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

François Robin*

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 single models.

JEL Classification: C51, C53, C11, E17
Keywords: Google Trends, nowcasting, trends, e-commerce, distance selling, Big Data, Bayesian averaging, variable selection, lasso

** Banque de France, Economic Surveys Department, Directorate General Statistics (francois.robin@banque-france.fr)*

The author would like to thank in particular those who took part in an internal research workshop at the Banque de France, Patrick Kwok for his regular follow-up, contributions and support for this work. A big thank you also to Martial Ranvier and Valérie Chauvin for their helpful comments and valuable methodological input. Lastly, thank you to François Guinouard, Léon Ipdjian and Mathilde Gerardin; and to François Brunet, who proposed the topic.

Translated from the original version: "Utilisation de Google Trends dans les enquêtes mensuelles sur le Commerce de Détail de la Banque de France"

To cite this article: Robin, F. (2018). Use of Google Trends Data in Banque de France Monthly Retail Trade Surveys. *Economie et Statistique / Economics and Statistics*, 505-506, 35–63.
https://doi.org/10.24187/ecostat.2018.505d.1965

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) provides monthly B2C (Business-to-Consumer) online retail sales data.[1] However, data releases are too late to be included in the first publication of the monthly retail trends survey, which instead uses estimates of the data.

Until now, the paucity of historical time series data limited the range of possible forecasting methods. The autoregressive model used until now may now be complemented by using exogenous data available at the time of carrying out estimations: quantitative indices for traditional retail trends (from the monthly retail trends survey) and Google Trends data. Estimation of FEVAD data for month M takes place during the survey period (at the beginning of month $M + 1$). Quantitiative monthly survey indices M are at this time under construction, while monthly Google Trends data for month M are final. These estimations fall fully within the scope of a nowcasting exercise.

This development runs into two issues. First, Google Trends provides a range of explanatory variables, from which those most suitable must be selected. The "adaptive lasso" machine learning-based approach developed by Zou (2006) addresses the twin challenges posed by a lack of historical FEVAD time series data (dating back to 2012 only) and the huge range of possible Google queries. Second, given the number of available models with exogenous variables from different sources, it is helpful to confirm whether the combination of models can produce better output. This topic has been the subject of much debate, as outlined in Bec & Mogliani (2015).

Following a review of the relevant literature, the second part looks at the datasets in greater detail, including an overview of the retail monthly survey data, FEVAD data and Google Trends data. Due to the particular nature and lack of clarity around the methodology of construction of Google Trends data, robustness checks and automated error corrections linked to breaks in series are required. The third part addresses the choice of models, looking at how stationarity is managed in time series, then at the stages of the model testing process. The fourth part examines the results and how they are interpreted. The final part concludes.

Literature Review

Google Trends Data

Available almost in real time, Google Trends indices show the evolution in queries made by users of the Google search engine over time. These indices represent an information flow and a source of big data. While there is no record of their use by public sector institutions in recurring studies, they have been the subject of a number of publications. Research by Ettredge *et al.* (2005) and Askistas & Zimmerman (2009), which look at the unemployment rate forecasting using keywords used in Google searches, indicate the potential benefit of such indices. Choi & Varian (2009, 2011) are more cautious about input from Google Trends. However, their literature review contains a number of papers using Google searches, mainly in the field of epidemiology; the tool used at the time and developed by Google (Google Flu) was discontinued on 20 August 2015, in light of shortcomings already highlighted by Bortoli & Combes (2015).

These tools are operated entirely by Google: the construction methodology is vague, which presents further risks for users. Methodological changes to Google Trends may involve breaks in series. Furthermore, the introduction of new actors has an impact on how user queries are formulated. McLaren & Shanbhogue (2011) warn about the mechanical fall in popularity of certain requests in their application to unemployment (e.g. in France, when the ANPE became "*Pôle emploi*" after the restructuring of the employment agency and unemployment insurance, then the ANPE and Assedic, respectively).

Selection of Variables

Machine learning methods offer a solution to variable selection, in particular the adaptive lasso approach developed by Zou (2006). As a reminder, the standard lasso[2] function introduced by Tibshirani (1996) is:

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

1. *According to the FEVAD, online retail as a share of total retail (excluding food and drink, in line with the scope of the Banque de France retail trends survey) was 7% in 2013, 8% in 2014 and 9% in 2015.*
2. *Lasso stands for* Least Absolute Shrinkage and Selection Operator.

In a lasso regression, the same penalty function λ is applied to all variables. Zou (2006) proposes adjusting the penalty based on the variables in the adaptive lasso (*adalasso*) :

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

The adaptive lasso is a weighted lasso. Its Oracle properties, as demonstrated by Zou (2006), offer the adaptive lasso two advantages over the standard lasso. The first is the consistency in its variable selection, i.e. the best sub-set of variables (from the initial set) is chosen; which is not always the case with a standard lasso (see Zou, 2006). The other Oracle property is the consistency of parameter estimation (asymptotic convergence of the estimator in normal distribution).

While Zou (2006) defines individual penalties as $\widehat{w}=1/\left|\widehat{\beta}\right|^{\gamma}$, with $\widehat{\beta}$ the ordinary least squares estimator and $\gamma>0$ (in practice, $\gamma\in\{0.5\ ;1\ ;2\}$), an alternative approach involves using the estimator from the ridge regression,[3] introduced by Hoerl & Kennard (1970), to define the vector of individual penalties. Its use helps prevent errors in estimation of penalties due to multicolinearity in the regressors.

The adaptive lasso is optimised in two stages. First, the individual penalties are obtained from a ridge regression:

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

The penalty value κ is then obtained by leave-one-out cross-validation (Hyndman & Athanasopoulos, 2018).[4] Then, $\widehat{w}=\widehat{\beta}_{ridge}$ results in the lasso function (for which the penalty λ is also optimised by leave-one-out cross-validation):

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

The advantage of the adaptive lasso is its large-scale operation (greater number of variables than observations, i.e. to the size of the temporal window in this case). It is also considered parsimonious. Both of these properties address the twin challenges posed by the large number of possible Google queries and the short historical time series for FEVAD releases.

Combination of Models or Global Model?

Three individual models have been used: the Google Trends model, the retail model from the retail trends survey, and the SARIMA model that has been used up to now.[5]

Bec & Mogliani (2015) document the most common methods of combining data. In their view, Bates & Granger (1969) were the first to support the aggregation of forecasts from different models. Subsequently, Diebold (1989) recommends the use of a single model, combining multiple heterogeneous data sources. More recently, Huang & Lee (2010) argue that a global model with sound specifications is preferable. Moreover, Clements & Galvão (2008) and Kuzin *et al*. (2013) argue in favour of aggregation for empirical applications. Bec and Mogliani (2015) find that aggregation performs better when forecasting movements in consumption indices. The test designed by Diebold & Mariano (1995), whose null hypothesis is that two forecasts generated by different models are not significantly different, is a critical indicator when opting for a model.

This paper seeks to contribute to the debate around a new application by comparing output from a combination of models with that from a global model with the same specifications as the individual models, in this case the adaptive lasso applied to all regressors simultaneously (Google Trends, retail indices and SARIMA). De Gooijer & Hyndman (2006) highlight the benefits of aggregation, in particular comprehensibility where aggregated models can be easily interpreted. Here, aggregation applies to three individual models, each with their own effects:

- The SARIMA model reproduces the past time series pattern;
- The retail model exploits traditional retail data;
- The data of interest is extracted from the model based on Google Trends indices.

The issue is weighting each forecast:

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

There are a number of possible approaches to aggregation: from the most straightforward, such

3. Ridge and lasso regressions are penalised by L2 and L1 norms respectively.
4. Specifically, the validation sample is made up of one observation; the training sample is made up of the $n-1$ other observations (for sample size n). The n values for κ, obtained for each training sample (each minimising the RMSE) give a mean to obtain the final value of κ.
5. The retail model is an adaptive lasso function for which explanatory variables are quantitative retail indices.

as weighting by the mean ($\gamma = \mu = \vartheta = 1/3$) or by the inverse of errors – in-sample or out-of-sample (see Aiofli & Timmerman, 2006), to the most elaborate. For example, Bayesian inference, based on Bayes' theorem[6] (see Marin & Robert, 2010), determines the probability of an event from prior measured events. Bayesian statistics, commonly used for small sample sizes, produces methods of classification, or aggregation in this case. Hoeting *et al.* (1999) highlight the effectiveness of Bayesian aggregation. Zeugner (2011) developed an R package on this subject. The purpose is to test models of a given category M and weight them according to their probability of being the correct model. The category M is that for linear models. Usually, the large number of models complicates Bayesian aggregation (see Hoeting *et al.*, 1999). This is not the case here: with three regressors (for the Google Trends, retail trends and SARIMA model estimates), eight linear models are possible. By denoting data as D and a given model as $M_j\left(1 \leq j \leq 8\right)$ un modèle donné, le théorème de Bayes donne Bayes theorem gives:

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

Both terms of the numerator used to measure *a posteriori*[7] probability are as follows:

- $P\left(M_j\right)$ corresponds to the *a priori*[8] probability that model M_j is the correct one;

- $P\left(D|M_j\right) = \int pr\left(D|\beta_j, M_j\right) pr\left(\beta_j|M_j\right) d\beta_j$ with β_j the parameters of the model: $\beta_j = \left\{\gamma_j, \mu_j, \vartheta_j\right\}$ estimé sur le modèle M_j. Here, we are interested in the values for the parameters.

Specifically, the values of coefficients obtained in each model M_j for category M are weighted by the probability that each model M_j is the correct one: $\gamma = \sum_i \gamma_i P(M_i \mid D)$ with $\gamma_i = E(\gamma \mid D, M_i)$ the value of the coefficient in model M_i. The same applies to μ and ϑ.

Data

The Monthly Retail Trends Survey

One of the monthly trend surveys undertaken by the Banque de France covers the retail sector.[9] The survey tracks changes in sales including tax for a sample population of 6,800 (divided among more than 4,000 businesses); each month, the response rate is approximately 90%. Each entity provides its total sales figure and the respective shares of its main products (where it is not a "single-good" retailer). Individual data are then grouped according to characteristics common to retail businesses: by method of distribution (physical: small traditional retailer, large specialist and chain retailer, hypermarkets and supermarkets, department store and variety store and mail-order store: distance selling) and by product (e.g. household appliances, shoes, etc.). Quantitative indices are established for these groupings (e.g. small traditional furniture retailers).

Construction of Quantitative Survey Indices

Each sales index Y from the survey is constructed as follows (with X the relevant sales figure):

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

All quantities for the above equation apply to the same companies. Under the survey methodology, sales samples are "balanced", i.e. the scope of X_M is the same as for X_{M-12}. In other words, the same companies are measured in the case of both sales values. This approach prevents extreme variations not representative of the sample (outliers). The closure (or opening) of a store is the most common extreme event that is problematic for the survey: the resultant reduction (respectively increase) in sales is offset by opposite movements for all of its competitors, which will not be fully captured by the sample. Furthermore, it is easier to track a store closure than a store opening (store or new brand not yet included in the sample), which would present a measurement bias in the case of unbalanced data.

Availability of Indices

Only a portion of indices cross-referencing products and physical distribution systems (in-store sales) are measured. This is due to the lack of an adequate sample size and for data

6. *The theorem is commonly formulated as:* $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})}$ *with P measuring probability,* A *and* B *two events.*
7. *The* a posteriori *probability is obtained using in-sample data.*
8. *There are a number of ways to obtain* a priori *probabilities, as shown in Zeugner (2011). In our case, a number of tests were carried out (e.g. prior binomial, uniformity, deterministic, etc.) with no significant effect on output.*
9. *The latest survey results are available at (French only): https://www.banque-france.fr/statistiques/chiffres-cles-france-et-etranger/enquetes-de-conjoncture/conjoncture-commerce-de-detail.*

Table 1
Mean and Standard Deviation of Quantitative Survey Indices for Physical Retail Sales (Retail Trends Survey)

	Small traditional retailers	Large specialist retailers and chains	Hypermarkets and supermarkets	Department stores and variety stores	All physical sales
Total industrial products excl. cars					91.2 17.5
Shoes	91.2 21.1	94.3 27.0			
Consumer electronics	81.5 30.3	98.1 46.3	82.5 33.1		
Household appliances	96.1 15.6	103.6 15.4	97.6 23.9		
Furniture	105.1 18.4	108.4 20.5	126.1 38.7		
Clothing	102.1 28.4	101.6 28.8	99.0 19.1	87.4 24.3	

Reading Note: An empty cell denotes the absence of the indicator for the cross-reference in question. The mean values and standard deviations calculated over the period January 2012 - December 2017 are indicated on the first and second line.
Sources: Banque de France DGS SEEC.

protection reasons (see empty cells in Table 1). These indices have been available since 1990, whereas those for remote selling have been available since 2012. This paper examines raw indices (see the section on Models). Table 1 shows the quantitative survey indices for physical sales of products covered by the FEVAD data releases.[10]

FEVAD

Online retail data are not collected directly. The Federation of E-Commerce and Distance Selling (*Fédération du e-commerce et de la vente à distance* – FEVAD) has been providing the Banque de France with monthly aggregated sales data for its largest members since January 2012. There are currently around 70 members, changing over time. For the survey, these data are used to construct sales indices (defined above) applied to remote selling. In line with the survey methodology, sales data for month M and the revised panel figure for month $M-12$ are released every month. These releases concern total sales ("total industrial goods excluding cars") and those for five products: household appliances, textiles (clothing and household textiles – hereafter referred to as clothing), shoes (including leather goods), consumer electronics and furniture (furniture only). As the total covers more than the five products combined, its sales figure is higher than that for all five product sales combined. On average (for the time series history), sales for the five products account for 68% of total sales. Furthermore, Table 2 gives the share (in %) of remote selling for each product as captured by FEVAD.

Approximation of FEVAD Data with Retail Survey Data

Data taken from the survey are used for each of the six estimations (indices for total sales by remote selling and for the five products by remote selling). Figure I presents the indices for each distribution channel for consumer electronics (physical and remote sales).

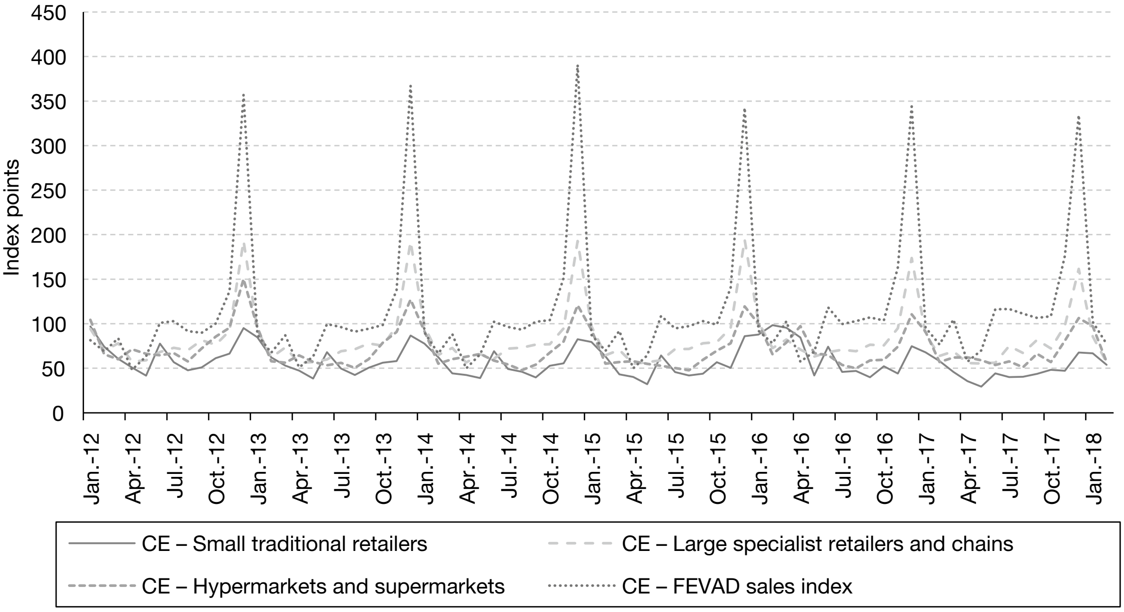
For consumer electronics, the December sales peak obtains for all distribution channels. The correlations[11] between FEVAD sales data and physical retail sales indices for consumer electronics (as a %) complement the information on the graph (Table 3).

The correlation between the remote sales index and that for large specialist and chain retailers prompts us to use physical sales data to estimate FEVAD data.

Generally, approximating these data allows us to observe straightforward economic processes. For example, over the long term, a substitution effect can be observed through

10. NB: products other than shoes, consumer electronics, household appliances, furniture and clothing make up the total.
11. Correlation is measured for the differentiated indices on a monthly basis, in line with data used for modelling (see below).

Figure I
Raw Indices for the Various Consumer Electronics (CE) Distribution Channels



Sources: FEVAD, Banque de France DGS SEEC.

Table 2
Share of Remote Sales for Each Product

Product	Remote sales weighting (in %)
Shoes	11
Consumer electronics	23
Household appliances	18
Furniture	13
Clothing	13
Total	10

Reading Note: According to the FEVAD, remote sales represent 11% of shoe sales in 2017.
Source: FEVAD.

Table 3
FEVAD Sales Index Correlations with Traditional Retail Indices from the Consumer Electronics (CE) Survey

Distribution channels	Correlation with remote sales index (in %)
CE – Small traditional retailers	44
CE - Large specialist retailers and chains	96
CE – Hypermarkets and supermarkets	48

Notes: Correlations are calculated for the period 01/2012 - 01/2018.
Sources: Banque de France DGS SEEC, FEVAD.

a reduction in sales at physical retail outlets; the corollary is an increase in remote sales. On the other hand, in the short term, an increase (or decrease) in physical sales may predict an increase (decrease, respectively) in remote sales: such collective movements reflect an increase in household consumption.

Google Trends

Google Trends provide monthly indices for terms queried via the Google search engine by users. Developed by Google using a methodology that has not been made public, indices are created by user-defined fields based on geography (in this case, France), time period (series date back no further than 2004), frequency (in this case, monthly) and belonging to a category (e.g. "Shopping", see below). Available where search volumes are "sufficient" (as defined by Google), these indices are made up of whole values between 0 and 100 and are produced for samples of all completed searches. Aside from the vagueness of Google Trends' index construction methodology, some earlier points raised call for robustness tests to be carried out.

Google Sampling

Constructed from a random sample of searches, a Google Trends index will differ between two samples. Comparing the series for the same term, queried multiple times, helps to verify the robustness of the tool. To illustrate this, Table 4 provides the correlations obtained for two separate samples taken a few days apart (i.e. with constant Google and Google Trends methodologies, *a priori*).

This was repeated several times, without obtaining a rate of correlation below 90% for differentiated monthly indices. Under these conditions, the sampling method appeared sufficiently reliable to periodically query Google Trends indices. The impact of sampling on the output will be discussed in the relevant section.

Whole-Value Indices

The simultaneous extraction of Google Trends indices is subsequently problematic. When making a common extraction of indices (between two and five) using the tool, the value 100 is attributed to the index experiencing a peak in searches for the period under query; the maximum values for other indices are given as a proportion. Where search volumes differ significantly, indices for less popular queries take on a limited number of values – as they are made up of whole values – which do not fully capture their fluctuations. However, in a statistical model, the number of decimal points for variables and, more generally their precision, can have an influence on the final estimation, according to Kozicki & Hoffman (2004). In order to obtain the most precise values, each Google Trends series is extracted individually. Taken together, the latter two points – sampling and the fact that indices consist of whole values – do not facilitate precision in Google Trends data.

Category

The Google Trends tool lists Google queries by category, corresponding to the context in which the search is made.[12] The example of the "iPhone" query urges caution when extracting data (Figure II).

Table 4
Correlations Between Google Trends Indices Taken Several Days Apart

(In %)

Amazon	Cdiscount	Fnac	E. Leclerc	eBay
98.1	97.4	98.9	95.5	90.2

Notes: Correlations calculated for the differentiated indices on a monthly basis, from January 2004 to February 2018 (170 points).
Sources: Google Trends, Banque de France DGS DESS SEEC.

While the "Commercial and industrial markets" category is not useful for analysis of remote sales, the line chart underlines the importance of category selection: its maximum level, reached in September 2013, does not equate to an explosion in sales. In the absence of more information about the categories, all subsequent queries referred to in the paper belong to the "Shopping" category, most closely corresponding *a priori* to online retail.

Breaks in Series

While Google does not share a great deal of information regarding changes to its methodology in constructing indices, the extraction page includes two observations:

- "The feature for determining geographic position has been updated. This update was applied as of 1 January 2011."

- "Our system for collecting data has been updated. This update was applied as of 1 January 2016."

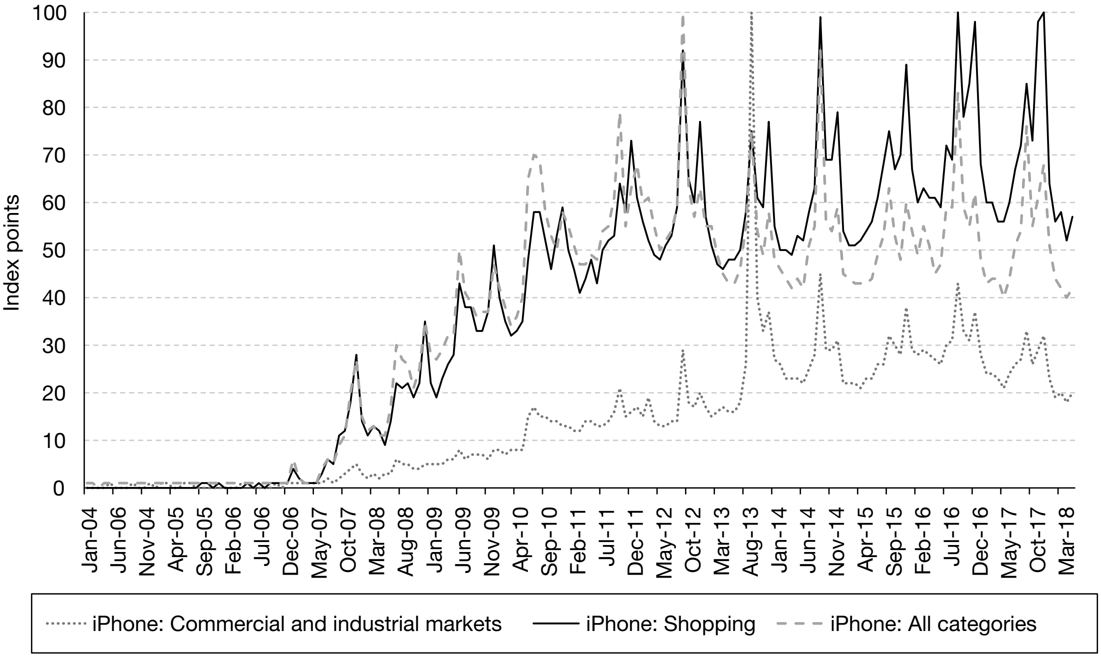
Users are therefore notified of major changes to the tool. In addition, these are in effect several months later. As FEVAD sales indices begin from January 2012, the second observation requires particular attention.[13]

Analysis of Google Trends using the X-13 method detects a greater number of outliers, in particular for January 2016. Due to the vagueness of the methodology for constructing Google Trend indices and their substantial number (more than 150) – likely to increase further with the growth of online retail – outliers are now treated systematically. Using the Google Trends index for Amazon as an example, the various steps can be explicitly set out. Here, a level shift is detected in January 2016; following evaluation, the series can be corrected (Figure III).

The first step in treatment is seasonal adjustment of the index, because two indices are used in detection (raw and seasonally-adjusted)

12. For example, "jaguar" may refer to the animal or the car manufacturer. Google queries are most likely listed in categories based on post-query browsing activity (i.e. websites visited after the query).
13. In order to improve the robustness of calculations, Google Trends indices have been extracted since January 2011. Although it is generally accepted that seasonal adjustment is not possible for historical time series of less than 3 years, adding one year of series data helps stabilise seasonally adjusted time series and, in so doing, improve outlier detection.

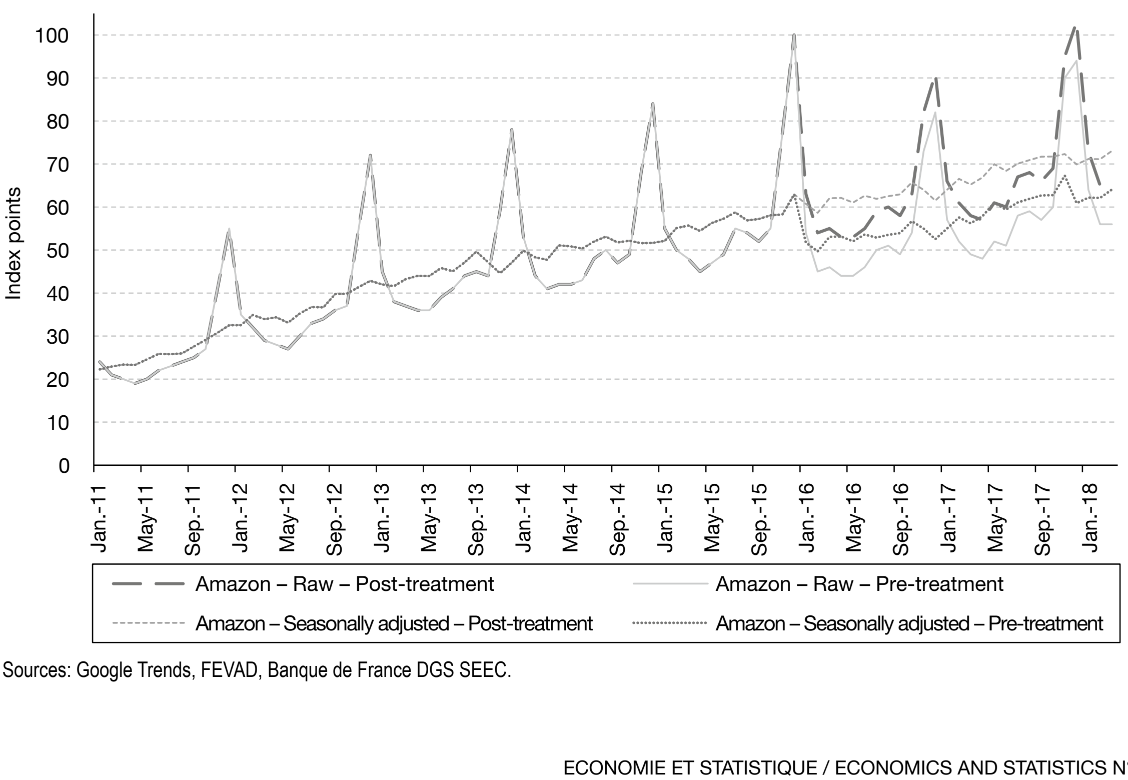
Figure II
Google Trends "iPhone" Queries



Sources: Google Trends.

in order to identify the maximum number of outliers. The nature of the outlier is then identified as a level shift, transitory change or additive outlier. In the case of Amazon – and outliers detected in January 2016 more generally (see Cdiscount, Appendix 1) – this is a level shift. Lastly, the extent of the break in series is estimated by the deviation between the January 2016 value on the seasonally adjusted series and the same truncated series forecast

Figure III
Treatment of the Outlier Detected on the Google Trends Amazon Index



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

in December 2015.[14] This adjustment is then applied to the rest of the series (unlike single outliers, which are treated *ad hoc*).

With the improving quality of series in quasi-real time, detecting outliers is less reliable in real time, i.e. for the latest series value: not adjusting too soon for the outlier enables more accurate classification,[15] thereby improving the precision of its estimation. The only outliers not treated are those reflecting the emergence of new queries (new company, new brand, etc., see shoes example below). The ever-changing nature of online retail also requires caution to be exercised when selecting queries.

Lists of Variables

On the one hand, the emergence of online retail has introduced new actors. In the case of shoes, for example, the three "pure players" (online-only retailers) dominating the French online retail market are relative newcomers (Figure IV).

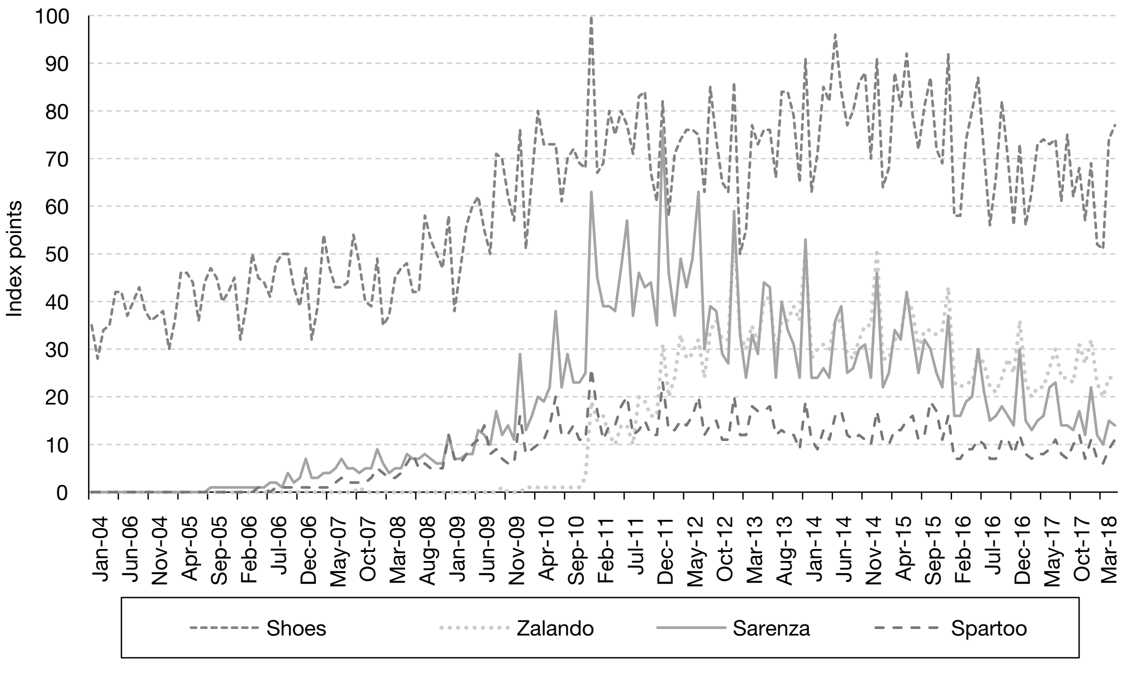
The movements in the "Shoes" index between 2004 and 2011 point to an emergence of online retail for shoes. The launch of Zalando in France in December 2010 is very clear on the graph (the index increases from 1 to 19 inside in the two months 11/2010 - 01/2011).

On the other hand, some erstwhile highly visible online retailers experience decline. In terms of household appliances, the Google Trends index for GrosBill serves as proof of this (Figure V).

While the query demonstrated a level of interest some years ago, this online electrical goods and consumer electronics retailer has lost market share by comparison with Boulanger, for example. Another example of the ever-changing face of online retail is the merger of Fnac and Darty: the related Google Trends index is now "Groupe Fnac Darty". In general, online retail has been in constant evolution, which Google Trends has managed to relay. For example, the fall in popularity in Google queries for one online retailer can be accompanied by an increase in queries for rival firms. In this space, it is essential to frequently review the variables used, particularly those applicable to online retailers for the various products. In order not to ignore the changing nature of search terms, it is possible to backward-extrapolate results with other variables through double collection (i.e. by testing the model on two sets of variables).

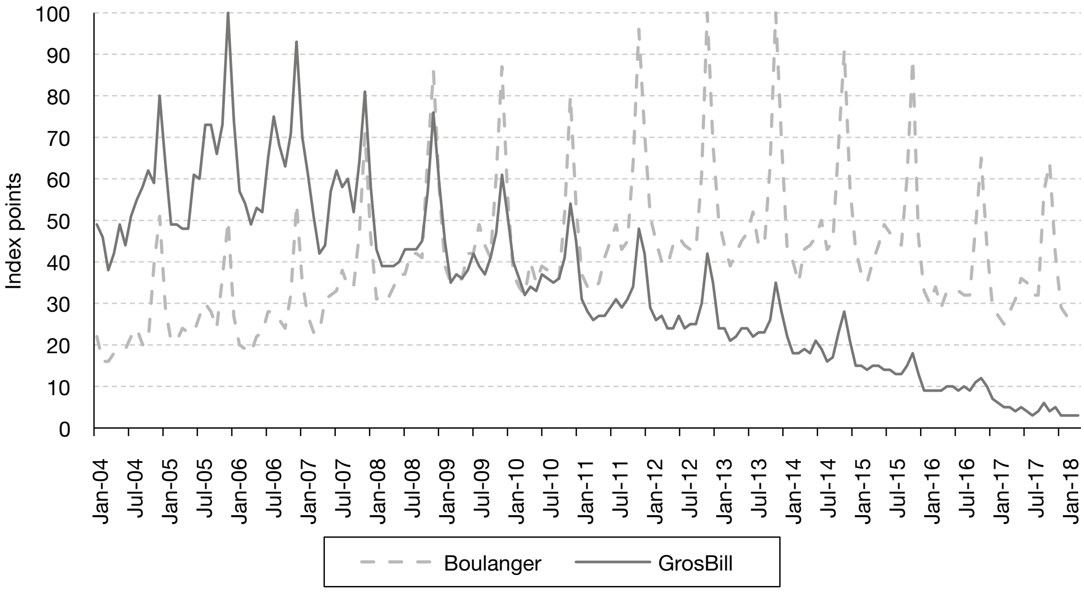
14. In this example, the level shift is estimated by adjusting for seasonal variations as the estimate provided by raw data appeared less consistent.
15. For example, a level shift can only be detected a posteriori*: when it appears, the outlier may be characterised (at best) as a single outlier before being reclassified as a level shift (following the appearance of further observations).*

Figure IV
Google Trends Indices for "Shoes"



Sources: Google Trends.

Figure V
Google Trends Indices for “Boulanger” and “GrosBill”



Sources: Google Trends.

However, one of the limits of this approach lies in the initial lists of variables (see Appendix 2). For the overall index, such pre-selection corresponds for the most part to the major online retailers in France. Pre-selections for the five products are a mix of pure players (e.g. Sarenza, in the case of shoes), major retailers (La Halle), generic terms (ladies’ shoes) and brands (Converse). Preliminary research, such as a document search for the products in question by appraising or visiting specialist websites, was carried out in order to predict behaviour prior to making a purchase. This resulted in the compilation of lists of heterogenous variables (see full table in Appendix 2).

Moreover, the trend in a website’s popularity is not necessarily the same as that for the Google Trends index, as not all internet users visit Google: internet browsing patterns change, in particular with the emergence of m-commerce,[16] where applications eliminate the need to use a search engine.

Models

Treatment of Stationarity and Seasonality

Most series are not stationary but are instead integrated to the order one: differentiation is required. This standard operation helps prevent spurious regressions (see Phillips, 1986), a common occurrence in time series regressions, which produce overly optimistic output reflected in an unusually high R^2 (see Granger & Newbold, 1974). Introducing a variable measuring the trend (Phillips & Perron, 1988) or autoregressive terms, also play a role.

In order to better measure trends in online retail, reference has been made to working with seasonally adjusted series. This solution has not been adopted. First, the short time series do not facilitate meaningful seasonal adjustment across all series,[17] chiefly for the initial estimations (36 points in the first iteration; more than 70 at present); particularly given that online retail itself has seen shifts in seasonality (Figure VI).

Figure VI represents the raw series for two sales indices for clothing (remote sales and small traditional retailers) and two product-related Google Trends queries: Kiabi and Zara. The remote sales index has seen changes in seasonality; for example in the early years, July figures were far higher than those for August. In 2015, the gap between both months narrowed and in 2016, the figures for July were lower. An overview of the series aptly demonstrates changes in seasonality. This phenomenon is

16. According to FEVAD, 36.6 million people in France shop online, of whom 9.3 million have made a purchase using their mobile phone in the past (2017).
17. More than 150 series are used to complete six estimations.

common to Google Trends indices. For example, in the case of Zara, the annual maximum is reached in January in the years 2013 to 2016; however, the value for November 2017 exceeds that for both January 2017 and January 2018. In addition, the Kiabi series does not exhibit any noticeable seasonality. Under such conditions, seasonal adjustment of multiple series becomes problematic. On the other hand, seasonal trends for the small traditional retailer index remain stable. Changes are more rare for well-established series (the index dates back to 1990). More generally, survey time series systematically pass seasonality tests (autocorrelation, Friedman, Kruskall-Wallis, spectral peaks, periodogram), which is not always the case for Google Trends series.

In addition, the latest values of a seasonally adjusted series are more likely to be revised in light of subsequent data releases (see Eurostat, 2018). At each FEVAD release, when a prior forecast can be evaluated, the most recent values of the seasonally adjusted series change, which can have a substantial impact on the model. The instability of seasonal adjustment on the most recent values is particularly pronounced for online retail series, notably due to poorly established seasonal trends and short historical time series. While the extent of instability from seasonal adjustments is on average 0.2 points between 01/2015 and 01/2018 for the large retailer index, it averages 1.6 for total remote sales (see Appendix 3) – the same order of magnitude as forecasting errors (see below). These arguments tend to favour a differentiated raw data model.

Process of Performance Estimation and Evaluation

Models

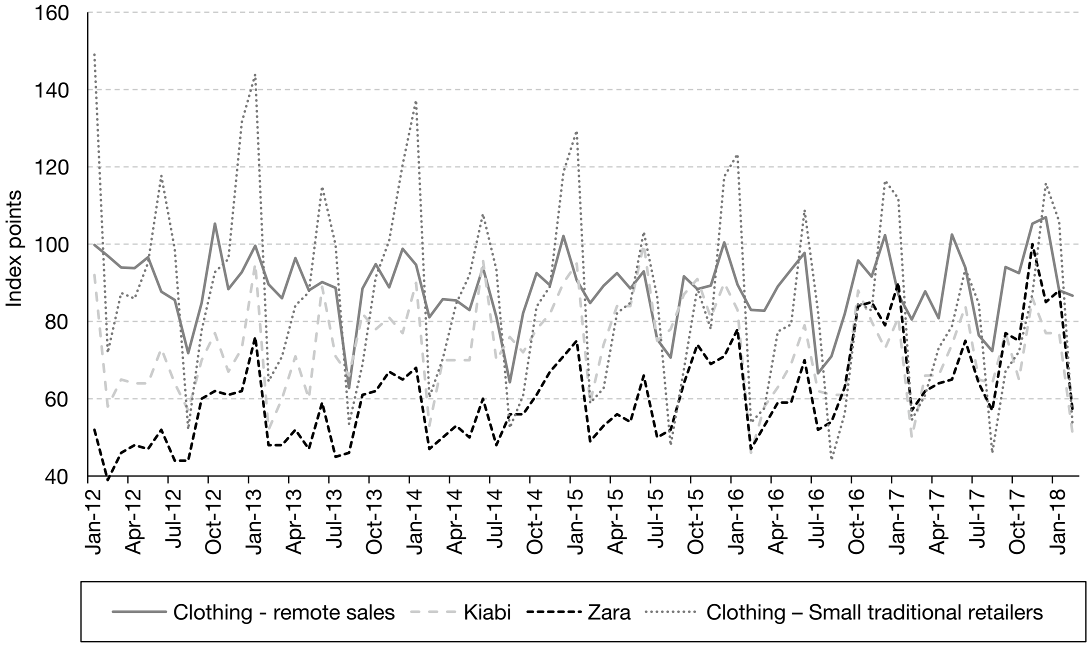
Until now, a SARIMA model has been used for each product. It is always updated and serves as a subsequent reference. Furthermore, the adaptive lasso is used in three models, implemented for each product:

- The Google Trends model, using Google Trends (see Appendix 4);

- The retail model, based on quantitative survey data for physical sales from retail surveys[18] (see Table 1);

18. To recap, for the overall index, sales indices for the five products are also used.

Figure VI
Indices for Clothing



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

- Global model, which is a selection of all available variables.[19]

In addition to exogenous variables, a trend and an autoregressive component are also included in the set of initial variables for these three models. The incorporation of the trend variable addresses the (*a priori* non-linear) rapid growth of online retail. As well as being an autoregressive component, it is the SARIMA model for the index, which becomes a variable potentially selected by the adaptive lasso algorithm, in the same way as the trend and exogenous variables (Google Trends and/or quantitative survey indices). Lastly, the fifth model is a Bayesian aggregation ("model combination") of SARIMA, Google Trends and retail trends models. The comparison of its output with that of the global model contribute to the debate regarding data combination.

Test Protocol

For each iteration of the test protocol, i.e. each month, the actual conditions are replicated. Specifically, the values of Google Trends data and quantitative survey data for physical sales for month M are available, which is not the case for FEVAD data.

Estimation takes place in two stages: the first involves modelling the index using an autoregressive process (SARIMA). As well as obtaining its own forecast, this action also helps determine the variable used in adaptive lasso models. In the second stage, the three variable selection models (Google Trends, retail trends and global) are formulated. The model combination can only be constructed after the SARIMA, Google Trends and retail trends models.

The model quality can be determined following the release of FEVAD data as the evaluation criterion adopted for the nowcasting process, is predictive capacity. The predictor is therefore the RMSFE (Root Mean Squared Forward Error), the standard deviation of forecasting errors, measuring the out-of-sample error. The RMSE (Root Mean Squared Error), measuring the in-sample error, is also provided as it can be used to ascertain the weighting in the model combination and identify any overfitting.

Furthermore, each month the estimation window for the models expands by one observation. Due to the limited sample sizes available, working with an expandable window rather than a rolling window for the sample estimation contributes to the models' stability. FEVAD data releases began in January 2012. Differentiation of data brings the series to February 2012. With a minimum time series history of three years required to ensure the robustness of the estimation, the initial forecast is that for February 2015.

Results

Only the results for the total will be set out in detail; those applicable to products will be summarised.

Total

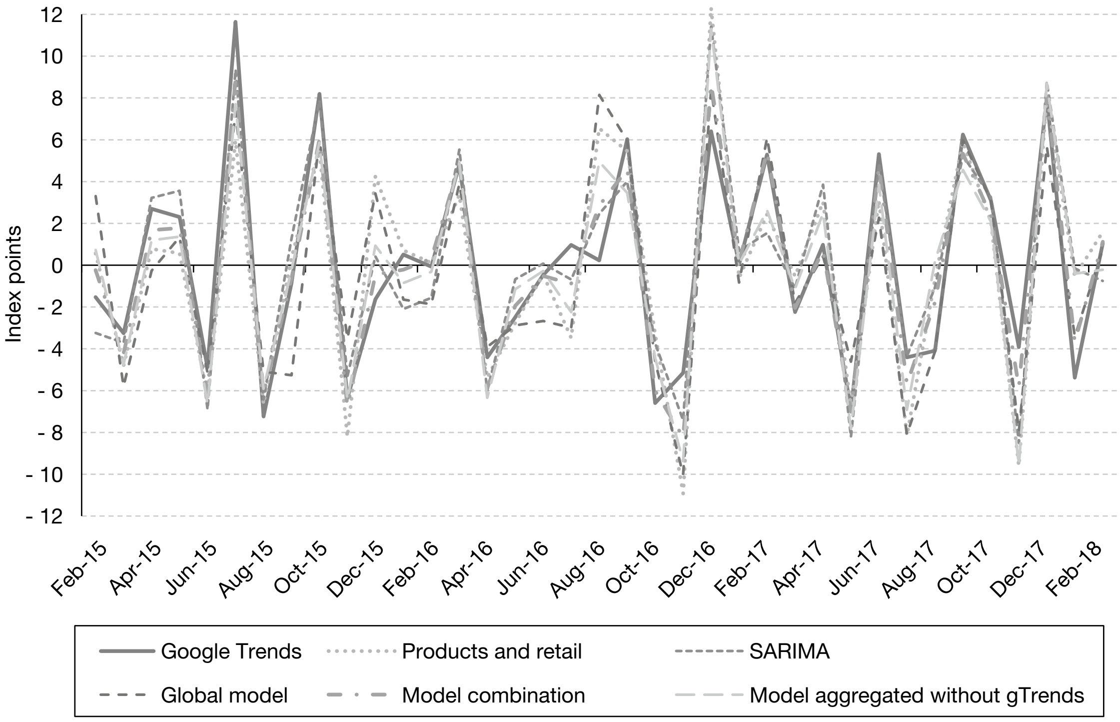
In line with the aim of the study, forecasting errors (out-of-sample) represent an important result (Figure VII).

Figure VII shows the forecasting errors for each model. The results are visually close: overall, the output from the Diebold-Mariano test (see Diebold & Mariano, 1995) does not conclude that the model forecasts are significantly different. The RMSFEs and average forecasting errors (in absolute terms) offer better insight into the forecast output, while the RMSEs attest to the responsiveness to in-sample data (Table 5).

For the purpose of the RMSFE, which remains the preferred indicator, the Google Trends model performs best with the model combination (4.8), for the Google dataset selected (deemed representative of simulations carried out, see Box). In this case, the poorer performance of the model combination without Google data justifies the use of Google Trends. The model combination also performs best in terms of the average of absolute errors. This error measurement is relevant, as one of the purposes of aggregation is also to minimise large forecasting errors. The result from individual models is relatively close. In terms of the RMSE, both models with all available information (the model combination and the global model) are a better fit for the sample data.

19. This model may, for some iterations, be identical to one of the individual models (e.g. if no Google Trends query is selected).

Figure VII
Forecasting Errors of Models in Estimation of Total Index



Reading Note: The forecasting error for August 2016 by the global model is 8.1 index points: following the release of FEVAD data, the observed value of the sales index for the total was higher (by 8 points) to the global model forecast.
Sources: Google Trends, FEVAD, Banque de France DGS SEEC.

Before comparing both models, it is worthwhile to detail the output from individual models in detail – in particular the Google Trends model, for which parsimony and stability must be ensured.

SARIMA Model

Serving as a reference in this paper and in the literature, the SARIMA model offers sound forecasting performance (RMSFE=5.0), despite a less impressive fit with sample data (RMSE=4.2). However, it emerges as less suitable than the other models. For example, in December 2016, exogenous data provide real information.

Google Trends Model

In line with test protocol, variable for the adaptive lasso are selected for each iteration. Model coefficients therefore change over time (Figure VIII). For greater clarity, 30 variables for estimation of total sales have been included in six graphs. In the secondary axis for each one, the change in the lasso penalty.

Table 5
RMSFE and Mean RMSE for Models in Estimation of the Total Index

Total	Google Trends	Retail	SARIMA	Global model	Model combination	Model combination without gTrends
RMSFE	4.8	5.2	5.0	5.5	4.8	5.0
Mean average forecasting error	3.9	4.0	3.9	4.5	3.8	3.9
Mean RMSE	3.3	3.7	4.2	2.3	2.6	2.8

Notes: The model combination without Google Trends corresponds to the aggregation of the retail and SARIMA models. It allows us to determine the input of Google Trends data. However, as the SARIMA variable is present in all retail models, the aggregation becomes less significant; it will therefore not be presented in results obtained for the products.
Sources: Google Trends, Banque de France DGS SEEC.

Box – **Sensitivity of Models to Google Sampling**

Google Trends variables may be modified from one month to the next, due to Google's sampling method. The standard deviations of RMSFEs, obtained from thirty simulations[(a)], of modules using Google Trends variables:

As can be seen in Table A, the impact of Google sampling is significant.

Furthermore, each simulation corresponds to the linked estimation of indices for the five products and for the total. However, some Google Trends variables are common to the total and one product. It was therefore not possible to extract a set of Google Trends variables for which the model results, in terms of RMSFE, are all at the median. Those set out in the main text of the paper correspond to one of the most representative simulations (for the six estimations), i.e. the RMSFE for the Google Trends models are very close to the median.

(a) Each simulation applies to the total and the individual products across 150 Google Trends series. As Google restricts bulk extraction of series, it is difficult to substantially increase the number of simulations.

Table A
Impact of Google Sampling on Output in Terms of Standard Deviation of RMSFE

	Google Trends	Global model	Model combination
Total	0.4	0.3	0.4

Reading Note: The standard deviation of RMSFEs obtained for the 30 simulations for the Google Trends model is 0.4.
Sources: Google Trends, FEVAD, Banque de France DGS SEEC.

The graphs in Figure VIII show the change over time in coefficients on the primary axis and that in the lasso penalty in the secondary axis. While it is not common to observe changes in the lasso penalty over time, as it is a different optimisation for each iteration, it helps explains the change in the number of variables selected: the lower the lasso penalty, the higher the number of Google Trends queries selected. With respect to the SARIMA variable, it was expected that its coefficient would be close to 1 as it corresponds to the autoregressive model for the variable. Moreover, changes in coefficients of Google Trends variables – highlighted with those of the lasso penalty – are stable, indicating that these variables model a portion of data not captured by the SARIMA component. The table of mean, minimum and maximum values obtained for each variable is included in Appendix 5.

With respect to selection, almost 9 variables are selected on average for each iteration, which is acceptable in light of the sample sizes (36 observations for the first iteration; 72 for the most recent). The most frequently selected Google Trends variables are eBay, PriceMinister, Groupon, Showroomprivé and Leroy Merlin (see Appendix 5).

Retail Model

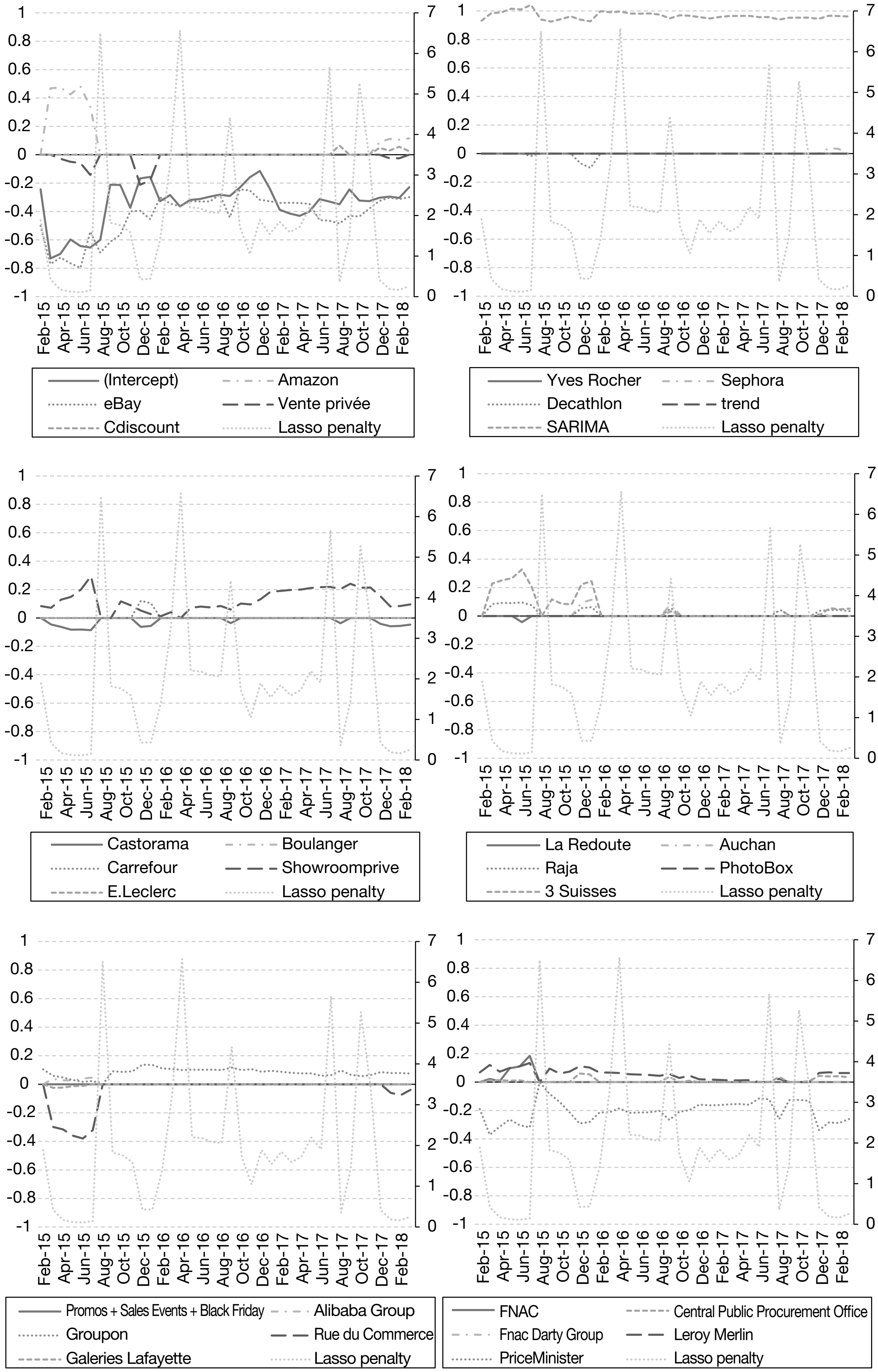
As well as the sales index for physical sales (cf. Table 1), remote sales indices for five products are also used. The sales for the five products contribute by design to the total sales figure. However, as all FEVAD data are released simultaneously, these indices extend to the latest observation using the SARIMA model. Movements in coefficients are detailed in Appendix 6. The model is parsimonious, selecting one to two variables in addition to the autoregressive component (SARIMA estimation). The remote sales index for clothing is selected systematically and logically – on average, the value of sales for clothing represents 22% of the total, the largest of the five products. Its out-of-sample output (RMSFE and mean absolute forecasting errors) is inferior to the Google Trends and SARIMA models (cf. Table 5). For in-sample output, it places between the Google Trends and SARIMA models.

Model Combination

The model combination offers the best forecasting performance under both indicators, RMSFE and the mean absolute error. Over time, it never produces the least accurate forecast. The weight of the models allows us to determine its stability (Figure IX).

Since the end of 2016, the weight of the Google Trends model has increased. On average, it is greater (0.55) than that of the other two models, SARIMA (0.21) and retail (0.18).

Figure VIII
Change in Coefficients for Google Trends Models and the Lasso Penalty



Sources: Google Trends, Banque de France DGS SEEC.

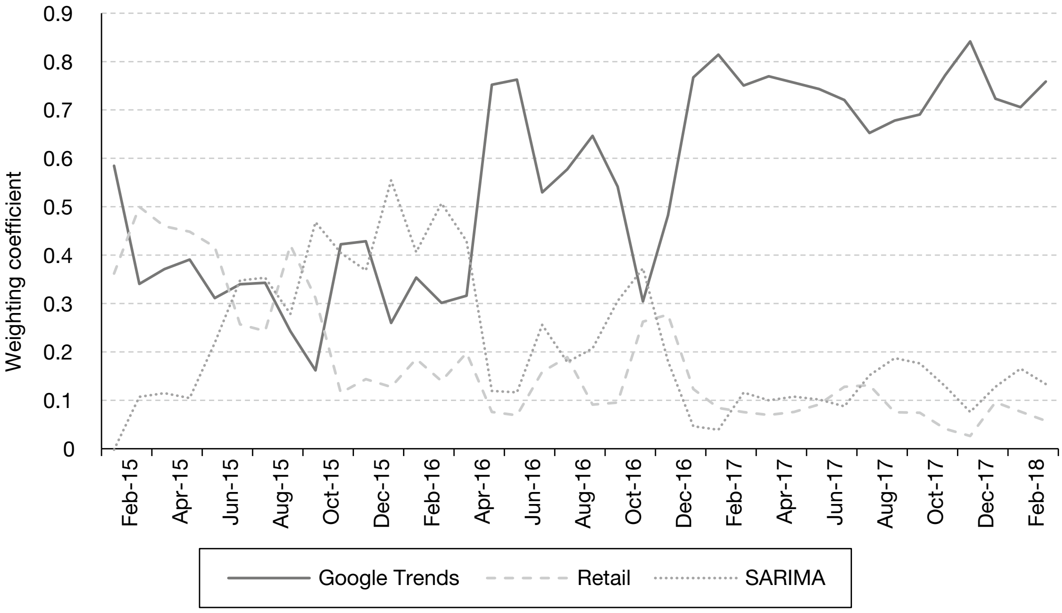
The movements in forecasting errors for the Google Trends, retail and SARIMA models highlight the change in weighting. While the forecasting errors for the three models are relatively close, which can be explained above all by the presence of the SARIMA variable in the Google Trends and retail models, some differences merit particular attention. For example, in October 2016, the Google Trends model had the largest weighting in aggregation, with 0.54, while that for the SARIMA and retail models were 0.31 and 0.10 respectively. For the FEVAD data release at the end of November 2016, it is possible to compare forecasts with their actual value. Figure VII, which sets out forecasting errors, shows that the Google Trends model is the least accurate of the three with an error of 6.0 index points, against 3.3 and 3.7 for the retail and SARIMA models. After the error was "learned", the weighting changed dramatically the following month: the weighting for Google Trends fell to 0.30, with 0.37 for the SARIMA model and 0.26 for the retail model.

Figure IX provides an occasion to detail the Bayesian aggregation formulas. As per the literature review, eight models are possible using three regressors (corresponding in this case to values estimated by Google Trends, retail and SARIMA models). Table 7 sets out the coefficients for each regressor in models M_i $(1 \leq i \leq 8)$ and probability $P(M_i \mid D)$ that each model M_i is the correct one. Lastly, the final column refers to the Bayesian model, whose coefficients are obtained by weighting those for models M_i by probabilities $P(M_i \mid D)$. The values for Table 6 are those for September 2016.

Values for the final column are consistent with Figure IX (September 2016). The Bayesian aggregation is incorporated in the machine learning algorithms; the in-sample error is used to determine the weightings. Figure X shows the movements in the RMSE for the various models.

For each iteration, it is possible to calculate the RMSE obtained for the model's estimation sample. The SARIMA model provides the greatest in-sample error over the period, in contrast to its high predictive capacity. Figure X also shows that the aggregation of data reduces the model's estimation sample error, by comparison with its components. Subsequently, where movements in the in-sample errors for the aggregated models (with and without Google Trends) are similar, the predictive performance is improved with the incorporation of Google Trends data. Lastly, the Google Trends model produces errors close to the other models, supporting the view that the number of variables selected by the adaptive lasso is suitable and that there is no overfitting. To be sure of this,

Figure IX
Change of Weightings in the Model Combination



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

Table 6
Detailed Calculation of the Weightings in Bayesian Aggregation (September 2016)

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

Sources: Google Trends, Banque de France DGS SEEC.

the in-sample errors (RMSE) may be compared with those produced out-of-sample (RMSFE) (cf. Table 5).

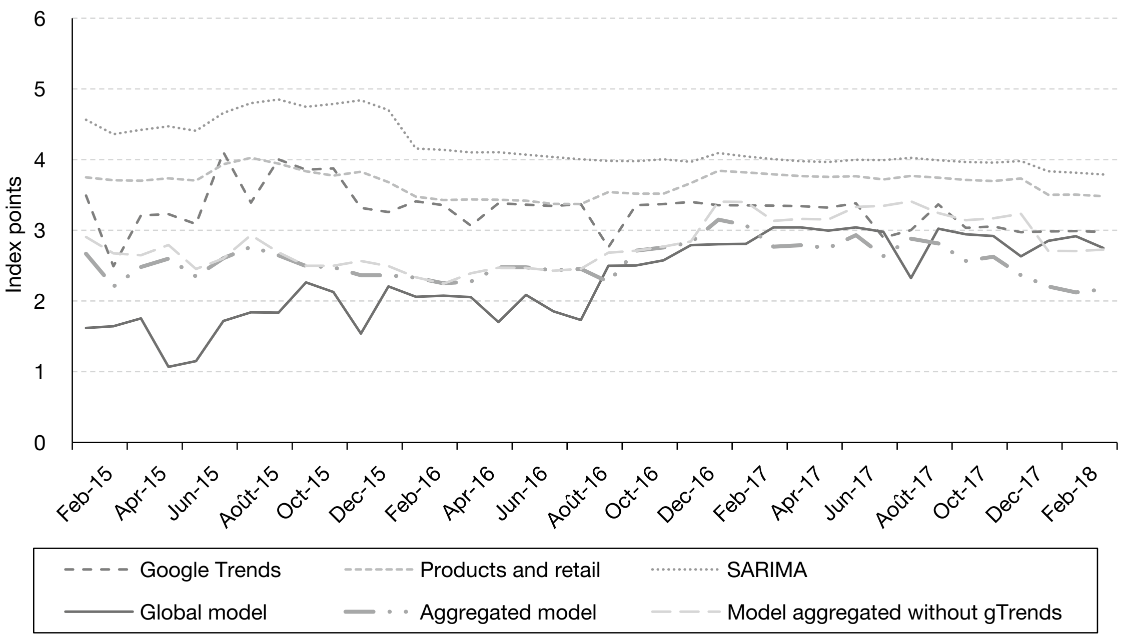
Logically, the forecasting errors are larger. The classifications of models are followed when changing from RMSE to RMSFE, except for the global model, whose out-of-sample error more than doubles.

Global Model

This phenomenon can likely be explained by overfitting. Although the adaptive lasso process is the same as for the Google Trends and retail models, the global model is less parsimonious: on average, 13 variables are selected, which is relatively high by comparison with the number of observations (36 at the first iteration). It selects more variables than the Google Trends and retail models combined. Specifically, over the test protocol period, 82% of variables selected in the global model are selected in one of the other two models; 12% are selected by the global model only and the remaining 6% refer to variables selected by the Google Trends or retail models but not by the global model. In summary, the selection of variables for the global models is too broad, which leads to overfitting. Indeed, movements in coefficients are less stable.

In the case of the overall index, the global model does not perform as well as the model

Figure X
Change in Model RMSEs for Estimation of the Total Sales Index



Reading Note: When estimating the models for the December 2015 forecast, the lowest RMSE was that of the global model (1.5). The highest RMSE was that of the SARIMA (4.8).
Sources: Google Trends, Banque de France DGS SEEC.

combination. In addition, the clarity of the model combination, although limited (output from the Diebold-Mariano test does not conclude that the forecasts of the three models are significantly different), remains better than that of the global model, in which changes in coefficients complicate interpretation. The model combination is therefore preferred. While the output obtained for the overall index has been set out in detail, those for the individual products are summarised below.

Products

Parsimony

The adaptive lasso aims to ensure that models are parsimonious. For each product, Table 7 shows the average number of variables selected per model (covered by the selection of variables).

The retail models are the most parsimonious; one survey variable is selected most frequently, in addition to the SARIMA component. The Google Trends models are less parsimonious; the number of variables selected remains correct in light of the sample sizes, with the possible exception of clothing.

In the Google Trends product models, in addition to the SARIMA component and the constant (systematically selected), the five most selected variables (from 38 iterations) are included in Table 8.

Table 8 illustrates the heterogeneity of the most selected Google queries in Google Trends models: items (ovens, televisions), brands (Cinna, Samsung), general queries (women's clothing, football boots), pure players (Spartoo, GrosBill) and remote retail specialists (3 Suisses). The variety of Google search engine user actions is well-captured here. Note that the trend variable is never selected.

In contrast to the overall index, the global model is more parsimonious for each product than the Google Trends model, thereby reducing one of the risks of overfitting. Table 9 below illustrates the mean values of RMSE.

As expected, the models with full information are better overall, in terms of the RMSE, than models with a single source of information (Google Trends, retail trends, SARIMA). The second finding from Table 9 is that the Google Trends is systematically a better fit for sample

Table 7
Number of Variables Selected by Model Using Adaptive Lasso and by Product

	Google Trends	Retail	Global model
Shoes	9.7	2.2	10.0
Furniture	10.3	2.4	10.9
Household appliances	9.0	2.1	5.9
Consumer electronics	8.8	2.0	10.5
Clothing	12.8	3.6	8.7

Sources: Google Trends, Banque de France DGS SEEC.

Table 8
Most Frequently Selected Variables in Google Trends Models, by Product

Shoes	Spartoo (38)	Sarenza (36)	Converse (36)	Dress shoes (32)	Football boots (28)
Furniture	Cinna (38)	Roset (33)	Wooden furniture (32)	Dresser + cupboard + cabinet (31)	IKEA (26)
Household appliances	Washing machine (37)	Oven (28)	Cooker (28)	Conforama (26)	GrosBill (24)
Consumer electronics	SLR digital camera (37)	Television (35)	JBL (35)	Sony (27)	Samsung Electronics (23)
Clothing	Suit (34)	Decoration (29)	Jennyfer (28)	Lingerie (27)	3 Suisses (25) and Women's clothing (25)

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

data than the retail and SARIMA models; which may be explained by the larger number of variables selected.

Predictive Capacity

While the Google Trends model was, on average, systematically better over the estimation period (based on RMSE) than the retail and SARIMA models, it is not the best for forecasting. Its overall predictive performance is broadly equal to that of the retail and SARIMA models (Table 10). More generally, the results obtained for the various products are mixed. The addition of exogenous data – Google Trends data or quantitative survey indices – does not reduce the forecasting error.

The retail model is the best-performing model for shoes; the Google Trends is slightly better, in terms of RMSFE, than the SARIMA model. For furniture, Google Trends offers the clearest input. “Consumer electronics” also recorded an improvement (by comparison with the SARIMA model) with exogenous data. However, for household appliances and clothing, their input did not improve the results (by comparison with the SARIMA model).

With respect to the combination of data, the results are also mixed. On the one hand, the combination of models offers improved forecasting output (RMSFE) than the global model, except for consumer electronics. On the other hand, the combination of data does not deliver the expected results. Based on RMSFE, the model combination is only better for furniture, the only product for which the Google Trends model outperforms the retail and SARIMA models. The in-sample performances affect the weighting of the models in aggregation. As the Google Trends model produces the best estimations (based on the mean RMSE, see Table 5), its weighting in aggregation is larger (Table 11).

The mean weighting in aggregation is based on the test protocol period, as is the case for the mean RMSE. Clothing is the sole product group for which the Google Trends model weighting is not the largest.

Table 9
Mean RMSE for Models in Estimation of Product Sales Indices

	Google Trends	Retail	SARIMA	Global model	Model combination
Shoes	8.2	10.5	10.9	7.8	7.6
Furniture	6.0	7.3	7.4	5.7	5.5
Household appliances	6.1	6.9	7.2	6.4	5.5
Consumer electronics	5.8	7.4	7.7	5.5	7.2
Clothing	5.3	6.0	6.3	5.9	4.3

Sources: Google Trends, Banque de France DGS SEEC.

Table 10
RMSFE and Standard Deviations related to Google Sampling in the Estimation of Product Sales Indices

	Google Trends		Retail	SARIMA	Global model		Model combination	
Shoes	13.2	0.3	12.7	13.6	13.8	0.4	13.4	0.2
Furniture	11.9	0.5	12.3	12.0	13.2	0.4	11.8	0.5
Household appliances	11.7	0.3	10.4	10.2	12.3	0.3	11.2	0.3
Consumer electronics	15.5	0.3	15.3	16.4	11.5	0.5	13.1	0.3
Clothing	9.8	0.3	10.1	9.2	15.2	0.5	9.7	0.2

Notes: Results included in the main text of the paper for the five products are from the same simulation as those for the total index; RMSFEs are very close to the median values obtained for the thirty simulations.
Reading Note: The RMFSE for the Google Trends model in estimation of the shoe sales index is 13.2; for the thirty simulations carried out to determine the sensitivity of results to Google sampling, the standard deviation is 0.3. The RMSFE of the retail model is 12.7 (and is not impacted by Google sampling).
Sources: Google Trends, Banque de France DGS SEEC.

Table 11
Weighting of Individual Models in the Model Combination

	Google Trends	Retail	SARIMA
Shoes	0.55	0.21	0.19
Furniture	0.71	0.09	0.13
Household appliances	0.47	0.14	0.33
Consumer electronics	0.48	0.20	0.27
Clothing*	0.58	- 0.50	0.80

* For clothing, the values predicted by the three models exhibit strong colinearity, poorly handled by Bayesian aggregation: the contribution of variables in "intermediate" models (see detailed calculation of weightings for the total, in the relevant section) are artificially overvalued; this has a knock-on effect on mean weightings for the model combination.
Sources: Google Trends, Banque de France DGS SEEC.

* *
*

Online retail is rapidly growing. Purchases made online account for a greater proportion of household consumption, and thus the Banque de France monthly retail trends survey. Against this backdrop, the estimation of sales figures released (belatedly) by FEVAD becomes a prominent question.

Up to now, this has been carried out using an autoregressive model. The research set out in this paper looks at the contribution of the exogenous data that are traditional retail indices for physical sales (in monthly retail trend surveys) and Google Trends indices. Each data source provides its own input. The common benefit of such data sources, namely being available before FEVAD releases, is ideal for nowcasting.

However, a new source of data (Google Trends) must be used with caution. Firstly, robustness tests prior to use have been necessary. A system of treating outliers was implemented. Such outliers are sometimes the result of methodological changes introduced by Google and for which little information is made available. The sensitivity of output to the sampling method used by Google prompts multiple simulations to increase the reliability of output. Secondly, it was necessary to reconcile the huge range of possible Google variables with the lack of historical FEVAD time series data (monthly releases date back to 2012). This twin constraint can be overcome by machine learning, using the adaptive lasso process (Zou, 2006). The selection of variables at each iteration, thereby minimising the risks associated with rapid developments in online retail and possible instability of corresponding keywords, as it is possible to backward-extrapolate output with other sets of variables. This way, the model is flexible and offers substantial adaptive capacity, which the ever-changing nature of the modelled phenomenon requires.

The question then arises as to how to exploit the complementarity of the various data sources. In this paper, Bayesian aggregation of single models produces better results in terms of RMSFE, than the global model (adaptive lasso applied to all variables simultaneously). The small size of estimation samples for the models may work against a model with many variables. For example, in the case of the overall index, overfitting is detected for the global model. In addition, aggregation offers clarity in the combination of models, which is useful in production.

In general, the contribution of exogenous data remains mixed. It is clearer for the overall index than the index for products. FEVAD releases are developed from a sample of 70 of its largest respondents (in terms of sales). The number of respondents is therefore lower for product groups; this substantiates the results that are most impactful and thus the most difficult to comprehend. The forecasting error for sales is therefore two to three times greater for products than for the total.

Lastly, one of the possible causes of mixed results lies with model selection. While they meet many of the constraints posed, seasonality is not always fully taken into account. Due to the short time series, the model does not operate on seasonally adjusted series, unlike the standard econometric approach; here, the presence of SARIMA estimation for explanatory variables seeks to capture seasonality.

However, this method overlooks the differences in seasonality between endogenous and exogenous variables.

With longer time series, seasonality in online retail time series should stabilise, offering the opportunity to refine output with other models. As well as the possibility of working on seasonally adjusted time series, combining RegARIMA modelling, for which the residual specification is more applicable, with variable selection methods may prove valuable and is currently not addressed in the existing body of research. ☐

BIBLIOGRAPHY

Aiofli, 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-humblot.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

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 and 2017). *Chiffres clés*.
https://www.fevad.com/wp-content/uploads/2016/09/Plaquette-Chiffres-2016_Fevad_205x292_format-final_bd.pdf
https://www.fevad.com/wp-content/uploads/2018/06/Chiffres-Cles-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

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 Econometric*, 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

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

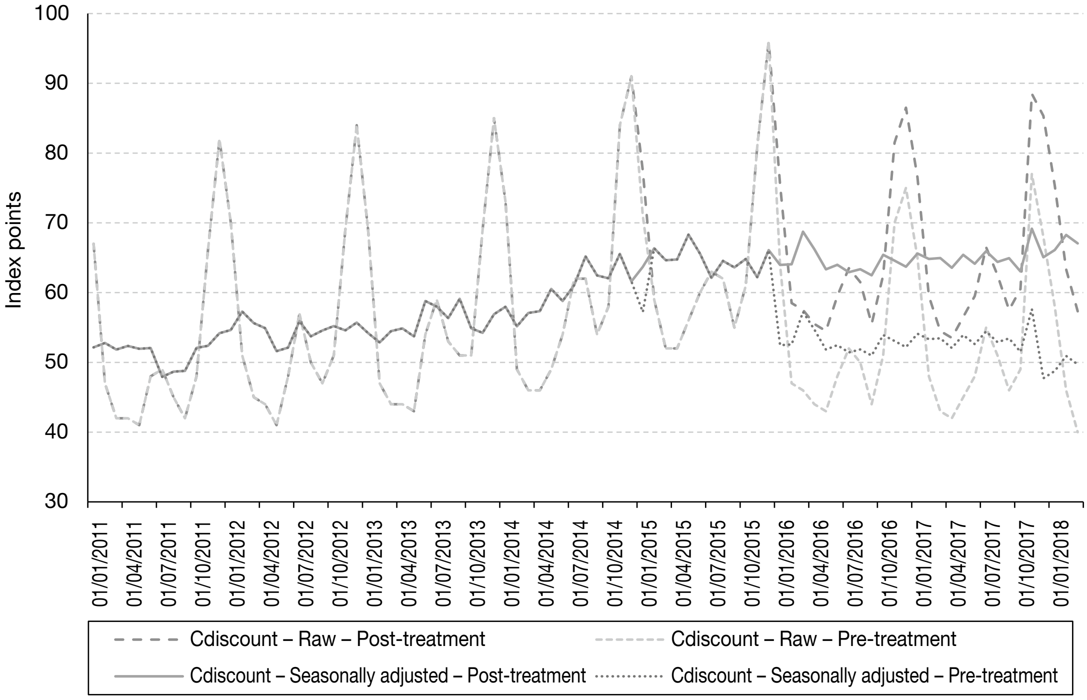
Zeugner, S. (2011). *Bayesian Model Averaging with BMS*.
https://cran.r-project.org/web/packages/BMS/vignettes/bms.pdf

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

APPENDIX 1

TREATMENT OF OUTLIERS - THE EXAMPLE OF CDISCOUNT

Figure A1
Treatment of the Break in Series of the Google Trends Index for Cdiscount



Note: The treatment applied to the Cdiscount series is analogous with that for Amazon.
Sources: Google Trends, Banque de France DGS SEEC.

APENDIX 2

LIST OF VARIABLES BY PRODUCT

Table A2
Set of Initial Variables for Each Estimation

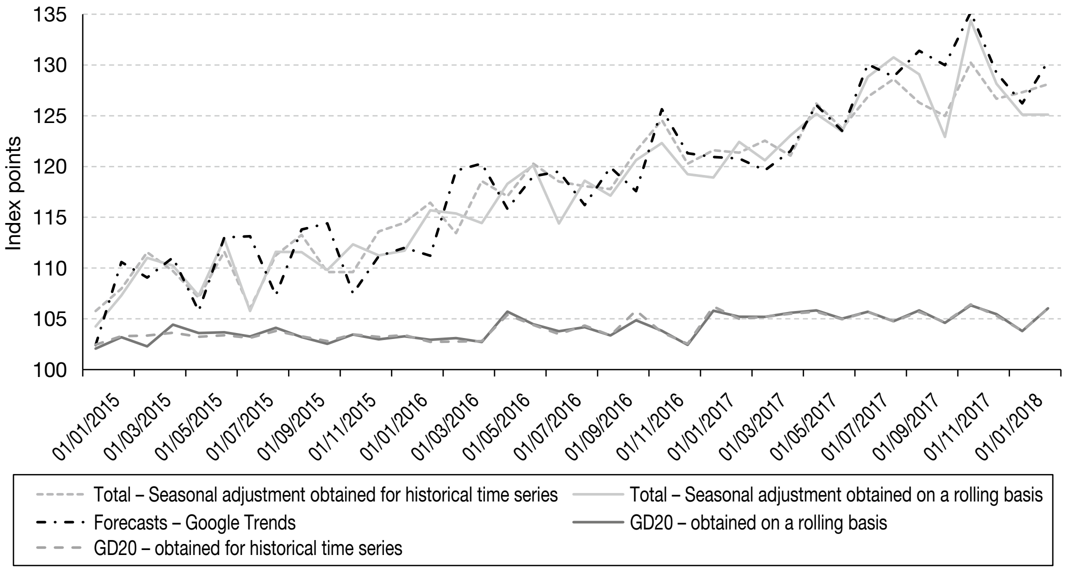
Total	Amazon, eBay, Vente privée, Cdiscount, Fnac, Fnac Darty Group, PriceMinister, Leroy Merlin, UGAP, Castorama, Boulanger, Carrefour, Showroomprive, E. Leclerc, La Redoute, Auchan, Raja, Rue du commerce, 3 Suisses, Promos + Sales Events + Black Friday, Alibaba, Groupon, PhotoBox, Galeries Lafayette, Yves Rocher, Sephora, Decathlon
Clothing	Vertbaudet, Kiabi, H&M, C&A, Jules, home textile, Zara, Suit, Underwear, 3 Suisses, Devred, Robe, Etam, La Redoute, Jeans + Chinos + Trousers, Coat + Jacket, Jacket, Women's clothing, ASOS, Maisons du Monde, Lingerie, Jennyfer, Clothing, Galeries Lafayette, Bonobo, Brandalley, Camaïeu, Showroomprive, Vente Privée, curtain, Blanche Porte, bedsheet, Cushion, Homemaison, Underwear, La Halle, Decathlon
Consumer electronics	iPhone, Apple, Cdiscount, PC Gaming, iPad, Telephone + smartphone , FNAC, Television, Boulanger, Sony, LDLC Pro, Amazon, Phillips, LG Group, Samsung Electronics, Darty, Tablet, Speaker, SLR digital camera, Laptop computer + PC, Bose, JBL, Fnac Darty Group, Soundbar, Camera, Marshall, Samsung Group
Shoes	Shoes, Shoe, Belt, Leather goods, Boots, Sport shoes, Vans, Converse, Zalando, Spartoo, Sarenza, Showroomprive, Prada, Escarpin, Adidas Stan Smith, Women's shoes, Pumps, Men's shoes, Timberland, Football boots, Children's shoes, San Marina, Eram, Dress shoes, J.M Weston, Chaussea, Bexley, Gémo, Handbag, La Halle, Nike shoes
Household appliances	Clubic, Boulanger, Cdiscount, Oven, Fridge, Washing Machine, Darty, Bosch, Electrolux, Conforama, Amazon, Cooker, Électro-dépôt, Brandt, Microwave oven, Fnac Darty Group, Vacuum Cleaner, Whirlpool Corporation, Mistergooddeal, GrosBill, Pulsat, Ubaldi, But
Furniture	But, Legallais, Kitchen, Raja, Staples, Roche Bobois, Castorama, Conforama, Vega, Bureau, Furniture, Leroy Merlin, Ikea, Knives, Cupboard + shelves, Maisons du Monde, Cinna, Wooden furniture, Dresser + cupboard + cabinet, Roset, Table + chair + sofa, Armchair

Reading Note: “+” here denotes a Google Trends index with combined queries.
Sources: Google Trends, Banque de France DGS SEEC.

APPENDIX 3

INSTABILITY OF LATEST OBSERVATIONS IN SEASONAL ADJUSTMENT

Figure A3-I
Stability of Seasonal Adjustment for the Most Recent Observation



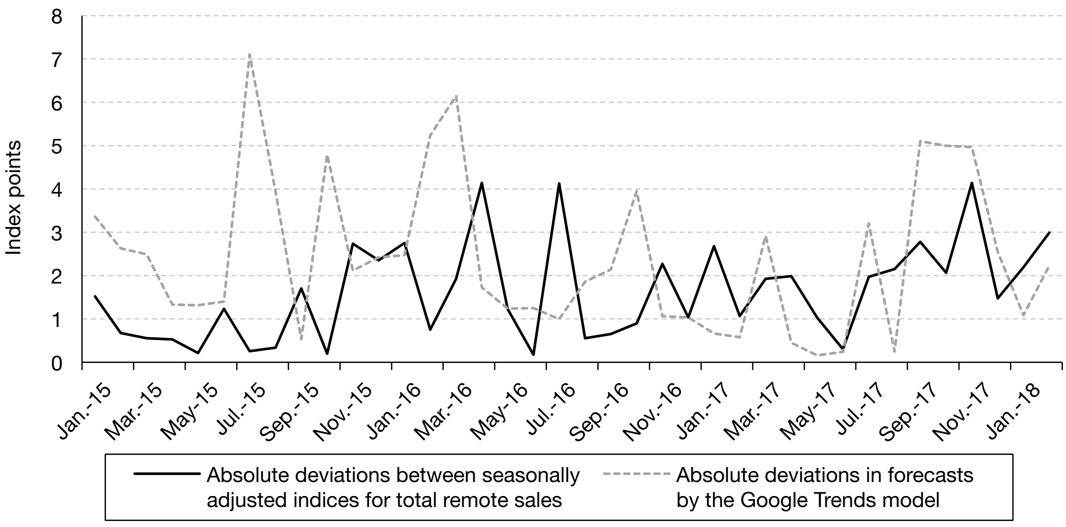
Sources: Google Trends, FEVAD, Banque de France DGS SEEC.

Two groups of series are distinguished. The first applies to remote selling/remote sales and consists of three series: an aggregate remote sales index for which seasonal adjustment takes account of the whole data, a “rolling” aggregate remote sales index, implemented on a rolling basis – i.e. each observation is the last in the seasonally adjusted series obtained using the truncated index on that date, and the Google Trends forecasts index.

The second group is applied to large retailers: the mean absolute deviation between two seasonally adjusted indices for large retailers (obtained using all data versus those available on a rolling basis) is 0.2 (0.05 where the size is related to the amplitude, defined as the largest variation in the reference series – seasonal adjustment obtained for historical data); against 1.6 (0.29 related to amplitude) for both seasonally adjusted indices that make up the total for remote sales.

While revisions to the most recent observations of seasonally adjusted series are not unknown (see Eurostat, 2018), the magnitude observed here presents a challenge: both of these indices vary by proportions similar to the forecasting errors of the models: movements in errors between, on the one hand, the two seasonal adjustments, and on the other hand, forecasts from the Google Trends model and seasonally adjusted series based on historical data, account for this (Figure A3-II).

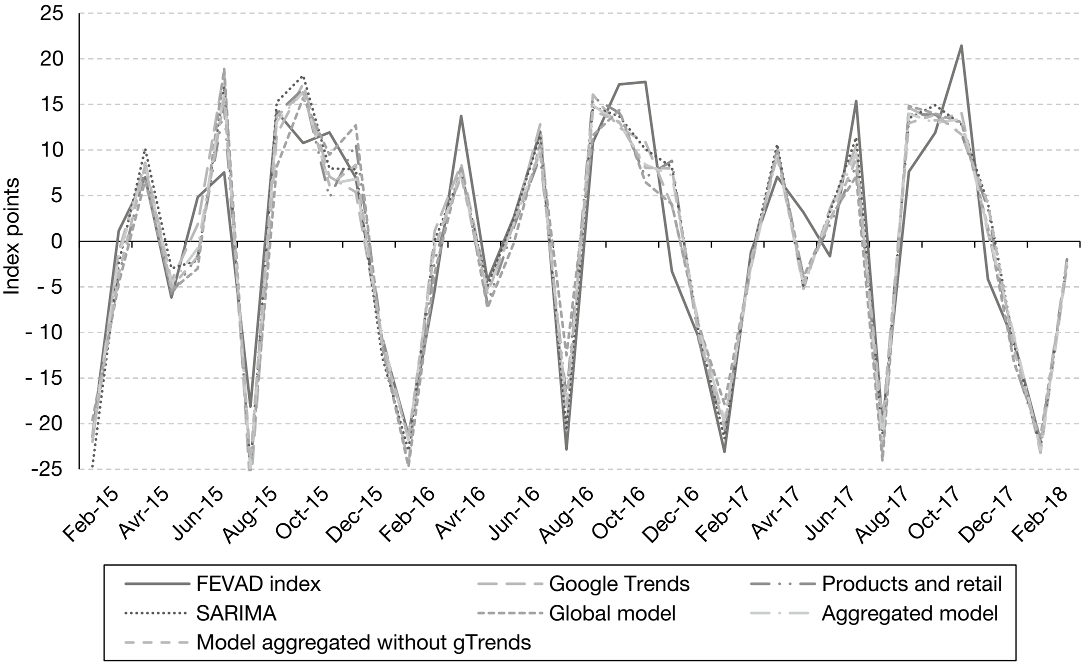
Figure A3-II
Absolute Deviations with Seasonal Adjustment for the Historical Time Series



Reading Note: The seasonally adjusted estimation using the data available at present, varies by 4.1 index points from that carried out in quasi-real time. The Google Trends forecast for July 2016 is even closer to the obtained value for the seasonally adjusted index than that obtained using the data available in July 2016.
Sources: Google Trends, FEVAD, Banque de France DGS SEEC.

MODEL FORECASTS FOR THE TOTAL SALES INDEX

Figure A4
Forecasts from the Various Models in Estimation of the Total



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

APPENDIX 5

DESCRIPTION OF VARIABLE SELECTION IN THE GOOGLE TRENDS MODEL FOR TOTAL SALES ESTIMATIONS

Table A5
Descriptive Statistics for Variables Selected in Estimation of the Total

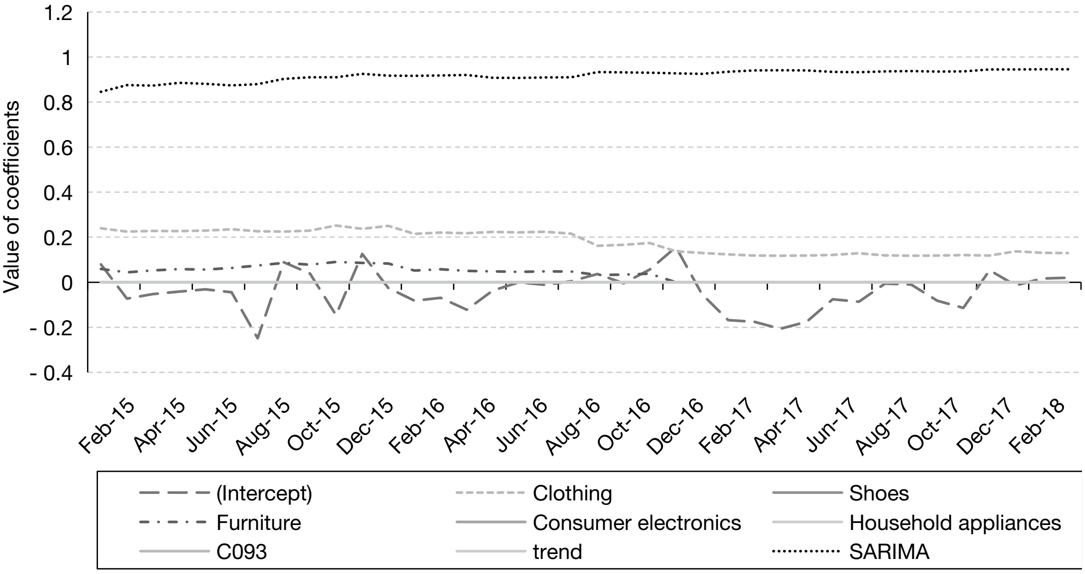
	Mean	Minimum	Maximum	Selections
Amazon	0.07	0.00	0.48	9
eBay	- 0.43	- 0.80	- 0.24	38
Vente-privee.com	- 0.02	- 0.21	0.00	8
Cdiscount	0.01	0.00	0.07	5
FNAC	0.01	0.00	0.18	5
Fnac Darty Group	0.00	0.00	0.00	0
PriceMinister	- 0.21	- 0.37	0.00	37
Leroy.Merlin	0.05	0.00	0.13	32
Central Public Procurement Office	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 + Sales Events + Black Friday	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
Galeries 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
Decathlon	0.00	- 0.10	0.00	3
Trend	0.00	0.00	0.00	0
SARIMA	0.97	0.93	1.04	38

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

APPENDIX 6

CHANGE IN COEFFICIENTS IN THE RETAIL TRENDS MODEL IN THE ESTIMATION OF TOTAL SALES

Figure A6
Change in Coefficients for the Retail Model in the Estimation of the Total



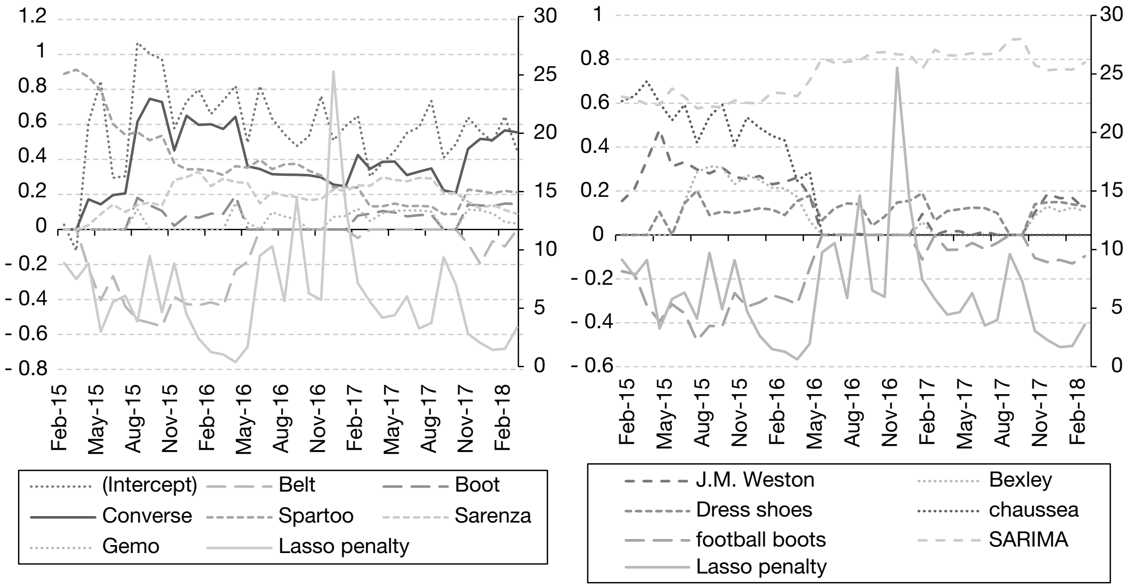
Note: Only selected variables are not shown as a solid line.
Sources: FEVAD, Banque de France DGS SEEC.

APPENDIX 7

STABILITY OF THE GOOGLE TRENDS MODEL IN ESTIMATION OF THE SHOES INDEX

Similar to the figure VIII showing the change in Google Trends model coefficients for the estimation of the total sales index, the lasso penalty is on the secondary axis. Figure A7 below presents only the most often selected variables (at least eight times out of 38 iterations) over the period.

Figure A7
Change in Google Trends Model Estimation, Coefficients of the Index for Shoes



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