Non-nested national and European classification
On the Prodcom classification
01 · The Annual Production Survey
02 · Non-nested hierarchies
03 · Disclosure control
01 The Annual Production Survey
The Annual Production Survey (EAP)

Stakes

- Frequency: Annual
- Field: All French companies
- Sample: 35 000 companies
- Collection mode: Internet
The Annual Production Survey (EAP)

Stakes

2 goals:
- Identify companies main activities
- Produce statistics at a national and european level
The Annual Production Survey (EAP)

Stakes

2 goals:
- Identify companies' main activities
- Produce statistics at a national and European level

EAP  →  fuels  →  Prodcom

- CPF4-5
- Prodfra

- Prodcom4
- Prodcom
## The Annual Production Survey (EAP)

### Example of disseminated results by CPF4-5

<table>
<thead>
<tr>
<th>CPF4-5</th>
<th>Label</th>
<th>Number of legal units</th>
<th>2017 sales (in millions of euros)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3240</td>
<td>Manufacture of games and toys</td>
<td>194</td>
<td>321 381</td>
</tr>
<tr>
<td>3250A</td>
<td>Manufacture of medical, surgical and dental equipment</td>
<td>3251</td>
<td>5 383 356</td>
</tr>
<tr>
<td>3250B</td>
<td>Manufacture of glasses</td>
<td>106</td>
<td>735 022</td>
</tr>
<tr>
<td>3511</td>
<td>Production of electricity</td>
<td>c</td>
<td>c</td>
</tr>
</tbody>
</table>

*c = confidential*
The Annual Production Survey (EAP)

Example of disseminated results by Prodfra

<table>
<thead>
<tr>
<th>Prodfra</th>
<th>Label</th>
<th>Number of legal units</th>
<th>2017 sales (in millions of euros)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3240320000</td>
<td>Manufacture of puzzles</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>3240410000</td>
<td>Manufacture of playing cards</td>
<td>8</td>
<td>8 569</td>
</tr>
<tr>
<td>3240421000</td>
<td>Manufacture of pool tables and its accessories</td>
<td>16</td>
<td>11 294</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c = confidential
Non-nested hierarchies
Nested product classifications?

Manufacture of contact lenses

325041300B
Nested product classifications?

Manufacture of contact lenses

325041300B

Prodcom4

CPF4-5
Nested product classifications?

- Manufacture of contact lenses

3250B

Prodcom4

CPF4-5

325041300B
Nested product classifications?

- Manufacture of contact lenses

3250B

<table>
<thead>
<tr>
<th>Product classification</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>3250B</td>
<td>Prodcom</td>
</tr>
<tr>
<td>0B</td>
<td>CPF4-5</td>
</tr>
<tr>
<td>3250</td>
<td>Prodcom4</td>
</tr>
<tr>
<td>4130</td>
<td></td>
</tr>
</tbody>
</table>
Nested product classifications?

Manufacture of contact lenses

3250B

CPF4-5

Prodcom

Prodcom4

325041300B

Prodfra
Non-nested national and European classification

Nested product classifications?
- Manufacture of contact lenses

3250B

2 hierarchies:
- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
Nested product classifications?

What should be disseminated in 2020
Nested product classifications?

What should be disseminated in 2020

CPF4-5 < Prodfra

Prodcom4 < Prodcom < Prodfra
Nested product classifications?

What is disseminated so far
Models : another hierarchy

An extract of the questionnaire
### Models: another hierarchy

- **An extract of the questionnaire**

<table>
<thead>
<tr>
<th>Prodfra - Product</th>
<th>Manufactured outside the company</th>
<th>Manufactured by the company (on a national territory)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1 = inputs not provided by the company</td>
<td>M2 = inputs provided by the company</td>
</tr>
<tr>
<td>2042185000 - Toothpastes</td>
<td>..........k€</td>
<td>..........k€</td>
</tr>
<tr>
<td>2042191500 - Soaps</td>
<td>..........k€</td>
<td>..........k€</td>
</tr>
</tbody>
</table>
## Models: another hierarchy

What is disseminated so far

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M5</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodfra</td>
<td>M2+M3+M4</td>
</tr>
</tbody>
</table>
**Models : another hierarchy**

**What is disseminated so far**

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M5</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodfra</td>
<td>M2+M3+M4</td>
</tr>
</tbody>
</table>

\[ \text{M2+M3+M4+M5} < (\text{M2+M3+M4, M5}) \rightarrow \text{MF hierarchy} \]
## Models: another hierarchy

What will be disseminated in 2020

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M5</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodfra</td>
<td>M2+M3+M4</td>
</tr>
</tbody>
</table>
## Models: another hierarchy

**What will be disseminated in 2020**

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M5</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodfra</td>
<td>M2+M3+M4</td>
</tr>
</tbody>
</table>

**MF**
Models: another hierarchy

What will be disseminated in 2020

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M5</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodfra</td>
<td>M2+M3+M4</td>
</tr>
</tbody>
</table>

Non-nested national and European classification
03 Disclosure control
Disclosure control

Linked tables

Deletion of the following cells:

- $\leq 2$ companies
- 1 company accounts for 85% of the sales
- induced secret
Disclosure control

Linked tables

Deletion of the following cells:

- $\leq$ 2 companies
- 1 company accounts for 85% of the sales
- induced secret

in the contingency table CPF4-5*Prodfra*MF*ME
Disclosure control

Linked tables

in the contingency table CPF4-5*Prodfra*MF*ME

\[
\begin{align*}
245 & \times 4000 \times 2 \times 2 \\
\approx 4 \text{ millions}
\end{align*}
\]

Hierarchies:
- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
- M2+M3+M4+M5 < (M2+M3+M4, M5) \rightarrow MF
- M3+M4+M5 < (M3+M4, M5) \rightarrow ME
Disclosure control

Linked tables

→ Too big for Tau-Argus 😞

in the contingency table CPF4-5*Prodfra*MF*ME

245 * 4000 * 2 * 2 ≃ 4 millions

Hierarchies:
- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
- M2+M3+M4+M5 < (M2+M3+M4, M5) → MF
- M3+M4+M5 < (M3+M4, M5) → ME
Disclosure control

How to slice this table?

Let's slice $\text{CPF4-5*Prodfra*MF*ME}$ by $\text{CPF4-5}$!
Disclosure control

- How to slice this table?

Let’s slice CPF4-5*Prodfra*MF*ME by CPF4-5!

2 things to consider beforehand:

- Totals?

Hierarchies:
- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
- M2+M3+M4+M5 < (M2+M3+M4, M5) → MF
- M3+M4+M5 < (M3+M4, M5) → ME
Disclosure control

How to slice this table?

Let’s slice CPF4-5*Prodfra*MF*ME by CPF4-5!

2 things to consider beforehand:

- Totals? → Not a problem

Hierarchies:

- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
- M2+M3+M4+M5 < (M2+M3+M4, M5) → MF
- M3+M4+M5 < (M3+M4, M5) → ME
Disclosure control

How to slice this table?

Let’s slice CPF4-5*Prodfra*MF*ME by CPF4-5!

2 things to consider beforehand:

- **Totals?** → Not a problem
- **Disjunctions** between CPF4-5 and Prodcom4

E.g.:

<table>
<thead>
<tr>
<th>Prodcom4</th>
<th>CPF4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1721</td>
<td>1721A</td>
</tr>
<tr>
<td></td>
<td>1721B</td>
</tr>
<tr>
<td></td>
<td>1721C</td>
</tr>
</tbody>
</table>

Hierarchies:
- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
- M2+M3+M4+M5 < (M2+M3+M4, M5) → MF
- M3+M4+M5 < (M3+M4, M5) → ME

Not a problem
Disclosure control

How to slice this table?

Let's slice CPF4-5*Prodfra*MF*ME by CPF4-5!

2 things to consider beforehand:

- **Totals**? → Not a problem
- **Disjunctions** between CPF4-5 and Prodcom4

E.g.:

<table>
<thead>
<tr>
<th>Prodcom4</th>
<th>CPF4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1721</td>
<td>1721A</td>
</tr>
<tr>
<td></td>
<td>1721B</td>
</tr>
<tr>
<td></td>
<td>1721C</td>
</tr>
</tbody>
</table>

→ Have to be in the same slice

Hierarchies:
- CPF4-5 < Prodfra
- Prodcom4 < Prodcom < Prodfra
- M2+M3+M4+M5 < (M2+M3+M4, M5) → MF
- M3+M4+M5 < (M3+M4, M5) → ME
Disclosure control
- It works!
Disclosure control

- It works!

- 215 chunks of anonymized tables

- 1 big table

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M5</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>M2+M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom4</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M5</td>
</tr>
<tr>
<td>Prodcom</td>
<td>M3+M4+M5</td>
</tr>
<tr>
<td>Prodfra</td>
<td>M2+M3+M4</td>
</tr>
</tbody>
</table>

- 8 disseminated tables
# Disclosure control
- Information loss

<table>
<thead>
<tr>
<th>Classification</th>
<th>Nb hidden cells / Total nb of cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Until 2020</td>
</tr>
<tr>
<td>Prodfra</td>
<td>65 %</td>
</tr>
<tr>
<td>CPF4-5</td>
<td>6 %</td>
</tr>
</tbody>
</table>
Conclusion

Cell suppression method for non-nested classifications
Conclusion

Cell suppression method for non-nested classifications

- It is possible with:
  - No aggregated totals
  - A small disjunction
Conclusion

- Cell suppression method for non-nested classifications

  • It is possible with:
    - No aggregated totals
    - A small disjunction

  • The information loss is:
    - limited for Prodfra
    - quite significant for CPF4-5
Conclusion

Cell suppression method for non-nested classifications

- It is possible with:
  - No aggregated totals
  - A small disjunction

- The information loss is:
  - Limited for Prodfra
  - Quite significant for CPF4-5

- The EAP team will use these results to decide the evolution of the French dissemination
Thanks for your attention!

Any questions?
Join us on:

insee.fr  

Maxime Beauté
maxime.beaute@insee.fr
Maël Buron
mael-luc.buron@insee.fr
Insee, France