pour se prémunir contre toute erreur d'offre) ainsi que d'un regroupement (le résultat d'une offre de couverture à un prix plus élevé convenu) pour le groupe de traitement pendant la période de l'entente.

Perdiguero et Jiménez (2021<sup>[70]</sup>) utilisent l'approche des doubles différences pour tester « l'effet lundi ». Les opérateurs espagnols de carburants de qualité supérieure ont conjointement réduit les prix de l'essence le lundi pour faire baisser les prix officiels des carburants automobiles, qui étaient perçus le lundi, afin que l'Espagne ne figure pas en tête du classement européen des prix. Les auteurs ont constaté que cet effet « lundi » a soudainement cessé, lorsque le gouvernement a modifié la réglementation pour utiliser les prix hebdomadaires moyens au lieu des prix du lundi. La variable dépendante est le prix de détail de l'essence. Le groupe de traitement est constitué de stations-service de marque de l'Espagne

métropolitaine qui ont été potentiellement touchées par « l'effet lundi ». Les groupes de contrôle sont les stations-service des îles Canaries (une région d'Espagne) et celles qui sont sans marque ou exploitées par des détaillants indépendants. La période concernée est celle de l'entente (avant la modification de la réglementation) et celle d'après l'entente (après la modification de la réglementation).

## Fractions d'offre historiques

Chassang et al. (2022<sup>[71]</sup>) examinent les fractions d'offres concurrentielles historiques à l'aide de données provenant de deux enchères avec prix de retrait secret portant sur des marchés de travaux publics au Japon, sans entente avérée. Les auteurs calculent la fraction d'offre concurrentielle pour chaque offre. Ils calculent la fraction d'offre concurrentielle comme étant la propre offre du soumissionnaire moins le minimum des autres offres, divisée par le prix de réserve. Une valeur négative signifie que le soumissionnaire a remporté l'enchère, tandis qu'une valeur positive signifie que le soumissionnaire n'a pas remporté l'enchère. Les auteurs examinent ensuite la distribution de ces fractions d'offres concurrentielles. Une absence de masse d'offres autour de zéro indique un comportement suspect, car cela suppose qu'aucune offre n'était proche de l'offre gagnante et que l'écart entre l'offre gagnante et les offres restantes était considérable. Cette approche de la fraction de l'offre historique est également utilisée et décrite par (Ortner et al., 2022<sup>[72]</sup>) dans le contexte des filtres d'entente asymptotiquement sûrs (c'est-à-dire des filtres que les entreprises concurrentes passent avec une probabilité proche de un).

## Méthodes de réseau

Wachs et al. (2019<sup>[73]</sup>) utilisent une approche de réseau d'apprentissage automatique non supervisé pour révéler les points chauds où les entreprises interagissent plus fréquemment et où la collusion peut être plus probable et plus facile à maintenir. Cette approche a été appliquée dans d'autres contextes, comme celui de la corruption. Les auteurs appliquent cette approche à deux marchés : (1) L'entente avérée sur le lait dans les écoles de l'Ohio ; et (2) les marchés publics attribués en République de Géorgie entre 2011 et 2016, sans entente avérée. Les auteurs ont constaté que dans l'entente sur le lait de l'Ohio, les entreprises participant à l'entente interagissent fréquemment en étant relativement isolées des entreprises concurrentielles. En revanche, pour les données sur les marchés publics de Géorgie, les groupes ayant des interactions cohésives et exclusives ont des prix plus élevés et une variance plus faible dans les offres et les prix.

## Méthodes probabilistes

Signor et al. (2021<sup>[74]</sup>) adoptent une méthode probabiliste en utilisant les données d'enchères de Petrobras, la compagnie pétrolière publique brésilienne, dans le cas de « l'Opération Car Wash ». Il s'agissait de projets de construction pour lesquels Petrobras a lancé un appel d'offres. Le « Club des 16 », groupe d'entreprises de construction, a truqué des offres pour des dizaines de projets d'infrastructures publiques. Elles ont participé à certains des appels d'offres de Petrobras. Les auteurs examinent la probabilité qu'un ensemble particulier d'offres se produise par hasard. Les auteurs calculent un ratio d'offre, qui est l'offre divisée par l'estimation préalable à l'appel d'offres (qui a été fournie par un expert au moment de l'appel d'offres). Les auteurs se penchent ensuite sur la distribution de ces ratios dans des appels d'offres « honnêtes » de référence (appels d'offres non collusoires, c'est-à-dire qu'aucun membre du « Club des 16 » n'a participé à ces appels d'offres). Les auteurs calculent enfin la probabilité qu'un appel d'offres donné ait fait l'objet d'une collusion sur la base des ratios d'offres observés. Les auteurs ont constaté que leur méthode semblait dépasser les résultats présentés par (Imhof, 2018<sup>[75]</sup>) et (Bajari and Ye, 2003<sup>[19]</sup>) (bien que cette conclusion provienne de la comparaison des mesures de performance des études respectives, plutôt que de l'application de ces approches à leurs propres données). D'autres approches

probabilistes sont fournies par (Signor et al., 2017<sup>[76]</sup>) et (Signor et al., 2020<sup>[77]</sup>) et appliquées dans la même affaire « Operation Car Wash ».

## Méthodes de discontinuité de régression

Kawai et al. (2022<sup>[15]</sup>) adoptent un modèle de discontinuité de la régression pour identifier l'avantage de l'entreprise en place et la rotation des offres en utilisant les données de l'entente avérée sur le lait dans les écoles de l'Ohio analysée par (Porter and Zona, 1997<sup>[17]</sup>) et les données des enchères avec prix de retrait secret portant sur des marchés de travaux publics au Japon utilisées par (Chassang, 2022<sup>[71]</sup>) dans lesquelles aucune entreprise n'a encore été accusée de collusion (toutefois, des articles universitaires précédents laissaient entendre que la collusion existait). Un modèle de discontinuité de régression identifie un effet de traitement moyen en considérant les différences dans une variable de résultat (dans ce cas, le statut d'entreprise en place ou le niveau de retard des entreprises qui n'ont pas remporté beaucoup d'enchères dans un passé récent) pour les observations qui sont très légèrement au-dessus ou au-dessous d'un seuil de traitement (dans ce cas, gagner ou perdre une offre). Étant donné que les anneaux d'enchères adoptent souvent des systèmes de rotation ou accordent la priorité aux entreprises en place dans l'attribution des projets, la rotation des offres et l'avantage donné à l'entreprise en place sont souvent considérés comme des indicateurs de collusion. Cependant, il existe des explications non collusives basées sur les coûts : la rotation des offres peut survenir dans un contexte de concurrence si les coûts marginaux augmentent avec le retard accumulé par certaines entreprises, et l'avantage de l'entreprise en place peut s'expliquer par des asymétries de coûts entre les entreprises concurrentes ou par l'apprentissage par la pratique.

Par conséquent, pour dissocier ces effets, les auteurs mettent l'accent sur les enchères où les offres gagnantes et perdantes sont très proches (c'est-à-dire qu'ils isolent les gagnants et les perdants marginaux). Indépendamment du statut d'entreprise en place ou du retard accumulé par certaines entreprises dans leur taux d'offres gagnantes, les gagnants et les perdants devraient être aussi aléatoires que pour les offres fermées. Ainsi, si les gagnants marginaux sont plus souvent des concurrents que des entreprises en place, cela peut attester d'une rotation anticoncurrentielle des offres, tandis que si les gagnants marginaux sont plus souvent des entreprises en place que des concurrents, cela peut indiquer l'avantage de l'entreprise en place. Sur la base des données concernant l'entente sur le lait de l'Ohio, les auteurs constatent que le gagnant marginal est plus susceptible d'être une entreprise en place, ce qui suppose un avantage illégal pour celle-ci. Grâce aux données sur les marchés publics japonais, les auteurs possèdent des preuves potentielles d'une combinaison de l'avantage illégal de l'entreprise en place (observant que les gagnants marginaux sont plus souvent des entreprises en place) et de la rotation illégale des offres (notant que les gagnants marginaux ont un retard accumulé plus faible).

## Méthodes de rupture structurelle

Crede (2019<sup>[78]</sup>) adopte un filtre d'entente à rupture structurelle en utilisant les prix à la production domestiques mensuels pour les pâtes alimentaires, en présence d'ententes avérées en Italie et en Espagne et des données hors entente en France. Comme pour l'utilisation des fonctions d'offres, l'auteur estime le processus de génération de données (DGP) sous-jacent, qui peut être considéré comme une fonction du prix. L'auteur estime le DGP à l'aide d'une régression de forme réduite qui inclut des variables qui déterminent les prix des pâtes, comme divers types de coûts de production, des approximations de la demande et des contrôles pour les ruptures structurelles qui sont induites par l'activité potentielle des ententes. L'auteur vérifie s'il existe des ruptures structurelles dans le processus de génération des données, c'est-à-dire des variations inexplicables des prix au cours de la période, qui seraient très probablement dues aux ententes. L'auteur estime que sa méthode de rupture structurelle permet d'identifier correctement les ententes en Italie et en Espagne et l'absence d'entente en France.

Alors que d'autres approches utilisent également les ruptures structurelles en comparant les données de la période de l'entente avec celles de la période hors entente, en utilisant les dates de l'entente indiquées de manière exogène, cette méthode est la seule qui teste les ruptures structurelles de manière endogène (c'est-à-dire sans connaître la date précise de l'entente). C'est pourquoi, elle peut servir à un filtrage ex ante.

L'auteur compare cette approche aux filtres de variance, en estimant un filtre de coefficient de variance (Abrantes-Metz et al., 2006<sup>[20]</sup>) et un filtre de variance GARCH (Bolotova, Connor and Miller, 2008<sup>[54]</sup>). Ces filtres de variance ne sont pas aussi performants (c'est-à-dire qu'ils ne permettent pas d'identifier les ententes). Mais l'auteur considère que cela peut être dû à la courte durée de l'entente qui a connu des hausses et des baisses de prix brutales, ce qui signifie qu'il n'a pas été possible d'observer une variation réduite des prix résultant d'une collusion stable.

Crede (2016<sup>[79]</sup>) fournit une brève introduction sur les filtres d'entente. Il explique la raison d'être des filtres de variance des prix, ainsi que leurs limites potentielles (dans le cas d'une entente où les prix sont augmentés, la mesure de variance des prix pourrait aussi s'accroître, et l'entente serait considérée à tort comme non collusive). L'auteur explique ensuite que le test des ruptures structurelles peut être une solution à ce problème (car l'augmentation du prix au-dessus du coût provoquera une rupture structurelle inexpliquée dans la relation entre le coût et le prix). Cet article n'applique pas un test de rupture structurelle à des données réelles, mais explique plutôt le raisonnement intuitif sous-jacent.

## Regroupements suspects dans les données spatiales géographiques

Heijnen et al. (2015<sup>[47]</sup>) identifient des regroupements suspects de stations-service dans une approche spatiale, en utilisant des données quasi quotidiennes sur les prix de détail du marché néerlandais de l'essence, sans s'appuyer sur une affaire ou une entente avérée. Les auteurs utilisent le coefficient de variation, défini comme l'écart-type du prix de détail résiduel d'une entreprise (le prix de détail après contrôle des caractéristiques observées de la station-service) divisé par sa moyenne. Les auteurs considèrent comme suspects les 5 % de points de vente les plus bas sur la base du coefficient de variation. Les auteurs adoptent une approche en deux étapes. Ils commencent par identifier tout regroupement de ces points de vente suspects. Ensuite, ils identifient les régions les plus suspectes où une enquête antitrust pourrait être justifiée.

## Combinaison de plusieurs filtres

Fazekas et al. (2022<sup>[21]</sup>) combinent plusieurs filtres, tels que les méthodes basées sur la distribution des offres, les modalités de soumission des offres et la concentration du marché. Le tableau ci-dessous présente les filtres utilisés dans l'article, que les auteurs combinent à l'aide d'une approche d'apprentissage automatique supervisé par forêts aléatoires. Les auteurs associent les cas avérés d'entente aux données des marchés publics de plusieurs secteurs et pays afin de créer un ensemble de données final, et appliquent l'approche des forêts aléatoires à cet ensemble de données. Il ressort qu'aucun filtre unique ne fonctionne dans la majorité des cas et qu'une approche multi-filtres est la plus efficace. Cet article s'appuie sur les travaux des mêmes auteurs dans (Czibik, Tóth and Fazekas, 2015<sup>[80]</sup>) et (Fazekas and Tóth, 2016<sup>[22]</sup>).

**Tableau A 0.1. Indicateurs testés**

Catégorie	No	Indicateur	Niveau d'observation	Description
Prix	1	Différence entre le prix le plus bas et le deuxième prix le plus bas	Contrat	Différence relative entre le prix le plus bas et le deuxième prix le plus bas (différences de 1 %, 5 %, 10 % etc.)

	2	Fourchette de prix relative	Contrat	La fourchette de prix relative basée sur le prix de l'offre la plus basse et la plus haute est inférieure à 10 % ou supérieure à 90 % de la distribution.
	3	Écart-type relatif	Contrat	L'écart-type relatif des prix des offres est inférieur à 10 % ou supérieur à 90 % de la distribution.
	4	Prix relatif	Contrat	Rapport entre le prix final et le prix estimé
	5	Loi de Benford	Période de marché	Les premiers chiffres des prix contractuels d'un marché donné au cours d'une période donnée suivent-ils la loi de Benford ?
<b>Modalités de soumission des offres</b>	6	Offre unique	Contrat	Contrat recevant une seule offre lors de la procédure d'appel d'offres.
	7	Soumissionnaires manquants	Activité de l'entreprise	Le nombre d'acheteurs uniques/de villes/de régions/de sociétés de codes de produits qui soumettent une offre.
	8	Sous-traitance	Contrat	Si un contrat comporte un sous-traitant.
	9	Consortium	Contrat	Si l'offre gagnante était un consortium.
	10	Position de rupture	Période de marché	La présence ou non d'entreprises en position de rupture sur un marché et une période donnés.
	11	Probabilité de gagner	Période de marché	La probabilité moyenne que les entreprises ont de gagner sur un marché et une période donnés.
<b>Structure du marché</b>	12	Structure de marché concentrée	Période de marché	Variation de l'indice HHI pendant et après la période d'entente
	13	Structure de marché stable	Période de marché	Variations absolues moyennes des parts de marché pendant et après la période d'entente

Source : (Fazekas et al., 2022<sup>[21]</sup>)



## Annexe B. Méthodes d'apprentissage automatique

Il existe trois grands types d'approches d'apprentissage automatique : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement. La plupart des articles universitaires sur les filtres d'ententes utilisent des approches d'apprentissage supervisé, certains se servent des approches d'apprentissage non supervisé et très peu, voire aucun, utilisent des approches d'apprentissage par renforcement.

Le filtrage des ententes est généralement considéré un *problème de classification*, c'est-à-dire que l'enquêteur souhaite estimer si une enchère est collusive ou non, sur la base des caractéristiques de cette enchère. Les économistes abordent généralement ce problème à l'aide d'une analyse de régression, en l'occurrence une régression logit, étant donné la variable de résultat binaire (collusive ou non collusive). Cependant, l'apprentissage automatique offre des approches supplémentaires. Deng (2017<sup>[81]</sup>), Varian (2014<sup>[82]</sup>) et Athey & Imbens (2019<sup>[83]</sup>) fournissent une synthèse utile de ces approches et expliquent la pertinence des techniques d'apprentissage automatique pour l'économie. Il existe également des manuels qui décrivent ces approches, tels que ceux de (James et al., 2013<sup>[84]</sup>) (moins élaboré) et (Hastie et al., 2009<sup>[28]</sup>) (plus élaboré). Torres Berru et al. (2019<sup>[85]</sup>) proposent également une analyse documentaire des techniques d'intelligence artificielle (telles que la classification, la régression, le regroupement, la prédiction, la détection des valeurs aberrantes et la visualisation) utilisées pour détecter d'autres types de corruption dans les marchés publics (tels que la corruption, la collusion, le détournement de fonds, la malversation, la fraude, l'abus de pouvoir discrétionnaire, le favoritisme et le népotisme).

La présente annexe décrit plusieurs approches d'apprentissage automatique utilisées pour filtrer les ententes. Elle expose aussi l'apprentissage supervisé et non supervisé et décrit les techniques utilisées dans la littérature académique récente sur les filtres des ententes. Chaque méthode s'accompagne d'une brève explication, suivie de la citation d'études qui ont adopté chaque approche.

Lorsque les auteurs comparent des approches d'apprentissage automatique supervisé, ils utilisent généralement les mêmes données pour chaque modèle (notamment, les mêmes statistiques de filtres sous-jacentes, comme divers types de filtres basés sur la distribution des offres et des prix). Lorsque l'on compare des algorithmes d'apprentissage automatique, il faut prendre en compte aussi bien la précision (la capacité à détecter la collusion) que la complexité (le temps de calcul). À titre d'exemple, les forêts aléatoires sont l'un des modèles d'apprentissage automatique les plus simples et ont donc un temps de calcul plus faible, tandis que la méthode LASSO est plus longue car elle nécessite le calcul d'un terme de pénalité. La méthode d'ensemble est l'une des plus longues car elle doit calculer des pondérations optimales. Cependant, la méthode d'ensemble peut s'avérer très précise<sup>34</sup>.

### Régression linéaire

La régression linéaire (Hastie et al., 2009, pp. 44-66<sup>[28]</sup>) modélise la relation entre une variable de sortie et au moins une variable d'entrée. Souvent, la régression linéaire est modélisée à l'aide d'une approche des moindres carrés, qui minimise la somme des résidus au carré (où le résidu est la différence entre la valeur prédite et la valeur de sortie réelle). La régression linéaire est une méthode couramment utilisée par les

économistes. De nombreux éléments de l'apprentissage automatique supervisé peuvent donc être considérés davantage comme une évolution et non pas comme une révolution (Abrantes-Metz and Metz, 2018<sup>[23]</sup>). Elle peut être classée dans la catégorie de l'apprentissage automatique car, à mesure que de nouvelles données d'entraînement sont obtenues, la fonction linéaire se met à jour avec ces nouvelles informations.

## Régression logistique

La régression logistique (Hastie et al., 2009, pp. 119-128<sup>[28]</sup>) modélise la probabilité qu'un résultat appartienne à une certaine catégorie (ainsi, la variable de sortie est comprise entre 0 et 1), subordonnée à au moins une variable d'entrée. Par exemple, la variable de sortie pourrait indiquer la probabilité que des soumissions soient concertées.

(Silveira et al., 2022<sup>[60]</sup>) utilisent la régression logistique.

## Méthode de régression ridge

La méthode de régression ridge (Hastie et al., 2009, pp. 61-67<sup>[28]</sup>) est similaire à la régression linéaire, bien qu'elle réduise les coefficients de régression en imposant une pénalité sur leur taille. En l'absence de terme de pénalité, par exemple (c'est-à-dire qu'il est fixé à zéro), cette approche se résume alors aux moindres carrés ordinaires (régression linéaire). Dans la régression linéaire, en présence de nombreuses variables, il peut y avoir des coefficients corrélés, qui entraînent une variance élevée, les coefficients positifs et négatifs s'annulant mutuellement. Le fait d'imposer des contraintes de taille sur les coefficients, comme dans le cas de la régression ridge, vise à résoudre ce problème.

(Silveira et al., 2022<sup>[60]</sup>) utilisent la méthode de régression ridge.

## Méthode de l'opérateur de sélection et de retrait le moins absolu (LASSO)

L'opérateur de sélection et de retrait le moins absolu (LASSO) (Hastie et al., 2009, pp. 68-69<sup>[28]</sup>) est une méthode de retrait, tout comme la régression ridge, le terme de pénalité étant la principale distinction entre les deux approches. Alors que la régression ridge interagit le terme de pénalité avec le coefficient au carré, la régression LASSO interagit le terme de pénalité avec la valeur absolue du coefficient. Autrement dit, avec un terme de pénalité suffisamment petit, sous l'approche LASSO, certains coefficients auront une valeur de zéro (en fait un poids nul), tandis que dans une régression ridge, les coefficients n'auront jamais une valeur de zéro.

(Imhof, 2018<sup>[75]</sup>), (Huber and Imhof, 2019<sup>[26]</sup>), (Wallimann, Imhof and Huber, 2020<sup>[58]</sup>), (Imhof and Wallimann, 2021<sup>[25]</sup>), et (Silveira et al., 2022<sup>[60]</sup>) utilisent la méthode LASSO.

## Réseaux neuronaux

Les réseaux neuronaux (Hastie et al., 2009, pp. 389-416<sup>[28]</sup>) sont essentiellement des modèles statistiques non linéaires. Les réseaux neuronaux constituent la base de l'apprentissage profond, qui est un sous-domaine de l'apprentissage automatique. Les réseaux neuronaux peuvent être considérés comme une généralisation non linéaire d'une fonction linéaire. Le modèle accepte des entrées, qu'il fait passer à travers une ou plusieurs couches de neurones (qui peuvent être considérées comme les processeurs du modèle), puis prédit une sortie. Si une seule couche de neurones est présente, alors il s'agit juste d'un modèle linéaire standard.

(Imhof, 2018<sup>[75]</sup>), (Huber and Imhof, 2019<sup>[26]</sup>), (Huber, Imhof and Ishii, 2020<sup>[43]</sup>), (Wallimann, Imhof and Huber, 2020<sup>[58]</sup>), (Imhof and Wallimann, 2021<sup>[25]</sup>), (Rodríguez et al., 2022<sup>[61]</sup>) et (Silveira et al., 2022<sup>[60]</sup>) utilisent des réseaux neuronaux (quoique dans le cadre d'une approche d'ensemble). (Wachs and Kertész, 2019<sup>[73]</sup>) utilisent une approche « réseau » non supervisée. (Huber and et Imhof, 2021<sup>[24]</sup>) utilisent une approche « réseau neuronal à convolution ».

## Arbres de régression

Les arbres de régression (Hastie et al., 2009, pp. 307-308<sup>[28]</sup>) sont basés sur un partitionnement récursif binaire qui est un processus itératif qui divise les données en partitions et divise ensuite chaque partition en groupes plus petits. Ainsi, les arbres de régression sont basés sur des arbres de décision (qui sont en fait des instructions if-else). Les arbres de régression, contrairement aux arbres de décision, sont pertinents lorsque la sortie est continue.

(Imhof, 2018<sup>[75]</sup>), (Huber and Imhof, 2019<sup>[26]</sup>), (Huber, Imhof and Ishii, 2020<sup>[43]</sup>), et (Wallimann, Imhof and Huber, 2020<sup>[58]</sup>) utilisent des arbres de régression ensachés (bien que tous dans le cadre d'une approche d'ensemble).

## Machines à vecteurs de support

Les machines à vecteurs de support (Hastie et al., 2009, pp. 417-458<sup>[28]</sup>) peuvent servir pour des problèmes de classification, notamment pour déterminer si une soumission est ou non concertée. Les hyperplans de séparation optimale s'appliquent lorsque deux classes sont linéairement séparables. Les machines à vecteurs de support s'appliquent lorsque ces classes se chevauchent et ne sont pas linéairement séparables. L'approche de la machine à vecteurs de support crée une limite non linéaire pour classer les entrées. La machine à vecteurs de support sélectionnera la limite qui maximise la marge (qui est la différence entre la limite et l'observation la plus proche d'une classe donnée) des deux classes.

(Imhof and Wallimann, 2021<sup>[25]</sup>) et (Rodríguez et al., 2022<sup>[61]</sup>) utilisent une approche de machine à vecteurs de support.

## Forêts aléatoires

Les forêts aléatoires (Hastie et al., 2009, pp. 587-604<sup>[28]</sup>) sont une approche d'apprentissage automatique utile pour résoudre des problèmes de classification. Elle peut donc prédire une variable de résultat binaire basée sur un certain nombre de « caractéristiques » (également appelées variables explicatives). L'approche s'appuie sur le concept d'arbres de décision (qui sont en fait des instructions if-else).

L'approche des forêts aléatoires comporte les étapes suivantes. Tout d'abord, créer un certain nombre de  $n$  ensembles de données à partir de l'ensemble de données d'apprentissage en sélectionnant de manière aléatoire des observations (lignes) à partir des données (c'est ce qu'on appelle l'autoamorçage). Les ensembles de données sélectionnés au hasard auront le même nombre de lignes que l'original. Deuxièmement, sélectionner au hasard un certain nombre de caractéristiques (ou variables explicatives) pour chacun de ces ensembles de données. Troisièmement, créer un arbre de décision pour chaque ensemble de données auto-amorcées et combinaison de sélection de caractéristiques des deux premières étapes. Quatrièmement (et enfin), exécuter chaque nouvelle valeur dans les données de test, à travers tous ces arbres de décision (c'est-à-dire à travers les 'nœuds de décision') et obtenir un résultat (1 ou 0) (à partir des 'nœuds feuilles'). Ensuite, prendre la valeur majoritaire et l'attribuer comme résultat (c'est ce qu'on appelle « l'agrégation »). La combinaison de l'autoamorçage et de l'agrégation est désignée sous le nom de « ensachage ».

La forêt aléatoire contient deux éléments aléatoires : (i) la création aléatoire de nouveaux ensembles de données à partir des données d'apprentissage lors de l'étape d'amorçage ; et (ii) la sélection aléatoire des caractéristiques (ou variables explicatives) dans l'étape de sélection des caractéristiques. Elle est appelée « forêt » car elle se sert de plusieurs arbres de décision pour remédier aux limites de l'approche par arbre de décision et améliorer la précision des résultats.

(Imhof, 2018<sup>[75]</sup>), (Huber and Imhof, 2019<sup>[26]</sup>), (Huber, Imhof and Ishii, 2020<sup>[43]</sup>), (Wallimann, Imhof and Huber, 2020<sup>[58]</sup>), (Imhof and Wallimann, 2021<sup>[25]</sup>), (Rodríguez et al., 2022<sup>[61]</sup>), (Silveira et al., 2022<sup>[60]</sup>) et (Fazekas et al., 2022<sup>[21]</sup>) utilisent une approche de forêts aléatoires (bien que certaines ne fassent partie que d'une approche d'ensemble).

## Ensemble

L'approche d'ensemble (Hastie et al., 2009, pp. 605-624<sup>[28]</sup>) vise à construire un modèle qui combine les atouts de plusieurs modèles plus « simples ». En fait, elle reprend plusieurs des modèles « simples » ci-dessus (la régression logistique, LASSO, les réseaux neuronaux, les forêts aléatoires, etc.) et exécute les données de test à travers chacun d'eux, en obtenant la prédiction de chaque modèle, puis en les combinant (à l'aide des pondérations) en vue d'obtenir une prédiction finale.

(Imhof, 2018<sup>[75]</sup>), (Huber and Imhof, 2019<sup>[26]</sup>), (Huber, Imhof and Ishii, 2020<sup>[43]</sup>), (Wallimann, Imhof and Huber, 2020<sup>[58]</sup>), (Imhof and Wallimann, 2021<sup>[25]</sup>) et (Rodríguez et al., 2022<sup>[61]</sup>) utilisent une approche d'ensemble.

# Notes de fin de document

<sup>1</sup> Voir le projet « Computational Antitrust » de l'Université de Stanford, lancé en janvier 2021, qui rassemble des autorités de la concurrence et des universitaires de plusieurs disciplines, y compris le droit, l'économie et l'informatique, <https://law.stanford.edu/codex-the-stanford-center-for-legal-informatics/computational-antitrust-project/>

<sup>2</sup> L'utilisation du filtrage dans le contrôle des concentrations a été partiellement examinée au cours du récent Forum mondial sur la concurrence, lors des sessions consacrées à l'analyse économique dans les enquêtes sur les concentrations (OCDE, 2020<sup>[90]</sup>) et au contrôle des concentrations sur des marchés dynamiques (OCDE, 2020<sup>[89]</sup>).

<sup>3</sup> (Harrington, 2021<sup>[98]</sup>) définit quatre conditions requises pour le succès d'une entente : (i) une compréhension commune entre les entreprises (condition de coordination) de la manière de ne pas se faire concurrence ; (ii) un accord collusif qui encourage ses membres à s'y conformer (condition de stabilité interne) ; (iii) un moyen de limiter l'offre de tiers non membres de l'entente (condition de stabilité externe) ; et (iv) une manière d'éviter la détection de l'entente et sa pénalisation par l'autorité de la concurrence et les clients (condition d'évitement de la répression). Harrington conclut que la collusion est difficile mais gérable.

<sup>4</sup> Voir : <https://www.youtube.com/watch?v=lovsp5aHcuU> (2:40:12)

<sup>5</sup> Il est prévu que l'OCDE organise une table ronde sur « L'avenir des programmes de clémence » en juin 2023. L'OCDE avait déjà organisé une table ronde sur le thème « Défis et coordination des programmes de clémence » en juin 2018 (OCDE, 2018<sup>[91]</sup>). Marvão et Spagnolo (Marvão and Spagnolo, 2016<sup>[93]</sup>) fournissent également une synthèse de certaines études théoriques et empiriques sur la clémence.

<sup>6</sup> Entre 1990 et 2016, les ventes nominales affectées par des ententes internationales injustifiables ont excédé les 50 trillions USD. Les surfacturations brutes des ententes ont dépassé 1,5 trillion USD. Plus de 100 000 entreprises ont été reconnues responsables d'entente internationale sur les prix (Connor, 2016<sup>[13]</sup>) Voir également la Recommandation de l'OCDE concernant une action efficace contre les ententes injustifiables <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0452#backgroundInformation>.

<sup>7</sup> Une entente stratégique peut réduire, mais non éliminer, l'efficacité de la plupart des filtres comportementaux. Les ententes se forment généralement afin d'en tirer un bénéfice en augmentant le prix. Une entente peut essayer de réduire la probabilité de détection en augmentant le prix lentement, mais cela suppose toutefois d'abandonner une part du bénéfice. Une entente n'irait jamais jusqu'à éliminer

totalemment la probabilité d'une détection, car cela supposerait de ne pratiquer aucune hausse du prix. (Harrington and Imhof, 2022<sup>[10]</sup>).

<sup>8</sup> « *Le perfectionnement et l'usage plus répandu des outils logiciels de filtrage conduiront les participants à des ententes à peaufiner les techniques de soumission concertée afin de les rendre invisibles à ces outils. Réciproquement, le perfectionnement des méthodes de soumission concertée exigera le développement de meilleurs outils de filtrage. En conséquence, nous sommes au début d'une autre joute entre les autorités de la concurrence et les membres d'ententes* », (Lianos, 2021<sup>[37]</sup>).

<sup>9</sup> Une fonction d'offre précise les variables qui déterminent la valeur des offres. Elle est généralement estimée à l'aide d'une analyse de régression, la valeur de l'offre étant la variable dépendante et les facteurs qui déterminent la valeur de l'offre (tels que les coûts d'un projet) étant les variables indépendantes.

<sup>10</sup> Cette classification est en partie fondée sur une présentation récente de Jens Roat Kultima, de l'Autorité danoise de la concurrence et de la consommation, lors de la conférence CMA Data, Technology and Analytics 2022 sur « la détection de collusion dans les marchés publics à l'aide de méthodes informatiques », <https://www.youtube.com/watch?v=lovsp5aHcuU> (2:38:33)

<sup>11</sup> Beth et Gannon résument l'étendue des approches de filtrage des ententes en évoquant : (1) les filtres de vente par rapport aux filtres d'approvisionnement ; (2) les filtres structurels par rapport aux filtres comportementaux ; (3) les filtres accompagnés d'antécédents (c'est-à-dire les croyances ou connaissances préalables désignées dans la littérature statistique sous le nom d'antécédents) par rapport aux filtres sans antécédents ; et (4) les filtres d'appels d'offres par rapport aux filtres de prix de marché affichés (Beth and Gannon, 2022, pp. 78-80<sup>[11]</sup>).

<sup>12</sup> (Harrington and Imhof, 2022<sup>[10]</sup>) fournissent deux exemples plus anciens d'anomalies. Premièrement, les entreprises ne facturent pas systématiquement certains prix, notamment lorsque les teneurs de marché du Nasdaq évitent les cotations d'offre et de demande se terminant par un huitième impair, règle simple pour soutenir une majoration prix-coût plus élevée (Christie and Schultz, 1994<sup>[103]</sup>). Deuxièmement, les entreprises qui pratiquent des prix plus bas lorsque le coût est plus élevé. Au rang de celles-ci, on peut citer certains fournisseurs de lait qui ont soumis des offres à des tarifs bas lors des appels d'offres pour le lait destiné aux écoles des districts les plus éloignés de leurs usines (et dont les coûts de transport étaient par conséquent plus élevés) (Porter and Zona, 1999<sup>[18]</sup>).

<sup>13</sup> Épisode 14 : Cartel Screening and Machine Learning (Harrington & Imhof) <https://www.youtube.com/watch?v=NVholi8mFys> (18:13)

<sup>14</sup> « *Afin de comprendre pourquoi cela pourrait être le cas, imaginons le marché 'idéal' pour la collusion : deux entreprises, des produits homogènes, une demande stable, aucun gros acheteur, un excédent de capacité, etc. Même en supposant que ce marché serait certainement repéré par un outil d'enquête structurel, ma première conviction est qu'une très grande fraction de ces marchés fonctionne sans ententes ... Il existe de multiples équilibres – certains impliquent une collusion, d'autres non – et des variables non observées peuvent avoir une influence sur la question de savoir si des entreprises choisissent un équilibre collusif* » (Harrington, 2006<sup>[7]</sup>).

<sup>15</sup> « *Le fait de ne pas trouver de preuves d'une collusion peut être dû à une définition erronée du modèle de collusion ; par exemple, nous nous sommes concentrés sur le mauvais équilibre collusif* » (Harrington, 2006<sup>[7]</sup>). Un filtre d'entente ne vaut souvent que ce que valent les données sur lesquelles il a été formé. Huber et al. ont découvert un nombre accru de faux négatifs lorsqu'ils ont utilisé un modèle formé sur les données d'enchères japonaises afin de l'appliquer à des données d'enchères suisses. La raison tient au

fait que les statistiques de filtrage et le coefficient de variation étaient très similaires pour des soumissions collusives suisses et pour des soumissions japonaises de pleine concurrence, ce qui signifiait que le modèle n'était pas capable d'identifier effectivement la collusion dans les données suisses (Huber, Imhof and Ishii, 2020<sup>[43]</sup>)

<sup>16</sup> Il s'agit d'outils basés sur des algorithmes, qui apprennent de manière autonome à partir de données sur la passation de marchés publics ou sur le marché, si et comment des fournisseurs s'entendent de manière illégale, en détectant des différences dans la dispersion des prix collusifs et des prix concurrentiels (<https://martinhuber.shinyapps.io/carteldetection/>).

<sup>17</sup> C'est un exemple de logiciel développé dans le secteur privé, qui utilise réseau neuronal d'intelligence artificielle afin de détecter des ententes et des comportements anticoncurrentiels, dans le cadre d'un programme de conformité de l'entreprise ([www.dlapiper.com/fr/france/insights/publications/2022/02/lawand-aiscension-ai-tool-to-ensure-effective-risk-management-detect-anti-competitive-practices/](http://www.dlapiper.com/fr/france/insights/publications/2022/02/lawand-aiscension-ai-tool-to-ensure-effective-risk-management-detect-anti-competitive-practices/))

<sup>18</sup> La loi de 2018 qui a modifié le code pénal (Crimes Legislation Amendment (Powers, Offences and Other Measures) Act 2018) (« la Loi CLAPOOM ») permet de collecter, d'utiliser et de divulguer des informations à caractère personnel qui peuvent être importantes à des fins d'intégrité ([https://www.aph.gov.au/Parliamentary\\_Business/Bills\\_LEGislation/Bills\\_Search\\_Results/Result?bld=r5838](https://www.aph.gov.au/Parliamentary_Business/Bills_LEGislation/Bills_Search_Results/Result?bld=r5838))

<sup>19</sup> « Outre l'existence d'un site web central, sur lequel sont publiés tous les avis d'appels d'offres pour des marchés publics, le format ou la 'lisibilité par machine' des informations est également essentiel pour l'analyse quantitative des données sur la passation des marchés publics. Bien que l'accès aux données grâce à une interface de programmation d'application (API) ou un téléchargement direct de la base de données soient le meilleur mode d'accès, les pays divulguent pour la plupart des fichiers structurés au format xml ou html/pdf [...]. Bien que les fichiers xml puissent être facilement convertis dans une base de données analysable, les informations contenues dans les fichiers html et pdf sont souvent difficiles à extraire (Fazekas and Tóth, 2016<sup>[22]</sup>)

<sup>20</sup> « Des bases de données publiquement disponibles sur les marchés publics ne contiennent généralement pas d'informations précises sur les prix des offres et d'autres informations pertinentes pour les filtres d'ententes, notamment l'identité des sous-traitants. L'incidence importante des données manquantes et le manque courant d'identifiants des acteurs du marché sont autant d'obstacles supplémentaires » (Fazekas et al., 2022<sup>[21]</sup>)

<sup>21</sup> « La CNMC construit sa propre base de données sur la passation des marchés publics, sur la base du téléchargement sélectionné de certaines données [figurant sur la plateforme de dématérialisation des marchés publics], qui filtre et nettoie automatiquement les erreurs manifestes, et catégorise les données par niveaux de qualité » (traduction) (Campuzano, 2021<sup>[40]</sup>)

<sup>22</sup> (Fazekas and Tóth, 2016<sup>[22]</sup>) fixent les objectifs d'un cadre quantitatif de détection de la collusion (appliqué au système de passation des marchés publics suédois, mais valable en général), comme suit : (1) constituer un pipeline de données automatisé, actualisé en continu, alimenté par des fournisseurs de données sur la passation de marchés publics, afin de garantir la qualité de la base de données et la disponibilité des données en temps opportun ; (2) collecter les informations manquantes sur certains marchés à haut risque sélectionnés, afin d'améliorer la qualité des données et de déployer des indicateurs pour qu'ils fournissent leur plein potentiel ; (3) collecter et corrélérer des données supplémentaires avec les informations fournies par les registres sur la passation des marchés publics, en particulier avec des

informations provenant du registre des sociétés, des informations financières et des informations sur la propriété ; (4) réaliser des contrôles fondés sur les risques, en utilisant les indicateurs identifiés pour des marchés au titre desquels il existe des données de qualité suffisante ; (5) commencer à collecter systématiquement des informations sur les ententes dans la passation de marchés publics, afin de les entrer dans le système d'indicateurs en vue d'améliorer ce système.

<sup>23</sup> Plusieurs projets visent à rassembler des données sur la passation de marchés publics dans un format standardisé, notamment le projet EU DIGIWHIST (lanceur d'alerte numérique). Le projet collecte et analyse les données sur la passation des marchés publics dans 35 juridictions, y compris tous les membres de l'Union européenne, voir <https://digiwhist.eu/about-digiwhist/>

<sup>24</sup> (Fazekas and Tóth, 2016<sup>[22]</sup>) ont fait des recommandations politiques afin d'améliorer la qualité des données sur la passation de marchés publics, applicables au système suédois de passation de ces marchés, qui peuvent s'appliquer à un plus grand nombre de juridictions. Ces recommandations sont les suivantes : (1) garantir un processus uniforme de saisie des données et sauvegarder la qualité des données, en utilisant idéalement la plateforme centrale de dématérialisation des marchés publics et sous la supervision d'une autorité publique dédiée ; (2) introduire des formulaires standard définissant le contenu minimum des données requises ; (3) exiger la collecte d'informations sur l'exécution des marchés, notamment la modification du marché, la valeur finale totale du marché et la date effective d'achèvement, ainsi que des informations sur des facteurs soutenant l'analyse du risque de collusion, notamment sur les groupements d'entreprises et la sous-traitance.

<sup>25</sup> « *Un économiste peut être parfaitement positionné afin d'identifier des variables à inclure dans X afin d'obtenir une prédiction utile. On peut soutenir que le risque majeur est le risque de surapprentissage, le risque que l'ordinateur identifie une connexion fautive qui peut se produire dans l'ensemble de données de formation et considère qu'elle se produira toujours. Le meilleur moyen de lutter contre ce surapprentissage est d'appliquer la même sorte de théorie économique que celle qu'utilise un économétricien* », (Abrantes-Metz and Metz, 2018<sup>[23]</sup>)

<sup>26</sup> « *Le coût marginal de copie du code est proche de zéro. Les problèmes que les autorités de la concurrence rencontrent de part le monde sont très similaires, et, dans le cas d'entreprises multinationales, nous contrôlons, réglémentons et nous attaquons souvent à un comportement identique ou similaire. Dans la mesure où les autorités de la concurrence peuvent se partager le code – pour les pipelines de données, l'extraction de données, les logiciels/outils, et l'analyse – elles peuvent bénéficier de certaines des mêmes forces numériques que celles qui remodelent les marchés* », (Hunt, 2022<sup>[38]</sup>).

<sup>27</sup> « *Les données synthétiques [...] aident à former des algorithmes d'apprentissage automatique qui nécessitent un volume immense de données de formation étiquetées, ce qui peut être coûteux ou s'accompagner de restrictions à l'utilisation des données* ». Toutefois, « *il convient d'évaluer les garanties en matière de protection de la vie privée, afin de veiller à ce que les données synthétiques en résultant ne soient pas des données à caractère personnel. Cette garantie de protection de la vie privée évalue la mesure dans laquelle les personnes concernées peuvent être identifiées dans les données synthétiques et combien de données nouvelles sur ces personnes concernées seraient révélées en cas d'identification réussie* », [https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data\\_en](https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en)

<sup>28</sup> Episode 14 : Cartel Screening and Machine Learning (Harrington & Imhof) <https://www.youtube.com/watch?v=NVholi8mFys> (37:57)



<sup>29</sup> Les informations sensibles de l'autorité sont « *des informations créées ou détenues par une autorité qui ne sont pas dans le domaine public, si l'autorité n'est pas tenue d'une obligation légale de non-divulgence mais si ces informations sont considérées comme confidentielles ou sensibles par l'autorité. [...] l'analyse et la compilation d'informations publiques par une autorité peuvent transformer les informations publiques en 'informations confidentielles de l'autorité', par exemple en procédant à l'analyse de données complexes* » (OCDE/RIC, 2021<sup>[46]</sup>)

<sup>30</sup> « *En effet, la preuve d'une collusion n'implique pas nécessairement la preuve d'une infraction au droit de la concurrence* » (Harrington, 2006<sup>[71]</sup>)

<sup>31</sup> Il a été soutenu qu'une preuve indirecte devrait être encore moins retenue dans une procédure administrative que dans une procédure pénale, étant donné que la norme de preuve est plus clémentine dans les procédures administratives (qui exigent généralement l'existence de preuves substantielles), et sachant que la décision de première instance est généralement rendue par une autorité de la concurrence et non par un tribunal qui suit une procédure plus rigoureuse (OCDE, 2006<sup>[49]</sup>), contribution du BIAC.

<sup>32</sup> Jens Roat Kultima, Autorité danoise de la concurrence et de la consommation, lors de la conférence CMA Data, Technology and Analytics 2022 sur « la détection de la collusion dans les marchés publics à l'aide de méthodes informatiques », <https://www.youtube.com/watch?v=lovsp5aHcuU> (2:38:33).

<sup>33</sup> (Ishii, 2014<sup>[62]</sup>) figure à l'0 bien que le document date de 2014. En effet, l'0 n'inclut que les articles universitaires de la période 2015 à 2022.

<sup>34</sup> Episode 14 : Cartel Screening and Machine Learning (Harrington & Imhof) <https://www.youtube.com/watch?v=NVholi8mFys> (22:39)

# Bibliographie

- Abrantes-Metz, R. (2012), "Why and How to Use Empirical Screens in Antitrust Compliance", *CPI Antitrust Chronicle* 1, <http://cendoc.sc.gob.sv/textocompleto/927.pdf>. [33]
- Abrantes-Metz, R. et al. (2006), "A variance screen for collusion", *International Journal of Industrial Organization* vol. 24, n° 3, pp. pp. 467-486, <https://www.sciencedirect.com/science/article/pii/S016771870500158X>. [20]
- Abrantes-Metz, R. and A. Metz (2020), "Why Screening Is a 'Must Have' Tool for Effective Antitrust Compliance Programs", [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3734004](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3734004). [34]
- Abrantes-Metz, R. and A. Metz (2019), "The Future of Cartel Deterrence and Detection", *CPI Antitrust Chronicle, January*, <https://papers.ssrn.com/abstract=3360615>. [14]
- Abrantes-Metz, R. and A. Metz (2018), "Can Machine Learning Aide in Cartel Detection?", *Antitrust Chronicle, Competition Policy International*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3291633](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3291633). [23]
- Abrantes-Metz, R. and A. Metz (2012), "How far can screens go in distinguishing explicit from tacit collusion? New evidence from the LIBOR setting", *CPI Antitrust Chronicle*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2021515](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2021515). [55]
- Athey, S. and G. Imbens (2019), "Machine learning methods that economists should know about", *Annual Review of Economics*, Vol. n° 11, pp. pp. 685-725, <https://www.annualreviews.org/doi/full/10.1146/annurev-economics-080217-053433>. [83]
- Bajari, P. and L. Ye (2003), "Deciding between competition and collusion", *Review of Economics and statistics* vol. 85, n° 4, pp. pp. 971-989, <https://direct.mit.edu/rest/article/85/4/971/57436/Deciding-Between-Competition-and-Collusion>. [19]
- Barto, A. and T. Dietterich (2004), "Reinforcement learning and its relationship to supervised learning", *Handbook of learning and approximate dynamic programming*, <https://web.engr.oregonstate.edu/~tgd/publications/Barto-Dietterich-03.pdf>. [30]
- Bergman, M. et al. (2020), "Interactions across firms and bid rigging", *Review of Industrial Organization* vol. 56, n° 1, pp. pp. 107-130, <https://link.springer.com/article/10.1007/s11151-018-09676-0>. [65]
- Bergman, M. et al. (2015), "Using spatial econometrics to test for collusive behavior in procurement auction data". [66]

- Beth, H. and O. Gannon (2022), “Cartel screening—can competition authorities and corporations afford not to use big data to detect cartels?”, *Competition Law & Policy Debate* vol. 7, n° 2, pp. pp.77-88, <https://www.elgaronline.com/view/journals/clpd/7/2/article-p77.xml>. [11]
- Bharadwaj, B. et al. (2022), “Game, Set and Fuzzy match”, <https://www.compasslexecon.com/game-set-and-fuzzy-match/>. [41]
- Bolotova, Y., J. Connor and D. Miller (2008), “The impact of collusion on price behavior: Empirical results from two recent cases”, *International Journal of Industrial Organization*, Vol. vol. 26, n° 6, pp. pp. 1290–1307, <https://www.sciencedirect.com/science/article/pii/S0167718708000039>. [54]
- Buccirossi, P., G. Di Pierro and L. Giangregorio (n.d.), “Detecting Bid-Rigging in the “Big Data Era””, [http://ippa.org/images/BOOKS/IPPC8/Chapter5\\_Buccirossi\\_Di-Pierro\\_Giangregorio.pdf](http://ippa.org/images/BOOKS/IPPC8/Chapter5_Buccirossi_Di-Pierro_Giangregorio.pdf). [104]
- Bureau de la concurrence Canada (2022), *Outil d'évaluation du degré de risque de collusion*, <https://www.canada.ca/fr/bureau-concurrence/nouvelles/2022/06/specialistes-de-lapprovisionnement-utilisez-notre-outil-devaluation-du-degre-de-risque-de-collusion-pour-protoger-vos-contrats-contre-le-truquage-d.html>. [39]
- Campuzano, S. (2021), “Riesgos y oportunidades de la inteligencia artificial desde la perspectiva de la competencia. Un análisis desde la CNMC”, *Boletín Económico de la ICE*, <https://doi.org/10.32796/bice.2021.3137.7259>. [40]
- Chassang, S. (2022), *Robust Screens for Noncompetitive Bidding in Procurement Auctions*, pp. pp. 315-346, <https://doi.org/10.3982/ECTA17155>. [71]
- Christie, W. and P. Schultz (1994), “Why do NASDAQ market makers avoid odd-eighth quotes?”, *The Journal of Finance* vol. 49, n° 5, pp. pp. 1813-1840, <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1994.tb04782.x>. [103]
- Clark, R., D. Coviello and A. Leverano (2020), “Complementary bidding and the collusive arrangement: Evidence from an antitrust investigation”, *ZEW Discussion Papers* 20, <https://madoc.bib.uni-mannheim.de/57783/1/dp20052.pdf>. [69]
- Clemens, G. (2017), “Raising Rivals’ Costs Through Cartel Detection - Why Downstream Buyers Rather Face an Upstream Cartel than Downstream Competition”, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2946114](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2946114). [102]
- Connor, J. (2016), *The Private International Cartels (PIC) Data Set: Guide and Summary Statistics, 1990-July 2016*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2821254](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2821254). [13]
- Contrôleur européen de la protection des données (n.d.), *Publications du CEPD*. [45]
- Crede, C. (2019), “A structural break cartel screen for dating and detecting collusion”, *Review of Industrial Organization*, Vol. vol. 54, n° 3, pp. pp. 543-574, <https://link.springer.com/content/pdf/10.1007/s11151-018-9649-5.pdf>. [78]
- Crede, C. (2016), “Getting a fix on price-fixing cartels”, *Significance* vol. 13, n° 1, pp. Pp. 38-41, <https://rss.onlinelibrary.wiley.com/doi/pdfdirect/10.1111/j.1740-9713.2016.00882.x>. [79]

- Czibik, Á., B. Tóth and M. Fazekas (2015), “How to Construct a Public Procurement Database from Administrative Records. With examples from the Hungarian public procurement system of 2009–2012”, [https://www.govtransparency.eu/wp-content/uploads/2017/11/GTI\\_publicprocurement\\_techreport\\_171109.pdf](https://www.govtransparency.eu/wp-content/uploads/2017/11/GTI_publicprocurement_techreport_171109.pdf). [80]
- de Andrade Lima, R. and G. Resende (2021), “Using the Moran’s I to detect bid rigging in Brazilian procurement auctions”, *The Annals of Regional Science* vol. 66, n° 2, pp. pp. 237-254, <https://link.springer.com/article/10.1007/s00168-020-01018-x>. [68]
- Deng, A. (2020), “Algorithmic Collusion and Algorithmic Compliance: Risks and Opportunities”, *The Global Antitrust Institute Report on the Digital Economy* 27, <https://gaidigitalreport.com/wp-content/uploads/2020/11/Deng-Algorithmic-Collusion-and-Algorithmic-Compliance.pdf>. [101]
- Deng, A. (2017), “An Antitrust Lawyer’s Guide to Machine Learning”, *Antitrust* vol. 33, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3082514](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3082514). [81]
- Deng, A. (2017), “Cartel detection and monitoring: a look forward”, *Journal of Antitrust Enforcement* vol. 5, n° 3, pp. pp. 488-500, <https://academic.oup.com/antitrust/article-pdf/5/3/488/21390250/jnw017.pdf>. [29]
- Doane, M. et al. (2015), “Screening for collusion as a problem of inference”, *Oxford handbook of international antitrust economics* n° 2, pp. Pp. 523-553, <http://www.competitioneconomics.com/wp-content/uploads/2013/05/Screening-for-Collusion-as-a-Problem-of-Inference.pdf>. [100]
- Fazekas, M. et al. (2022), “Public procurement cartels: A systematic testing of old and new screens”, *Government Transparency Institute*, [http://www.govtransparency.eu/wp-content/uploads/2022/03/GTI-WP-Cartel\\_20220304-1.pdf](http://www.govtransparency.eu/wp-content/uploads/2022/03/GTI-WP-Cartel_20220304-1.pdf). [21]
- Fazekas, M. and B. Tóth (2016), “Assessing the potential for detecting collusion in Swedish public procurement”, *Konkurrensverket, Stockholm*. [22]
- Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press. [64]
- Harrington, J. (2021), “Cartel screening is for companies, law firms, and economic consultancies, not just competition authorities”. [31]
- Harrington, J. (2021), “The Practical Requirements of a Successful Cartel”, *Research Handbook on Cartels (Peter Whelan, dir. pub.)*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3798852](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3798852). [98]
- Harrington, J. (2015), “Thoughts on Why Certain Markets are More Susceptible to Collusion and Some Policy Suggestions for Dealing with Them”, *OECD Expert background paper (Global Forum on Competition)*, [https://joeharrington5201922.github.io/pdf/OECD\\_Background%20Paper\\_Harrington.pdf](https://joeharrington5201922.github.io/pdf/OECD_Background%20Paper_Harrington.pdf). [32]
- Harrington, J. (2008), “Detecting cartels”, *Handbook of Antitrust Economics* p. 213, pp. pp. 215-258, [http://mis.kp.ac.rw/admin/admin\\_panel/kp\\_lms/files/digital/SelectiveBooks/Economics/Handbook%20of%20Antitrust%20Economics.%20Edited%20\[Buccrossi,%20Paolo.\].pdf.pdf#page=236](http://mis.kp.ac.rw/admin/admin_panel/kp_lms/files/digital/SelectiveBooks/Economics/Handbook%20of%20Antitrust%20Economics.%20Edited%20[Buccrossi,%20Paolo.].pdf.pdf#page=236). [99]

- Harrington, J. (2006), "Behavioral screening and the detection of cartels", *European competition law annual*, pp. 51-68, <https://joeharrington5201922.github.io/pdf/Florence.pdf>. [7]
- Harrington, J. and D. Imhof (2022), "Cartel Screening and Machine Learning", *Stanford Computational Antitrust*, <https://www-cdn.law.stanford.edu/wp-content/uploads/2022/08/harrington-imhof-2022.pdf>. [10]
- Hastie, T. et al. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: springer, <https://link.springer.com/content/pdf/10.1007/978-0-387-84858-7.pdf>. [28]
- Heijnen, P., M. Haan and A. Soetevent (2015), "Screening for collusion: a spatial statistics approach", *Journal of Economic Geography* vol. 15, n° 2, pp. pp. 417-448, <https://academic.oup.com/joeg/article/15/2/417/928498>. [47]
- Huber, M. and D. et Imhof (2021), "Deep learning for detecting bid rigging: Flagging cartel participants based on convolutional neural networks", *arXiv preprint arXiv:2104.11142*. [24]
- Huber, M. and D. Imhof (2019), "Machine learning with screens for detecting bid-rigging cartels", *International Journal of Industrial Organization* n° 65, pp. pp. 277-301. [26]
- Huber, M., D. Imhof and R. Ishii (2020), "Transnational machine learning with screens for flagging bid-rigging cartels", *Université de Fribourg*, [https://doc.rero.ch/record/329575/files/WP\\_SES\\_519.pdf](https://doc.rero.ch/record/329575/files/WP_SES_519.pdf). [43]
- Hunt, S. (2022), "The technology-led transformation of competition and consumer agencies: the Competition and Markets Authority's experience", [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1082753/Stefan\\_discussion\\_paper.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082753/Stefan_discussion_paper.pdf). [38]
- Imhof, D. (2018), *Empirical Methods for Detecting Bid-rigging Cartels*, Université Bourgogne Franche-Comté, <https://tel.archives-ouvertes.fr/tel-01963076/document>. [75]
- Imhof, D. (2017), "Simple statistical screens to detect bid rigging", *Université de Fribourg*. [57]
- Imhof, D., Y. Karagök and S. Rutz (2018), "Screening for Bid Rigging—Does It Work?", *Journal of Competition Law & Economics* vol. 14, n° 2, pp. pp. 235-261. [27]
- Imhof, D. and H. Wallimann (2021), "Detecting bid-rigging coalitions in different countries and auction formats", *International Review of Law and Economics*. [25]
- Ishii, R. (2014), "Bid roundness under collusion in Japanese procurement auctions", *Review of Industrial Organization*, Vol. vol. 44, n° 3, pp. pp. 241-254, <https://link.springer.com/content/pdf/10.1007/s11151-013-9408-6.pdf>. [62]
- James, G. et al. (2013), *An introduction to statistical learning*, Springer, <https://link.springer.com/book/10.1007/978-1-4614-7138-7>. [84]
- Jiménez, J. and J. Perdiguero (2012), "Does rigidity of prices hide collusion?", *Review of Industrial Organization* vol. 41, n° 3, pp. pp. 223-248, <https://link.springer.com/content/pdf/10.1007/s11151-012-9337-9.pdf>. [56]
- Johnson, J. and D. Sokol (2020), "Understanding AI Collusion and Compliance", *Cambridge Handbook of Compliance*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3413882](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413882). [35]

- Kawai, K. (2022), *Detecting Large-Scale Collusion in Procurement Auctions*, [97]  
<http://www.journals.uchicago.edu/doi/10.1086/718913>.
- Kawai, K. (2022), *Using Bid Rotation and Incumbency to Detect Collusion: A Regression Discontinuity Approach*, [15]  
<https://doi.org/10.1093/restud/rdac013>.
- Kultima, J. (2022), “Collusion detection in public procurement using computational methods”, [96]  
*Danish Competition and Consumer Authority: Competitive Markets and Consumer Welfare* n° 56 (avril), [https://www.en.kfst.dk/media/cnldn11q/bid-viewer\\_56\\_seneste.pdf](https://www.en.kfst.dk/media/cnldn11q/bid-viewer_56_seneste.pdf).
- Kurdoglu, B. and E. Yucel (2022), *A Cointegration-based cartel screen for detecting collusion*, [63]  
[https://mpa.ub.uni-muenchen.de/113888/1/MPRA\\_paper\\_113888.pdf](https://mpa.ub.uni-muenchen.de/113888/1/MPRA_paper_113888.pdf).
- Laitenberger, U. and K. Huschelrath (2011), “The adoption of screening tools by competition authorities”, [95]  
*Antitrust Chronicle* septembre,  
<https://www.competitionpolicyinternational.com/assets/0d358061e11f2708ad9d62634c6c40ad/HuschelrathSEP-112.pdf>.
- Lamontanaro, A. (2019), “Bounty Hunters For Algorithmic Cartels: An Old Solution for a New Problem”, [94]  
*Fordham Intell. Prop. Media & Ent. LJ* vol. 30, n° 4,  
<https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1760&context=iplj>.
- Lianos, I. (2021), “Computational Competition Law and Economics - Inception Report”, [37]  
<https://www.epant.gr/en/enimerosi/publications/research-publications/item/1414-computational-competition-law-and-economics-inception-report.html>.
- Lundberg, J. (2017), “On cartel detection and Moran’s I”, [67]  
*Letters in Spatial and Resource Sciences* vol. 10, n° 1, pp. pp. 129-139, <https://link.springer.com/article/10.1007/s12076-016-0176-4>.
- Marvão, C. and G. Spagnolo (2016), “Cartels and leniency: Taking stock of what we learnt”, [93]  
*Handbook of Game Theory and Industrial Organization*, Vol. II,  
<https://www.econstor.eu/bitstream/10419/204750/1/site-wp0039.pdf>.
- Nicholls, R. (2021), “Regtech as an antitrust enforcement tool”, [92]  
*Journal of Antitrust Enforcement* vol. 9, n° 1, pp. pp. 135-151, <https://academic.oup.com/antitrust/article-pdf/9/1/135/37173888/jnaa011.pdf>.
- OCDE (2022), “OECD Competition Trends 2022”, [12]  
<https://www.oecd.org/daf/competition/oecd-competition-trends-2022.pdf>.
- OCDE (2021), “Les programmes de conformité au droit de la concurrence”, [36]  
*Comité de la concurrence de l’OCDE*, [https://one.oecd.org/document/DAF/COMP/WP3\(2021\)1/fr/pdf](https://one.oecd.org/document/DAF/COMP/WP3(2021)1/fr/pdf).
- OCDE (2021), “OECD Business and Finance Outlook 2021”, [1]  
[https://www.oecd-ilibrary.org/sites/ba682899-en/1/3/5/index.html?itemId=/content/publication/ba682899-en&csp\\_02d27ef0d7308d76a010fd2a9882228f&itemIGO=oecd&itemContentType=book](https://www.oecd-ilibrary.org/sites/ba682899-en/1/3/5/index.html?itemId=/content/publication/ba682899-en&csp_02d27ef0d7308d76a010fd2a9882228f&itemIGO=oecd&itemContentType=book).
- OCDE (2021), *Recommandation sur la transparence et l’équité procédurale dans la mise en œuvre du droit de la concurrence*, [53]  
<https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0465>.
- OCDE (2020), “Digital Evidence Gathering in Cartel Investigations”, [6]  
[https://one.oecd.org/document/DAF/COMP/LACF\(2020\)2/en/pdf](https://one.oecd.org/document/DAF/COMP/LACF(2020)2/en/pdf).

- OCDE (2020), "Economic analysis in merger investigations", [90]  
<https://www.oecd.org/daf/competition/economic-analysis-in-merger-investigations-2020.pdf>.
- OCDE (2020), "Merger Control in Dynamic Markets", [89]  
<https://www.oecd.org/daf/competition/merger-control-in-dynamic-markets-2020.pdf>.
- OCDE (2019), *La norme de contrôle appliquée par les tribunaux dans les affaires de concurrence*, [51]  
<https://www.oecd.org/fr/daf/concurrence/la-norme-de-contrôle-appliquée-par-les-tribunaux-dans-les-affaires-de-concurrence.htm>.
- OCDE (2019), *L'Accès au dossier et protection de l'information confidentielle*, [52]  
<https://www.oecd.org/fr/concurrence/laccesaudossieretprotectiondelinformationconfidentielle.htm>.
- OCDE (2018), "Défis et coordination des programmes de clémence", [91]  
[https://one.oecd.org/document/DAF/COMP/WP3\(2018\)1/fr/pdf](https://one.oecd.org/document/DAF/COMP/WP3(2018)1/fr/pdf).
- OCDE (2018), *OECD-BWB Workshop on Complex Cartel Case Management*, [2]  
<https://www.oecd.org/daf/competition/oecd-bwb-workshop-on-complex-cartel-case-management.htm>.
- OCDE (2018), "Workshop on Cartel Screening in the Digital Era", [3]  
<https://www.oecd.org/competition/workshop-on-cartel-screening-in-the-digital-era.htm>.
- OCDE (2017), "Algorithms and Collusion: Competition Policy in the Digital Age", [4]  
<http://www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm>.
- OCDE (2015), *Recommandation sur les marchés publics*, [105]  
<https://www.oecd.org/fr/gov/commande-publique/Recommandation-OCDE-sur-les-marchés-publics.pdf>.
- OCDE (2013), "Ex officio cartel investigations and the use of screens to detect cartels", [5]  
<http://www.oecd.org/daf/competition/exofficio-cartel-investigations.htm>.
- OCDE (2012), *Recommandation sur la lutte contre les soumissions concertées dans les marchés publics*, [42]  
<https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0396>.
- OCDE (2009), *Lignes directrices pour la lutte contre les soumissions concertées dans les marchés publics*, [9]  
<https://www.oecd.org/fr/daf/concurrence/42340181.pdf>.
- OCDE (2006), *Prosecuting Cartels Without Direct Evidence*, [49]  
<http://www.oecd.org/daf/competition/prosecutionandlawenforcement/37391162.pdf>.
- OCDE (n.d.), *Glossary of Statistical Terms*. [44]
- OCDE/RIC (2021), *Report on International Co-operation in Competition Enforcement*, [46]  
<http://www.oecd.org/competition/oecd-icn-report-on-international-cooperation-in-competition-enforcement-2021.htm>.
- Ortner, J. et al. (2022), "Screening adaptive cartels" n° 2020-59, [72]  
[https://people.bu.edu/jortner/index\\_files/screening\\_adaptive\\_cartels.pdf](https://people.bu.edu/jortner/index_files/screening_adaptive_cartels.pdf).
- Pavanelli de Lorenzi, C. (2022), *CADE eyes development of gun-jumping detection tool*. [88]

- Perdiguero, J. and J. Jiménez (2021), “Price coordination in the spanish oil market: the monday effect”, *Energy Policy* vol. 149, n° 112016, <https://www.sciencedirect.com/science/article/pii/S0301421520307278>. [70]
- Porter, R. and J. Zona (1999), “Ohio school milk markets: An analysis of bidding”, *The RAND Journal of Economics*, Vol. n° 30/2, pp. pp.2 63-288, <https://www.jstor.org/stable/2556080>. [18]
- Porter, R. and J. Zona (1997), “Ohio school milk markets: An analysis of bidding”, [https://www.nber.org/system/files/working\\_papers/w6037/w6037.pdf](https://www.nber.org/system/files/working_papers/w6037/w6037.pdf). [17]
- Porter, R. and J. Zona (1993), “Detection of bid rigging in procurement auctions”, *Journal of political economy* vol. 101, n° 3, pp. Pp. 518-538, <https://www.journals.uchicago.edu/doi/abs/10.1086/261885>. [16]
- Ragazzo, C. (2012), *Screens in the gas retail market: the Brazilian experience*, <https://www.competitionpolicyinternational.com/screens-in-the-gas-retail-market-the-brazilian-experience/>. [48]
- Rodríguez, M. et al. (2022), “Collusion detection in public procurement auctions with machine learning algorithms”, *Automation in Construction* vol. 133, n° 104047, <https://www.sciencedirect.com/science/article/pii/S0926580521004982>. [61]
- Schrepel, T. and T. Groza (2022), “The Adoption of Computational Antitrust by Agencies: 2021 Report”, *Stanford Computational Antitrust*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4142225](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4142225). [87]
- Signor, R. (2021), “Collusion detection in infrastructure procurement: A modified order statistic method for uncapped auctions”, *IEEE Transactions on Engineering Management*, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9336335>. [74]
- Signor, R. et al. (2020), “Detection of collusive tenders in infrastructure projects: learning from operation car wash”, *Journal of Construction Engineering and Management*, Vol. vol. 146, n° 1, <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CO.1943-7862.0001737>. [77]
- Signor, R. et al. (2017), “Collusive bidding in Brazilian infrastructure projects”, *Proceedings of the Institution of Civil Engineers-Forensic Engineering* vol. 170, n° 3, pp. pp. 113-123, <https://espace.curtin.edu.au/bitstream/handle/20.500.11937/59452/257611.pdf?sequence=2>. [76]
- Silveira, D. et al. (2021), “Cartel screening in the Brazilian fuel retail market”, *Economia* vol. 22, n° 1, pp. pp. 53-70, <https://www.sciencedirect.com/science/article/pii/S1517758021000011>. [59]
- Silveira, D. et al. (2022), “Won’t get fooled again: A supervised machine learning approach for screening gasoline cartels”, *Energy Economics* vol. 105, n° 105711, <https://www.sciencedirect.com/science/article/pii/S0140988321005594>. [60]
- Torres Berru, Y. et al. (2019), “Artificial Intelligence techniques to detect and prevent corruption in procurement: a systematic literature review”, *International Conference on Applied Technologies*, pp. pp. 254-268, [https://link.springer.com/chapter/10.1007/978-3-030-42520-3\\_21](https://link.springer.com/chapter/10.1007/978-3-030-42520-3_21). [85]
- Vadász, P. et al. (2017), *Chapter 16 Identifying Illegal Cartel Activities from Open Sources*, Springer, <http://197.156.112.159/bitstream/handle/123456789/862/Babak%20Akhgar.pdf?sequence=1&isAllowed=y>. [86]



- Varian, H. (2014), "Big data: New tricks for econometrics", *Journal of Economic Perspectives*, [82]  
Vol. vol. 28, n° 2, pp. pp. 3-28, <https://pubs.aeaweb.org/doi/pdf/10.1257%2Fjep.28.2.3>.
- von Bonin Andreas, S. (2020), *The Use of Artificial Intelligence in the Future of Competition Law Enforcement*, [50]  
<https://doi.org/10.1093/jeclap/lpaa077>.
- Wachs, J. and J. Kertész (2019), "A network approach to cartel detection in public auction markets", *Scientific Reports* vol. 9, n° 1, pp. pp. 1-10, [73]  
<https://www.nature.com/articles/s41598-019-47198-1>.
- Wallimann, H., D. Imhof and M. Huber (2020), "A machine learning approach for flagging incomplete bid-rigging cartels", *arXiv preprint arXiv:2004.05629*, [58]  
<https://arxiv.org/pdf/2004.05629.pdf>.
- Zlatcu, I. and M. Suciu (2017), "The role of economics in cartel detection: A review of cartel screens", *Journal of Economic Development, Environment and People* vol. 6, n° 3, [8]  
pp. pp. 15-26, [http://jedep.spiruharet.ro/RePEc/sph/rjedep/JEDEP22\\_2Zlatcu\\_p16-26.pdf](http://jedep.spiruharet.ro/RePEc/sph/rjedep/JEDEP22_2Zlatcu_p16-26.pdf).