

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

François Robin

Question clé

Les données Google Trends restituent le nombre de recherches effectuées par les internautes sur le moteur Google pour un thème donné. Elles sont gratuites et disponibles en quasi temps réel, soit des caractéristiques *a priori* intéressantes. Cependant, apportent-elles une information complémentaire sur le chiffre d'affaires du e-commerce du mois passé par rapport à des données d'enquête et à l'évolution attendue du fait de l'historique des séries ?

Méthodologie

La méthodologie de construction des indices Google Trends est opaque. Cela nécessite de corriger automatiquement les valeurs aberrantes. Ensuite, l'immense champ de variables possibles invite à la sélection, en recourant à l'« *adaptive Lasso* ». Enfin, utiliser trois sources d'information permet de confronter la combinaison de modèles simples, *via* l'agrégation Bayésienne, au modèle global.

Principaux résultats

- Trois modèles simples nous permettent de fournir une prévision du e-commerce à horizon d'un mois : le SARIMA et deux modélisations *adaptive Lasso*, utilisant d'une part les données Google Trends et d'autre part les indices quantitatifs conjoncturels issus de l'enquête mensuelle de conjoncture du commerce de détail. La combinaison de ces trois modèles et le modèle global (*adaptive Lasso* appliqué à toutes les variables) disposent de toute l'information.
- Pour chacune des six estimations (total, chaussures, EGP, habillement, meubles et électroménager), les prévisions des cinq modèles sont proches. L'apport des données exogènes est plus net pour le total.
- Le modèle Google Trends est parcimonieux et stable, en termes de sélection et de coefficients. Cependant, les résultats des modèles utilisant les Google Trends sont sensibles à l'échantillonnage opéré par Google pour leur construction.
- La combinaison de modèles est globalement meilleure que le modèle global.

Performance des modèles dans l'estimation du e-commerce total

| Total | Google Trends | CD | SARIMA | Modèle global | Combinaison de modèles |
|--|---------------|-----|--------|---------------|------------------------|
| Erreur quadratique moyenne des erreurs de prévisions | 4.8 | 5.2 | 5.0 | 5.5 | 4.8 |

Note : une erreur quadratique moyenne plus faible indique une meilleure performance.

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

Message

Gratuites et disponibles en quasi temps réel, les données Google Trends présentent des caractéristiques intéressantes pour prévoir le e-commerce à horizon d'un mois. Néanmoins, leur utilisation nécessite une analyse préalable des thèmes de recherche adéquats, l'automatisation de la correction des valeurs aberrantes et la multiplication des recherches de termes identiques pour en fiabiliser les résultats. En particulier du fait de la variabilité du commerce en ligne et de l'adaptation des habitudes de consultation sur internet, les données Google Trends sont utiles mais pas décisives pour prévoir l'activité du e-commerce à court terme.