BANQUE DE FRANCE
EUROSYSTÈME

Use of Google Trends Data to Improve Survey Measures of Retail Trade

Presented by Valerie Chauvin, written by Francois Robin

- Context
  - Retail trade survey
  - e-commerce in the survey
  - Late release of FEVAD data
- Google Trends & Aggregated model
  - Google Trends series
  - Selecting variables
  - Aggregating
- Conclusion
  - Encouraging results
  - Limits

Retail trade survey

- Available on the internet :

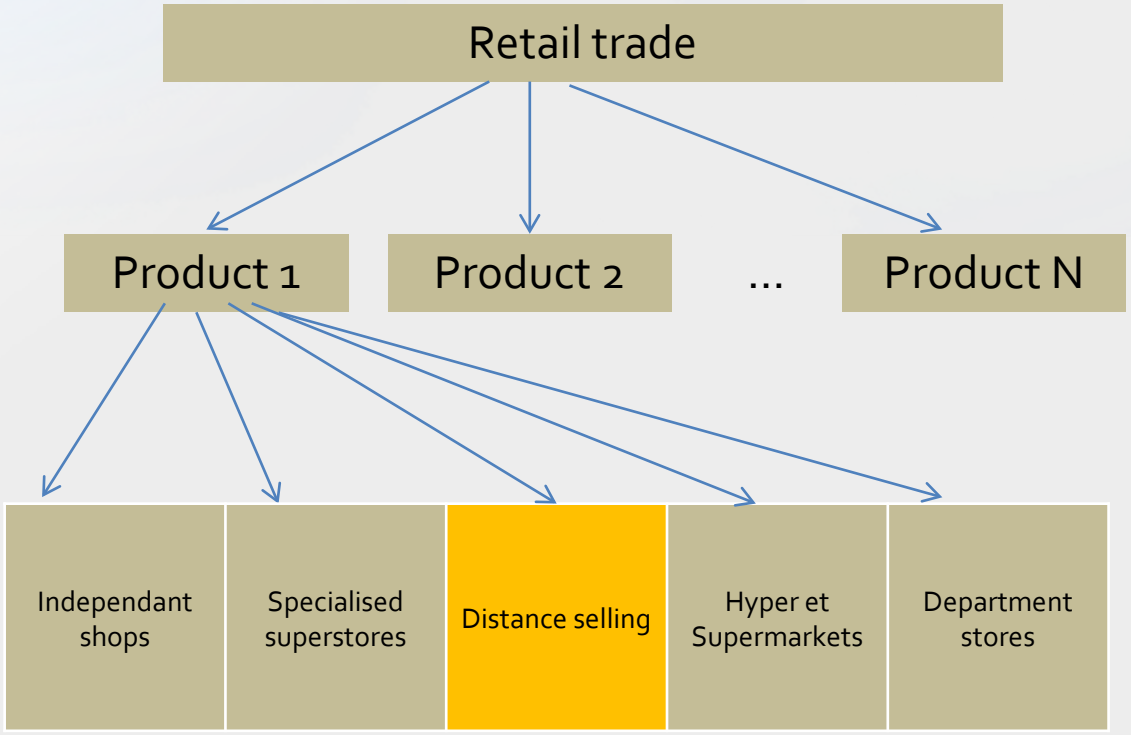
https://www.banque-france.fr/statistiques/chiffres-cles-france-et-etranger/enquetes-de-conjoncture/conjoncture-commerce-de-detail

- Quantitative index based on :
  - Y-o-y growth of turnover of a balanced panel
  - Declared by distribution channel (hyper/supermarkets, chain stores, independant shops,…)
  - And by products (Do it yourself, furniture, appliances, clothing,…)

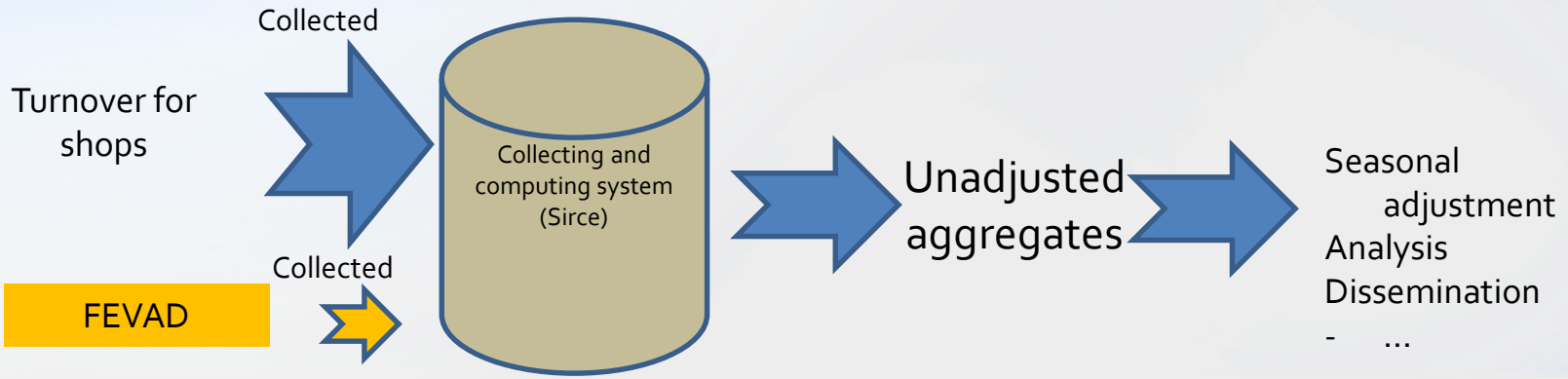
- Incorporated by INSEE in quarterly national accounts

E-commerce in the survey

By product and by distribution channel



E-commerce in the survey



Cooperation with FEVAD

- Well-known professional federation .
- Preliminary work (data consistency, trend over time)
- e-commerce: fast-changing population
- High cost to follow the players using standard panel practices

FEVAD turnover compiled late

Fevad turnover : gathered for different products and total sales

Consumer electronics

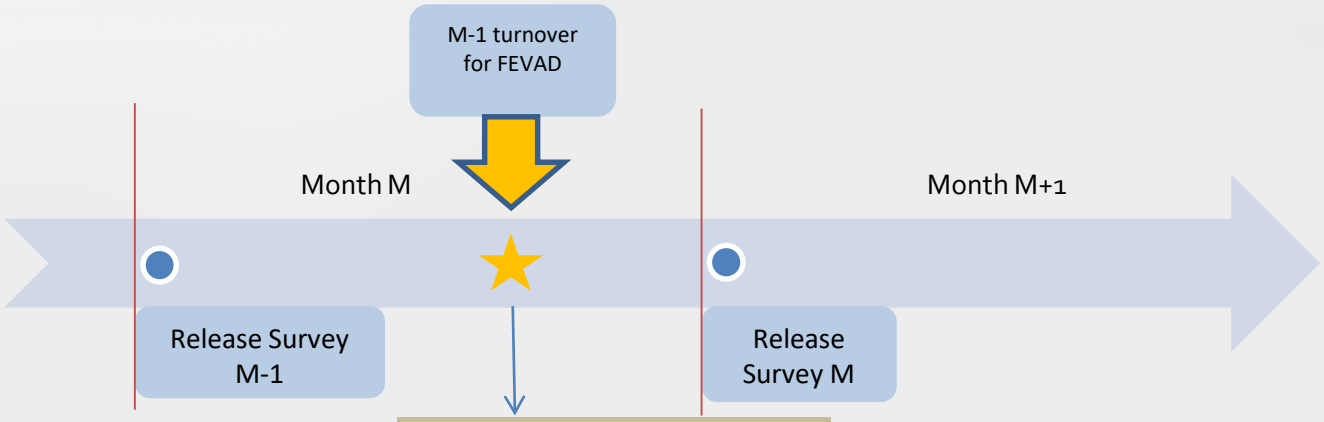
Household Appliances

Total distance sales

Clothing

Furniture

Footwear



- **FEVAD turnover for M-1 included in the second release**
- **6 FEVAD aggregates estimated for the first release**

Google Trends Data

- Indices computed from samples
- Monthly data

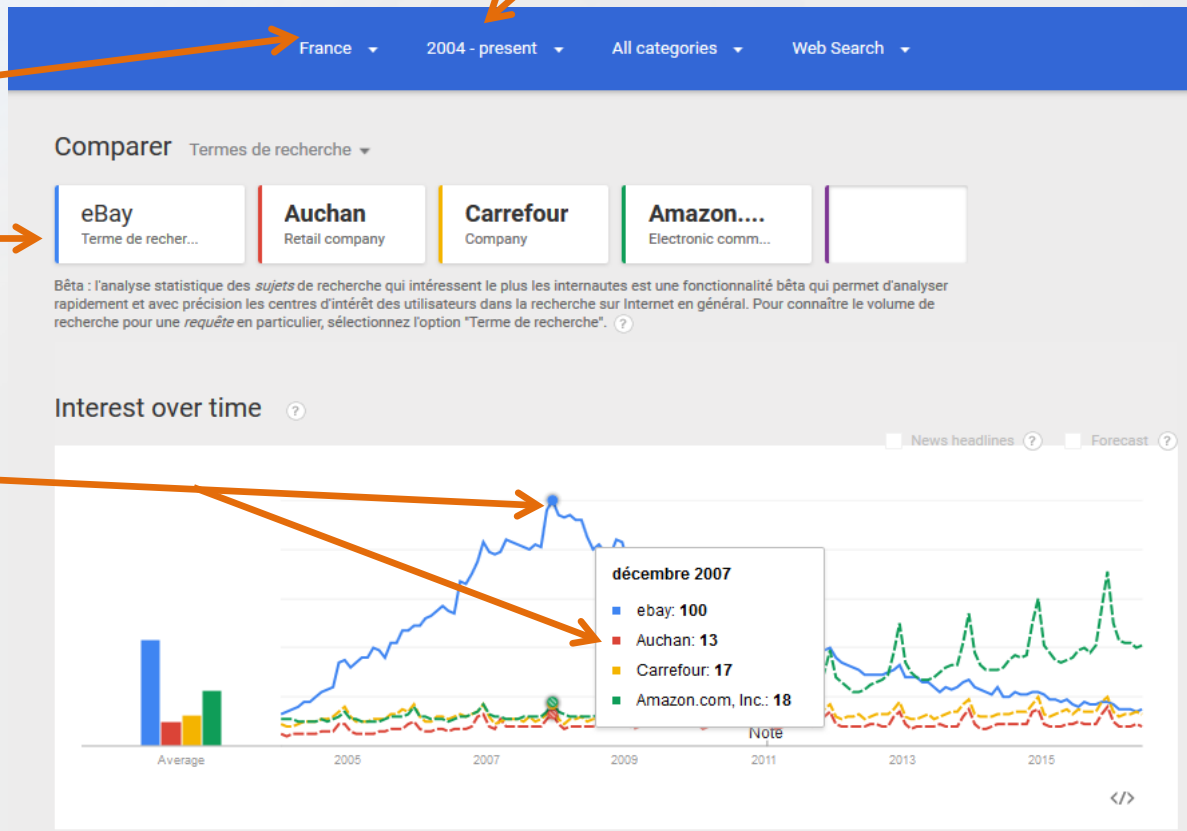
Back to 2004 at the earliest

Geographic scope of searches

5 searches maximum

Relative maximum
Set at 100

Rounded values
[*Kozicki & Hoffman*]



- Also used [*Choi H. et Varian H. (2011), Askistas N. & Zimmerman K. F. (2009)*]

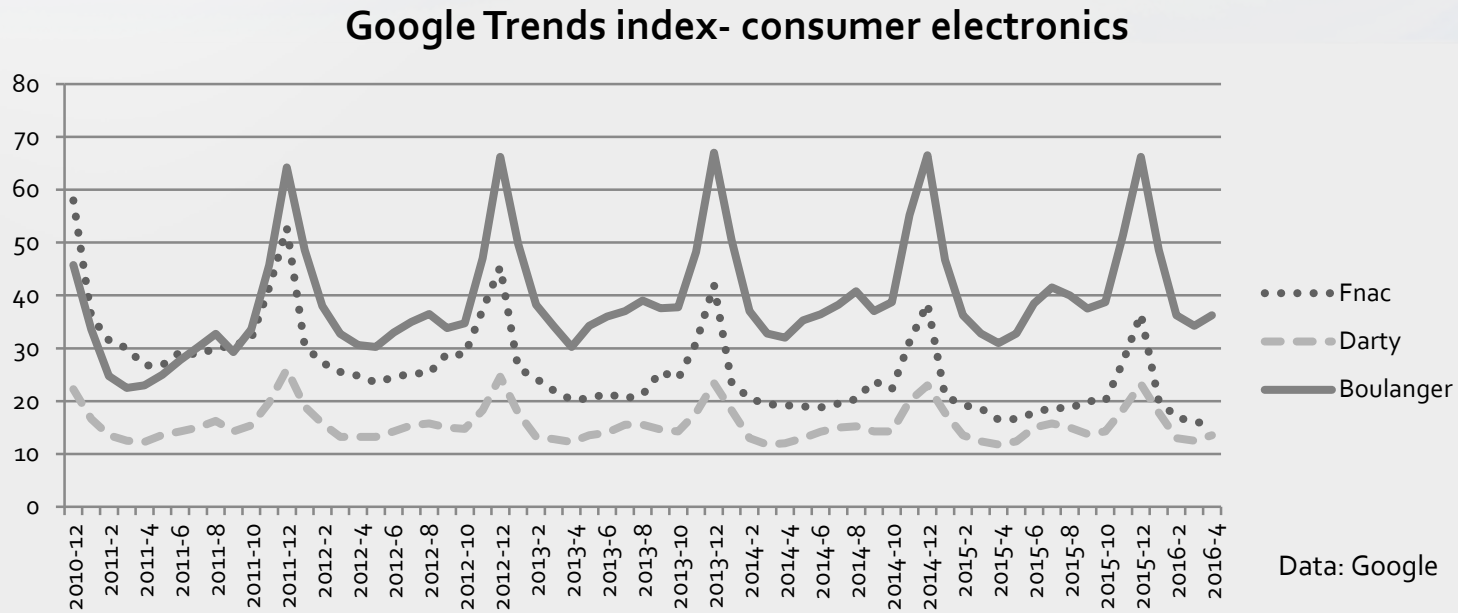
Robustness

- Correlations between two drawings:

Amazon	Cdiscount	Fnac	eBay	Carrefour
99,96%	99,50%	99,95%	99,98%	99,21%

Données Google

- Sectoral seasonality:



Modelling

$$Y = \alpha + X'\beta + \delta t + \gamma I_{mois} + \varepsilon$$

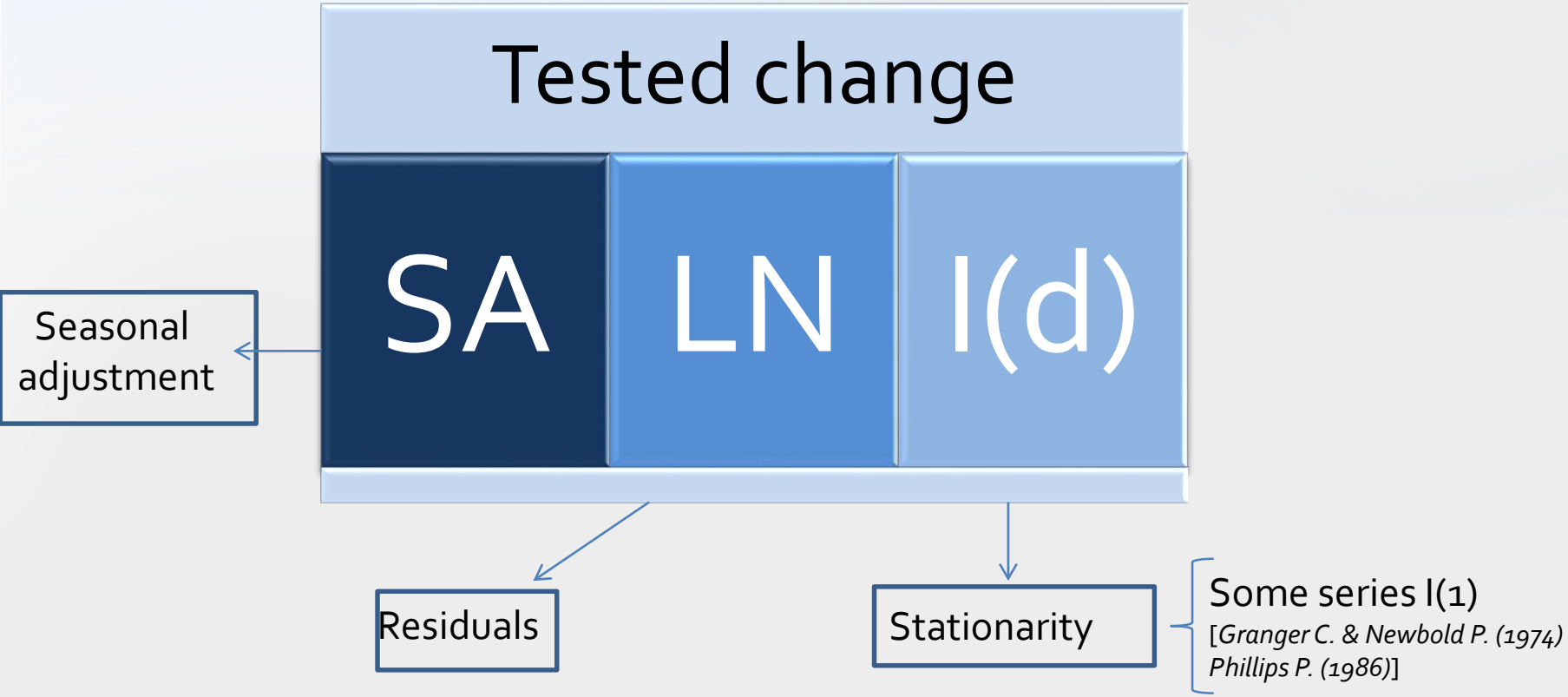
- VAR, ARIMAX, standard regressions,..

- Y: index for e-turnover by product
- X: Google Trends series or index for turnover in traditional distribution channel
- t: trend
- γI_{mois} : seasonal dummies

- Monthly data from January 2012 to May 2017

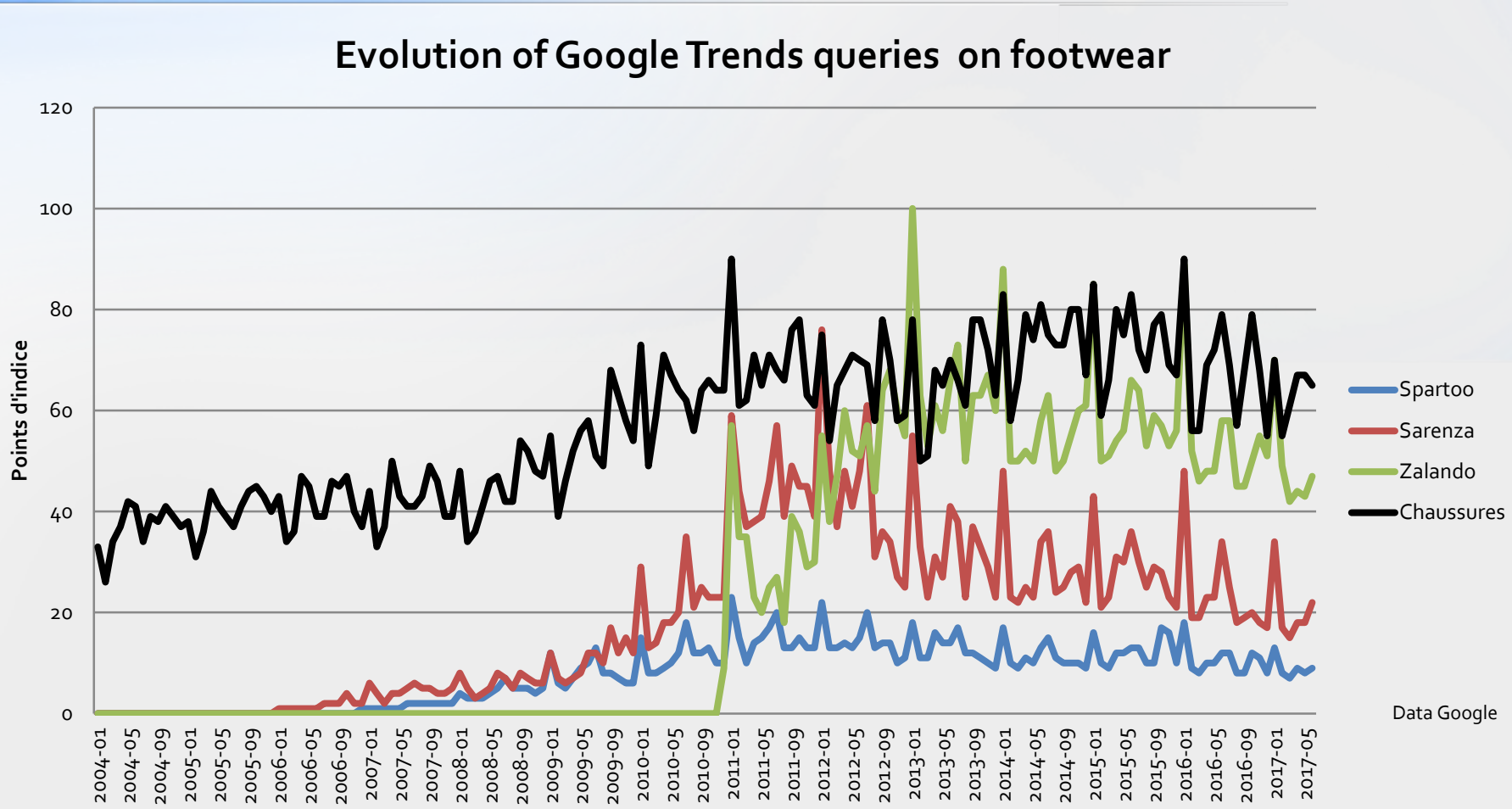
Which data?

$$Y = \alpha + X'\beta + \delta t + \gamma I_{mois} + \varepsilon$$



BANQUE DE FRANCE
EUROSYSTÈME

Emerging phenomenom



Selecting variables

For each iteration of the backtest :

step 0 : Constant only

steps i : inclusion(*Forward*) or
Suppression (*Backward*)
of the variable optimising the selection criterium (BIC)

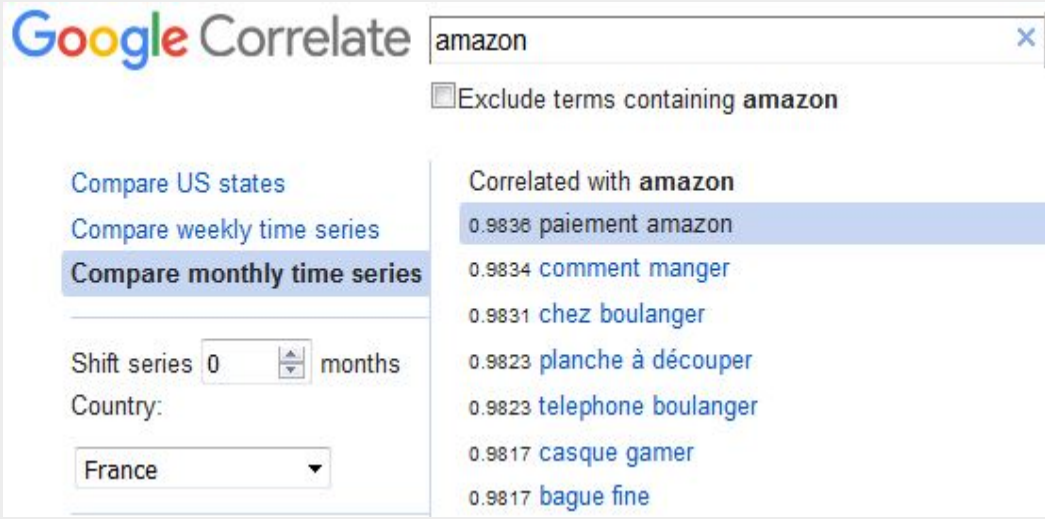
Step n : Final model

Chosen variables:
Significant, among the
Most often selected

Stepwise Regression

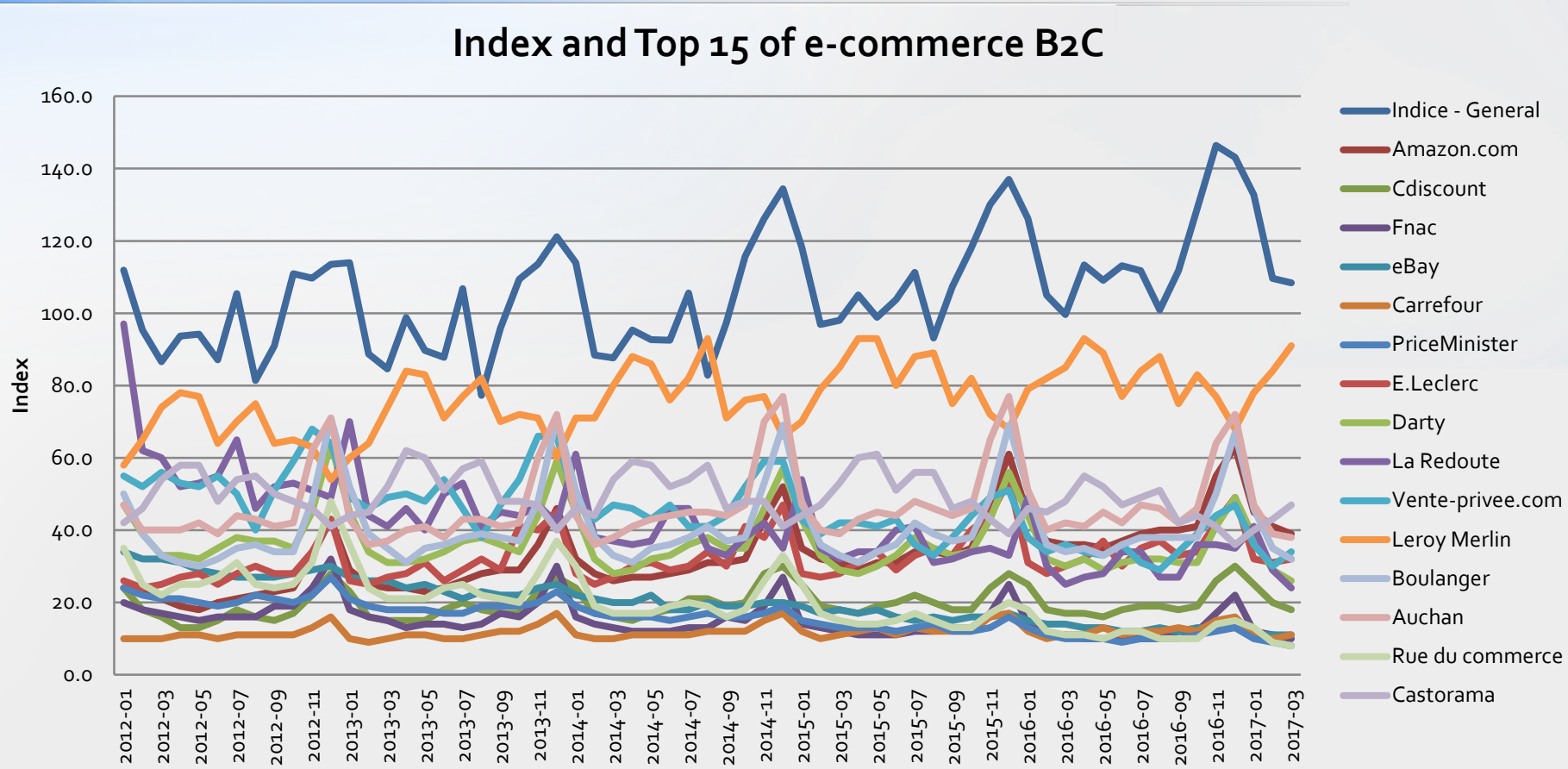
Google Correlate

PCA



BANQUE DE FRANCE
EUROSYSTÈME

Variables for total sales



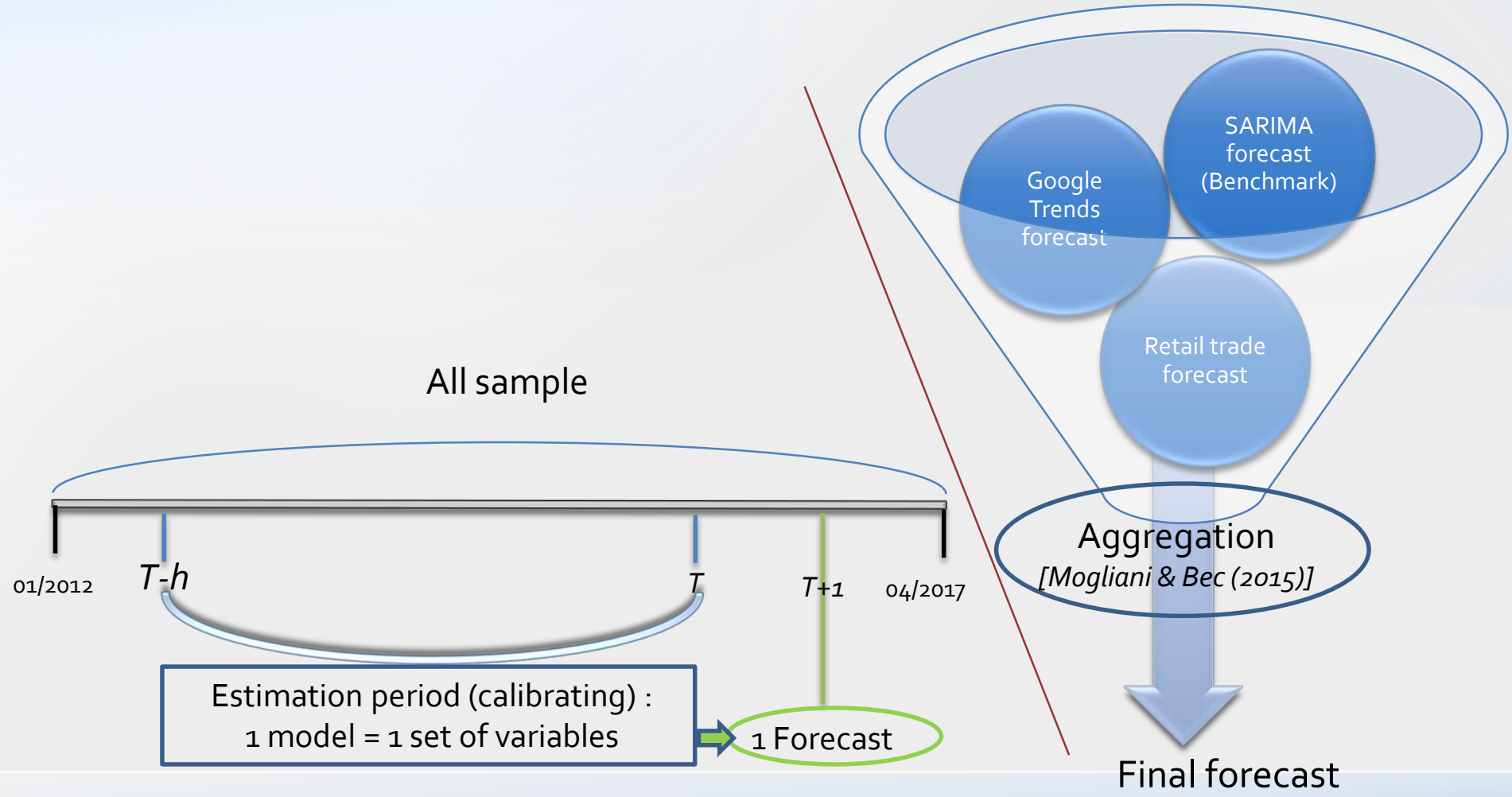
Too many variables → PCA ?

Data: Google-FEVAD
Computation: SEEC

Chosen GT queries by products

Footwear	Furniture	Clothing	Consumer electronics t	Household appliances
Spartoo	Ikea	Kiabi	cDiscount	GrosBill
Zalando	Leroy Merlin	H&M	Amazon	Carrefour
Sarenza	Conforama	Robe + Jupe	Deezer	Amazon
Chaussures	Maisons du monde	Jeans + Pantalon	FNAC	Darty
Botte	Meuble	La Redoute		Boulanger
Sandale				

Methodology



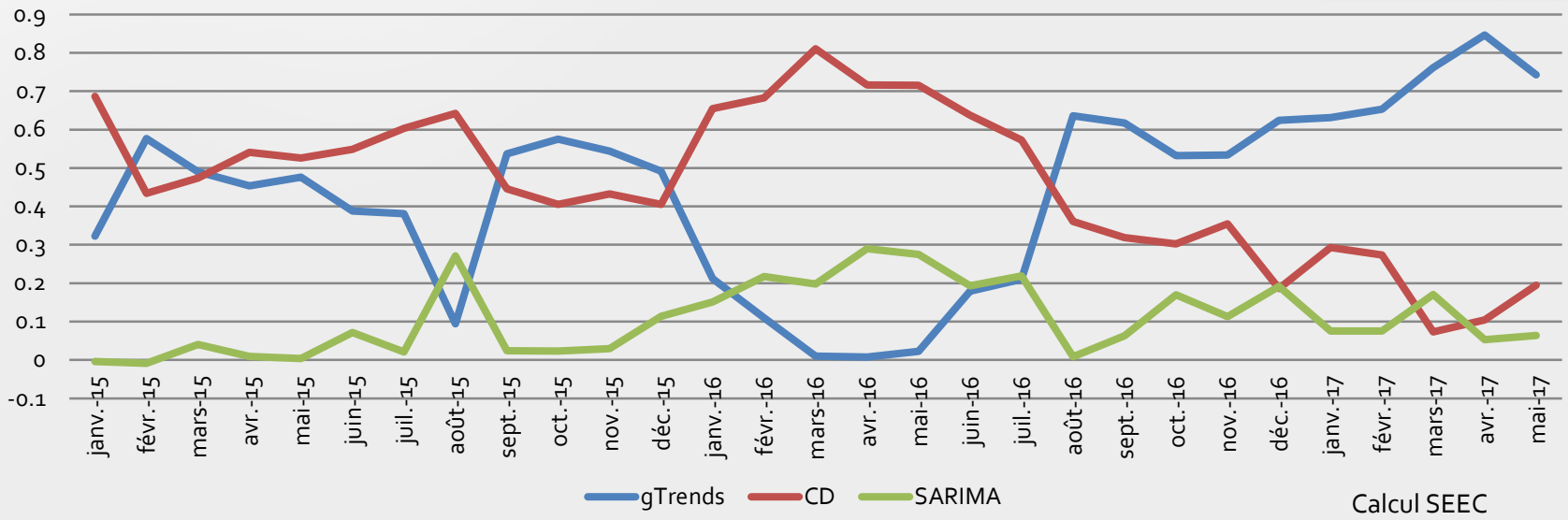
Bayesian aggregation

[*Zeugner (2011), Hoeting, Madigan, Raftery & Volinsky (1999)*]

$$Y = \alpha + \theta \hat{Y}_{gTrends} + \mu \hat{Y}_{CD} + \gamma \hat{Y}_{SARIMA} + \varepsilon$$

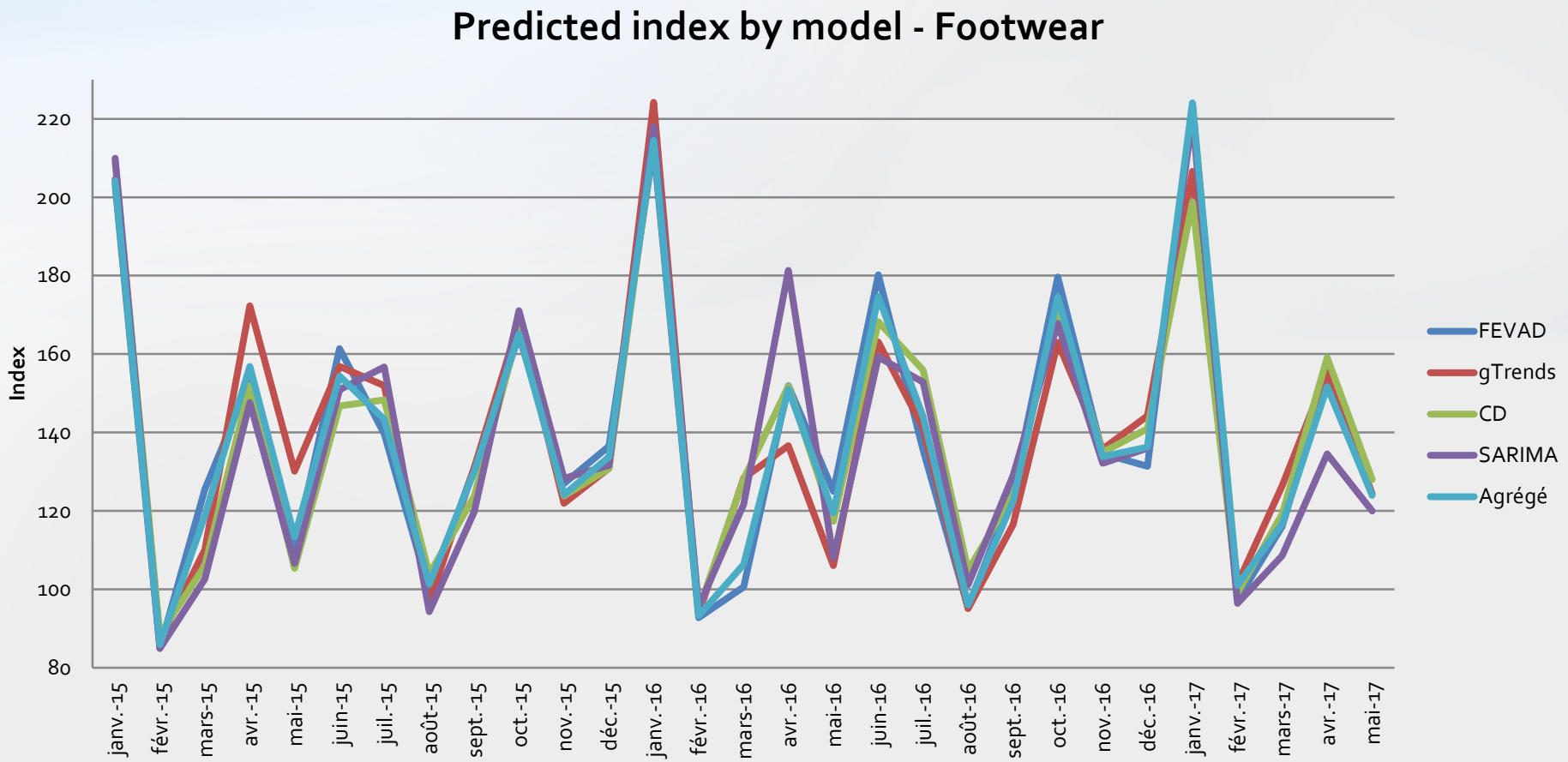
$$\hat{\theta} = \sum_{k=0}^{7} P(M_k|D)\hat{\theta}_k$$

Evolution of the weights – Total sales



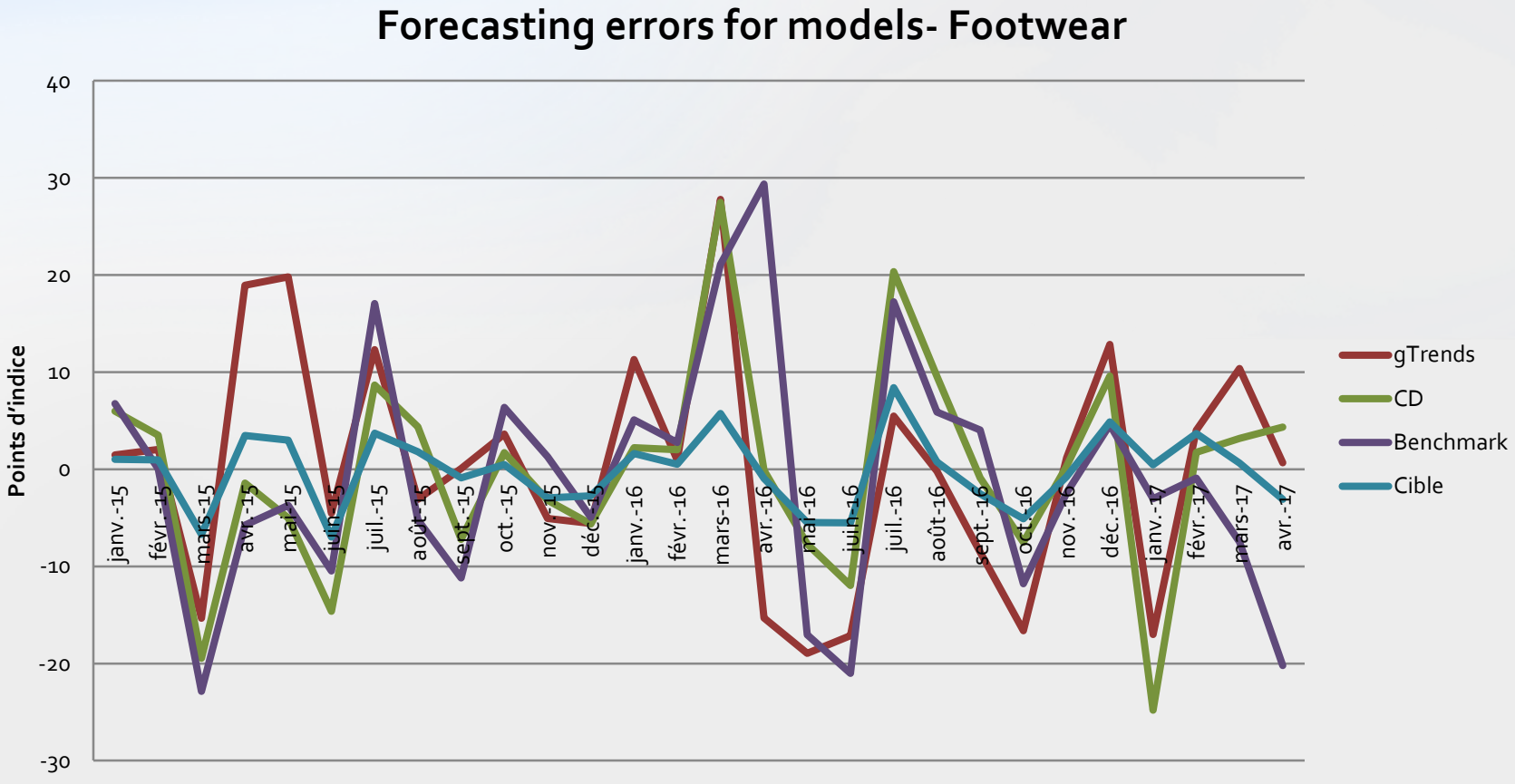
Calcul SEEC

Encouraging results: the case for footwear



Data: Google-FEVAD-BdF
Computation: SEEC

Footwear



Data: Google-FEVAD-BdF
Computation: SEEC

BANQUE DE FRANCE
EUROSYSTÈME

Summarised results: unadjusted series

RMSFE	Google Trends	CD	ARIMA	Cible
Total	8,3	**3,4***	4,3	1,7
Footwear	12,2	**10,7**	12,7	3,9
Furniture	10,1	**9,5**	10,9	7,9
Clothing	**5,3**	6,4	5,6	2,4
Consumer electronics	**12,9**	15,3	14,1	9,5
Household appliances	8,5	**7,2**	8,4	4,5

Data: Google-FEVAD-BDF
Computation: SEEC

Limits

- Google Trends Data out of the reach of statisticians:
  - Unknown counting method
  - Evolutions in search engine

- Fast-changing phenomenon
  - Players quickly in and out
  - Link between internet search and purchase
  - ➢ Models need to be revised (every year)

- Short backtesting period (24 months)

BANQUE DE FRANCE
EUROSYSTÈME

Conclusion

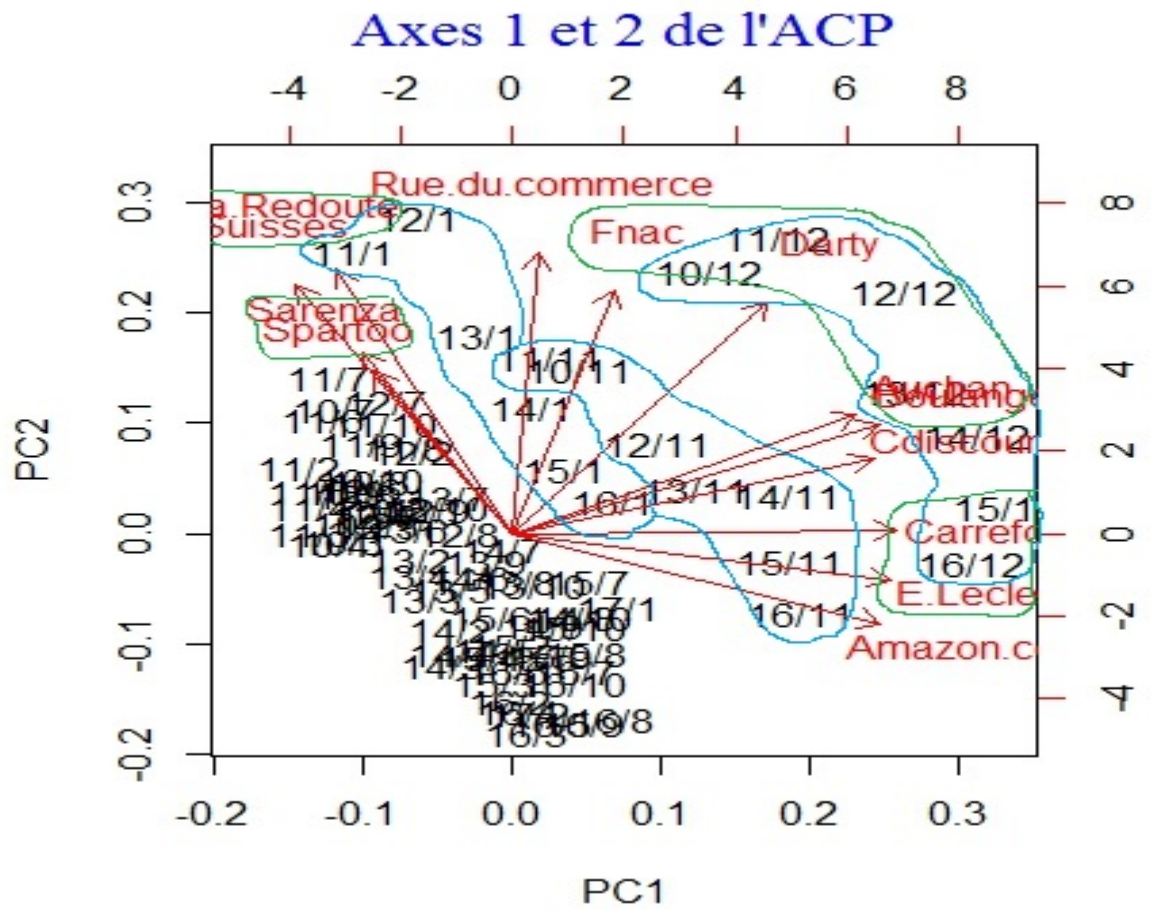
- Positive results
- Multidimensional information
- Use in production

BANQUE DE FRANCE
EUROSYSTÈME

Annexes

Robustesse

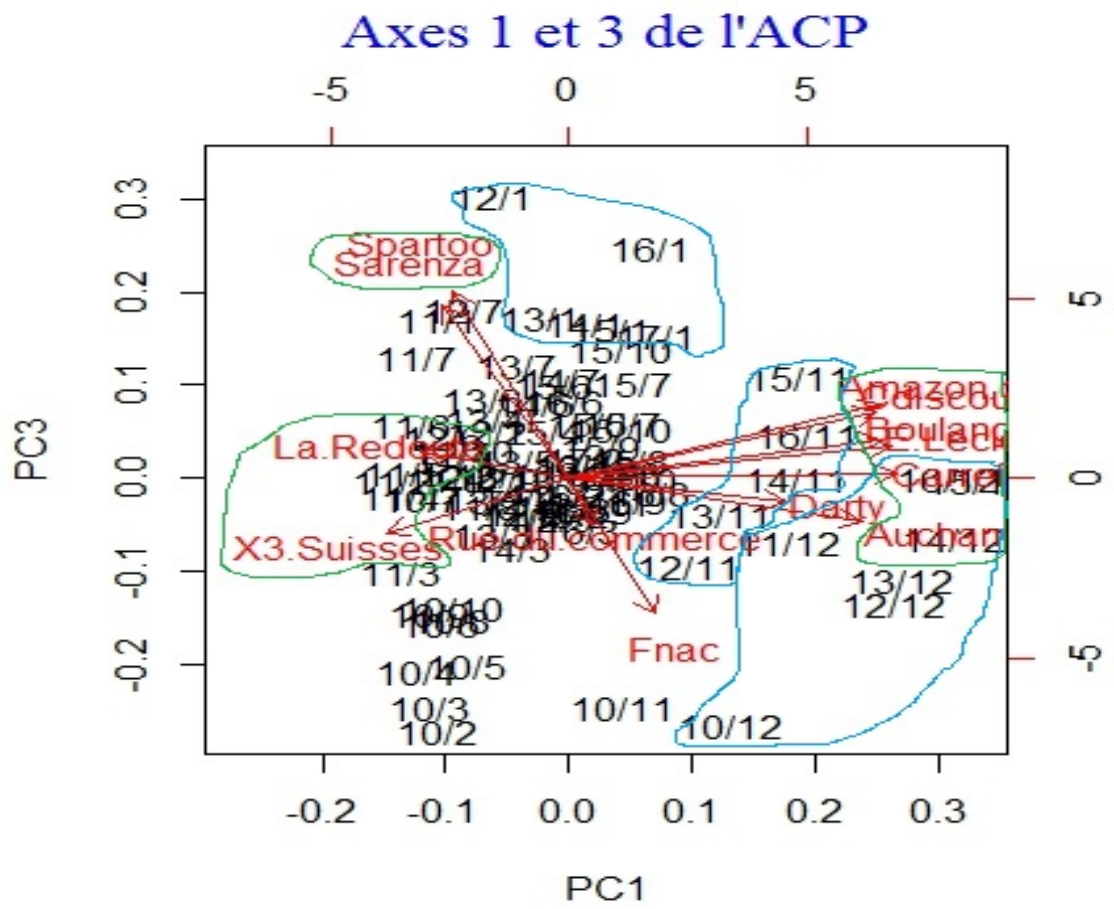
Requêtes Google Trends
Spartoo Sarenza
Carrefour Auchan E.Leclerc
Darty Boulanger FNAC
Amazon cDiscount RueDuCommerce
La Redoute 3 Suisses



Données Google
Calcul SEEC

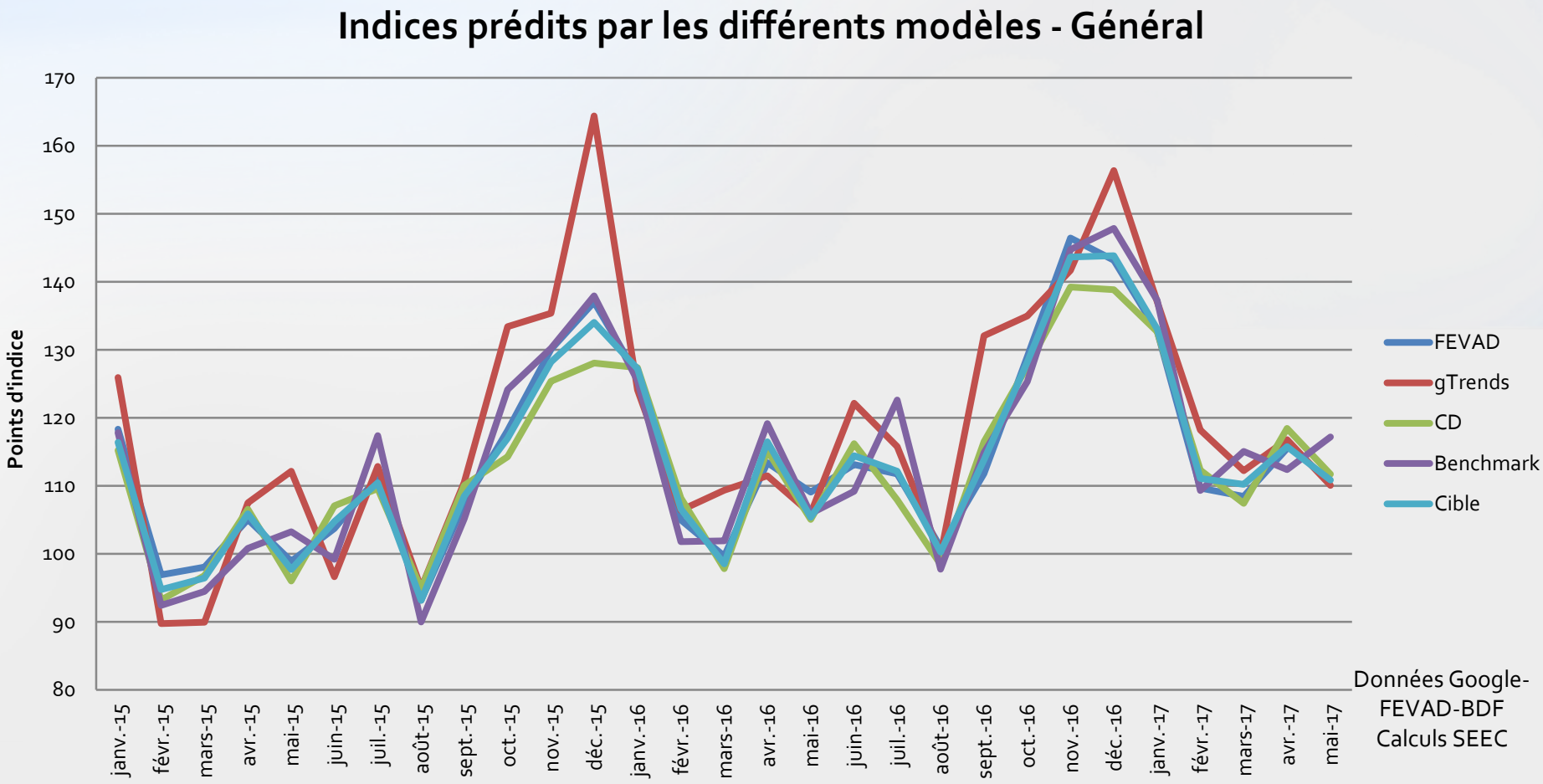
Robustesse

- Corrélations stables dans le temps
- Partition par produit

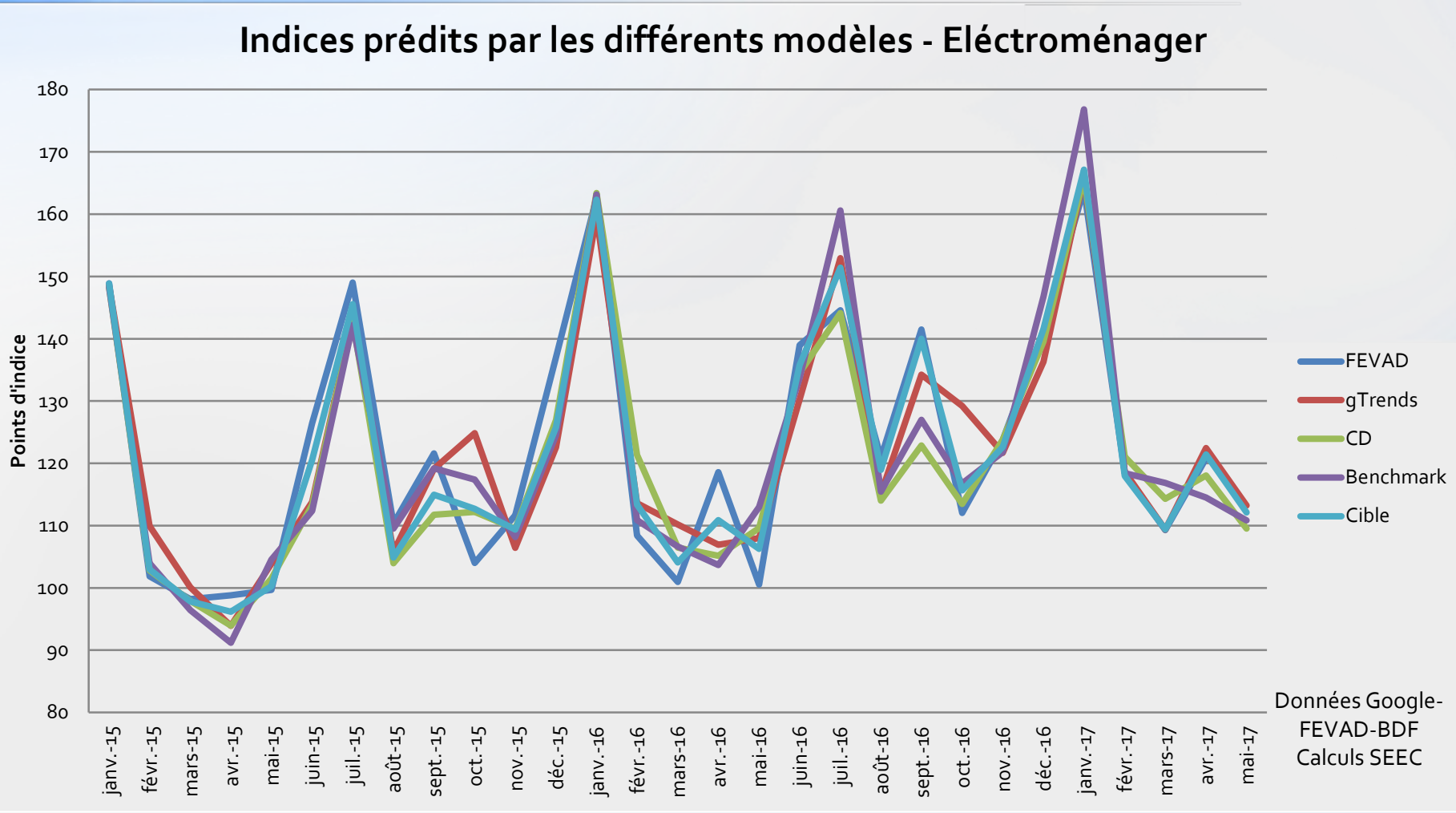


Données Google
Calcul SEEC

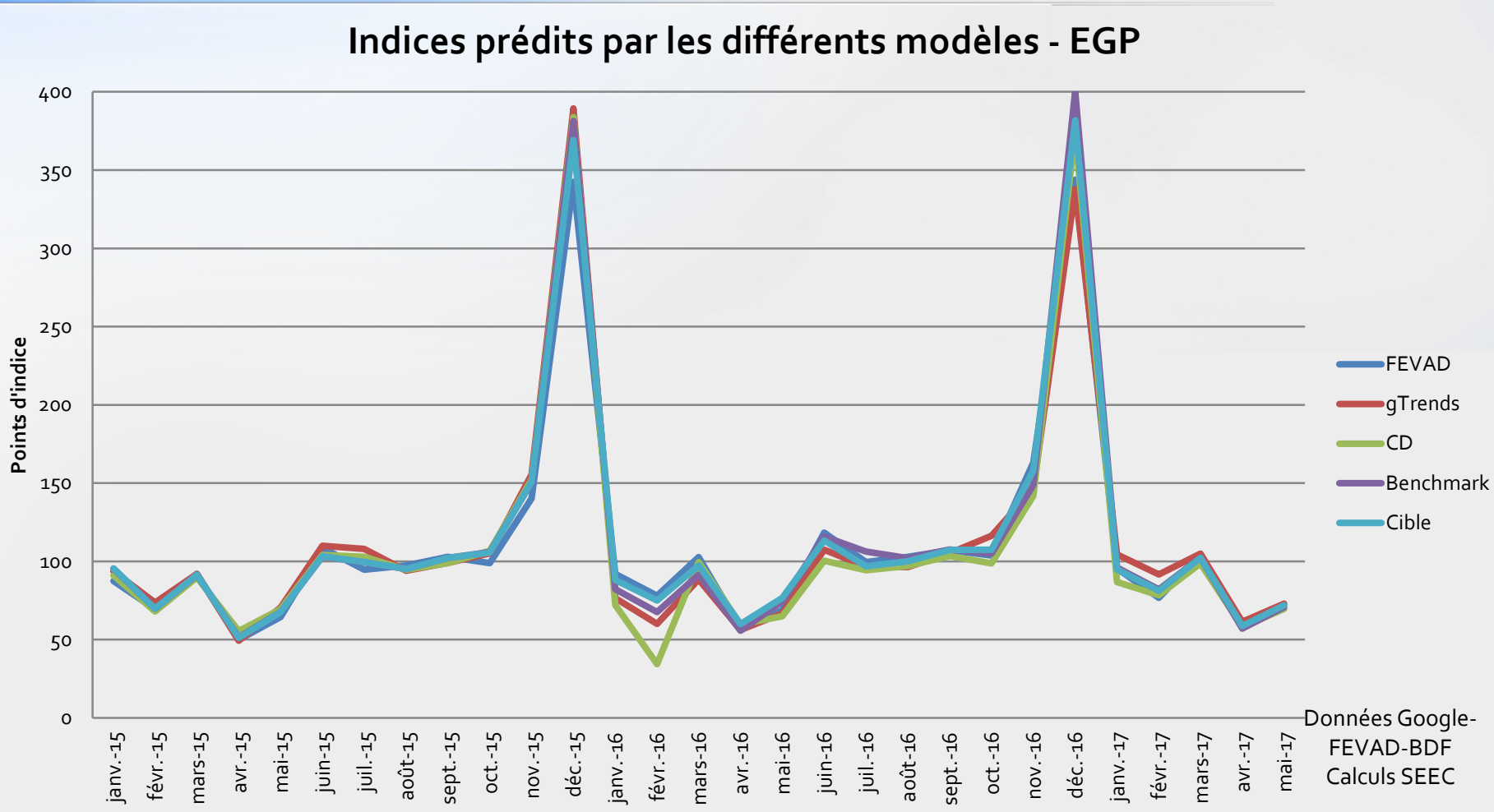
Général


Indices prédits par les différents modèles - Général
Points d'indice
170
160
150
140
130
120
110
100
90
80
janv.-15
févr.-15
mars-15
avr.-15
mai-15
juin-15
juil.-15
août-15
sept.-15
oct.-15
nov.-15
déc.-15
janv.-16
févr.-16
mars-16
avr.-16
mai-16
juin-16
juil.-16
août-16
sept.-16
oct.-16
nov.-16
déc.-16
janv.-17
févr.-17
mars-17
avr.-17
mai-17
FEVAD
gTrends
CD
Benchmark
Cible
Données Google-
FEVAD-BDF
Calculs SEEC

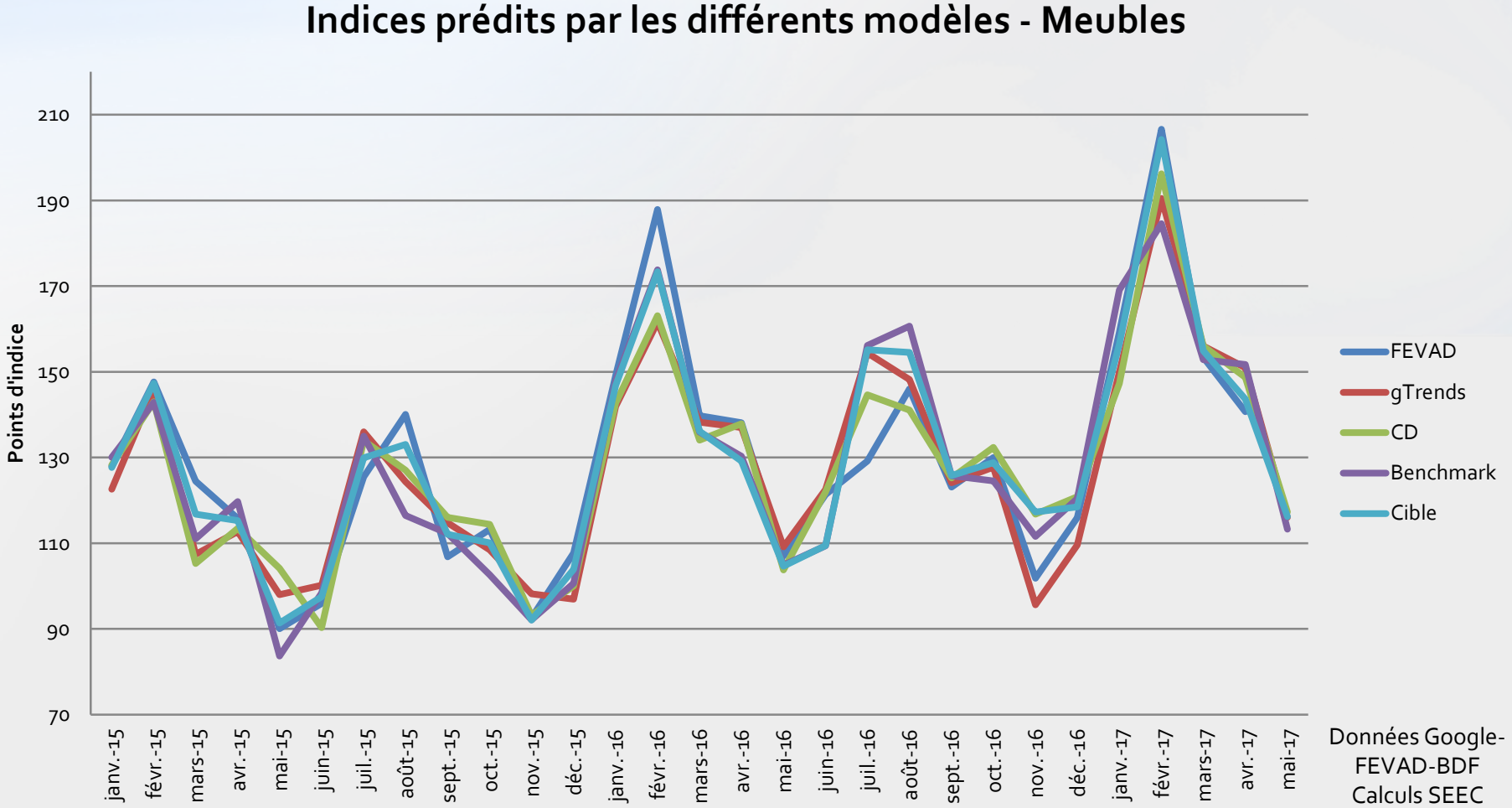
Électroménager



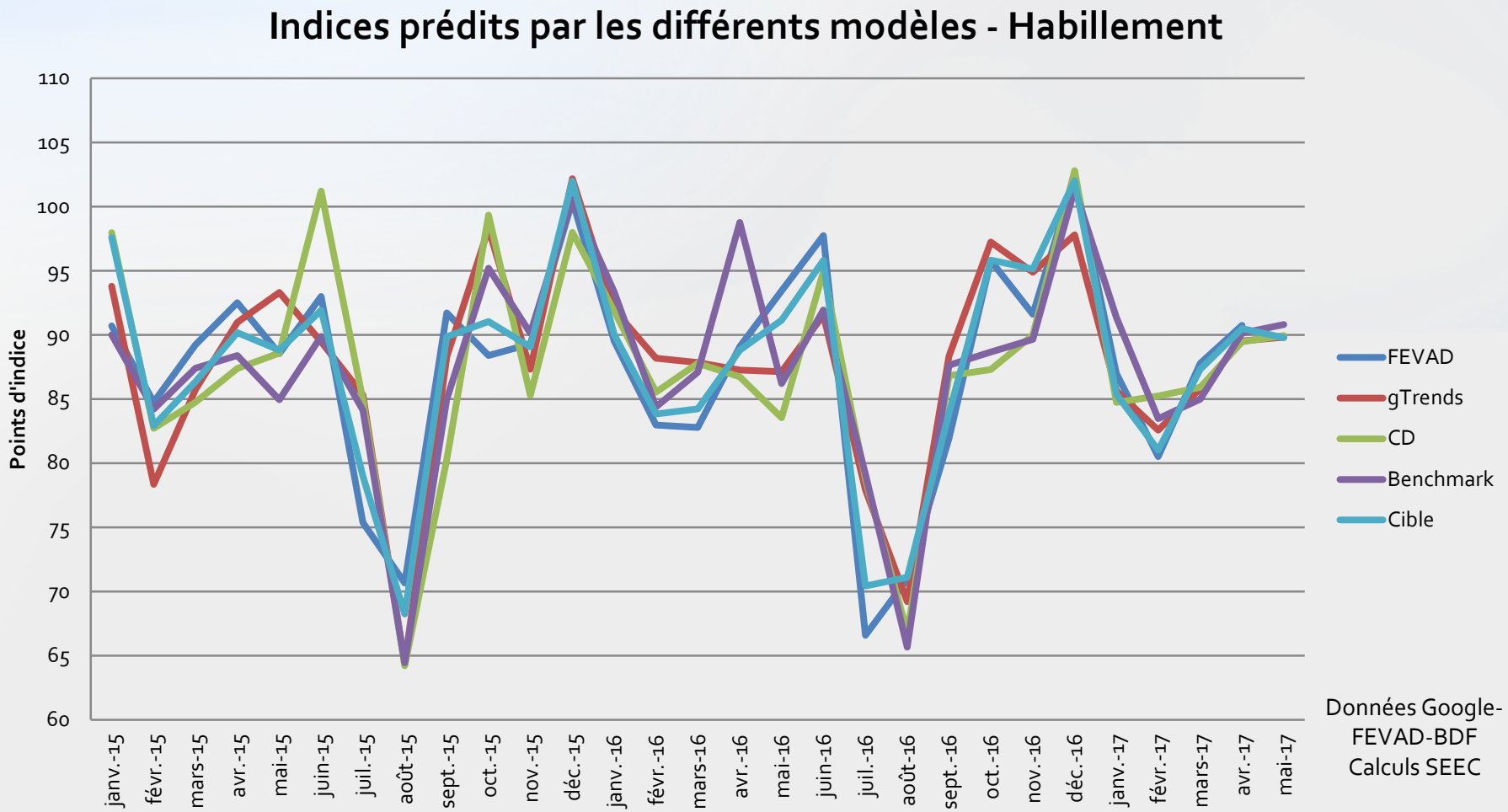
Électronique grand public



Meubles


Indices prédits par les différents modèles - Meubles
Points d'indice
210
190
170
150
130
110
90
70
janv.-15
févr.-15
mars-15
avr.-15
mai-15
juin-15
juil.-15
août-15
sept.-15
oct.-15
nov.-15
déc.-15
janv.-16
févr.-16
mars-16
avr.-16
mai-16
juin-16
juil.-16
août-16
sept.-16
oct.-16
nov.-16
déc.-16
janv.-17
févr.-17
mars-17
avr.-17
mai-17
FEVAD
gTrends
CD
Benchmark
Cible
Données Google-FEVAD-BDF
Calculs SEEC

Habillement



Données Google-FEVAD-BDF
Calculs SEEC

BANQUE DE FRANCE
EUROSYSTÈME

Résultats majeurs : Données log

RMSFE	Google Trends	CD	ARIMA	Cible
General	8,5	**2,4***	4,3	1,6
Chaussures	13,2	**11,2**	12,7	4,7
Meubles	10,7	**8,6**	10,9	7,4
Habillement	**5,6**	6,0	**5,6**	2,2
EGP	**9,1**	11,4	14,1	6,6
Électroménager	7,9	**7,6**	8,4	5,3

Données Google-FEVAD-BDF
Calculs SEEC

BANQUE DE FRANCE
EUROSYSTÈME

Résultats majeurs : Données CVS

RMSFE	Google Trends	CD	ARIMA	Cible
General	4,6	**3,3***	4,3	1,7
Chaussures	**10,7**	11,4	12,7	5,0
Meubles	9,2	**8,2**	10,9	7,6
Habillement	**5,1**	5,5	5,6	2,8
EGP	**9,4**	14,0	14,1	11,6
Électroménager	**8,0**	8,3	8,4	7,3

Données Google-FEVAD-BDF
Calculs SEEC

BANQUE DE FRANCE
EUROSYSTÈME

Résultats majeurs : Données différenciées

RMSFE	Google Trends	CD	ARIMA	Cible
General	8,7	**2,7***	4,3	1,9
Chaussures	16,2	14,8	**12,7**	7,2
Meubles	14,4	12,0	**10,9**	10,7
Habillement	8,9	9,2	**5,6**	5,5
EGP	14,3	17,8	**14,1**	13,3
Électroménager	12,1	9,7	**8,4**	5,9

Données Google-FEVAD-BDF
Calculs SEEC

Méthodologie Google ?