

Utilisation de Google Trends dans les enquêtes mensuelles sur le Commerce de Détail de la Banque de France

Use of Google Trends Data in Banque de France Monthly Retail Trade Surveys

François Robin*

Résumé – Dans le cadre du partenariat la liant à la Banque de France, la Fédération du e-commerce et de la vente à distance (FEVAD) fournit mensuellement le chiffre d'affaires réalisé en e-commerce auprès des particuliers, depuis 2012. Dans l'attente des livraisons, la Banque de France procède à des estimations, dont l'enjeu est renforcé par la croissance du e-commerce. Le modèle autorégressif (SARIMA(12)) utilisé jusqu'ici peut désormais être complété par d'autres modèles statistiques s'appuyant sur des données exogènes grâce à un historique plus long de données. Cet article détaille les différents choix opérés conduisant à la prévision finale : transformation des données, modèles à sélection de variables et stratégie pour la prévision. Les requêtes Google notamment, mesurées par Google Trends, permettent d'améliorer la capacité prédictive du modèle final, obtenu en combinant les modèles simples.

Abstract – Under its partnership with the Banque de France, the Federation of E-Commerce and Distance Selling (Fédération du e-commerce et de la vente à distance - FEVAD) has provided monthly consumer online retail sales data since 2012. Pending the release of new data, the Banque de France carries out estimations, a task complicated by the growth of online retail. The autoregressive model (SARIMA(12)) used up to now can now be complemented by other statistical models that draw on exogenous data with a longer historical time series. This paper details the system of choices that results in the final forecast: data conversion, variable selection methods and forecasting approaches. In particular, Google queries, as measured by Google Trends, help enhance the predictive accuracy of the final model, obtained by combining models.

Codes JEL / JEL Classification : C51, C53, C11, E17

Mots-clés : Google Trends, *nowcasting*, conjoncture, e-commerce, vente à distance, Big Data, *Bayesian averaging*, sélection de variables, lasso

Keywords: Google Trends, *nowcasting*, trends, e-commerce, distance selling, Big Data, *Bayesian averaging*, variable selection, lasso

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

* Banque de France, Service des Enquêtes Économiques de Conjoncture, Direction Générale des Statistiques (francois.robin@banque-france.fr)

L'auteur remercie tout particulièrement les participants à un atelier interne de recherche de la Banque de France, Patrick Kwok d'avoir continuellement suivi, contribué et soutenu ces travaux. Un très grand merci aussi à Martial Ranvier et Valérie Chauvin pour leurs remarques pertinentes et leurs conseils méthodologiques précieux. Enfin, merci à François Guinouard, Léon Ipdjian et Mathilde Gerardin ; et à François Brunet, qui est à l'origine du sujet.

Reçu le 9 août 2017, accepté après révisions le 14 janvier 2019

Dans le cadre du partenariat la liant à la Banque de France, la Fédération du e-commerce et de la vente à distance (FEVAD) fournit mensuellement le chiffre d'affaires réalisé en e-commerce¹ « B2C » (*Business to Consumer*). Cependant, ces livraisons sont trop tardives pour être intégrées à la première publication de l'enquête mensuelle de conjoncture (EMC) du Commerce de détail, pour laquelle ces données sont estimées.

Jusqu'à présent, la faible profondeur historique des données restreignait le champ des possibles techniques en termes de prévision. Le modèle autorégressif utilisé jusqu'ici peut maintenant être complété par des modèles utilisant des données exogènes disponibles au moment de l'estimation : les indices quantitatifs conjoncturels du commerce traditionnel (issus de l'enquête mensuelle de conjoncture du Commerce de Détail) et les données Google Trends. Concrètement, l'estimation des données de la Fédération du e-commerce et de la vente à distance (FEVAD) du mois M a lieu en période d'enquête (au début du mois $M + 1$). Les indices quantitatifs conjoncturels du mois M sont alors en cours de construction, et les données Google Trends du mois M sont définitives. Ces estimations rentrent pleinement dans le cadre d'un exercice de *nowcasting*.

Cette évolution rencontre deux problèmes. Premièrement, Google Trends fournit de nombreuses variables explicatives, parmi lesquelles les meilleures doivent être sélectionnées. Une méthode de *machine learning*, le « lasso adaptatif », développée par Zou (2006), répond à la contrainte duale du sujet, à savoir combiner la faible profondeur de l'historique des données FEVAD (les données commencent en 2012) avec l'immense champ des requêtes Google possibles. Ensuite, disposant de plusieurs modèles avec des variables exogènes d'origines différentes, il est intéressant de vérifier si la combinaison de modèles permet d'obtenir de meilleurs résultats. Ce sujet est largement débattu, comme l'expliquent Bec et Mogliani (2015).

Après la première partie situant l'article dans la littérature, la deuxième partie décrit les données, avec la présentation de l'enquête Commerce de Détail (CD), des données FEVAD et des données Google Trends. Les particularités et l'opacité de la méthodologie de construction de ces dernières nécessitent la mise en place de tests de robustesse et de corrections automatisées, liées aux ruptures de séries. La troisième partie porte sur les choix de modélisation. Y sont abordés successivement le

traitement de la stationnarité des séries, puis le processus complet de test des modèles. La quatrième partie est dédiée aux résultats et à leur interprétation. La dernière partie présente la conclusion.

Revue de littérature

Données Google Trends

Disponibles en quasi temps réel, les indices Google Trends fournissent l'évolution temporelle des requêtes effectuées par les utilisateurs sur le moteur de recherche Google. Ils constituent un flux d'information et une source de données massives. N'ayant pas trouvé de trace de leur utilisation par les institutions publiques pour des travaux récurrents, ils ont cependant fait l'objet de plusieurs publications. Les travaux d'Ettredge *et al.* (2005) ou ceux d'Askistas & Zimmerman (2009), consacrés à la prédiction du taux de chômage en utilisant des mots-clés recherchés dans Google, montrent l'intérêt de ces indicateurs. Choi & Varian (2009, 2011) se montrent plus mitigés quant à l'apport des Google Trends. Par ailleurs, leur revue de littérature recense de nombreuses publications utilisant les recherches Google, principalement dans le domaine de l'épidémiologie ; l'outil alors utilisé et développé par Google (Google Flu) a été supprimé (le 20 août 2015), suite à des défaillances déjà mentionnées par Bortoli & Combes (2015).

Ces outils sont complètement pilotés par Google : la méthodologie de construction est opaque, comportant de nouveaux risques pour les utilisateurs. Les changements de méthodologie de Google Trends sont susceptibles d'engendrer des ruptures de séries. Par ailleurs, l'émergence de nouveaux acteurs a un impact sur la formulation des requêtes par les utilisateurs. McLaren & Shanbhogue (2011) alertent quant aux baisses mécaniques de certaines requêtes dans leur application au chômage (par exemple en France, la baisse de la requête « ANPE » au profit de « Pôle emploi » suite à la restructuration de l'ANPE et des Assedic).

Sélection de variables

Les méthodes de *machine learning* apportent une solution à la sélection de variables, et plus

1. D'après la FEVAD, la part du e-commerce dans le commerce de détail (hors alimentaire, conformément au champ de l'enquête de conjoncture Commerce de Détail de la Banque de France) était de 7 % en 2013, 8 % en 2014 et 9 % en 2015.

particulièrement le lasso adaptatif développé par Zou (2006). Pour rappel, l'équation du lasso² classique introduit par Tibshirani (1996) est :

$$\hat{\beta}_{lasso} = \operatorname{argmin}_{\beta} Y - \sum_{j=1}^p x_j \beta_j^2 + \lambda \sum_{j=1}^p |\beta_j|, \lambda \geq 0$$

Dans la régression lasso, la même pénalité λ est appliquée à toutes les variables. Zou (2006) propose d'adapter la pénalité en fonction des variables dans le lasso adaptatif (*adlasso*) :

$$\hat{\beta}_{adlasso} = \operatorname{argmin}_{\beta} Y - \sum_{j=1}^p x_j \beta_j^2 + \lambda \sum_{j=1}^p w_j |\beta_j|, \begin{cases} \lambda \geq 0 \\ w_j \geq 0 \end{cases}$$

Le lasso adaptatif est un lasso pondéré. Ses propriétés d'Oracle, démontrées par Zou (2006), lui confèrent deux avantages sur le lasso classique. La première est la consistance de sa sélection de variables, i.e. le meilleur sous-ensemble de variables (parmi le jeu de variables initial) est choisi ; ce qui n'est pas toujours le cas d'un lasso classique (cf. Zou, 2006). La seconde propriété d'Oracle est la consistance de l'estimation paramétrique (convergence asymptotique en loi normale de l'estimateur).

Si Zou (2006) définit les pénalités individuelles $\hat{w}_j = 1/|\hat{\beta}_j|^\gamma$, avec $\hat{\beta}$ l'estimateur des moindres carrés ordinaires et $\gamma > 0$ (en pratique, $\gamma \in \{0.5 ; 1 ; 2\}$), une alternative consiste à utiliser l'estimateur issu de la régression ridge³, introduite par Hoerl & Kennard (1970), pour définir le vecteur de pénalités individuelles. Son utilisation permet notamment d'éviter une mauvaise estimation des pénalités due à la présence de multi-colinéarité parmi les régresseurs.

L'optimisation du lasso adaptatif se fait donc en deux étapes. Premièrement, les pénalités individuelles sont déduites d'une régression ridge :

$$\hat{\beta}_{ridge} = \operatorname{argmin}_{\beta} Y - \sum_{j=1}^p x_j \beta_j^2 + \kappa \sum_{j=1}^p \beta_j^2, \kappa \geq 0$$

La valeur de la pénalité κ est alors obtenue par validation croisée « *leave one out* »⁴ (Hyndman & Athanasopoulos, 2018). Ensuite, $\hat{w}_j = \hat{\beta}_{ridge}$ mène à l'équation du lasso (dont la pénalité λ est également optimisée par validation croisée « *leave one out* ») :

$$\hat{\beta}_{adlasso} = \operatorname{argmin}_{\beta} Y - \sum_{j=1}^p x_j \beta_j^2 + \lambda \sum_{j=1}^p |\beta_j| / \hat{w}_j, \lambda \geq 0$$

L'avantage du lasso adaptatif est de fonctionner en grande dimension (nombre de variables supérieur au nombre d'observations, i.e. à la taille de la fenêtre temporelle dans notre cas). Il est aussi réputé pour sa parcimonie. Ces deux propriétés répondent aux deux contraintes du grand nombre de requêtes Google possibles et des courts historiques des livraisons FEVAD.

Combinaison de modèles ou modèle global ?

Trois modèles individuels ont été mis en œuvre : le modèle Google Trends, le modèle CD⁵ issu de l'enquête commerce de détail, et le modèle SARIMA qui était utilisé jusqu'à présent.

Bec & Mogliani (2015) recensent les méthodes les plus fréquentes en termes de combinaison d'information. Selon eux, Bates & Granger (1969) sont les premiers à se positionner, en faveur de l'agrégation de prévisions issues de modèles distincts. Plus tard, Diebold (1989) encourage l'utilisation d'un seul modèle, combinant différentes sources d'information hétérogènes. Plus récemment, Huang & Lee (2010) affirment qu'un modèle global est meilleur s'il est bien spécifié. Par ailleurs, Clements et Galvão (2008) ou Kuzin *et al.* (2013) tranchent en faveur de l'agrégation lors d'applications empiriques. Dans l'exercice de prévision de l'indice de consommation, Bec & Mogliani (2015) ont de meilleurs résultats avec l'agrégation. Le test de Diebold et Mariano (1995), dont l'hypothèse nulle est que deux prévisions issues de modèles différents ne sont pas significativement différentes, est un indicateur déterminant pour favoriser un modèle.

Cet article vise à enrichir ce débat d'un nouveau cas d'application en comparant les résultats de la combinaison de modèles avec ceux d'un modèle global dont la spécification est la même que pour les modèles individuels, soit le lasso adaptatif appliqué à tous les régresseurs simultanément (Google Trends, indices CD et SARIMA). De Gooijer & Hyndman (2006) mettent en avant les bénéfices de l'agrégation, notamment sa lisibilité lorsque les modèles agrégés sont facilement interprétables. Ici, l'agrégation concerne les trois modèles individuels, chacun portant des effets qui lui sont propres :

- le modèle SARIMA reproduit le schéma passé de la série ;
- le modèle CD exploite les données issues du commerce de détail traditionnel ;
- l'information d'internet est extraite du modèle basé sur les indices Google Trends.

2. Lasso est l'acronyme de Least Absolute Shrinkage and Selection Operator.

3. La régression ridge et la régression lasso sont des régressions pénalisées de normes respectivement L2 et L1.

4. Concrètement, l'échantillon de validation est constitué d'une observation ; celui d'apprentissage, des $n-1$ autres observations (pour un échantillon de taille n). Les n valeurs de κ , obtenues sur chaque échantillon d'apprentissage (chacune minimisant la RMSE) font l'objet d'une moyenne pour obtenir la valeur finale de κ .

5. Le modèle CD est un lasso adaptatif dont les variables explicatives sont les indices quantitatifs du commerce de détail.

La question est de pondérer chaque prévision :

$$\widehat{Y}_{t+1} = \gamma \widehat{Y}_{t+1}^{SARIMA} + \mu \widehat{Y}_{t+1}^{gTrends} + \vartheta \widehat{Y}_{t+1}^{CD}$$

Plusieurs stratégies d'agrégation sont possibles, des plus simples, telles que la pondération par la moyenne ($\gamma = \mu = \vartheta = 1/3$) ou par l'inverse des erreurs – *in-sample* ou *out-of-sample* (cf. Aiofli & Timmerman, 2006) –, aux plus élaborées. Par exemple, l'inférence bayésienne, dont le fondement est le théorème de Bayes⁶ (cf. Marin & Robert, 2010), déduit la probabilité d'un événement à partir d'autres événements déjà évalués. La statistique bayésienne, particulièrement utilisée lorsque les échantillons sont petits, débouche sur des méthodes de classification, ou d'agrégation ici. Hoeting *et al.* (1999) mettent en avant les bonnes performances de l'agrégation bayésienne. Zeugner (2011) a développé un package R sur le sujet. L'idée est de tester les modèles d'une classe M donnée et de les pondérer selon leurs probabilités d'être le bon modèle. La classe M est celle des modèles linéaires. Habituellement, le grand nombre de modèles rend difficile l'agrégation bayésienne (cf. Hoeting *et al.*, 1999). Ce n'est pas le cas ici : avec trois régresseurs (correspondant aux estimations des modèles Google Trends, CD et SARIMA), huit modèles linéaires sont possibles. En notant D les données et M_j ($1 \leq j \leq 8$) un modèle donné, le théorème de Bayes donne :

$$P(M_j|D) = \frac{P(D|M_j)P(M_j)}{\sum_{1 \leq i \leq 8} P(D|M_i)P(M_i)}$$

Il s'agit de préciser les deux termes du numérateur pour évaluer la probabilité *a posteriori*⁷ :

- $P(M_j)$ correspond à la probabilité *a priori*⁸ que le modèle M_j soit le bon ;

- $P(D|M_j) = \int pr(D|\beta_j, M_j) pr(\beta_j|M_j) d\beta_j$ avec β_j les paramètres du modèle : $\beta_j = \{\gamma_j, \mu_j, \vartheta_j\}$ estimé sur le modèle M_j . Ici, les paramètres sont la quantité d'intérêt.

Concrètement, les valeurs des coefficients obtenus dans chaque modèle M_j de la classe M sont pondérées par la probabilité que chaque modèle M_j soit le bon : $\gamma = \sum \gamma_i P(M_i|D)$ avec $\gamma_i = E(\gamma|D, M_i)$ la valeur du coefficient dans le modèle M_i . Il en est de même pour μ et ϑ .

Données

L'Enquête mensuelle de conjoncture Commerce de Détail

Une des enquêtes mensuelles de conjoncture réalisées par la Banque de France est consacrée

au commerce de détail⁹ (CD). Elle suit l'évolution des chiffres d'affaires (CA) TTC des 6 800 entités de l'échantillon (réparties selon plus de 4 000 entreprises) ; chaque mois, le taux de réponse est d'environ 90 %. Chaque entité fournit son CA total et la part des principaux produits (si elle n'est pas « mono produit »). Ces données individuelles sont ensuite regroupées selon des caractéristiques communes aux entreprises : par mode de distribution (physique : petit commerce traditionnel, grande surface spécialiste et succursaliste, hyper et supermarchés, grand magasin et magasin populaire ; et à distance : vente à distance) et par produit (e.g. électroménager, chaussures, etc.). Les indices quantitatifs conjoncturels sont établis pour ces agrégats (par exemple : les petits commerces traditionnels vendant du mobilier).

Construction des indices quantitatifs conjoncturels

Chaque indice de chiffre d'affaires Y issu de l'enquête est ainsi construit (avec X le montant de chiffre d'affaires associé) :

$$Y_M = Y_{M-12} \frac{X_M}{X_{M-12}}$$

Toutes les grandeurs de l'équation précédente concernent les mêmes entreprises. Selon la méthodologie de l'enquête, les chiffres d'affaires sont cylindrés, i.e. le périmètre utilisé pour X_M est le même que pour X_{M-12} . Autrement dit, ce sont les mêmes entreprises qui sont considérées dans ces deux (sommés de) chiffres d'affaires. Le cylindrage évite les variations extrêmes non représentatives de l'échantillon (points aberrants). La fermeture (ou l'ouverture) d'un magasin est le phénomène extrême le plus fréquent auquel l'enquête est confrontée : la baisse (respectivement hausse) de chiffre d'affaires qui en découle est compensée par des évolutions opposées dans l'ensemble de ses concurrents, ensemble qui ne sera pas pris en compte dans sa totalité dans l'échantillon. Par ailleurs, il est plus probable de capter une fermeture qu'une ouverture de magasin (magasin ou nouvelle enseigne non encore entrés dans

6. L'écriture commune du théorème est : $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$ avec P la mesure de probabilité, A et B deux événements.

7. La probabilité *a posteriori* est déterminée à l'aide des données in sample.

8. Il y a différentes manières de déterminer les probabilités *a priori*, comme le montre Zeugner (2011). Dans notre cas, plusieurs tests ont été faits (prior binomial, uniforme, déterministe, etc.) sans affecter significativement les résultats.

9. Les derniers résultats sont disponibles à cette adresse : <https://www.banque-france.fr/statistiques/chiffres-cles-france-et-etranger/enquetes-de-conjoncture/conjoncture-commerce-de-detail>.

l'échantillon), si bien que la mesure serait biaisée en l'absence de cylindrage.

Disponibilité des indices

Seule une partie des indicateurs croisant produits et circuits de distribution physiques (vente en magasin) sont calculés, faute d'un échantillon suffisant et pour des raisons de confidentialité des données (cf. cases vides dans le tableau 1). Ces indices sont disponibles depuis 1990 à la différence de celui de la vente à distance, qui ne l'est que depuis 2012. Cet article porte sur les indices bruts (cf. la partie consacrée à la modélisation). Le tableau 1 présente les indices quantitatifs conjoncturels des ventes physiques des produits concernés par les livraisons de la FEVAD¹⁰.

La FEVAD

Le champ du e-commerce n'est pas suivi par une collecte directe. La FEVAD (Fédération du e-commerce et de la Vente à distance) livre à la Banque de France les chiffres d'affaires (CA) mensuels agrégés de ses adhérents les plus importants, depuis janvier 2012. Ils sont environ 70 et leur liste évolue. Dans l'enquête, ces données participent à la construction des indices de CA (définis supra) appliqués à la vente à distance (VAD). Conformément à la méthodologie de l'enquête, les CA du mois M et leur révision à échantillon constant du mois $M-12$ sont livrés chaque mois. Ces livraisons concernent le CA total (« total

produits industriels hors automobiles ») et ceux de cinq produits : électroménager, textile (habillement et textile de maison – dénommé habillement dans la suite), chaussures (maroquinerie incluse), électronique grand public (EGP) et mobilier (meubles essentiellement). Le périmètre du total étant plus vaste que celui des cinq produits réunis, son CA est supérieur à la somme de ceux des produits. En moyenne (sur l'historique des livraisons), les CA des cinq produits représentent 68 % du CA total. Le tableau 2 donne la part (en %) de la VAD dans chaque produit selon la FEVAD.

Rapprochement des données FEVAD avec celles de l'EMC

Pour chacune des six estimations (des indices de CA total VAD et des cinq produits VAD), les données issues de l'enquête sont utilisées. La figure I présente les indices des différents circuits de distribution dans l'EGP (ventes physiques et VAD).

Dans l'EGP, le pic des ventes en décembre est commun aux différents circuits de distribution. Les corrélations¹¹ entre l'indice de CA de la FEVAD et les indices de CA des ventes physiques pour l'EGP (en %) complètent l'information graphique (tableau 3).

10. NB : d'autres produits que les chaussures, l'EGP, l'électroménager, les meubles et l'habillement s'ajoutent pour constituer le total.

11. Ces corrélations sont calculées sur les indices différenciés sur un mois, conformément aux données utilisées lors de la modélisation (cf. infra).

Tableau 1

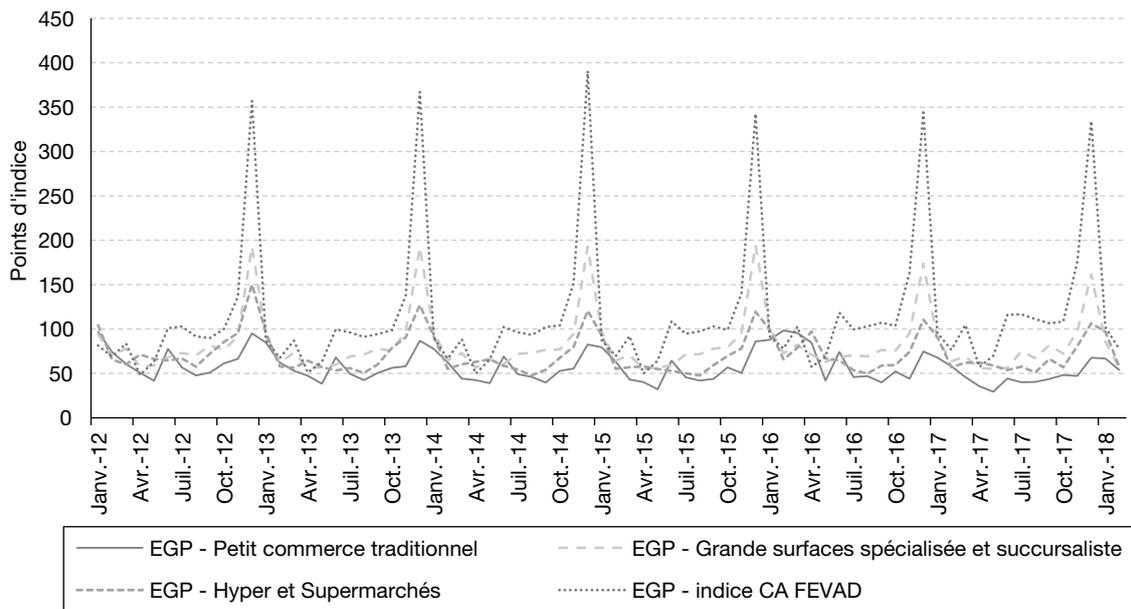
Moyennes et écart-types des indices quantitatifs conjoncturels des ventes physiques issus de l'enquête mensuelle de conjoncture du Commerce de Détail

	Petit commerce traditionnel	Grande surface spécialiste et succursaliste	Hyper et supermarché	Grand magasin et magasin populaire	Ensemble des ventes physiques
Total Produits industriels hors automobiles					91.2 17.5
Chaussures	91.2 21.1	94.3 27.0			
EGP	81.5 30.3	98.1 46.3	82.5 33.1		
Électroménager	96.1 15.6	103.6 15.4	97.6 23.9		
Meubles	105.1 18.4	108.4 20.5	126.1 38.7		
Habillement	102.1 28.4	101.6 28.8	99.0 19.1	87.4 24.3	

Lecture : une case vide signifie l'absence d'indicateur pour le croisement concerné. Les moyennes et écarts-types calculés sur la période janvier 2012 - décembre 2017 sont présentés respectivement en première et seconde ligne.

Source : Banque de France DGS SEEC.

Figure 1
Indices (bruts) des différents canaux de distribution dans l'électronique grand public (EGP)



Source : FEVAD, Banque de France DGS SEEC.

Tableau 2
Part de la vente à distance dans chaque produit

Produit	Poids de la VAD (en %)
Chaussures	11
EGP	23
Électroménager	18
Meubles	13
Habillement	13
Total	10

Lecture : selon la FEVAD, la VAD représente 11 % des ventes de chaussures en 2017.

Source : FEVAD.

Tableau 3
Corrélations de l'indice de chiffre d'affaires FEVAD avec les indices du commerce traditionnel issus de l'enquête dans l'électronique grand public

Canaux de distribution	Corrélation avec l'indice VAD (en %)
Petit commerce traditionnel	44
Grandes surfaces spécialisées et succursalistes	96
Hyper et Supermarchés	48

Note : ces corrélations sont calculées sur la période 01/2012-01/2018.
Source : Banque de France DGS SEEC, FEVAD.

La corrélation entre l'indice de CA VAD et celui des grandes surfaces spécialisées et succursalistes incite à utiliser les données des ventes physiques pour estimer les données FEVAD.

D'une manière générale, rapprocher ces données permet l'observation de mécanismes économiques simples. Par exemple, à long terme, un effet de substitution se traduit par une baisse des ventes dans les points de ventes physiques ; le corollaire est une hausse des ventes à distance. En revanche, à court terme, une hausse (ou baisse) des ventes physiques peut annoncer une hausse (baisse, respectivement) des ventes à distance : ces évolutions communes traduisent celle de la consommation des ménages.

Google Trends

Google Trends fournit les indices mensuels des termes recherchés sur le moteur de recherche Google par les utilisateurs. Élaborés par Google selon une méthodologie non publique, ces indices sont constitués selon les champs (définis par l'utilisateur) géographique (la France ici), temporel (l'historique maximum remonte à 2004), fréquentiel (mensuel ici) et d'appartenance à une catégorie (par exemple « Shopping », cf. infra). Disponibles si le volume des recherches est « suffisant » (au sens de Google), ces indices sont constitués de valeurs entières comprises entre 0 et 100 et portent sur des échantillons des recherches totales effectuées. Outre le fait que la méthodologie de construction des indices Google Trends est opaque, certains des points précédents appellent à effectuer des tests de robustesse.

L'échantillonnage de Google

Construit sur un échantillon aléatoire des recherches, un indice Google Trends diffère entre deux tirages. Comparer les séries d'un même terme, requêtées plusieurs fois, participe à vérifier la robustesse de l'outil. À titre illustratif, le tableau 4 donne les corrélations obtenues pour deux tirages distincts réalisés à quelques jours d'intervalle (i.e. à méthodologies Google et Google Trends constantes, *a priori*)

Cette expérience a été répétée de nombreuses fois, sans obtenir de corrélation inférieure à 90 % sur les indices différenciés sur un mois. Dans ces conditions, la méthode d'échantillonnage paraît suffisamment fiable pour requêter régulièrement les indices Google Trends. L'impact de l'échantillonnage sur les résultats sera discuté plus loin.

Indices de valeurs entières

Ensuite, l'extraction simultanée des indices Google Trends pose problème. Effectivement, lors d'une extraction commune d'indices (entre deux et cinq) *via* l'outil, la valeur 100 est attribuée à l'indice connaissant le pic de recherches sur l'historique requêté ; les maxima des autres indices sont au *pro rata*. Si les volumes de recherches diffèrent significativement, les indices des requêtes moins populaires prennent un nombre restreint de valeurs – du fait d'être constitués de valeurs entières –, qui reflètent mal leurs variations. Or, dans un modèle statistique, le nombre de décimales des variables, et plus généralement la précision des variables, peut avoir une influence sur l'estimation finale selon Kozicki & Hoffman (2004). Afin d'avoir les valeurs les plus précises possibles chaque série Google Trends est donc extraite individuellement. Les deux derniers points couplés – l'échantillonnage et le fait que les indices soient constitués de valeurs entières – ne favorisent pas la précision des données Google Trends.

Tableau 4
Corrélations entre indices Google Trends tirés à plusieurs jours d'intervalles

(En %)				
Amazon	Cdiscount	Fnac	E. Leclerc	eBay
98.1	97.4	98.9	95.5	90.2

Note : corrélations calculées sur les indices différenciés sur un mois allant de janvier 2004 à février 2018 (170 points).
Source : Google Trends, Banque de France DGS DESS SEEC.

Catégorie

L'outil Google Trends répertorie les requêtes Google par catégorie, correspondant au contexte dans lequel la recherche est faite¹². L'exemple de la requête « iPhone » appelle à la vigilance lors des extractions (figure II).

Si la catégorie « Marchés commerciaux et industriels » n'est pas utile pour étudier la VAD, la série tracée illustre l'importance du choix de la catégorie : son maximum, atteint en septembre 2013, n'est pas synonyme d'une explosion des ventes. En l'absence de plus d'information sur les catégories, toutes les requêtes utilisées dans la suite de l'article appartiennent à la catégorie « Shopping », correspondant *a priori* le mieux au e-commerce.

Ruptures de séries

Si Google communique peu sur l'évolution de la méthodologie de construction des indices, deux remarques figurent sur la page d'extraction :

- « La fonctionnalité de détermination de la position géographique a été améliorée. Cette mise à jour a été appliquée à partir du 1^{er} janvier 2011. »
- « Notre système de collecte de données a été amélioré. Cette mise à jour a été appliquée à partir du 01/01/2016. »

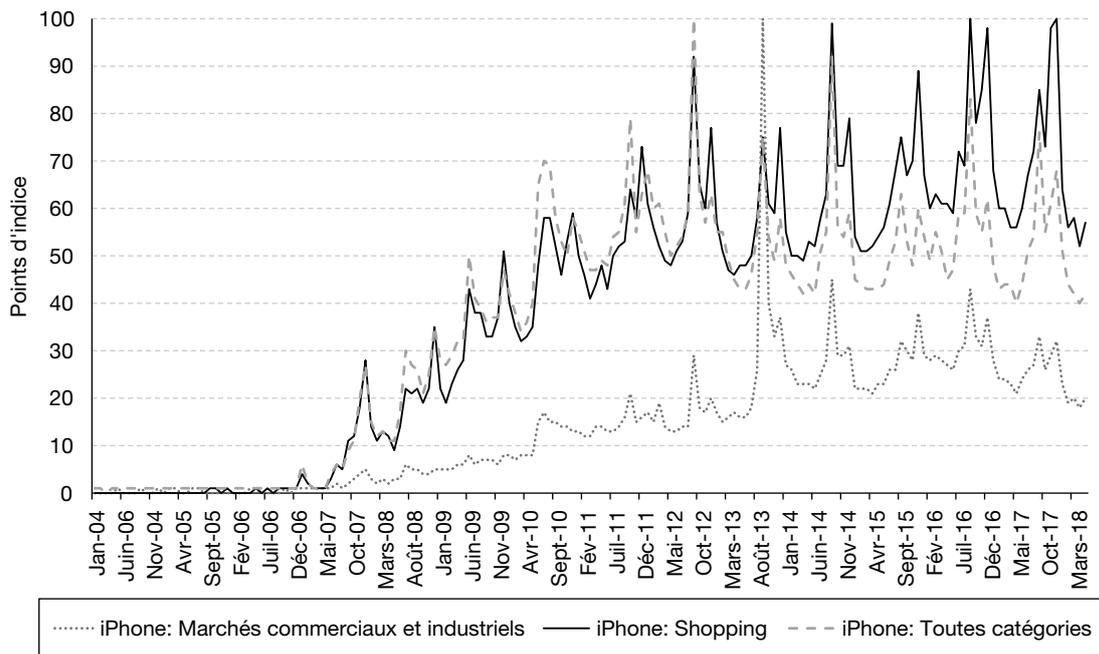
Les utilisateurs sont donc prévenus lors des principales modifications de l'outil. De plus, celles-ci sont effectives plusieurs mois après. Les indices de CA FEVAD démarrant en janvier 2012, la seconde remarque appelle une attention particulière¹³.

L'analyse des indices Google Trends à l'aide de la méthode X13 détecte un grand nombre de valeurs aberrantes, notamment en janvier 2016. Du fait de l'opacité de la méthodologie de construction des indices Google Trends et de leur nombre important (plus de 150) – susceptible d'augmenter avec le développement du e-commerce –, le traitement des

12. Par exemple, le terme « jaguar » peut référer à l'animal ou au constructeur automobile. Les requêtes Google sont probablement répertoriées dans les catégories grâce aux clics post requêtes (i.e. aux sites consultés suite à la requête).

13. Afin d'améliorer la robustesse des calculs, les indices Google Trends sont extraits depuis janvier 2011. S'il est globalement convenu qu'une désaisonnalisation ne peut se faire avec un historique inférieur à 3 ans, ajouter un an d'historique permet de stabiliser les séries CVS et, donc, d'améliorer la détection des valeurs aberrantes.

Figure II
Requêtes Google Trends « iPhone »

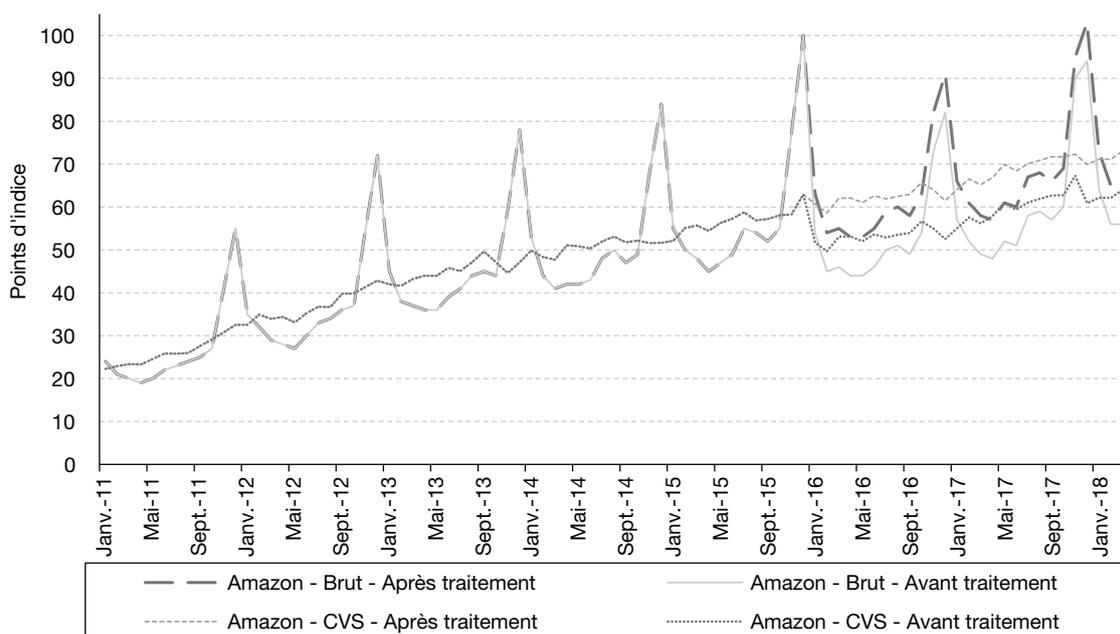


Source : Google Trends.

valeurs aberrantes a été systématisé. À travers l'exemple de l'indice Google Trends d'Amazon, ses étapes peuvent être explicitées. Ici, un saut de niveau est détecté en janvier 2016 ; après évaluation, la série peut être corrigée (figure III).

La première étape du traitement est la dessaisonnalisation de l'indice car, afin de capter un maximum de valeurs aberrantes, la détection s'opère sur les deux indices (brut et CVS). Ensuite, la nature de la valeur aberrante est déterminée : saut de niveau (*level shift*), variation

Figure III
Traitement de la valeur aberrante détectée sur l'indice Google Trends d'Amazon



Source : Google Trends, FEVAD, Banque de France DGS SEEC.

passagère (*transitory change*) ou valeur singulièrement aberrante (*additive outlier*). Dans le cas d'Amazon – et plus largement des valeurs aberrantes détectées en janvier 2016 (cf. Cdiscount, annexe 1) – il s'agit d'un saut de niveau. Enfin, l'ampleur de la rupture de série est estimée par l'écart entre le point de la série CVS¹⁴ de janvier 2016 et la prévision de la même série tronquée en décembre 2015. Cette correction est alors appliquée au reste de la série (contrairement aux valeurs aberrantes singulières dont le traitement est ponctuel).

Améliorant la qualité des séries en pseudo temps réel, la détection des valeurs aberrantes est moins fiable en temps réel, i.e. sur le dernier point de la série : du recul sur la valeur aberrante aide à la qualifier¹⁵ et améliore la précision de son estimation. Les seules valeurs aberrantes non traitées sont celles traduisant l'apparition de nouvelles requêtes (nouvel acteur, nouvelle marque, etc. ; cf. exemple de la chaussure *infra*). D'ailleurs, la mouvance du e-commerce appelle à la vigilance quant au choix des requêtes.

Listes de variables

D'une part, l'émergence du e-commerce s'accompagne de nouveaux acteurs. Pour les chaussures par exemple, les trois pure players (vendant uniquement sur internet) leaders du

e-commerce sur le marché français sont relativement récents (figure IV).

La croissance de l'indice « Chaussures » de 2004 à 2011 correspond à l'émergence des ventes de chaussures sur internet. Graphiquement, le lancement de Zalando en France, en décembre 2010, est très clair (l'indice passe de 1 à 19 en deux mois : 11/2010 - 01/2011).

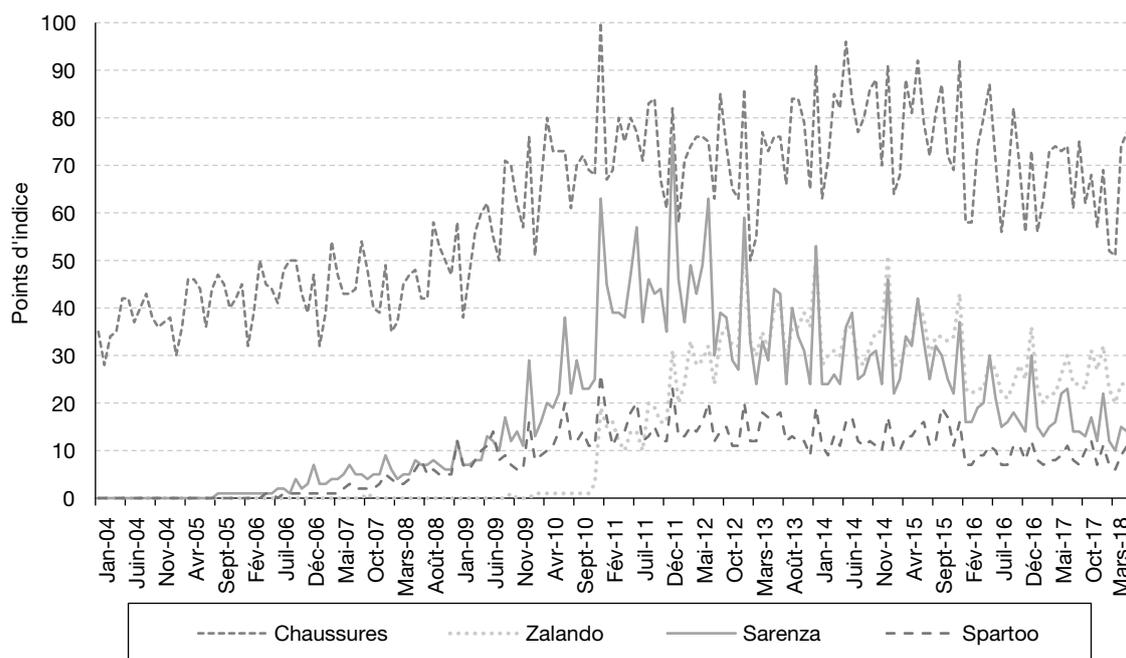
D'autre part, certains acteurs du e-commerce, présents au début, immergent. Dans le domaine de l'électroménager, l'indice Google Trends de GrosBill en témoigne (figure V).

Si la requête présentait un intérêt il y a quelques années, ce spécialiste du e-commerce pour l'électroménager et l'EGP a perdu des parts de marché, par rapport à Boulanger par exemple. Une dernière illustration de la mouvance des acteurs du e-commerce concerne la fusion d'entreprises telles que la Fnac et Darty : désormais, l'indice Google Trends associé est « Groupe Fnac Darty ». En général, le e-commerce connaît une évolution permanente, bien retranscrite par les

14. Dans cet exemple, l'estimation du saut de niveau est faite à l'aide des CVS car celle fournie par les données brutes semblait moins cohérente.

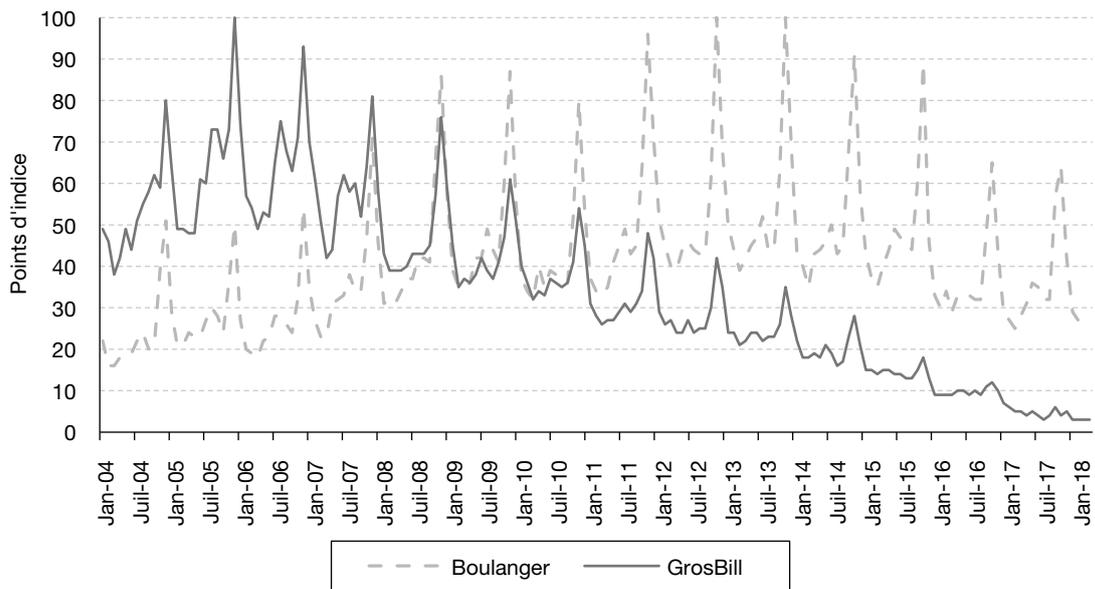
15. Par exemple, un saut de niveau ne peut être détecté qu'a posteriori : à son apparition, la valeur aberrante peut être (au mieux) qualifiée de valeur aberrante valeur aberrante singulière avant d'être requalifiée en saut de niveau (suite à l'apparition d'autres points).

Figure IV
Indices Google Trends liés à la chaussure



Source : Google Trends.

Figure V
Indices Google Trends de Boulanger et GrosBill



Source : Google Trends.

indices Google Trends. Par exemple, la baisse de popularité de la requête Google pour un acteur du e-commerce peut s'accompagner d'une hausse des requêtes pour des concurrents. Dans cet univers, il est impératif de régulièrement réviser les variables utilisées, en particulier celles concernant les acteurs du e-commerce pour les différents produits. Afin de ne pas négliger l'aspect évolutif des termes recherchés, il est possible de rétropoler les résultats avec d'autres variables en effectuant une double collecte (i.e. en testant le modèle sur deux jeux de variables).

Néanmoins, une des limites de l'approche réside dans les listes initiales de variables (elles sont présentées en annexe 2). Pour l'indice total, cette pré-sélection correspond essentiellement aux acteurs majeurs du e-commerce en France. Les pré-sélections des cinq produits sont un mélange de pure players (exemple : Sarenza, pour les chaussures), d'enseignes (La Halle), de termes génériques (chaussures femme) et de marques (Converse). Des travaux préliminaires, tels qu'une recherche documentaire sur les produits concernés par l'estimation ou la consultation de sites spécialisés, ont été menés afin d'anticiper les comportements préalables à un achat. Ils ont conduit à retenir des listes de variables hétérogènes (cf. tableau complet en annexe 2).

Du reste, la tendance de popularité d'un site internet n'est pas nécessairement la même que celle de l'indice Google Trends associé car

tous les internautes ne passent pas par Google : les modes de consultation des sites internet évoluent, notamment avec l'émergence du m-commerce¹⁶ où les applications évitent l'utilisation du moteur de recherche.

Modélisation

Traitement de la stationnarité et saisonnalité

La plupart des séries ne sont pas stationnaires mais plutôt intégrées d'ordre 1 ; la différenciation s'impose. Cette opération classique participe à éviter les régressions fallacieuses (cf. Phillips, 1986), un phénomène fréquent lors de régressions entre séries temporelles, traduit par des résultats trop optimistes signifiés par un R^2 anormalement élevé (cf. Granger & Newbold, 1974). L'introduction d'une variable mesurant la tendance (Phillips & Perron, 1988) ou de termes autorégressifs y participent aussi.

Dans le but de mieux mesurer la tendance du e-commerce, il a été question de travailler avec les séries corrigées des variations saisonnières (CVS). Cette solution n'a pas été retenue. D'abord, les historiques courts des séries

16. Selon la FEVAD, 36.6 millions de Français achètent sur internet, dont 9.3 millions ont déjà effectué un achat à partir de leur mobile (en 2017).

n'assurent pas une désaisonnalisation de qualité sur toutes les séries¹⁷, surtout lors des premières estimations (36 points à la première itération ; plus de 70 maintenant) ; d'autant que, du fait de son émergence, le e-commerce connaît des changements de saisonnalité (figure VI).

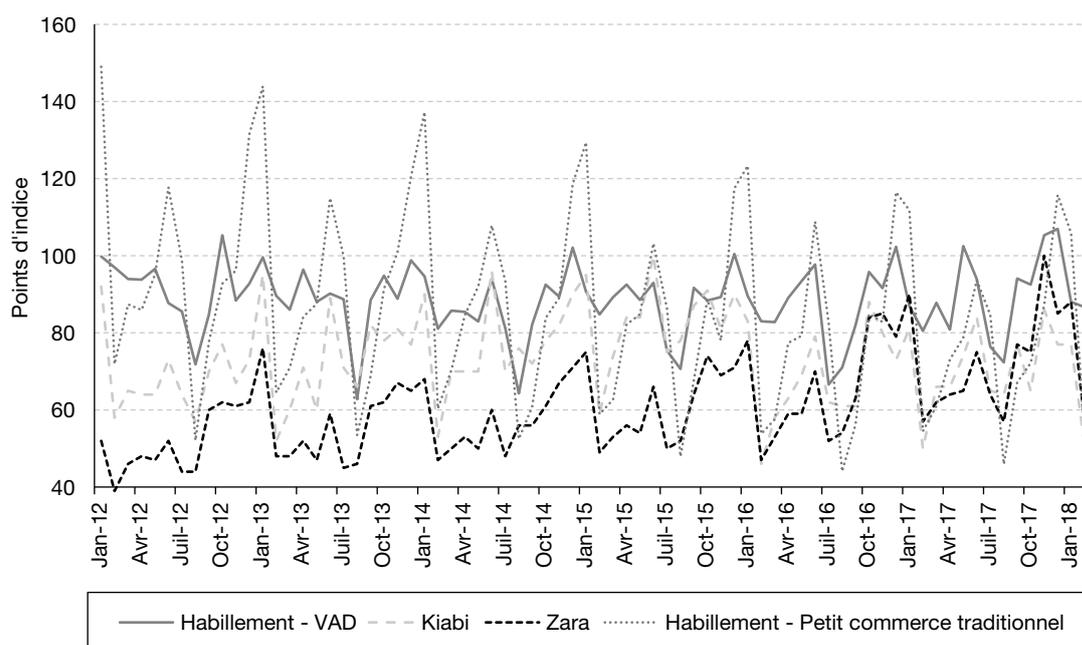
La figure VI représente les séries brutes de deux indices de CA de l'habillement (VAD et Petit commerce traditionnel) et deux requêtes Google Trends associées au produit : « Kiabi » et « Zara ». L'indice de CA de la VAD connaît des changements de saisonnalités : par exemple, dans les premières années, le mois de juillet est nettement au-dessus du mois d'août. En 2015, l'écart entre les deux mois se réduit et, en 2016, le mois de juillet est plus bas. L'allure générale de la série illustre bien les changements de saisonnalités. Ce phénomène est commun aux indices Google Trends. Par exemple, pour Zara, le maximum annuel est atteint au mois de janvier pour les années allant de 2013 à 2016 ; or la valeur du mois de novembre 2017 est supérieure à celles des mois de janvier 2017 et 2018. De même, la série de Kiabi ne présente pas une stabilité saisonnière remarquable. Dans ces conditions, la désaisonnalisation de nombreuses séries est de qualité incertaine. En revanche, les saisonnalités de l'indice du Petit commerce traditionnel sont stables. Les changements sont plus rares pour des séries

bien établies (l'indice a démarré en 1990). Plus généralement, les séries de l'enquête passent systématiquement les tests de saisonnalité (autocorrélation, Friedman, Kruskal-Wallis, pics spectraux, périodogramme), ce qui n'est pas toujours le cas des séries Google Trends.

En outre, les derniers points d'une série CVS sont les plus susceptibles d'être modifiés par l'apparition de nouveaux points (cf. Eurostat, 2018). À chaque livraison de la FEVAD, lorsqu'il est possible d'évaluer la prévision précédente, les derniers points de la série CVS changent ; pouvant influencer fortement le modèle. L'instabilité des CVS sur les derniers points est particulièrement forte pour les séries du e-commerce, notamment du fait de saisonnalités encore mal établies et des courts historiques. Alors que l'ampleur de l'instabilité des CVS est de 0.2 point en moyenne sur la période 01/2015-01/2018 pour l'indice de la grande distribution, elle est de 1.6 point en moyenne pour l'indice de CA total de la VAD (voir annexe 3), soit de l'ordre de grandeur des erreurs de prévisions (*infra*). Ces arguments conduisent à privilégier une modélisation des données brutes différenciées.

17. Plus de 150 séries sont utilisées pour réaliser les six estimations.

Figure VI
Indices issus de l'habillement



Source : Google Trends, FEVAD, Banque de France DGS SEEC.

Processus d'estimation et d'évaluation de la performance

Modèles

Jusqu'ici, un modèle SARIMA était utilisé pour chaque produit. Faisant office de référence par la suite, il est toujours mis à jour. Par ailleurs, le lasso adaptatif est utilisé dans trois modèles, implémentés sur chaque produit :

- modèle « Google Trends », utilisant les Google Trends (voir annexe 4) ;
- modèle « CD », basé sur les indices quantitatifs conjoncturels des ventes physiques issus de l'enquête CD¹⁸ (cf. tableau 1) ;
- modèle global, sélectionnant parmi toutes les variables disponibles¹⁹.

Outre les variables exogènes, une tendance et une composante autorégressive font aussi partie du jeu de variables initiales dans ces trois modèles. L'introduction de la tendance répond à la pleine croissance (*a priori* non linéaire) du e-commerce. Et, plus qu'une composante autorégressive, il s'agit de la modélisation SARIMA de l'indice, qui devient une variable potentiellement sélectionnée par l'algorithme du lasso adaptatif, au même titre que la tendance et les variables exogènes (indices Google Trends et/ou indices quantitatifs conjoncturels). Enfin, le cinquième modèle est le fruit de l'agrégation bayésienne (« Combinaison de modèles » dans la suite) des modèles SARIMA, Google Trends et CD. La confrontation de ses résultats avec ceux du modèle global contribue au débat sur la combinaison d'information.

Protocole de test

À chaque itération du protocole de test, i.e. chaque mois, les conditions réelles sont répliquées. Plus précisément, les valeurs des données Google Trends et des indices quantitatifs conjoncturels des ventes physiques issus de l'enquête du mois M sont disponibles, contrairement aux données FEVAD.

Concrètement, l'estimation se fait en deux étapes : la première consiste à modéliser l'indice par un processus autorégressif (SARIMA). Outre l'obtention de sa propre prévision, cette opération sert aussi à déterminer la variable utilisée dans les modèles de lasso adaptatif. Dans un second temps, les trois modèles à sélection

de variables (Google Trends, CD et global) sont élaborés. La combinaison de modèles ne peut être construite qu'après les modèles SARIMA, Google Trends et CD.

La qualité des modèles est jugée après livraison des données FEVAD puisque le critère de jugement, dans le cadre d'un exercice de *now-casting*, est la capacité de prévision. L'indicateur est donc la RMSFE (*Root Mean Squared Forward Error*), soit l'écart-type des erreurs de prévision, mesure de l'erreur *out-of-sample*. La RMSE (*Root Mean Squared Error*), mesure de l'erreur *in-sample*, est aussi présentée car elle aide à comprendre la pondération dans la combinaison de modèles et à repérer d'éventuels phénomènes de sur-apprentissage.

Par ailleurs, chaque mois, la fenêtre d'estimation des modèles s'agrandit d'un point. Du fait des courts échantillons à disposition, travailler avec une fenêtre extensible plutôt qu'avec une fenêtre glissante pour l'échantillon d'estimation participe à la stabilisation des modèles. Les données FEVAD sont livrées depuis janvier 2012. La différenciation des données mène à février 2012. Avec un historique minimal de 3 ans pour assurer la robustesse de l'estimation, la première prévision est celle de février 2015.

Résultats

Seuls les résultats pour le total seront détaillés, ceux concernant les produits seront exposés de façon plus synthétique.

Total

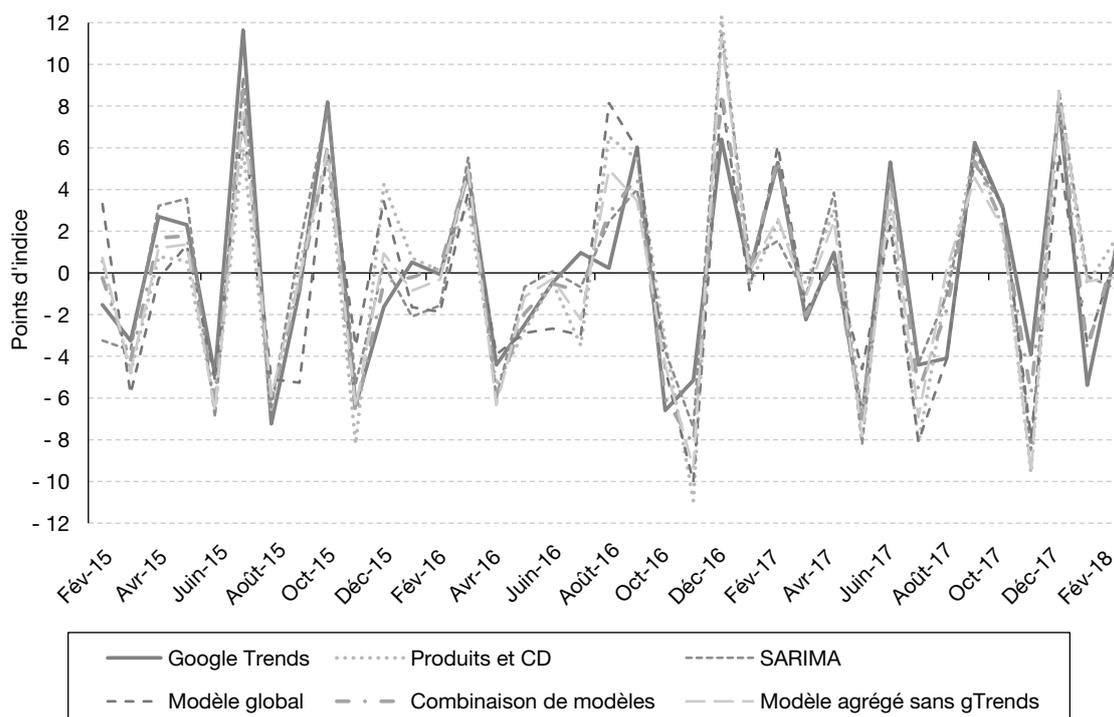
Conformément à l'objectif de l'étude, les erreurs de prévision (*out-of-sample*) constituent un résultat important (figure VII).

La figure VII montre les erreurs de prévision des différents modèles. Graphiquement, les résultats sont proches : globalement, les résultats du test de Diebold-Mariano (cf. Diebold & Mariano, 1995) ne permettent pas d'affirmer que les prévisions des modèles sont significativement différentes. Les RMSFE et les moyennes des erreurs de prévision (en absolu) permettent de mieux appréhender les résultats

18. Pour rappel, dans le cas de l'indice total, les indices de CA VAD des cinq produits sont aussi utilisés.

19. Ce modèle peut, à certaines itérations, être identique à l'un des modèles individuels (e.g. si aucune requête Google Trends n'est sélectionnée).

Figure VII
Erreurs de prévision des modèles dans l'estimation de l'indice total



Lecture : l'erreur de prévision pour le mois d'août 2016 du modèle global est de 8.1 points d'indice : après livraison des données FEVAD, la valeur de l'indice de CA du total était supérieure (de 8 points) à la prévision du modèle global.
Source : Google Trends, FEVAD, Banque de France DGS SEEC.

prédictifs, tandis que les RMSE témoignent de la capacité d'ajustement aux données *in-sample* (tableau 5).

Au sens de la RMSFE, qui reste l'indicateur privilégié, le modèle Google Trends est le plus performant avec la combinaison de modèles (4.8), pour le jeu de données Google retenu (jugé représentatif des simulations effectuées, voir encadré). Ici, les moins bonnes performances de la combinaison de modèles sans les données Google justifient l'apport de Google Trends. La combinaison de modèles est aussi meilleure au sens de la moyenne des erreurs absolues. Cette

mesure de l'erreur est intéressante car l'un des objectifs de l'agrégation est aussi de minimiser les gros écarts de prévision. Les résultats des modèles individuels sont relativement proches. En termes de RMSE, les deux modèles disposant de toute l'information (la combinaison de modèles et le modèle global) s'ajustent mieux aux données de l'échantillon.

Avant de comparer ces deux modèles, les résultats des modèles individuels méritent d'être détaillés ; particulièrement le modèle Google Trends qui doit apporter des garanties en termes de parcimonie et de stabilité.

Tableau 5
RMSFE et moyenne des RMSE des modèles dans l'estimation de l'indice total

Total	Google Trends	CD	SARIMA	Modèle global	Combinaison de modèles	Combinaison de modèles sans Google Trends
RMSFE	4.8	5.2	5.0	5.5	4.8	5.0
Moyenne des erreurs de prévision absolues	3.9	4.0	3.9	4.5	3.8	3.9
Moyenne des RMSE	3.3	3.7	4.2	2.3	2.6	2.8

Note : la combinaison de modèles sans Google Trends correspond à l'agrégation du modèle CD et du modèle SARIMA. Il permet notamment de juger l'apport des données Google Trends. Cependant, la variable SARIMA étant présente dans tous les modèles CD, l'agrégation perd de son sens ; ainsi, il ne sera pas présenté pour les résultats obtenus sur les produits.

Source : Google Trends, Banque de France DGS SEEC.

ENCADRÉ – Sensibilité des modèles à l'échantillonnage Google

Les variables Google Trends sont susceptibles d'être modifiées d'un mois sur l'autre du fait de l'échantillonnage de Google. En témoignent les écarts-types des RMSFE, obtenus sur trente simulations^(a), des modèles utilisant les variables Google Trends (tableau A) :

Par ailleurs, chaque simulation correspond à l'estimation conjointe des indices des cinq produits et du total. Or, certaines variables Google Trends sont communes au total et à un produit. Il n'a donc pas été possible d'extraire

un jeu de variables Google Trends dont les résultats des modèles, en termes de RMSFE, se situent tous à la médiane. Ceux présentés dans le corps de l'article correspondent à l'une des simulations les plus représentatives (pour les six estimations), i.e. les RMSFE des modèles Google Trends sont très proches de la médiane.

(a) Chaque simulation porte sur le total et les produits, soit 150 séries Google Trends. Google limitant l'extraction massive de séries, il est difficile d'augmenter considérablement le nombre de simulations.

Tableau A
Impact de l'échantillonnage Google sur les résultats en termes d'écart-type de RMSFE

	Google Trends	Modèle global	Combinaison de modèles
Total	0.4	0.3	0.4

Lecture : l'écart-type des RMSFE obtenues sur les 30 simulations pour le modèle Google Trends est de 0.4.
Source : Google Trends, FEVAD, Banque de France DGS SEEC.

Modèle SARIMA

Faisant office de référence comme habituellement dans la littérature, le modèle SARIMA présente de bonnes performances prédictives (RMSFE de 5.0), malgré un moins bon ajustement aux données de l'échantillon (RMSE de 4.2). Cependant, il lui arrive d'être moins bon que les autres modèles. Par exemple, en décembre 2016, les données exogènes apportent une réelle information.

Modèle Google Trends

Conformément au protocole de test, la sélection de variables du lasso adaptatif s'effectue à chaque itération. Donc, les coefficients du modèle évoluent dans le temps (figure VIII). Pour une meilleure lisibilité, les 30 variables de l'estimation du CA total ont été réparties sur six graphiques. En axe secondaire de chacun d'entre eux, l'évolution de la pénalité lasso.

Les graphiques de la figure VIII montrent l'évolution temporelle des coefficients selon l'axe primaire et celle de la pénalité lasso en axe secondaire. S'il n'est pas courant de regarder l'évolution de la pénalité lasso au cours du temps, puisqu'il s'agit d'une optimisation différente à chaque itération, celle-ci permet d'expliquer l'évolution du nombre de variables retenues : plus elle est faible, plus le nombre de requêtes Google Trends retenues est élevé.

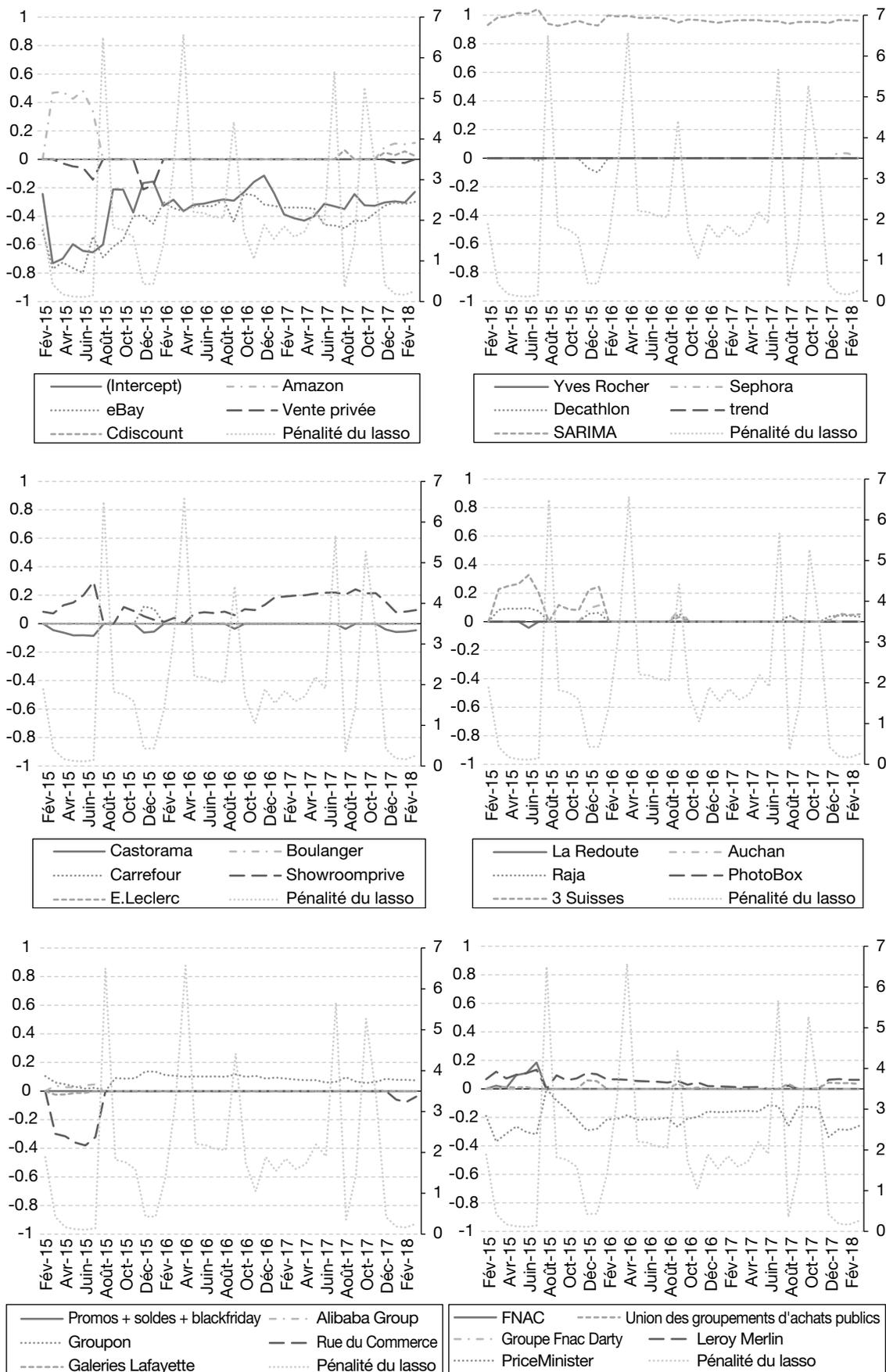
Ensuite, concernant la variable SARIMA, il était attendu que son coefficient soit proche de 1 puisqu'elle correspond à la modélisation autorégressive de la variable. Du reste, les évolutions des coefficients des variables Google Trends – mises en relief avec celle de la pénalité du lasso – sont stables ; signifiant que ces variables modélisent une partie de l'information non captée par la composante SARIMA. Le tableau des valeurs moyennes, minimums et maximums obtenues pour chaque variable est donné en annexe 5.

Concernant la sélection, près de 9 variables sont retenues en moyenne à chaque itération ; ce qui est acceptable étant donné la taille des échantillons (36 points à la première itération, 72 à la dernière). Les variables Google Trends les plus retenues sont eBay, PriceMinister, Groupon, Showroomprivé et Leroy Merlin (voir annexe 5).

Modèle commerce de détail (CD)

Outre l'indice de CA de l'ensemble des ventes physiques (cf. tableau 1), les indices de CA VAD des cinq produits sont utilisés. En effet, par construction, les CA des cinq produits contribuent au CA total. Cependant, toutes les données FEVAD étant livrées simultanément, ces indices sont prolongés au dernier point *via* une modélisation SARIMA. L'évolution des coefficients est donnée en annexe 6. Le modèle

Figure VIII
Évolution des coefficients du modèles Google Trends et de la pénalité du lasso



Source : Google Trends, Banque de France DGS SEEC.

est parcimonieux, sélectionnant une à deux variables en plus de la composante autorégressive (estimation SARIMA). L'indice de CA VAD de l'habillement est systématiquement et logiquement – en moyenne, le montant de CA de l'habillement représente 22 % du total, soit le plus gros des cinq produits – sélectionné. Ses résultats *out-of-sample* (RMSFE et moyenne des erreurs de prévision absolues) sont moins bons que les modèles Google Trends et SARIMA (cf. tableau 5). *In-sample* (RMSE), il se situe entre le modèle Google Trends et le modèle SARIMA.

Combinaison de modèles

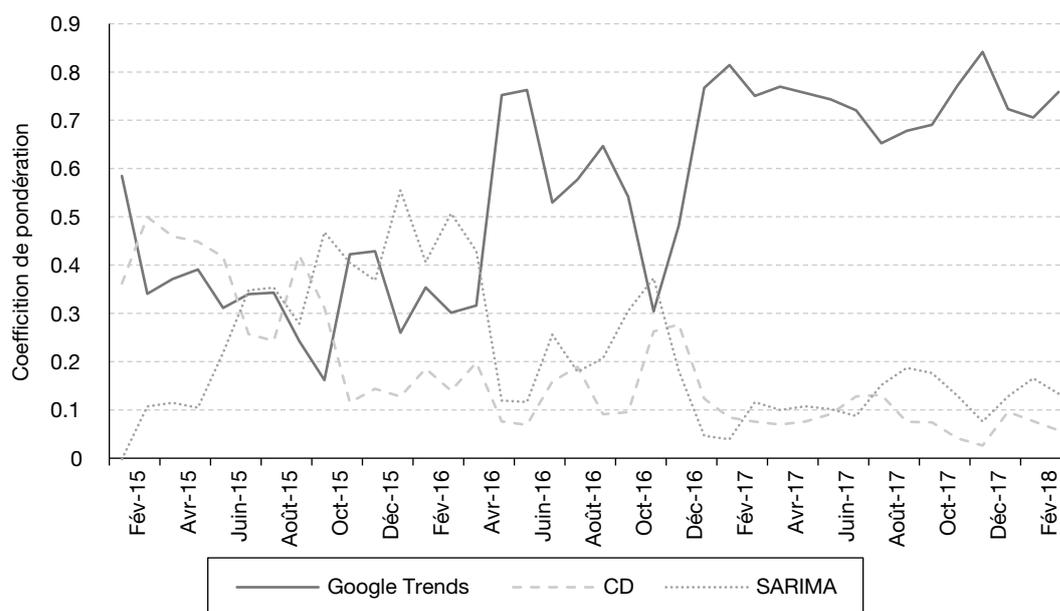
La combinaison de modèles présente les meilleures performances prédictives en moyenne selon les deux indicateurs, la RMSFE et l'erreur moyenne absolue. Au cours du temps, il n'a jamais la moins bonne prévision. Les poids des modèles permettent de juger sa stabilité (figure IX).

Depuis la fin 2016, le poids du modèle Google Trends augmente. En moyenne, il est supérieur (0.55) à celui des deux autres modèles, le SARIMA (0.21) et CD (0.18). L'évolution des erreurs de prévision des modèles Google Trends, CD et SARIMA éclaire celle des poids. Si les erreurs de prévision des trois modèles sont relativement proches, ce qui s'explique principalement par la présence de la variable SARIMA

dans les modèles Google Trends et CD, certaines différences méritent une attention particulière. Par exemple, au mois d'octobre 2016, le modèle Google Trends joue le rôle principal dans l'agrégation, avec un poids de 0.54, contre 0.31 pour le modèle SARIMA et 0.10 pour le modèle CD. Lors de la livraison des données FEVAD fin novembre 2016, il est possible de confronter les prévisions avec la valeur réelle. La figure VII, qui représentant les erreurs de prévision, montre que le modèle Google Trends est le moins bon des trois avec une erreur de 6.0 points d'indice, contre - 3.3 et - 3.7 pour les modèles CD et SARIMA. Une fois l'erreur apprise, le mois suivant, la pondération change drastiquement : le poids du modèle Google Trends passe à 0.30, contre 0.37 pour le modèle SARIMA et 0.26 pour le modèle CD.

Par ailleurs, la figure IX est l'occasion de détailler les formules de l'agrégation bayésienne. Conformément à la revue de littérature, huit modèles sont possibles à partir de trois régresseurs (correspondant ici aux valeurs estimées par les modèles Google Trends, CD et SARIMA). Le tableau 7 fournit les coefficients de chaque régresseur dans les modèles M_i ($1 \leq i \leq 8$) et la probabilité $P(M_i | D)$ que chaque modèle M_i soit le bon. Enfin, la dernière colonne correspond au modèle bayésien, dont les coefficients sont obtenus en pondérant ceux des modèles M_i par les probabilités $P(M_i | D)$. Les valeurs du tableau 6 sont celles de septembre 2016.

Figure IX
Évolution des poids dans la combinaison de modèles



Source : Google Trends, FEVAD, Banque de France DGS SEEC.

Tableau 6
Détail du calcul des poids dans l'agrégation bayésienne (en septembre 2016)

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	Modèle bayésien
Google Trends	0.96		0.69		0.78		0.82		0.65
CD			0.31	1.01		0.34	0.38		0.10
SARIMA		0.99			0.21	0.66	- 0.20		0.22
$P(M_i D)$	0.57	0.19	0.09	0.06	0.05	0.02	0.01	0.00	

Source : Google Trends, Banque de France DGS SEEC.

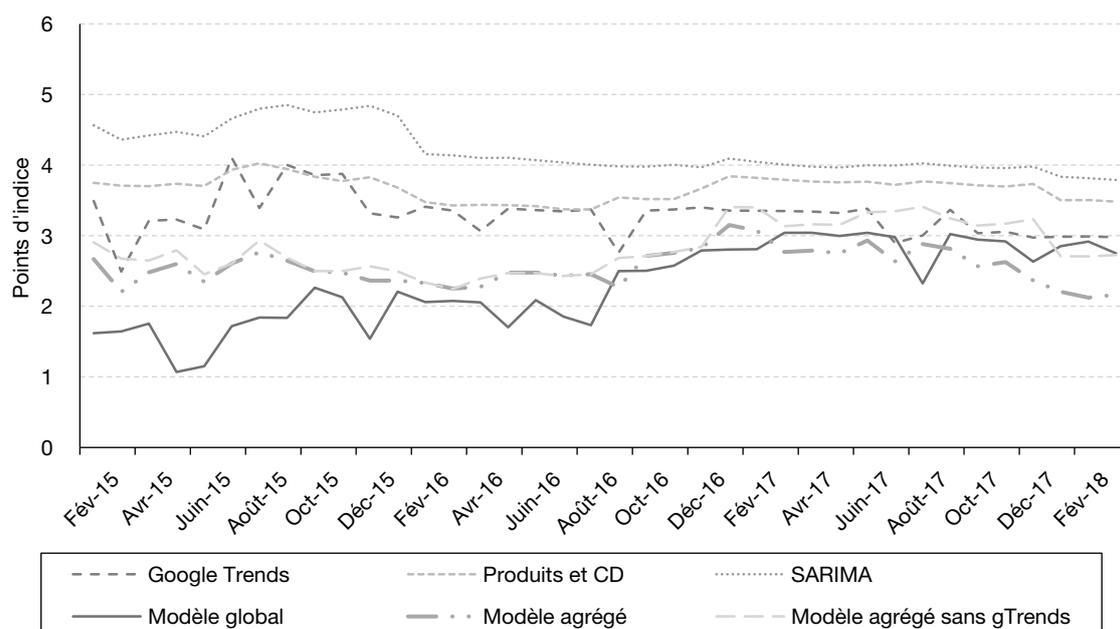
Les valeurs de la dernière colonne sont conformes à la figure IX (septembre 2016). L'agrégation bayésienne fait partie des algorithmes d'apprentissage statistique ; l'erreur *in-sample* contribue à la détermination des poids. La figure X représente l'évolution des RMSE.

À chaque itération, il est possible de calculer la RMSE réalisée sur l'échantillon d'estimation du modèle. Le modèle SARIMA présente sur toute la période la plus mauvaise erreur *in-sample*, contrastant avec ses bonnes capacités prédictives. Selon la figure X, l'agrégation d'information permet aussi de diminuer l'erreur sur l'échantillon d'estimation du modèle, par rapport à ses composantes. Ensuite, si les évolutions des erreurs *in-sample* des modèles agrégés (avec et sans Google Trends) sont

semblables, la performance prédictive est meilleure avec l'apport des données Google Trends. Enfin, le modèle Google Trends présente des erreurs proches des autres modèles, confortant l'idée que le nombre de variables sélectionnées par le lasso adaptatif est convenable et qu'il n'y a pas de sur-apprentissage. Pour s'en assurer, les erreurs *in-sample* (RMSE) peuvent être mises en perspective avec celles réalisées *out-of-sample* (RMSFE) (cf. tableau 5).

Logiquement, les erreurs de prévisions sont plus grandes. Ensuite, les classements des modèles sont respectés lors du passage de la RMSE à la RMSFE, sauf pour le modèle global : son erreur fait plus que doubler hors échantillon.

Figure X
Évolution des RMSE des modèles pour l'estimation de l'indice de CA total



Lecture : lors de l'estimation des modèles pour la prévision de décembre 2015, la RMSE la plus faible était celle du modèle global (1.5). La RMSE la plus élevée était celle du SARIMA (4.8).

Source : Google Trends, Banque de France DGS SEEC.

Modèle global

Ce phénomène s'explique probablement par un sur-apprentissage. Effectivement, bien que la procédure du lasso adaptatif soit la même que pour les modèles Google Trends et CD, le modèle global est moins parcimonieux : en moyenne, 13 variables sont sélectionnées, ce qui est un nombre relativement important par rapport au nombre de points observés (36 points à la première itération). Il retient plus de variables que les modèles Google Trends et CD cumulés. Plus précisément, sur la période du protocole de test, 82 % des variables sélectionnées dans le modèle global le sont dans l'un des deux autres modèles ; 12 % ne sont sélectionnées que dans le modèle global et les 6 % restant concernent les variables sélectionnées par les modèles Google Trends ou CD et non sélectionnées par le modèle global. En somme, la sélection de variables du modèle global est trop large ; engendrant un phénomène de sur-apprentissage. Du reste, l'évolution des coefficients est moins stable.

Dans le cas de l'indice total, le modèle global est moins performant que la combinaison de modèles. En outre, la lisibilité de la combinaison de modèles, bien que restreinte (les résultats du test de Diebold-Mariano ne permettant pas d'affirmer que les prévisions des trois

modèles sont significativement différentes), reste meilleure que celle du modèle global, dont l'évolution des coefficients le rend moins interprétable. L'utilisation de la combinaison de modèles est donc favorisée. Si les résultats obtenus pour l'indice total ont été largement détaillés, ceux des produits sont présentés plus succinctement.

Produits

Parcimonie

Le lasso adaptatif vise la parcimonie des modèles. Pour chaque produit, le tableau 7 présente le nombre moyen de variables retenues par modèle (concerné par la sélection de variables).

Les modèles CD sont les plus parcimonieux ; le plus souvent, une variable issue de l'enquête est sélectionnée, en plus de la composante SARIMA. Les modèles Google Trends sont moins parcimonieux ; le nombre de variables sélectionnées reste correct au vu de la taille des échantillons, à part peut-être pour l'habillement.

Dans les modèles Google Trends des produits, outre la composante SARIMA et la constante (systématiquement sélectionnées), les cinq variables les plus retenues (sur 38 itérations) sont recensées dans le tableau 8.

Tableau 7
Nombre de variables retenues par modèle utilisant le lasso adaptatif et par produit

	Google Trends	CD	Modèle global
Chaussures	9.7	2.2	10.0
Meubles	10.3	2.4	10.9
Électroménager	9.0	2.1	5.9
EGP	8.8	2.0	10.5
Habillement	12.8	3.6	8.7

Source : Google Trends, Banque de France DGS SEEC.

Tableau 8
Variables les plus retenues dans les modèles Google Trends, par produit

Chaussures	Spartoo (38)	Sarenza (36)	Converse (36)	Chaussures de ville (32)	Chaussures de foot (28)
Meubles	Cinna (38)	Roset (33)	Meuble en bois (32)	Buffet + commode + vaisselier (31)	IKEA (26)
Électroménager	Machine à laver le linge (37)	Four (28)	Cuisinière (28)	Conforama (26)	GrosBill (24)
EGP	Appareil photo reflex numérique (37)	Télévision (35)	JBL (35)	Sony (27)	Samsung Electronics (23)
Habillement	Costume (34)	Décoration (29)	Jennyfer (28)	Lingerie (27)	3 Suisses (25) et Vêtements femme (25)

Source : Google Trends, FEVAD, Banque de France DGS SEEC.

Le tableau 8 illustre l'hétérogénéité des requêtes Google le plus souvent sélectionnées dans les modèles Google Trends : articles (four, télévision), marques (Cinna, Samsung), requêtes générales (vêtements femme, chaussures de foot), *pure player* (Spartoo, GrosBill) et spécialiste de la VAD (3 Suisses). La diversité des comportements des utilisateurs du moteur de recherche Google est bien retranscrite ici. Notons que la variable de tendance n'est jamais sélectionnée.

Pour chacun des produits, contrairement au cas de l'indice total, le modèle global est plus parcimonieux que le modèle Google Trends, réduisant ainsi un des risques de sur-apprentissage. En premier lieu, le tableau 9 présente les moyennes des RMSE.

Conformément aux attentes, les modèles disposant de toute l'information sont globalement meilleurs, au sens de la RMSE, que les modèles à une seule source d'information (Google Trends, CD, SARIMA). Le second enseignement du tableau 9 est que le modèle Google Trends s'ajuste systématiquement mieux aux données de l'échantillon que le modèle CD et le modèle SARIMA ; ce qui peut s'expliquer par un nombre plus grand de variables retenues.

Capacités prédictives

Si le modèle Google Trends était, en moyenne, systématiquement meilleur sur la période d'estimation (selon la RMSE) que les modèles CD et SARIMA, il n'est pas meilleur pour la prévision. Ses performances prédictives sont globalement du même ordre de grandeur que celles des modèles CD et SARIMA (tableau 10). Plus généralement, les résultats obtenus sur les différents produits sont mitigés et contrastés. L'ajout des données exogènes – que ce soit les données Google Trends ou les indices conjoncturels quantitatifs – ne diminue pas l'erreur de prévision.

Le modèle le plus performant pour les chaussures est le modèle CD ; le modèle Google Trends est légèrement meilleur, au sens de la RMSFE, que le modèle SARIMA. Concernant les meubles, l'apport de Google Trends est plus net. L'EGP progresse (par rapport au modèle SARIMA) également avec les données exogènes. En revanche, dans les cas de l'électroménager et de l'habillement, leur apport n'améliore pas les résultats (par rapport au modèle SARIMA).

Tableau 9
Moyenne des RMSE des modèles pour l'estimation des indices de CA des produits

	Google Trends	CD	SARIMA	Modèle global	Combinaison de modèles
Chaussures	8.2	10.5	10.9	7.8	7.6
Meubles	6.0	7.3	7.4	5.7	5.5
Électroménager	6.1	6.9	7.2	6.4	5.5
EGP	5.8	7.4	7.7	5.5	7.2
Habillement	5.3	6.0	6.3	5.9	4.3

Source : Google Trends, Banque de France DGS SEEC.

Tableau 10
RMSFE et écart-types liés à l'échantillonnage Google pour l'estimation des indices de CA des produits

	Google Trends		CD	SARIMA	Modèle global		Combinaison de modèles	
Chaussures	13.2	0.3	12.7	13.6	13.8	0.4	13.4	0.2
Meubles	11.9	0.5	12.3	12.0	13.2	0.4	11.8	0.5
Électroménager	11.7	0.3	10.4	10.2	12.3	0.3	11.2	0.3
EGP	15.5	0.3	15.3	16.4	11.5	0.5	13.1	0.3
Habillement	9.8	0.3	10.1	9.2	15.2	0.5	9.7	0.2

Note : les résultats présentés dans le corps de l'article sur les cinq produits sont issus de la même simulation que ceux de l'indice total ; les RMSFE sont très proches des médianes obtenues sur les trente simulations.

Lecture : la RMSFE du modèle Google Trends pour l'estimation de l'indice de CA des chaussures est de 13.2 ; sur les trente simulations réalisées pour évaluer la sensibilité des résultats à l'échantillonnage Google, l'écart-type obtenu est de 0.3. La RMSFE du modèle CD est de 12.7 (et n'est pas impactée par l'échantillonnage Google).

Source : Google Trends, Banque de France DGS SEEC.

Tableau 11
Poids des modèles individuels dans la combinaison de modèles

	Google Trends	CD	SARIMA
Chaussures	0.55	0.21	0.19
Meubles	0.71	0.09	0.13
Électroménager	0.47	0.14	0.33
EGP	0.48	0.20	0.27
Habillement*	0.58	- 0.50	0.80

* Pour l'habillement, les valeurs prédites par les trois modèles présentent de fortes colinéarités, mal gérées par l'agrégation Bayésienne : les contributions des variables dans les modèles « intermédiaires » (cf. détail du calcul des poids dans le cas du total, dans la partie ad hoc) sont artificiellement surévaluées ; phénomène répercuté dans les poids moyens de la combinaison de modèles.
Source : Google Trends, Banque de France DGS SEEC.

Ensuite, vis-à-vis de la combinaison de l'information, les résultats sont aussi mitigés. D'une part, la combinaison de modèles présente de meilleurs résultats prédictifs (RMSFE) que le modèle global, excepté pour l'EGP. D'autre part, la combinaison d'information n'apporte pas les résultats escomptés. Selon la RMSFE, la combinaison de modèles n'est meilleure que dans le cas des meubles, seul produit pour lequel le modèle Google Trends surperforme les modèles CD et SARIMA. De fait, les performances *in-sample* influent sur la pondération des modèles dans l'agrégation. Et, le modèle Google Trends fournissant de meilleures estimations (selon la moyenne des RMSE, cf. tableau 5), son poids dans l'agrégation est important (tableau 11).

Comme pour les moyennes des RMSE, celles des poids dans l'agrégation sont établies sur la période du protocole de test. L'habillement est le seul produit pour lequel le poids du modèle Google Trends n'est pas le plus important.

* *
*

Le e-commerce est un phénomène en pleine expansion. De fait, les ventes réalisées sur internet prennent davantage de poids dans la consommation des ménages, et donc dans l'enquête mensuelle de conjoncture Commerce de Détail de la Banque de France. Dans ce contexte, l'estimation des chiffres d'affaires livrés (tardivement) par la FEVAD devient un sujet de premier plan.

Jusqu'ici, elle était le fruit d'un modèle auto-régressif. Les travaux présentés dans cet article étudient l'apport des données exogènes que sont les indices du commerce de détail traditionnel

des ventes physiques (issus de l'enquête mensuelle de conjoncture) et les indices Google Trends. Chaque source apporte une information qui lui est propre. L'avantage commun de ces sources de données, à savoir être disponible avant les livraisons de la FEVAD, est idoine pour l'exercice de *nowcasting*.

Cependant, une nouvelle source de données (Google Trends) doit être utilisée avec précaution. D'une part, des tests de robustesse préalables à son utilisation ont été nécessaires. Un traitement systématique des valeurs aberrantes a été mis en place. Ces valeurs aberrantes sont parfois dues à des changements méthodologiques opérés par Google et avec peu de renseignements. En outre, la sensibilité des résultats à la méthode d'échantillonnage de Google invite à multiplier les simulations pour fiabiliser les résultats. D'autre part, il a fallu coupler l'immense champ des variables Google possibles avec la faible profondeur de l'historique des données FEVAD (livraisons mensuelles depuis 2012). Cette contrainte duale trouve sa solution dans le domaine du *machine learning*, avec le lasso adaptatif (Zou, 2006). La sélection de variables opère à chaque itération, palliant ainsi le risque lié au dynamisme du e-commerce et à la possible instabilité des mots-clés correspondant puisqu'il est possible de rétropoler les résultats avec d'autres jeux de variables. Ainsi, le modèle est souple et présente une forte capacité d'adaptation, nécessitée par la mouvance du phénomène modélisé.

S'ensuit la question d'exploiter la complémentarité des différentes sources de données. Dans cette étude, l'agrégation bayésienne de modèles simples apporte de meilleurs résultats, en termes de RMSFE, que le modèle global (lasso adaptatif appliqué à toutes les variables simultanément). La petite taille des échantillons

d'estimation des modèles peut contribuer à défavoriser un modèle avec beaucoup de variables. Par exemple, dans le cas de l'indice total, un phénomène de sur-apprentissage est détecté pour le modèle global. De plus, l'agrégation apporte de la lisibilité dans la combinaison de modèles, utile en production.

En général, l'apport des données exogènes reste mitigé. Il est plus clair dans le cas de l'indice total que ceux des produits. Les livraisons de la FEVAD sont établies sur l'échantillon de ses 70 déclarants les plus gros (en termes de CA). Le nombre de déclarants est donc inférieur pour les produits ; ce qui participe à justifier les résultats plus heurtés et donc plus difficiles à appréhender. Ainsi, l'erreur de prévision sur le chiffre d'affaires est deux à trois fois plus forte pour les produits que pour le total.

Enfin, une des causes possibles des résultats mitigés réside dans les choix de modélisation. S'ils

répondent à de nombreuses contraintes du sujet, les effets saisonniers n'y occupent pas une place centrale. Du fait des courts historiques, la modélisation n'opère pas sur les séries corrigées des variations saisonnières, contrairement à l'approche économétrique classique ; ici, la présence de l'estimation SARIMA dans les variables explicatives vise à capter les effets saisonniers. Cette méthode néglige cependant les différences de saisonnalités entre variables endogènes et exogènes.

Avec des historiques plus longs, les saisonnalités des séries du e-commerce devraient se stabiliser, offrant l'opportunité d'affiner les résultats avec d'autres modélisations. Outre la possibilité de travailler sur les séries corrigées des variations saisonnières, coupler la modélisation Reg ARIMA, dont la spécification du résidu est plus pertinente, aux méthodes de sélection de variables peut s'avérer intéressant et manque aujourd'hui dans la littérature. □

BIBLIOGRAPHIE

Aioffi, M. & Timmerman, A. (2006). Persistence of forecasting performance and combination strategies. *Journal of Econometrics*, 135(1-2), 31–53. <https://doi.org/10.1016/j.jeconom.2005.07.015>

Askistas, N. & Zimmerman, K. F. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, 55(2), 107–120. <https://elibrary.duncker-humboldt.com/journals/id/22/vol/55/iss/1486/art/5561/>

Bates, J. & Granger, C. (1969). The combination of forecasts. *Operational Research Quarterly*, 20(4), 451–468. <https://doi.org/10.1057/jors.1969.103>

Bec, F. & Mogliani, M. (2015). Nowcasting French GDP in real-time with surveys and “blocked” regressions: Combining forecasts or pooling information? *International Journal of Forecasting*, 31(4), 1021–1042. <https://doi.org/10.1016/j.ijforecast.2014.11.006>

Bortoli, C. & Combes, S. (2015). Apports de Google Trends pour prévoir la conjoncture française : des pistes limitées. Insee, *Note de conjoncture*, mars 2015. https://www.insee.fr/fr/statistiques/fichier/1408926/mars2015_d2.pdf

Breiman, L. (1996). Stacked Regressions. *Machine Learning*, 24, 49–64. <https://statistics.berkeley.edu/sites/default/files/tech-reports/367.pdf>

Choi, H. & Varian, H. (2009). Predicting Initial Claims for Unemployment Benefits. Google, *Technical Report*. <https://static.googleusercontent.com/media/research.google.com/fr/archive/papers/initialclaimsUS.pdf>

Choi, H. & Varian, H. (2011). Predicting the Present with Google Trends. Google, *Technical Report*. <http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf>

Clements, M. & Galvão, A. (2008). Macroeconomic Forecasting With Mixed-Frequency Data: Forecasting Output Growth in the United States. *Journal of Business & Economic Statistics*, 26(4), 546–554. <https://doi.org/10.1198/073500108000000015>

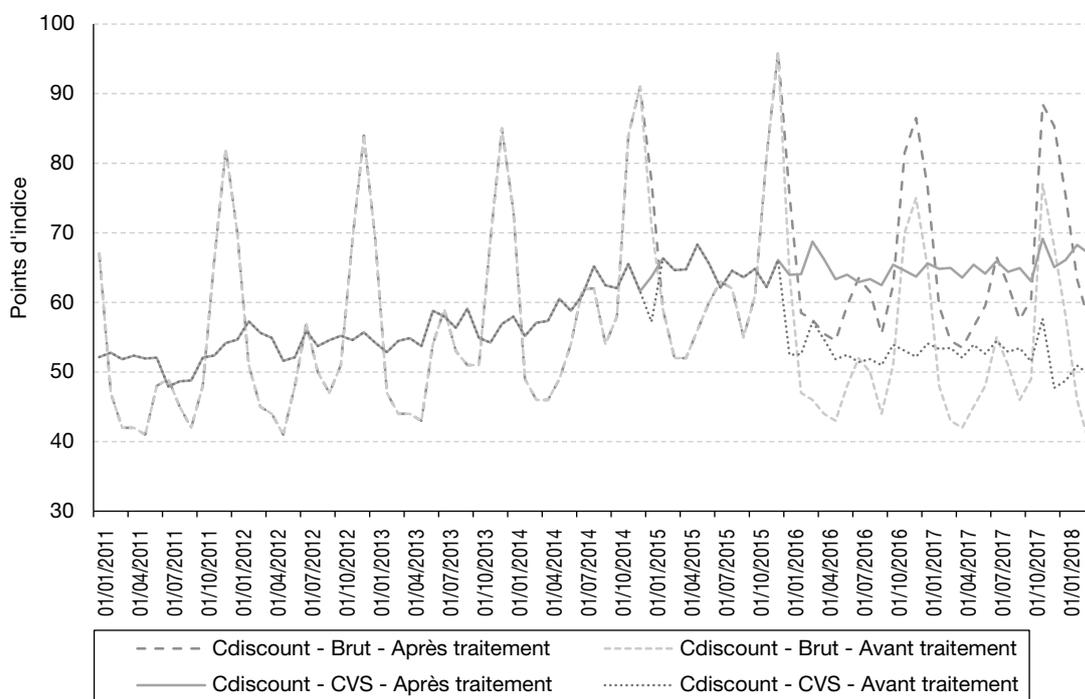
Diebold, F. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*, 5(4), 589–592. [https://doi.org/10.1016/0169-2070\(89\)90014-9](https://doi.org/10.1016/0169-2070(89)90014-9)

- Diebold, F. & Mariano, R. (1995).** Comparative Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144.
<https://doi.org/10.1198/073500102753410444>
- De Gooijer, J. & Hyndman, R. (2006).** 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), 443–473.
<https://doi.org/10.1016/j.ijforecast.2006.01.001>
- Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004).** Least angle regression. *The Annals of Statistics*, 32(2), 407–499.
<https://doi.org/10.1214/009053604000000067>
- Elliott, G., Rothenberg, T. & Stock, J. (1996).** Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64(4), 813–836.
<https://doi.org/10.2307/2171846>
- Ettredge, M., Gerdes, J. & Karuga, G. (2005).** Using Web-based Search Data to Predict Macroeconomic Statistics. *Communications of the ACM*, 48(11), 87–92.
https://www.researchgate.net/publication/200110929_Using_Web-based_search_data_to_predict_macro_economic_statistics
- Eurostat (2018).** *Handbook on Seasonal Adjustment*. Luxembourg: Publications Office of the European Union.
<https://ec.europa.eu/eurostat/documents/3859598/8939616/KS-GQ-18-001-EN-N.pdf>
- FEVAD (2016 et 2017).** *Chiffres clés*.
https://www.fevad.com/wp-content/uploads/2016/09/Plaque-Clés-2016_Fevad_205x292_format-final_bd.pdf
<https://www.fevad.com/wp-content/uploads/2018/06/Chiffres-Clés-2018.pdf>
- Granger, C. & Newbold, P. (1974).** Spurious Regressions in Econometrics. *Journal of Econometrics*, 2(2), 111–120.
[https://doi.org/10.1016/0304-4076\(74\)90034-7](https://doi.org/10.1016/0304-4076(74)90034-7)
- Hoerl, A. & Kennard, R. (1970).** Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67.
<http://www.jstor.org/stable/1267351?origin=JSTOR-pdf>
- Hoeting, J., Madigan, D., Raftery, A. & Volinsky, C. (1999).** Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4), 382–417.
<https://www.jstor.org/stable/2676803>
- Huang, H. & Lee, T. (2010).** To Combine Forecasts or to Combine Information? *Econometric Reviews*, 29(5-6), 534–570.
<https://doi.org/10.1080/07474938.2010.481553>
- Hyndman, R. & Athanasopoulos, G. (2018).** *Forecasting: principles and practice*. Melbourne, Australia : OTexts.
<https://otexts.org/fpp2/>
- Kozicki, S. & Hoffman, B. (2004).** Rounding Error: A Distorting Influence on Index Data. *Journal of Money, Credit and Banking*, 36(3), 319–338.
<https://www.jstor.org/stable/3838976>
- Kuzin, V., Marcellino, M. & Schumacher, C. (2013).** Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics*, 28(3), 392–411.
<https://doi.org/10.1002/jae.2279>
- Marin, J. & Robert, C. (2010).** Les bases de la statistique bayésienne. *Rapport des Universités Montpellier II et Dauphine – Crest Insee*.
<https://www.ceremade.dauphine.fr/~xian/mr081.pdf>
- McLaren, N. & Shanbhogue, R. (2011).** Using internet search data as economic indicators. Bank of England, *Quarterly Bulletin*, 51(2), 134–140.
<https://econpapers.repec.org/RePEc:boe:qbullt:0052>
- Phillips, P. (1986).** Understanding spurious regression in econometrics. *Journal of Econometrics*, 33(3), 311–340.
[https://doi.org/10.1016/0304-4076\(86\)90001-1](https://doi.org/10.1016/0304-4076(86)90001-1)
- Phillips, P. & Perron, P. (1988).** Testing for a Unit Root in Time Series Regression. *Biomètrika*, 75(2), 335–346.
<http://www.jstor.org/stable/2336182?origin=JSTOR-pdf>
- Tibshirani, R. (1996).** Regression shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.
<https://www.jstor.org/stable/2346178>
- Zeugner, S. (2011).** *Bayesian Model Averaging with BMS*.
<https://cran.r-project.org/web/packages/BMS/vignettes/bms.pdf>
- Zou, H. (2006).** The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.
<https://doi.org/10.1198/016214506000000735>

ANNEXE 1

TRAITEMENT DES VALEURS ABERRANTES – L'EXEMPLE DE CDISCOUNT

Figure A1
 Traitement de la rupture de série de l'indice Google Trends Cdiscount



Note : le traitement opéré sur la série Cdiscount est analogue à celui d'Amazon.
 Source : Google Trends, Banque de France DGS SEEC.

LISTE DES VARIABLES PAR PRODUIT

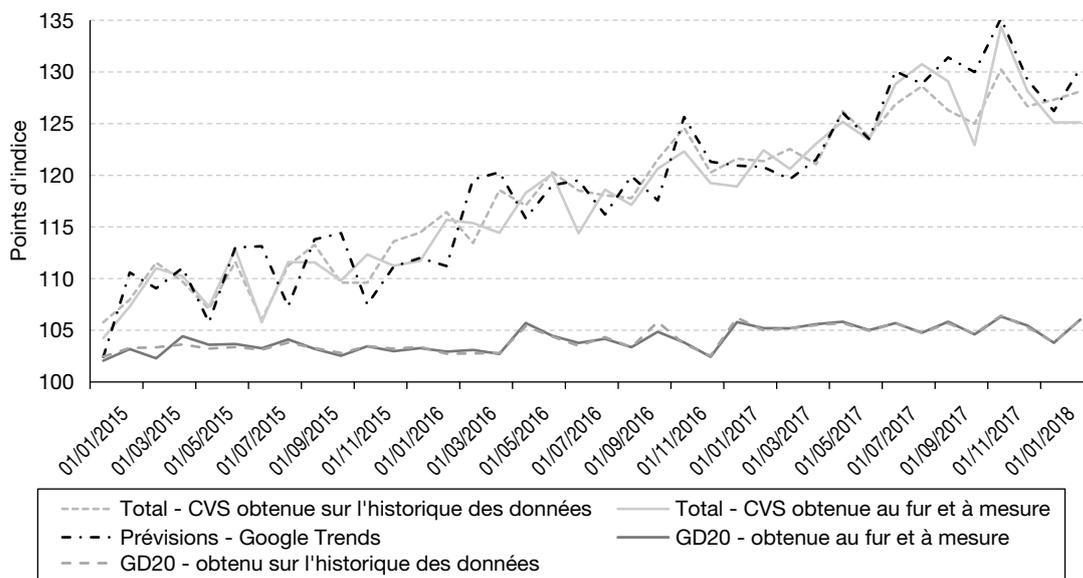
Tableau A2
Jeux initiaux de variables pour chaque estimation

Total	Amazon, eBay, Vente privée, Cdiscount, Fnac, Groupe Fnac Darty, PriceMinister, Leroy Merlin, UGAP, Castorama, Boulanger, Carrefour, Showroomprive, E. Leclerc, La Redoute, Auchan, Raja, Rue du commerce, 3 Suisses, Promos + soldes + blackfriday, Alibaba, Groupon, PhotoBox, Galeries Lafayette, Yves Rocher, Sephora, Decathlon
Habillement	Vertbaudet, Kiabi, H&M, C&A, Jules, Linge de maison, Zara, Costume, Culotte, 3 Suisses, Devred, Robe, Etam, La Redoute, Jeans + Chino + Pantalon, Manteau + Blouson, Veste, Vêtement femme, ASOS, Maisons du Monde, Lingerie, Jennyfer, Vêtement, Galeries Lafayette, Bonobo, Brandalley, Camaïeu, Showroomprive, Vente Privée, Rideau, Blanche Porte, Drap, Coussin, Homemaison, Sous-vêtement, La Halle, Decathlon
EGP	iPhone, Apple, Cdiscount, PC Gaming, iPad, Téléphone + smartphone, FNAC, Télévision, Boulanger, Sony, LDLC Pro, Amazon, Phillips, LG Group, Samsung Electronics, Darty, Tablette tactile, Enceinte, Appareil photographique reflex numérique, Ordinateur portable +PC, Bose, JBL, Groupe Fnac Darty, Barre de son, Appareil photo, Marshall, Groupe Samsung
Chaussures	Chaussures, Chaussure, Ceinture, Maroquinerie, Botte, Chaussures de sport, Vans, Converse, Zalando, Spartoo, Sarenza, Showroomprive, Prada, Escarpin, Adidas Stan Smith, Chaussures femme, Ballerine, Chaussures homme, Timberland, Chaussures de foot, Chaussures enfant, San Marina, Eram, Chaussures de ville, J.M Weston, Chaussea, Bexley, Gêmo, Sac à main, La Halle, Chaussures Nike
Électroménager	Clubic, Boulanger, Cdiscount, Four, Réfrigérateur, Machine à laver, Darty, Bosch, Electrolux, Conforama, Amazon, Cuisinière, Électro-dépôt, Brandt, Four à micro-ondes, Groupe Fnac Darty, Aspirateur, Whirlpool Corporation, Mistergooddeal, GrosBill, Pulsat, Ubaldi, But
Meubles	But, Legallais, Cuisine, Raja, Staples, Roche Bobois, Castorama, Conforama, Vega, Bureau, Meuble, Leroy Merlin, Ikea, Couteaux, Armoire + étagère, Maisons du Monde, Cinna, Meuble en bois, Roset, Buffet + commode + vaisselier, Table + chaise + canapé, Fauteuil

Lecture : l'utilisation du « + » permet de constituer un indice Google Trends correspondant au cumul des requêtes.
Source : Google Trends, DGS SEEC Banque de France.

ANNEXE 3

INSTABILITÉ DES DERNIERS POINTS D'UNE CVS

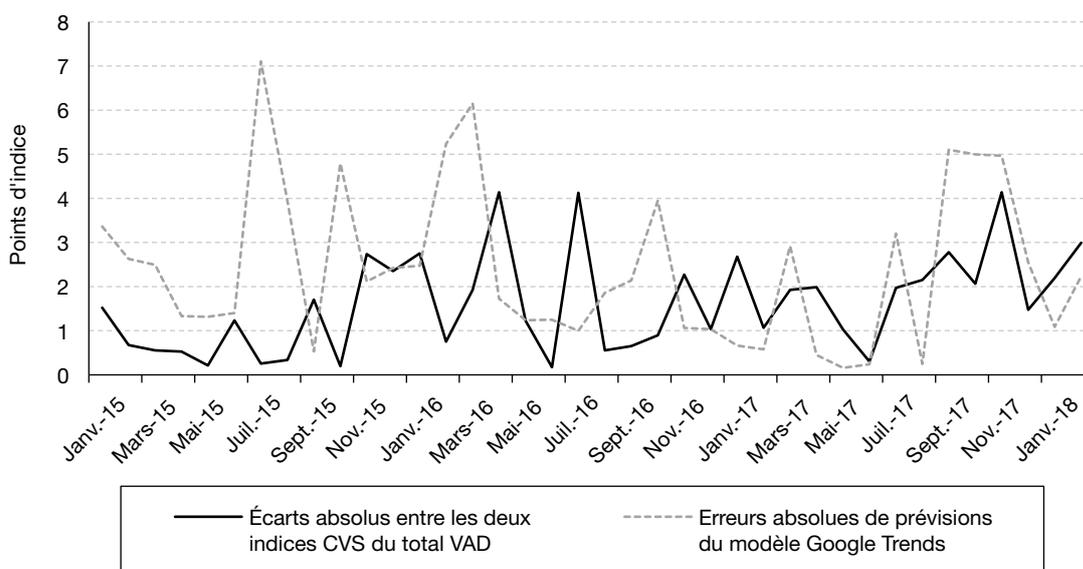
 Figure A3-I
Stabilité des CVS sur le dernier point


Source : Google Trends, FEVAD, Banque de France DGS SEEC.

Deux groupes de séries se distinguent. Le premier concerne la VAD et est constitué de trois séries : l'indice total VAD dont la dessaisonnalisation tient compte de l'ensemble des données, l'indice total VAD implémenté au fur et à mesure – i.e. chaque point est le dernier de la série CVS obtenue à partir de l'indice tronqué à cette date – et l'indice des prévisions Google Trends. Le second groupe s'intéresse à la grande distribution : l'écart moyen absolu entre les deux indices CVS de la grande distribution (obtenus avec toutes les données *versus* celles disponibles au fur et à mesure) est de 0.2 point (0.05 lorsque la grandeur est rapportée à l'amplitude, définie comme étant la plus

grande variation de la série de référence – la CVS obtenue sur l'historique des données) ; contre 1.6 point (0.29 rapporté à l'amplitude) pour les deux indices CVS du total VAD.

Si le phénomène des révisions des derniers points des CVS est connu (cf. Eurostat, 2018), l'ampleur constatée ici interpelle : ces deux indices diffèrent dans des proportions comparables aux écarts de prévision des modèles – les évolutions des écarts entre, d'une part, les deux CVS et, d'autre part, les prévisions issues du modèle Google Trends et la série CVS obtenue sur l'historique des données en rendent compte (figure A3-II).

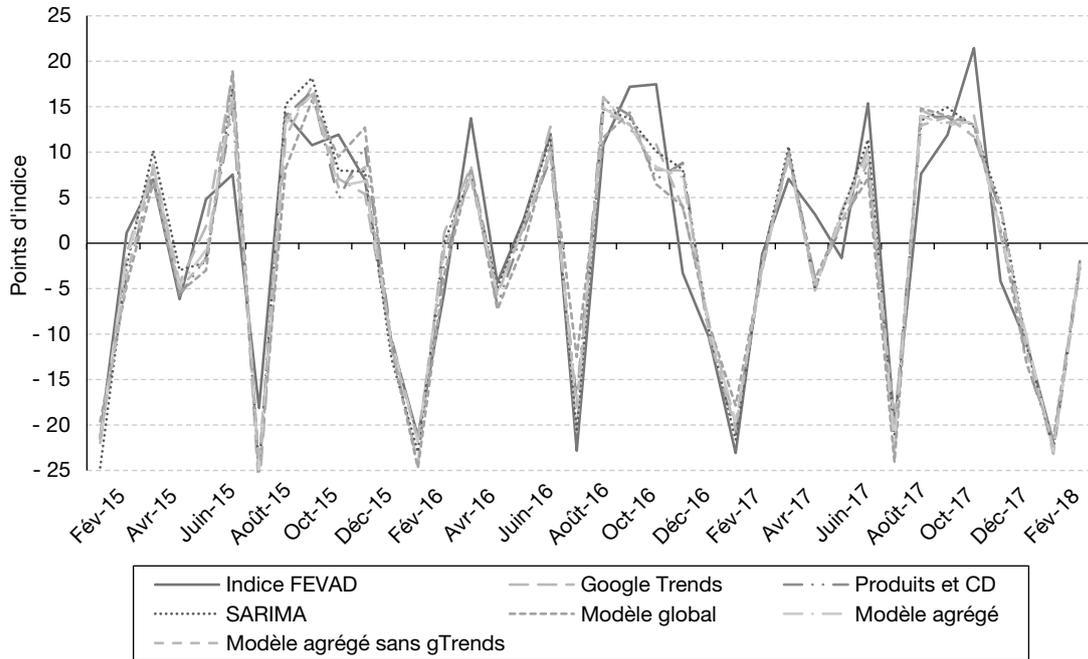
 Figure A3-II
Écarts absolus avec la CVS obtenue sur l'historique des données


Lecture : l'estimation CVS réalisée avec les données disponibles aujourd'hui, s'écarte de 4.1 points d'indice de celle réalisée en pseudo temps réel. La prévision Google Trends de juillet 2016 est même plus proche de la valeur de l'indice CVS obtenu aujourd'hui que celle obtenue avec les données disponibles en juillet 2016.

Source : Google Trends, FEVAD, Banque de France DGS SEEC.

PRÉVISIONS DES MODÈLES POUR L'INDICE DE CA TOTAL

Figure A4
Prévisions des différents modèles sur l'estimation du total



Source : Google Trends, FEVAD, Banque de France DGS SEEC.

ANNEXE 5

DESCRIPTIF DES SÉLECTIONS DE VARIABLES DANS LE MODÈLE GOOGLE TRENDS POUR L'ESTIMATION DU CA TOTAL

Tableau A5

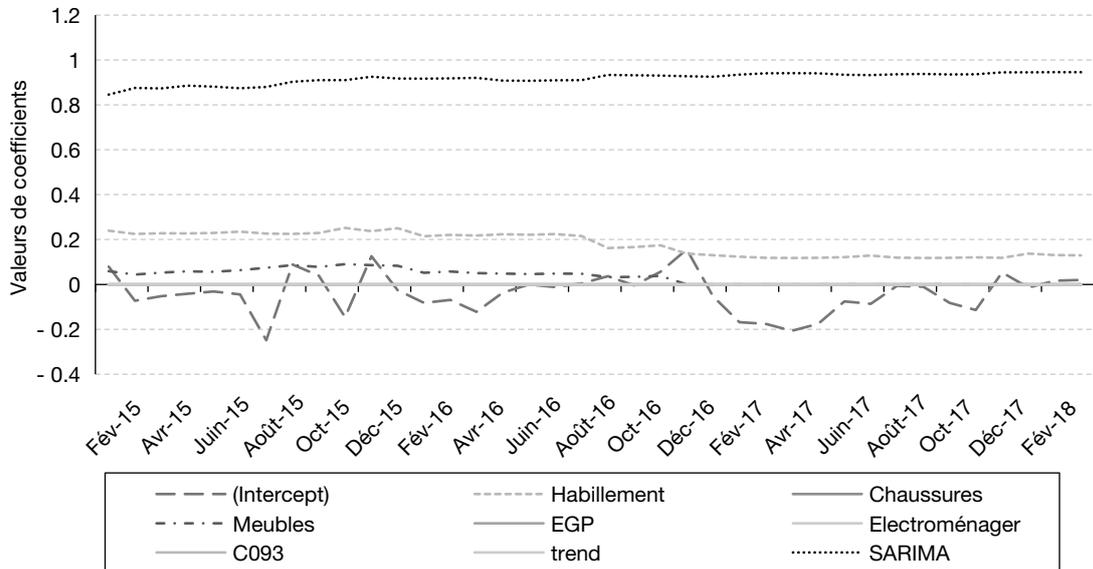
Statistiques descriptives des variables sélectionnées pour l'estimation du total

	Moyenne	Minimum	Maximum	Sélections
Amazon	0.07	0.00	0.48	9
eBay	- 0.43	- 0.80	- 0.24	38
Vente-privee.com	- 0.02	- 0.21	0.00	8
Cdiscount	0.01	0.00	0.07	5
FNAC	0.01	0.00	0.18	5
Groupe Fnac Darty	0.00	0.00	0.00	0
PriceMinister	- 0.21	- 0.37	0.00	37
Leroy.Merlin	0.05	0.00	0.13	32
Union des groupements d'achats publics	0.01	0.00	0.06	14
Castorama	- 0.02	- 0.09	0.00	13
Boulanger	0.00	0.00	0.00	0
Carrefour	0.01	0.00	0.12	2
Showroomprive.com	0.12	0.00	0.29	35
E.Leclerc	0.00	0.00	0.00	0
La Redoute	0.00	- 0.04	0.00	1
Auchan	0.01	0.00	0.11	5
Raja	0.02	0.00	0.10	13
Rue du Commerce	- 0.05	- 0.38	0.00	8
3 Suisses	0.06	0.00	0.33	15
Promos + soldes + blackfriday	0.00	0.00	0.00	0
Alibaba Group	0.00	0.00	0.05	7
Groupon	0.08	0.00	0.14	37
PhotoBox	0.00	0.00	0.00	0
Galerie Lafayette	0.00	- 0.02	0.00	4
Yves Rocher	0.00	0.00	0.00	0
Sephora	0.00	0.00	0.04	2
Décathlon	0.00	- 0.10	0.00	3
Trend	0.00	0.00	0.00	0
SARIMA	0.97	0.93	1.04	38

Source : Google Trends, FEVAD, Banque de France DGS SEEC.

ÉVOLUTION DES COEFFICIENTS DU MODÈLE CD DANS L'ESTIMATION DU CA TOTAL

Figure A6
Évolution des coefficients du modèles CD pour l'estimation du total



Note : seules les variables sélectionnées ne sont pas en trait plein.
Source : FEVAD, Banque de France DGS SEEC.

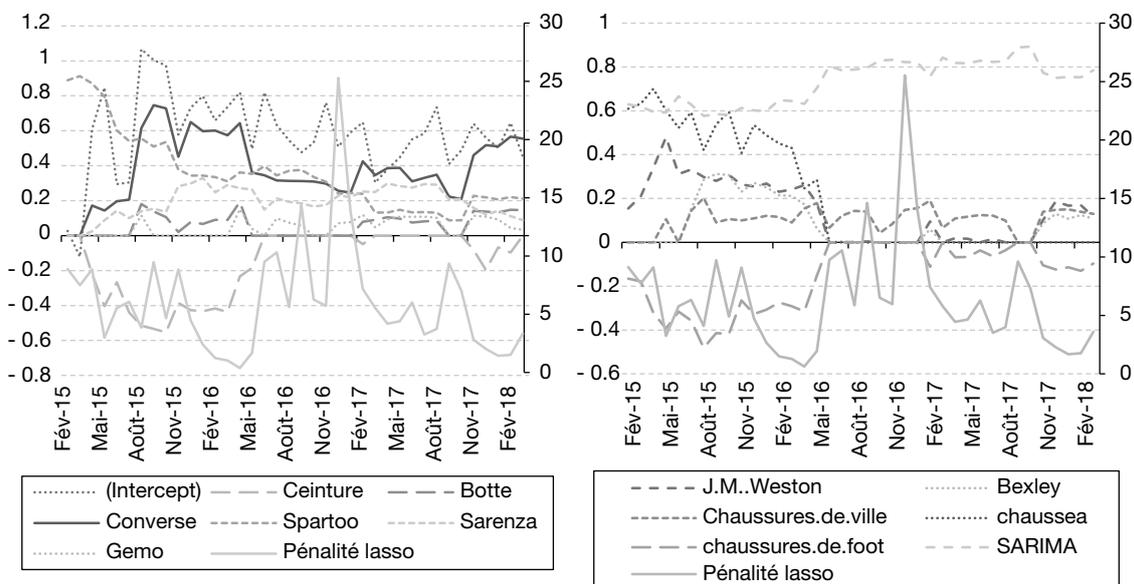
ANNEXE 7

STABILITÉ DU MODÈLE GOOGLE TRENDS DANS L'ESTIMATION DE L'INDICE DES CHAUSSURES

De manière analogue à la figure VIII présentant l'évolution des coefficients du modèle Google Trends pour l'estimation de l'indice de CA total, la pénalité lasso est en axe secondaire. Seules

les variables les plus présentes (retenues au moins 8 fois sur les 38 itérations) au cours du temps sont représentées sur la figure A7.

Figure A7
Évolution des coefficients du modèle Google Trends pour l'estimation de l'indice des chaussures



Source : Google Trends, Banque de France DGS DESS SEEC.

