"Papaye": new processing software for French PPI - data editing

Ideas borrowed to Edimbus and Canadian questionnaire

Jean-François Baron
Methodologist of PPI division

Insee
Mesurer pour comprendre

Structure of the presentation

A) Borrowing to the Canadian questionnaire (thanks!)

B) European guidance of selective editing with EDIMBUS

C) Rules of imputation for atypical price changes

A) The Canadian questionnaire and its explicit questions on the reasons of price change

Date of last reported price change: 2010-01	A	B	C1 to C4 Transaction description as specified above				Circle reasons for price change							Further explanation of price change (pertinent market information)
Date:	A	B	C1	C2	C3	C4	D							
2008-12							1	2	3	4	5	6	7	
2009-01							1	2	3	4	5	6	7	
2009-02							1	2	3	4	5	6	7	
2009-03							1	2	3	4	5	6	7	
2009-04							1	2	3	4	5	6	7	
2009-05							1	2	3	4	5	6	7	
2009-06							1	2	3	4	5	6	7	
2009-07							1	2	3	4	5	6	7	
2009-08							1	2	3	4	5	6	7	
2009-09							1	2	3	4	5	6	7	
2009-10							1	2	3	4	5	6	7	
2009-11							1	2	3	4	5	6	7	
2009-12							1	2	3	4	5	6	7	

Reasons for price change:
1. Material costs
2. Labour costs
3. Competitive factors
4. Physical content
5. Terms of sale
6. Exchange rate
7. Others - describe

Canadian PPI ask enterprises to describe reasons for price change:
1. Material costs
2. Labour costs
3. Competitive factors
4. Physical content
5. Terms of sale
6. Exchange rate
7. Others - describe

French now apply the same question to price changes higher than 10% in absolute terms.

A) On-line data entry: "Canadian questionnaire" for price changes > 10%

Confirmation of "pure" price change

Reasons for volume change

To analyse

PRIX DU PRODUIT

	Prix validés	Cliquez ici si vous souhaitez les modifier

Il y a de fortes variations de prix. Complétez ou modifiez le tableau si nécessaire.

	MAI 2013	JUIN 2013	JUILLET 2013	AOÛT 2013
Prix	235	236	300	235

Variations de prix pour les raisons suivantes :

		JUILLET 2013	AOÛT 2013
Coût matières premières, intrants		☐	☐
Coût de la main d'oeuvre		☐	☐
Facteurs concurrentiels		☐	☐
Qualité du produit, du service		☑	☑
Conditions de vente		☐	☐
Autres (précisez ci-dessous)		☐	☐

>>> AJOUTER UN COMMENTAIRE SUR CE PRODUIT

mix produits en juillet (exceptionnel)

Papaye software and selective editing

09/10/2013

Insee

A) Treatment of answer "6-others" to Canadian questionnaire (soon included in Papaye screen)

Id of series, month

Comment by the series respondant

Series prices, month by month

Decision by the hierarchy: if nothing is written, that's volume change

B) The European recommendations

1) The European Statistics Code of Practice and quality

- principle 8 of the Code of Practice :

"Appropriate statistical procedures, implemented from data collection to data validation, must underpin quality statistics"

- provide detailed high-quality data with less statistical burden

2) The EDIMBUS (EDiting and IMputation BUSiness surveys) project

-

ttp://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf

- The manual describes 5 different types of likelihood controls:
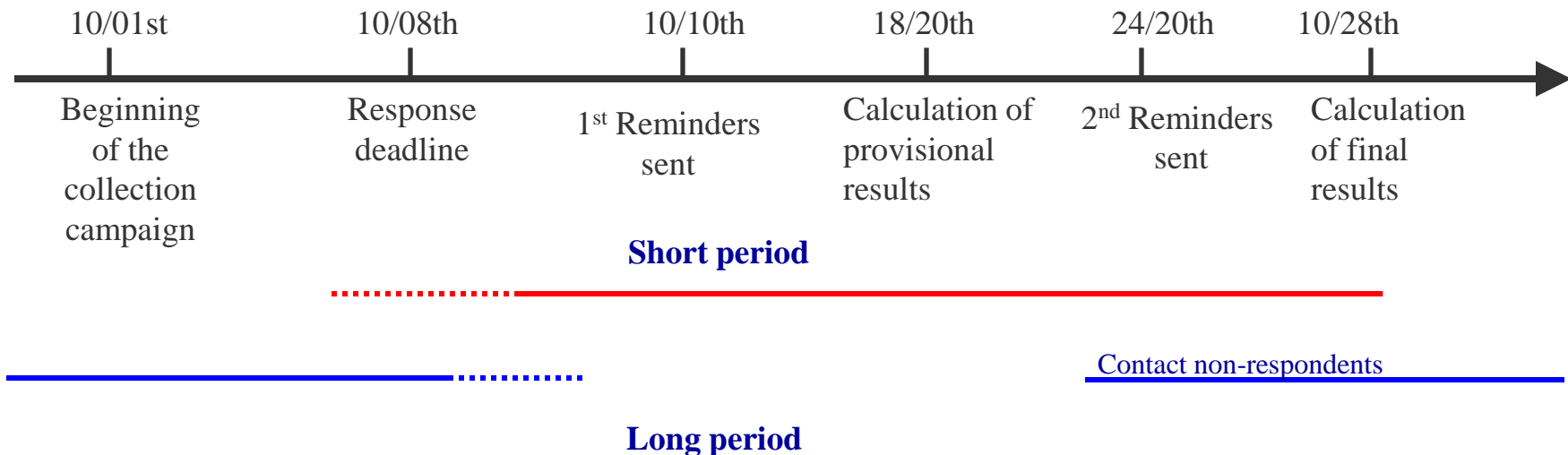
- influential errors;
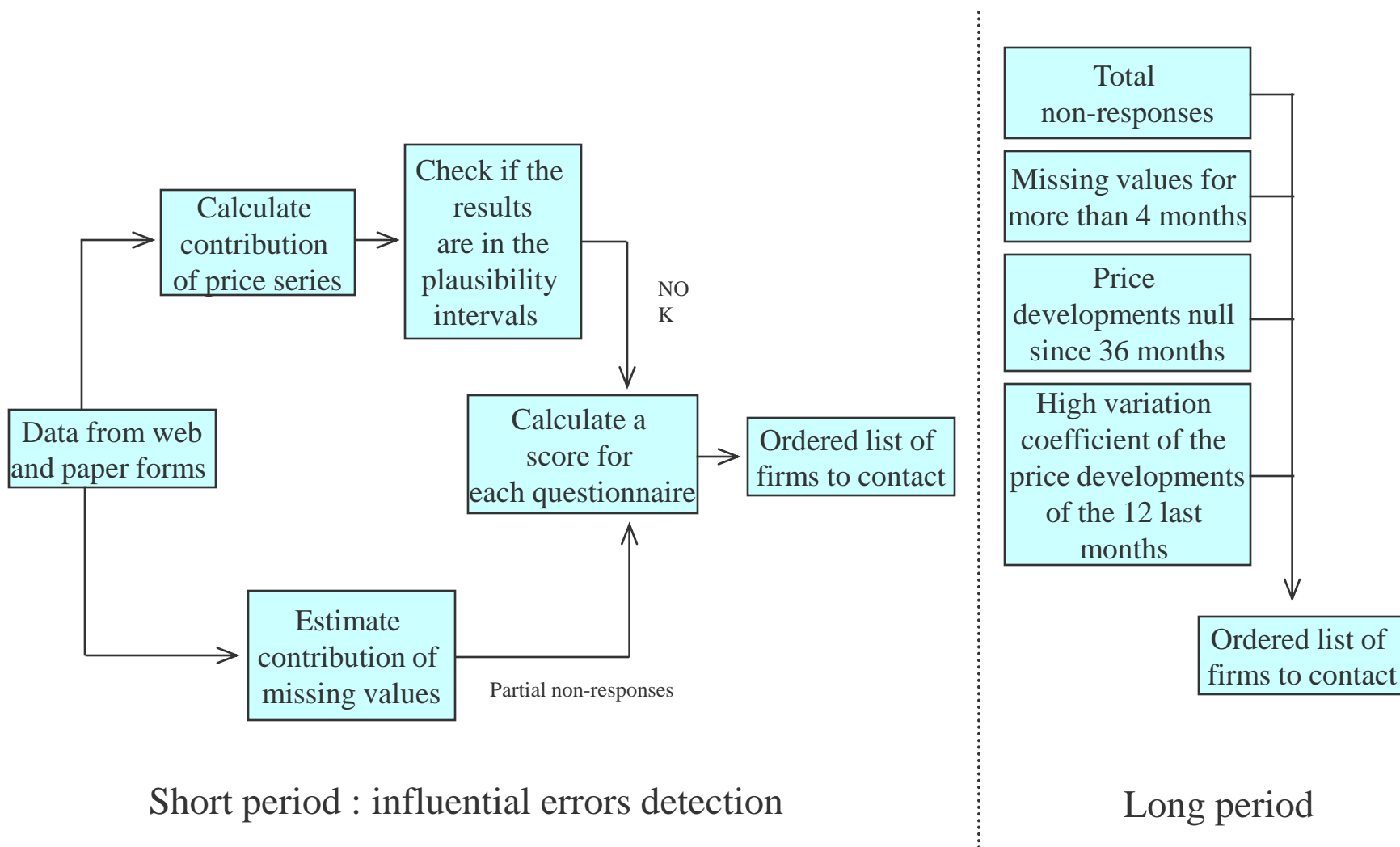- missing values;
- outliers;
- random errors;
- systematic errors.

Papaye software and selective editing

B) Data editing calendar

For instance : our calendar for this month

10/01st	10/08th	10/10th	18/20th	24/20th	10/28th
Beginning of the collection campaign	Response deadline	1st Reminders sent	Calculation of provisional results	2nd Reminders sent	Calculation of final results

Short period

Long period

Contact non-respondents

Insee

B) Macro-process diagram for selective editing

Short period : influential errors detection

Long period

Insee

B) Contribution calculation

Contribution :

$$c_k = w_k \underbrace{\frac{i_k^{t-1}}{I^{t-1}}}_{\text{adjusted weight}} \times \underbrace{\left(\frac{i_k^t}{i_k^{t-1}} - 1 \right)}_{\text{price development}}$$

For missing prices, the price development is estimated by the CPA4 index development.

The formula : the price development of an aggregate is the sum of its components' contributions

$$\frac{I^t}{I^{t-1}} - 1 = \sum_k w_k \frac{i_k^{t-1}}{I^{t-1}} \left(\frac{i_k^t}{i_k^{t-1}} - 1 \right)$$

I : index at CPA 4 digits level
i : elementary index of a price item
w : weight of the elementary index

Papaye software and selective editing
09/10/2013
Insee

B) List of contributions for an aggregate (C0811)

Évolution (%)	1,54	Taux de réponse pondéré (%)	68,29

Synthèse des commentaires

Description: Pierres ornementales ou de construction, calcaire industriel, gypse, craie et ardoise

Modifier la synthèse des commentaires | **Ouvrir la série**

Séries contributives

Séries contributives	Contribution	Évolution	Taux de réponse pondéré	Taux d'expertise	Commentaire
C_0811_SE_1	-0,04	-2,45			
C_0811_SE_2	0,14	4,56			
C_0811_SE_3	-0,03	-1,24			
C_0811_SE_4	0	0			
C_0811_SE_5	-0,02	-2,08			
C_0811_SE_6	0,01	3,58			
C_0811_SE_7	0,01	5,11			
C_0811_SE_8	0	1,64			
C_0811_SE_9	-0,02	-13,01			
C_0811_SE_10	-0,01	-6,23			
C_0811_SE_11	-0	-1,76			
C_0811_SE_12	0,19	2,21			
C_0811_SE_13	0,17	7,41			
C_0811_SE_14	0	0			
C_0811_SE_15	0	0			
C_0811_SE_16	0	0			
C_0811_SE_17	0	0			
C_0811_SE_18	0	0			
C_0811_SE_19	0,05	2,24			
C_0811_SE_20	0,11	16,62			
C_0811_SE_21	0,03	1,22			
C_0811_SE_22	0,01	0,36			
C_0811_SE_23	0	0,11			
C_0811_SE_24	0	0			

Small price development but large weight

Large price development but small weight

B) Selection of (possible) influential errors: calculation of a score

- score calculation based on contribution calculation

Objective : detect product that has a large contribution to the index CPA 4 digits level, in order to draw up a restricted and ordered list of firms to be contacted by data managers.

Influence measure of a price item = contribution to the index at CPA 4 digits level. Keep highest contributions (in absolute value) thanks to a plausibility threshold.

Marked with a red triangle in French software

Then, for each questionnaire, we calculate a score from the results obtained the step before.

$$S_t = \alpha_t \sum_{j \in A_t} \beta_j |g_{t,j}| + \sum_{k=1}^{3} \alpha_{t-k} \sum_{j \in A_{t-k}} \beta_j |g_{t-k,j}|$$

$A_t :$ *outlier values for t*

$\alpha_t :$ *relative weight for t*

$\beta_j :$ *relative weight of the series j price in the CPA 4 digits indicator*

$g_{t,j} :$ *corrected weight* $\times \left(\dfrac{i_k^t}{i_k^{t-1}} - trend \right)$

B) Respondants with highest scores on top of the list

Date of analysis

B) Influential errors?
(case 1 : the price is imputed)

Questionnaire de PA1EV pour la campagne de septembre 2013 - Série réponse SR_PA1EV_006

Informations générales

Description	BTA41-800 REF T87C0G PRODUIT SEMI-FINI	État	Active
Précisions	PRIX DE VENTE MOYEN MENSUEL HORS ...	Évolutions	

Données sur le(s) domaine(s)

Lots suivants

Terms of sale have changed

Valeur

Période	Valeur Brute	Valeur Travail	Pas de transaction	Coeff qualité	Origine		Valider	Valeur Retenue	Commentaire de gestion	Commentaire économique
Mai 2013	102.06		☐		⌄		☐	102.06		
Juin 2013	110.12		☐		⌄		☐	110.12		
Juil. 2013	85.1	85.1	☐	1.246	G ⌄	⚠	☑	85.1		Les conditions de ven
Août 2013	85.1		☐		⌄		☐	85.1		

Informations complémentaires : ⊙ Correspondant(s) : ⚲   [Ajouter une période]

Expertise et approfondissement

☐ Expertise

☐ Série plate
☐ Série en non-réponse répétée
☐ Série aux fluctuations atypiques

Commentaire du temps long

Lots suivants

[Enregistrer] [Effacer]

Quality adjustment

B) Influential errors?
(case 2 : the price is validated)

B) Detection and correction of a random error

B) Systematic errors

Price mistake the months before

Papaye software and selective editing **09/10/2013**

B) Treatment of a "products mix" for one month

Papaye software and selective editing   **09/10/2013**

Insee

B) Treatment of outliers

Compared to a plausibility threshold

Marked with a yellow triangle in French software Papaye

Papaye software and selective editing

09/10/2013

C) Rules of decision to accept or not the response given by the enterprise

	Series price development > 10%	Series price development < 10%
Influential questionnaire (= to analyse by data manager)	The questionnaire is in the list the data manager must analyse Reason of price change: - material costs, labour costs or competitive factors. The value is adjusted by software unless the variation is validated by data manager. Generally the data manager accepts. -Physical content or terms of sale. The value is corrected with a quality coefficient - Others – describe. The value is adjusted by software unless the "pure" price change is validated by data manager, according to the commentary of the enterprise.	The questionnaire is in the list that the data manager must analyse, but the enterprise does not have to answer to the Canadian questionnaire. If the price development is usual or the data manager knows the reason of the price development, he validates the value, otherwise he contacts the enterprise.
Non influential questionnaire	The questionnaire is **not** in the list, we trust in the enterprise answer. Reason of price change: - material costs, labour costs or competitive factors. The value is accepted by software. - Physical content or terms of sale. The value is corrected by software with a quality coefficient - Others – describe. The value is adjusted by software unless the data manager has a specific action.	The questionnaire is **not** in the list, the enterprise does not have to answer to the Canadian questionnaire.

C) Imputation

Missing value : estimated by software. Choice between 3 methods:

- imputation with stratum average price development

- imputation with another index (price series or aggregate)

- carrying forward of the last available price reported

By default: imputation with stratum average price development

Insee

"Papaye": new processing software for French PPI - data editing

Thank you for your attention!

Contact
M. Jean-François BARON
Tél. : 01 41 17 50 90
Courriel : jean-francois.baron@insee.fr

Insee