

Diffusion ESANE 2008 phase 1

Cnis 11 octobre 2010

J. M. Béguin



Diffusion Ésane : la situation à l'automne 2010

› Sur 2008 :

- Les premières publications, sur Insee.fr, de l'EAP 2008 ont eu lieu en *septembre 2009* et celles d'Ésane 2008 le *30/7/2010*.
- On est à la fois en retard et incomplet par rapport aux ambitions initiales :
 - EAP : uniquement les facturations
 - Ésane : 9 variables au niveau "groupe" (3 chiffres de la NAF cad 243 lignes du champ Ésane sur 272 groupes existants)

› Sur 2009 :

- On espère bien rattraper le retard en termes de champ
 - Ésane : 5 chiffres de la NAF (+ 3 chiffres x tailles)
- presque en termes de délais
 - EAP : la production est disponible depuis le 7 octobre 2010 (en facturations et en quantités physiques)
 - Ésane : fin premier trimestre 2011 sur toutes les variables



Objectif de l'exposé : quoi et comment

- › Comprendre quelles sont les données disponibles et comment elles sont désormais fabriquées
- › Comprendre en quoi elles diffèrent de ce qui était disponible précédemment (jusqu'en 2007)
- › Distinguer :
 - ce qui est consubstantiel à Ésane (phase 1)
 - ce qui est dû au retard sur 2008 qui a conduit à des décisions spécifiques sur cette année de démarrage

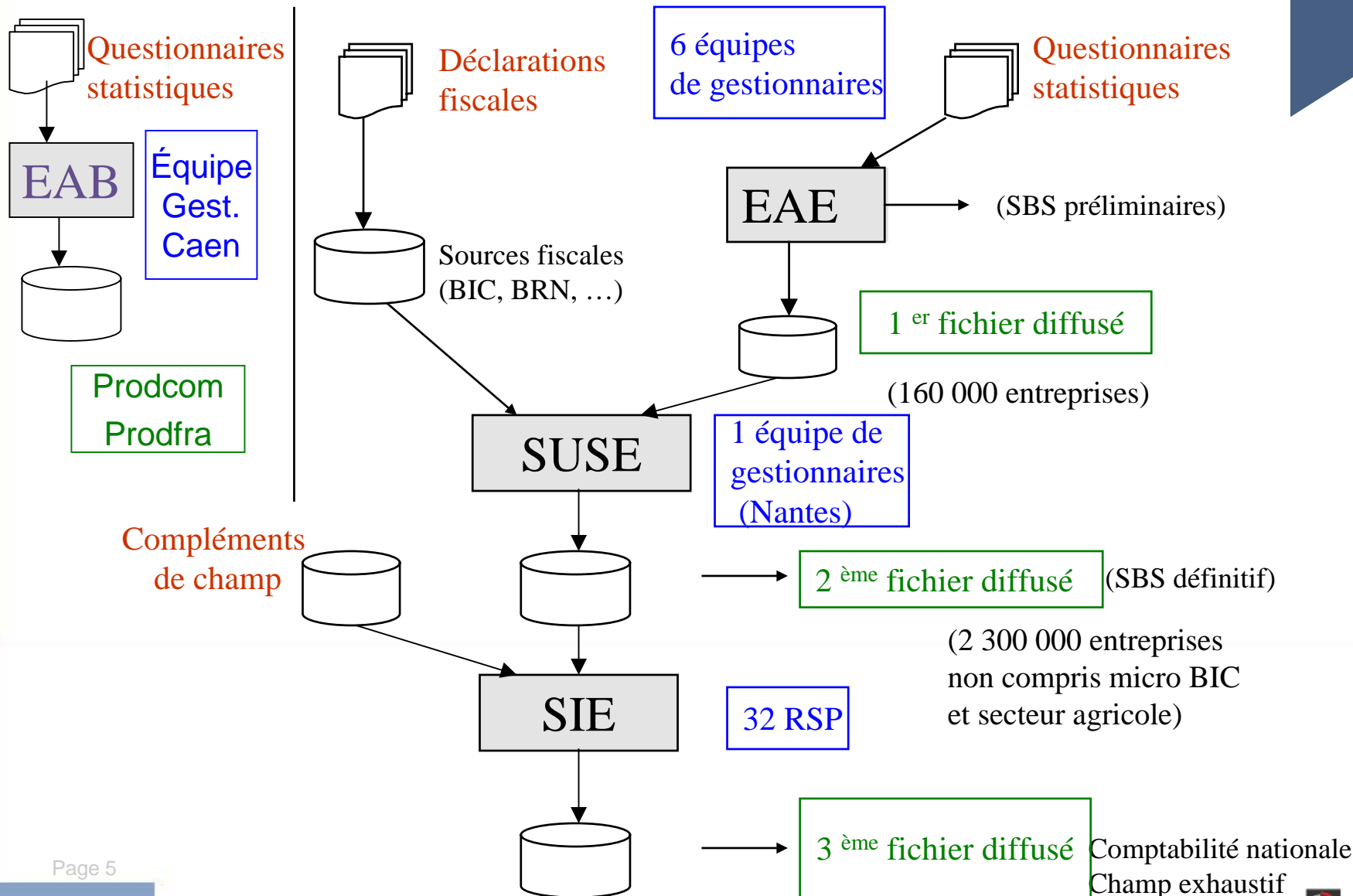


Plan de l'exposé

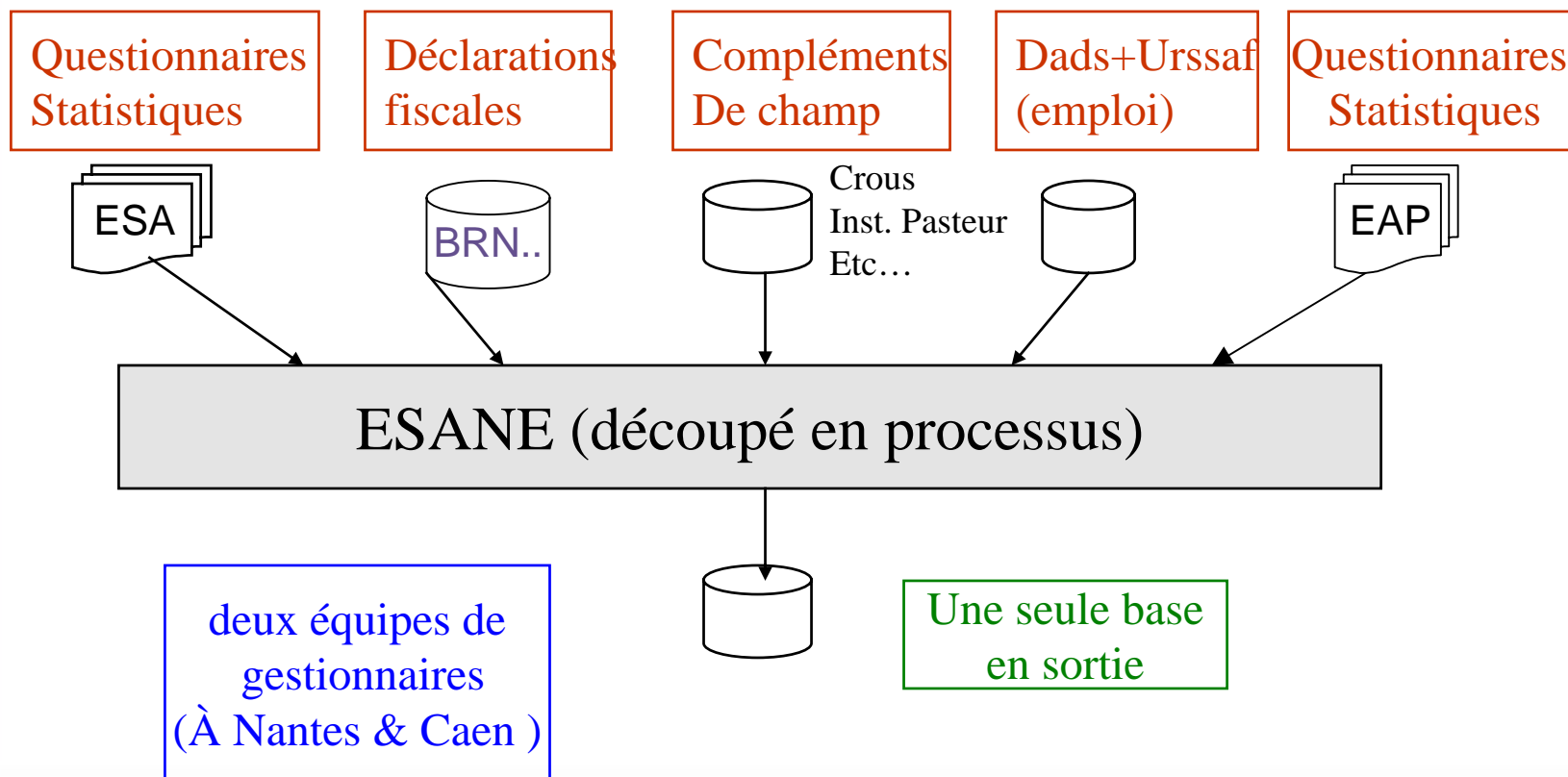
- ① L'architecture générale avant / après
- ② La structure des données (le « matériau » utilisé pour la production des résultats dans ESANE)
- ③ Les 15 principales innovations du système
- ④ Zoom sur 3 (champ / REDI / estimateurs composites)
- ⑤ Impact de ces innovations sur les résultats
- ⑥ La spécificité des traitements réalisés sur 2008
- ⑦ Les premiers résultats 2008
- ⑧ Les premiers résultats 2009



1 Le système précédent : architecture



① Le système Ésane : architecture



Là où on avait des systèmes indépendants (EAE, SUSE, SIE), on a maintenant des processus correspondants au traitement de chaque source (indépendamment les unes des autres) puis un processus spécifique final REDI : **réconciliation de données individuelles**.



② La structure des données

EAP ou ESA
(ventilation du CA)

Liasses fiscales

(Chiffres 2008,
hors DOM)

VA = 12 milliards

EAP ou ESA (ventilation du CA)		Liasses fiscales
CA ₁ , CA ₂ , ..., CA _n S ₁ , ..., S _p		CA _{liasse} , Var _{liasse} , ..., APE
	poids	
i	1	A 1
i	1	B 1
	30	A 1
j	40	C 1
	25	A 1
		Champ (ESA+EAP) (2 375 000) VA = 921 milliards
		Champ non échantillonné (569 000) VA = 90 milliards

exhaustif

(77 000)

VA = 641 milliards

Partie sondée

(79 000)

La forme des données est très particulière.
On ne dispose pas des mêmes données sur
toutes les unités. Ceci va avoir des consé-
quences sur les estimateurs.

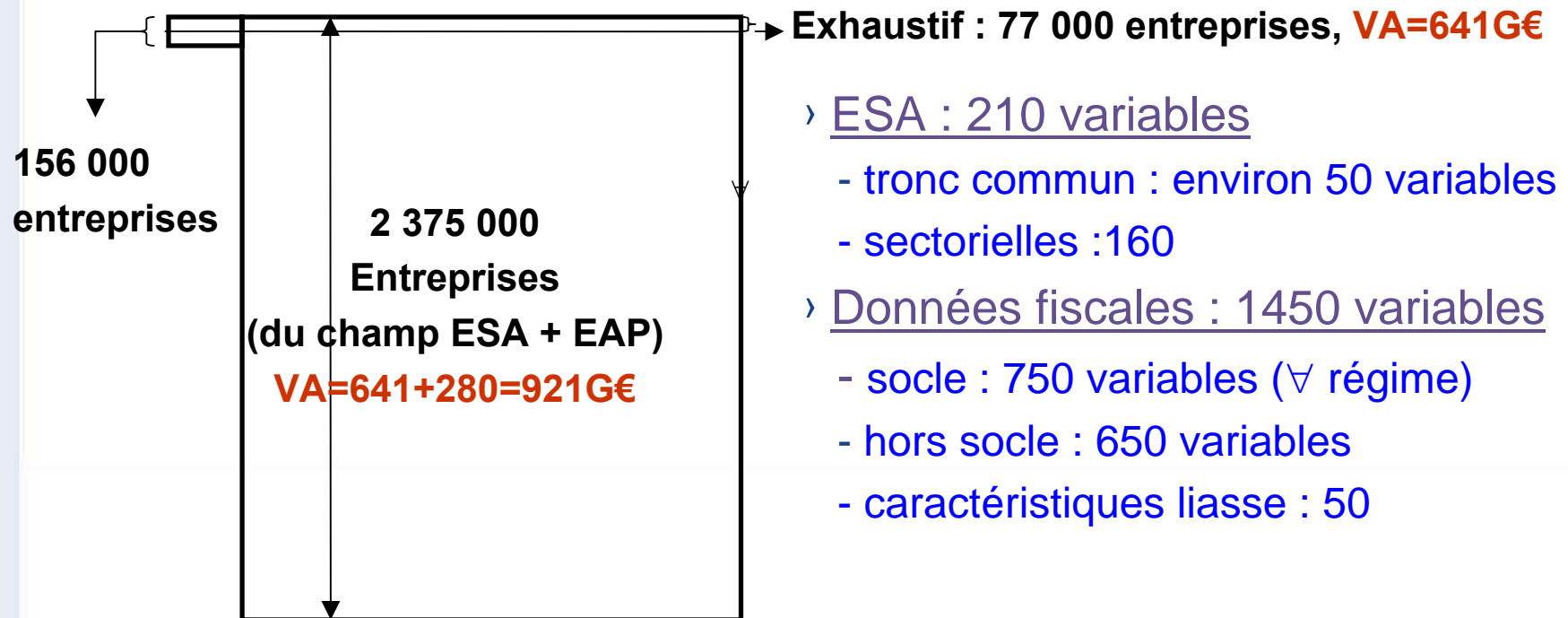
Le fait de prendre en compte les parties
sondées des enquêtes est **nouveau et**
complexe.

CA₁, ..., CA_n:
Ventilation du CA en
branches
(ventes biens, services
Marchandises, etc.)
S_i : autres variables
sectorielles (dt APE)



② Le volume et le poids des données dans la base de production

Échantillon – exhaustif – liasses à l'échelle : structure des données en Γ



Au total près de 4 milliards de données chaque année !

③ Une quinzaine d'innovations fortes (1/2)






- › Dans l'architecture des données
 - Utilisation de nouvelles sources administratives (emploi) *
 - Modification des questionnaires (ESA et EAP) *
 - Introduction des DOM dans le système (114 000 dont 9 500)
- › Dans le poste de travail des gestionnaires
 - Refonte complète du poste de travail gestionnaire
 - Aide à la codification des branches déclarées nouvelles par les entreprises dans l'ESA*
 - Organisation des contrôles (Traitement en priorité des macro-contrôles avant les micro contrôles)
 - Nouveau traitement des restructurations juridiques (généralisation de ce qui était fait au Sessi dans l'EAE industrie)



③ Une quinzaine d'innovations fortes (2/2)

› Innovations systémiques induisant des modifications dans les résultats

- Définition du champ du système a priori * 
- Validation des agrégats « branche »
- REDI (sur CA, puis emploi et données douanières)* 
- Nouveaux principes de redressements :
 - Calage du compte de résultat sur la ventilation du CA des enquêtes (primauté à l'enquête) (P1*)
 - Prise en compte des branches des « petites » (P2*)
 - Affectation des exercices décalés sur l'année où il y a plus de mois (Avant : année de clôture) (P3*)
 - Recalcul des exercices comptables sur 12 mois (P4*)
- Utilisation d'estimateurs « composites » (P5*) 

Plus le changement de nomenclature.....(2008)



③ Conséquences sur les données

- › Tout changement peut avoir un impact sur les données en sortie
- › Mais toutes les innovations n'ont pas le même impact sur les données
- › Les innovations purement "ergonomiques" en ont peu
- › Les modifications sur les "traitements" peuvent en avoir beaucoup
- › Les principales ayant des conséquences sur les données ont été indiquées par * dans la liste précédente

➔ Tout cela génère des ruptures de séries fortes

4 Zoom (1/3) ; champ : changement du mode de définition

- › Le champ est défini a priori à partir du répertoire SIRENE (en tenant compte des valeurs de divers codes de gestion du répertoire). Les unités statistiques restent des sociétés. (Unités légales)
- › Correspond aux **SNF-EI du secteur marchand participant au secteur productif, hors secteur financier et hors agriculture**
- › Des entreprises du champ sans liasse fiscale nombreuses (830 000) [on estime que 400 000 sont cessées et 430 000 vivantes pour lesquelles on impute une liasse]
- › Des liasses fiscales reçues pour des unités hors champ (600 000) [dont 250 000 agriculture, etc.... Le problème c'est : « n'est-ce pas une liasse écartée à tort ? »]



des difficultés d'imputation

Page 12



④ Zoom (2/3) : REDI sur le chiffre d'affaires

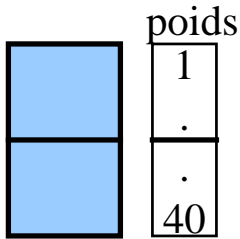
- › L'idée de départ : comparer l'information individuelle obtenue dans l'enquête (ESA ou EAP) et celle de la source fiscale concernant le chiffre d'affaires
- › Dans les cas de grandes divergences, opérer un contrôle et un redressement manuel qui peut modifier les données de la source fiscale (réelle ou imputée)
- › Pour le reste, opérer un traitement automatique
 - On va garder la ventilation du CA de l'enquête
 - Mais on va l'appliquer au CA de la liasse
- › Dans les deux cas, cela peut influencer sur le calcul de l'APE et donc nécessite son re-calcul
- › Les choses se compliquent quand l'une des 2 données est absente et donc extrapolée ou imputée....



④ Zoom (3/3) : estimateurs composites

Soit à estimer le CA (ou la VA...) du secteur S (échantillonné) :

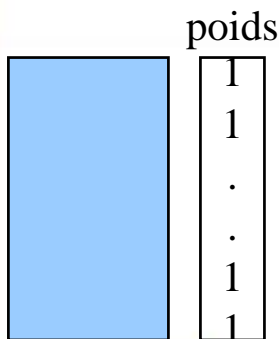
› Cas des ESA-EAP seuls (ou de l'EAE)



CA = somme pondérée des chiffres d'affaires des entreprises du secteur S

▪ **Nota : l'EAE sert entre autres à modifier l'APE. Il faut donc prendre l'APE résultant de l'observation du CA dans l'enquête.**

› Cas de Esane non échantillonné (ou de SUSE) (ex FICUS)



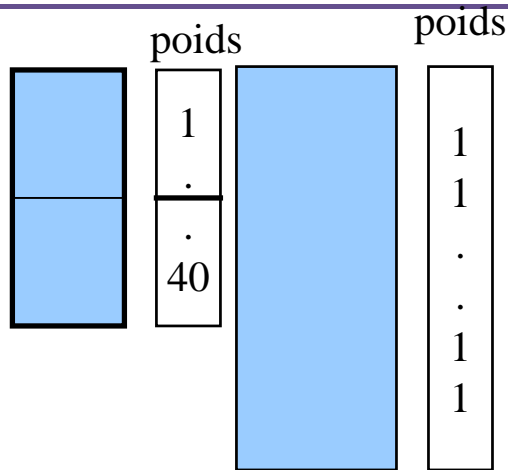
CA = somme non pondérée des chiffres d'affaires des entreprises du secteur S (source exhaustive)

▪ **Nota : au mieux on aura reporté l'APE résultant de l'observation du CA dans l'enquête des entreprises de l'échantillon EAE**



④ Zoom (3/3 suite) : estimateurs composites

› Cas de Esane complet



On a intérêt à se servir de la totalité de l'information dont on dispose grâce à l'information apportée par l'échantillon :

- entreprises "entrant" dans le secteur S (post REDI)
- entreprises "sortant" du secteur S (post REDI)
- Modifications apportées à la valeur du chiffre d'affaire fiscal par l'étape REDI (dilatation)

› la formule exacte (compliquée) est :

$$\sum_{i \in U} Y_i^{\text{IEG}} \mathbb{I}_{\text{APErep}=X}(i) + \sum_{i \in S} w_i (Y_i^{\text{final}} \mathbb{I}_{\text{active}}(i) \mathbb{I}_{\text{APEenq}=X}(i) - Y_i^{\text{IEG}} \mathbb{I}_{\text{APErep}=X}(i))$$



④ Zoom (3/3 suite) : estimateurs composites *explication* : 1) Apport de l'enquête

Avant contrôle ESA-Fisc :

	CA Fisc	APE Rép.	W_i
Soc. 1	20	A	45
Soc. 2	30	A	25
Soc. 3	30	B	20
Soc. 4	50	A	10

$$\begin{aligned}(CA)_A &= 900 + 750 + 0 + 500 \\ &= 2150\end{aligned}$$

Après REDI :

	CA Redi	APE Enq.	W_i
Soc. 1	20	A	45
Soc. 2	40	A	25
Soc. 3	30	A	20
Soc. 4	50	C	10

$$\begin{aligned}(CA)_A &= 900 + \mathbf{1000} + \mathbf{600} + \mathbf{0} \\ &= 2500\end{aligned}$$

On a une correction de 350 à appliquer au total du secteur



④ Zoom (3/3 suite) : estimateurs composites *explication* : 1) Apport de l'enquête

› Explicitation

- Soc 1 : cas le plus général et le plus fréquent ; ni son CA ni son code APE ne sont modifiés par les traitements des gestionnaires
- Soc 2 : le code APE de cette société n'est pas modifié mais par contre son CA total a été revu après la phase de confrontation entre les résultats individuels de l'ESA et de ses données fiscales (il passe de 30 à 40)
- Soc 3 : code APE modifié ; cette société « rentre » dans le secteur A.
- Soc 4 : code APE modifié ; cette société « sort » du secteur A.



④ Zoom (3/3 suite) : estimateurs composites *explication* : 2) Apport de la source fiscale

	CA Fisc	APE Rép.	CA Redi	APE Enq.	W_i
Soc. 1	20	A	20	A	1
Soc. 2	30	A	40	A	1
Soc. 3	30	B	30	A	1
Soc. 4	50	A	50	C	1
Soc. 5	30	A	30	A	1
Soc. 6	60	A	60	A	1
.	.	A	.	A	1
Soc. k	70	A	70	A	1
Total	X		X-10		k

- › On peut injecter dans la source fiscale les modifications intervenues sur l'échantillon
- › Mais toutes les sociétés ont un poids de 1.
- › On voit que la modification induite dans le total par l'injection n'est que de **-10** !
- › **C'est ce qu'on faisait avant dans FICUS**



④ Zoom (3/3 suite) estimateurs composites : *explication* : 3) Apport des deux (Ésane)

- › On voit que la bonne solution c'est de combiner les deux
- › Le bon estimateur sera donc **$X + 350$** . Le « 350 » prend en compte la totalité de l'information apportée par le travail sur échantillon
- › C'est exactement ce qu'exprime la formule ci-dessus
- › Si la variable et l'ensemble sur lequel on calcule un agrégat ont servi à caler les poids (ex : CA et groupe de la NAF 3 chiffres), alors $X=2150$ et $(CA)_{\text{secteur}} = 2500$

C'est intuitif mais ça complique quand même pas mal les estimations à la demande....



⑤ Impact de ces innovations sur les résultats

- › Le changement de nomenclature (et des questionnaires)
 - modification des contours yc dans les grands regroupements (industrie, commerce,...)
 - modification de la règle des donneurs d'ordre
 - 32,5 milliards € déplacés de la « production vendue de biens » vers les « ventes de marchandises » (Industrie -> Commerce) dans l'automobile
 - 30 milliards € de CA déplacés de l'industrie vers le commerce suite à des reclassements d'entreprises dont l'APE était erronée



⑤ Impact des innovations sur les résultats (suite)

› Les nouveaux principes de redressement

- des impacts qui se compensent souvent, dont le sens ne peut donc être anticipé et qui peuvent être forts sectoriellement
 - effet globalement faible
 - un même niveau G (114 – compta. nat.) peut varier au max de qq G€, soit en CA soit même en VA (cas de Airbus)
 - joue à la fois en branche et en secteur

› Les indicateurs composites

- Corrigent le nombre d'entreprises de 40 000 p/r enquêtes p ex
- Corrigent le CA total de 50 G€ environ (p/r enquêtes tj)



⑥ La spécificité des traitements réalisés sur 2008

On a privilégié un client : la comptabilité nationale

- › Les données n'ont été « contrôlées » par les gestionnaires qu'au niveau « groupe » (i.e. 3 caractères de la NAF)
 - soit 272 au lieu de 732 niveaux
 - aucun contrôle par taille d'unité légale
 - › Seules les variables comptables utiles à la comptabilité nationale ont été contrôlées ⇒ peu de variables disponibles
 - › Le manque de temps pour les gestionnaires a nécessité d'automatiser certains redressements ; il a fallu beaucoup de temps pour stabiliser certains algorithmes
 - › Aucune donnée « sectorielle » n'a été contrôlée
- ➔ ce qui n'a pas été contrôlé ne peut être publié**



⑦ Les premiers résultats 2008

1. Sur Insee.fr

- Comme on l'a vu, depuis le 30 juillet 2010, à l'adresse http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=esane

on a publié pour 243 groupes de la NAF, 9 variables (tableau):

Nombre d'entreprises

Effectif salarié en équivalent temps plein

Effectif salarié au 31 /12

Chiffre d'affaires HT

Valeur ajoutée

Frais de personnel

Excédent brut d'exploitation

Résultat courant avant impôt

Bénéfice ou perte

Une présentation sous forme de fiches thématiques sera disponible



Les premiers résultats 2008 (suite)

2. Sur Alisse

- Un « Alisse Rev. 2 » va ouvrir en octobre
- Pour les mêmes 243 groupes, on publiera a priori en novembre une dizaine de variables supplémentaires dont la liste est en cours de fixation:
 - **Au moins les CA branches**
 - **Plusieurs concepts de VA**
 - **Taux de marge**
 - **Pas de CA à l'export**

3. Dans les rapports sur les comptes (ou CCS?)

- Aux niveau G et quelquefois H figureront des données expertisées par les comptables nationaux (SGDG bien entendu)



⑧ Les premiers résultats 2009

1. EAP (i.e. Prodcom ou prodfra)

- Les résultats sont sur le site Insee.fr depuis le 7 octobre
 - Thème « industrie » ; données détaillées...;
- Contiennent les facturations et les données en quantités physiques auxquelles les professionnels étaient habitués et attachés

2. Ésane

- On espère pouvoir publier plus tôt : avril 2011 sur Alisse a priori aux niveaux prévus initialement
 - contrôle des données à un niveau plus fin (5 caractères de la NAF)
 - Faire un vrai contrôle REDI-CA (sur 3 semaines)
 - Faire le contrôle REDI-emploi
 - Traiter les parties « sectorielles »