

Journées de méthodologie statistique

**Champ et hors champ : impact de la
démographie des entreprises.
Cas de l'enquête Acemo-TPE de la DARES**

Ludovic VINCENT
DARES

Département salaires et conventions salariales

Le cadre de cette étude

- **Constater** le hors champ dans l'enquête Acemo sur les très petites entreprises (Acemo-TPE)
- Analyser l'**impact théorique** du hors champ dans les enquêtes
- Proposer une **méthode** générale de traitement du hors champ dans les enquêtes
- Appliquer la méthode à l'enquête **ACEMO-TPE**

Plan de la présentation

I. Le hors champ dans les enquêtes

- Qu'est-ce que le hors champ ?
- D'où vient le hors champ ?
- Quel est le problème ?

II. Le traitement du hors champ

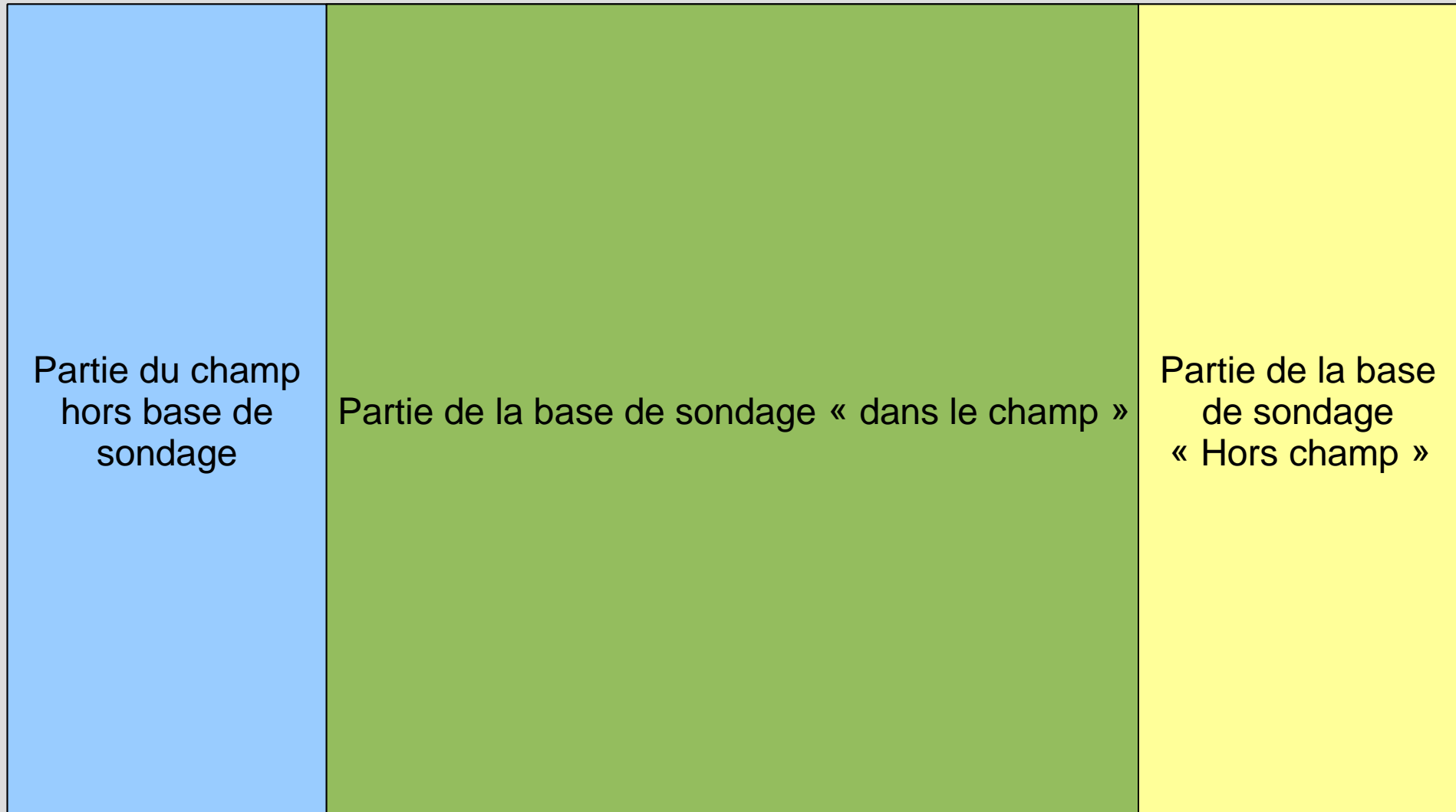
- Le hors champ dans d'autres enquêtes
- Les étapes pour traiter le hors champ
- Application à l'enquête Acemo-TPE 2013

III. Et si tous les non retours étaient indécis ?

Qu'est-ce que le hors champ ? (1/2)

- Tout ce qui n'est pas dans le champ...
- Ainsi, qu'est ce que le champ d'une enquête ?
 - Ensemble des unités statistiques concernées par l'enquête
 - Exemples :
 - *Tous les logements pour le recensement de la population*
 - Avoir une définition précise d'un logement
 - *Les entreprises de 10 salariés ou plus :*
 - Avoir une définition précise du nombre de salariés

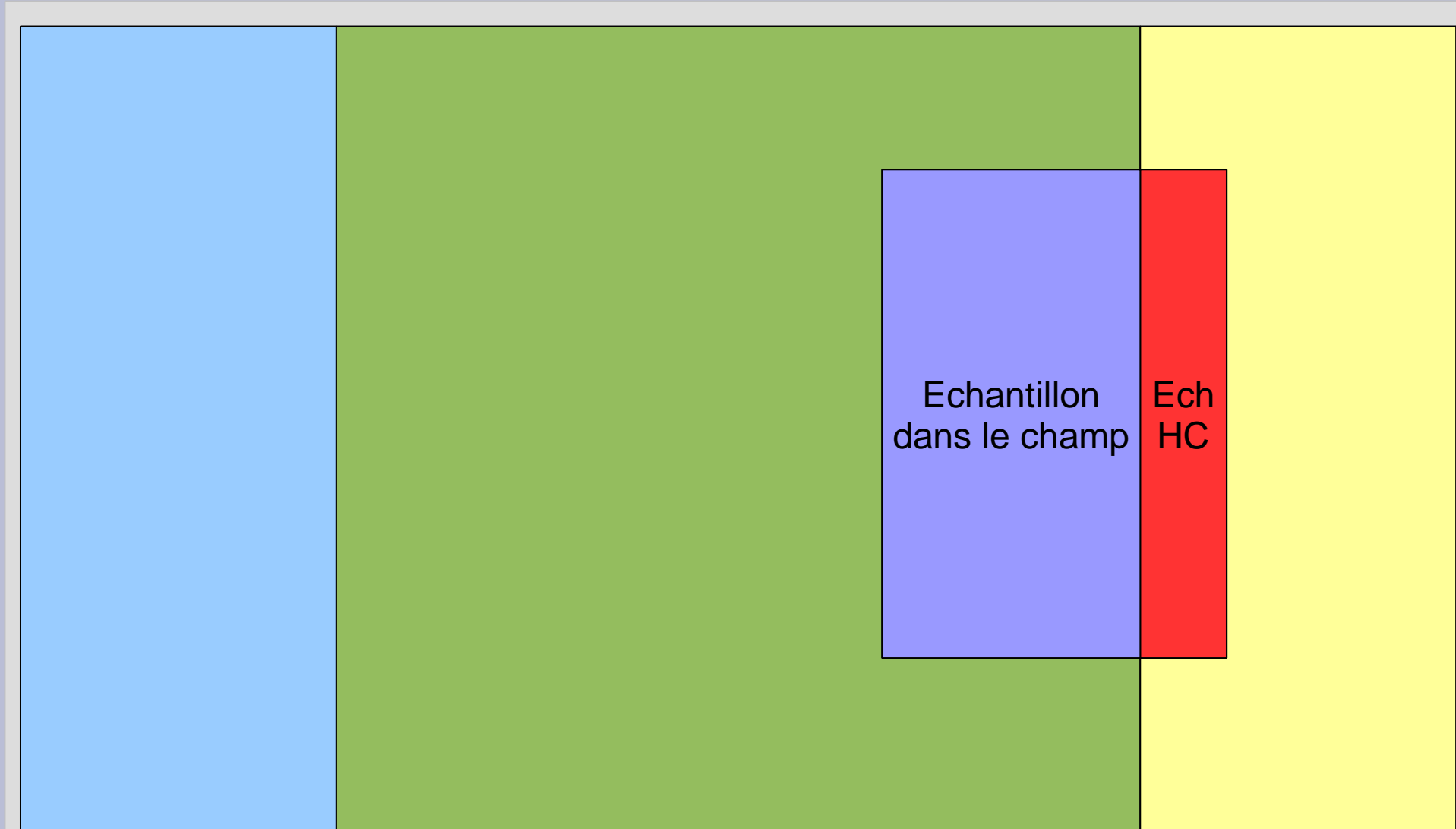
Qu'est-ce que le hors champ ? (2/2)



D'où vient le hors champ ? (1/1)

- Des imperfections dans les bases de sondage
 - Une forte volatilité dans le temps pour les valeurs des variables définissant le champ
 - Délai de mise à jour des bases de sondage
- Des problèmes de dates
 - Des bases de sondage antérieures aux dates de référence
 - Une collecte éloignée des dates de référence
 - Une collecte éloignée de la date de constitution de la base de sondage

Quel est le problème ? (1/2)



Quel est le problème ? (2/2)

- Des **valeurs** de variables d'intérêt potentiellement **différentes** pour les unités statistiques de part et d'autre de la frontière
- Un échantillon qui **biaise** les résultats :

$$\hat{Y}_{HT} = \sum_{i \in S} \frac{1}{\pi_i} * y_i = \sum_{i' \in S_C} \frac{1}{\pi_{i'}} y_{i'} + \sum_{i'' \in S_{HC}} \frac{1}{\pi_{i''}} y_{i''} = Y_C + Y_{CH} \quad (6)$$

$$Biais(\hat{Y}_{HT}) = E(\hat{Y}_{HT}) - Y_C = \sum_{k' \in U_C} \frac{1}{\pi_{k'}} y_{k'} * E(I_{k'}) + \sum_{k'' \in U_{HC}} \frac{1}{\pi_{k''}} y_{k''} * E(I_{k''}) - Y_C$$

$$Biais(\hat{Y}_{HT}) = Y_{HC} \quad (7)$$

Le hors champ dans d'autres enquêtes (1/2)

- L'Enquête Sectorielle Annuelle (ESA)
 - Sont présumées cessées – ou hors champ – les unités :
 - *non affiliées aux régimes d'imposition micro,*
 - *et sans liasse fiscale pendant trois années successives*
 - *et n'ayant effectué aucune déclaration de TVA l'année d'intérêt de l'enquête.*

Le hors champ dans d'autres enquêtes (2/2)

- L'enquête technologies de l'information et de la communication auprès des entreprises (TIC) et de commerce électronique
 - Sont considérées hors champ
 - *les unités déclarant un effectif inférieur à 10 personnes ou une activité qui aurait changé et ne serait plus dans le champ,*
 - *après la collecte, les unités cessées repérées dans les fichiers Sirius et de déclaration de TVA.*
 - 15% indécises et 56% hors champ
- L'enquête Innovation (2001)
 - Utilisation de source externe
 - Repondération des répondants

Les étapes pour traiter le hors champ (1/5)

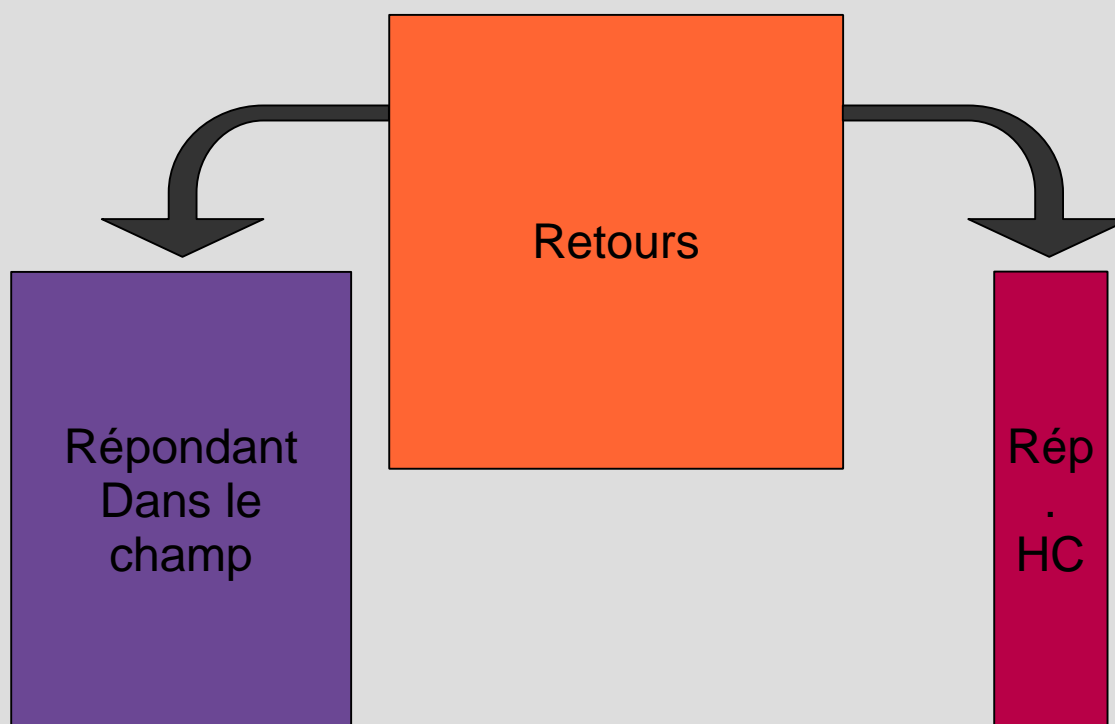
- Penser au hors champ dès la constitution de la base de sondage
- Apurer les questionnaires retournés au regard du hors champ
 - **Scinder les retours** en répondants du champ et répondants hors champ plutôt qu'en non répondants
- Déterminer le **statut des non retours** entre non répondants, hors champ et indécis
- Une approche originale : **le traitement les indécis**

Les étapes pour traiter le hors champ (2/5)

- Constituer précisément la base de sondage
 - Définition précise des contours du champ de l'enquête
 - *Date des effectifs, âge des interrogés...*
 - Fiabilité des informations permettant de définir le champ
 - *Variable peu remplie, ou méconnue (conventions collectives dans les DADS...)*
 - Base la plus à jour possible, par rapport aux dates de l'enquête (date de référence, collecte...)

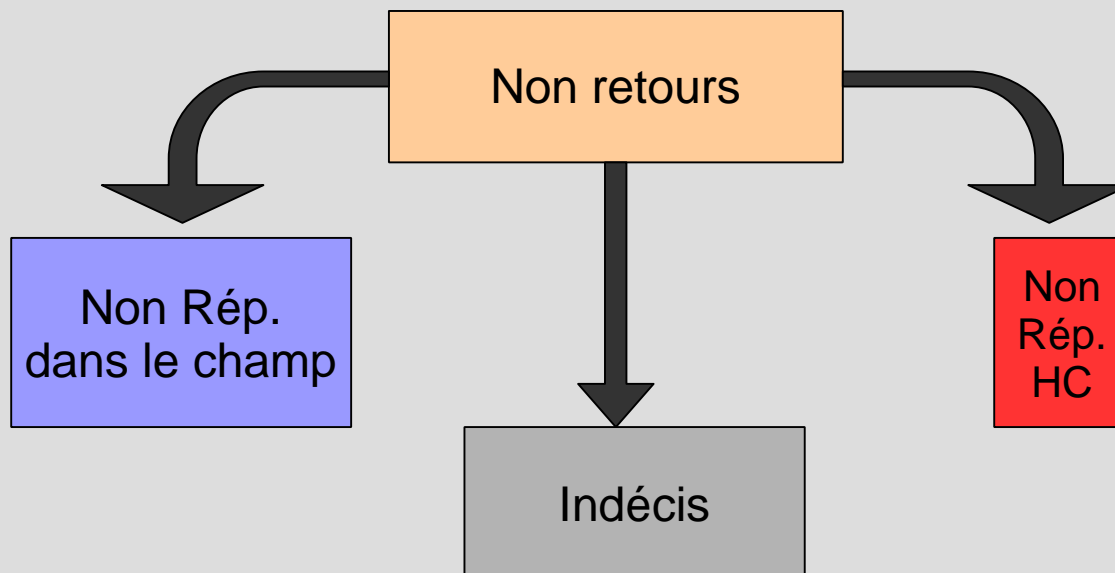
Les étapes pour traiter le hors champ (3/5)

- Apurer les questionnaires retournés au regard du hors champ
 - Statuer sur les questionnaires retournés
 - *grâce aux réponses aux questionnaires,*
 - *grâce à des informations auxiliaires*



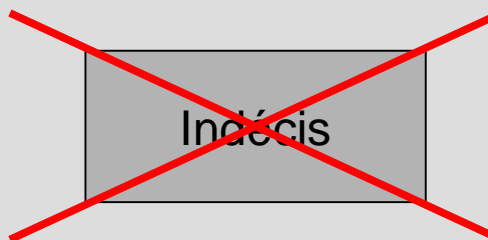
Les étapes pour traiter le hors champ (4/5)

- Déterminer le statut des questionnaires non retournés
 - Scinder les questionnaires en non répondants du champ et non répondants hors champ
 - Il peut rester une partie de questionnaires « **indécis** »



Les étapes pour traiter le hors champ (5/5)

- Une approche originale : le traitement des indécis



Méthode 1 : les indécis sont hors champ

Non Rép. Dans le champ	Non Rép. HC
------------------------------	-------------------

Méthode 2 : les indécis sont dans le champ

Non Rép. Dans le champ	Non Rép. · HC
---------------------------	------------------------

Méthode 3 : On répondère les « non indécis »

Non Rép. Dans le champ	Non Rép. HC
---------------------------	-------------------

Méthode 4 : On « prédit » les indécis

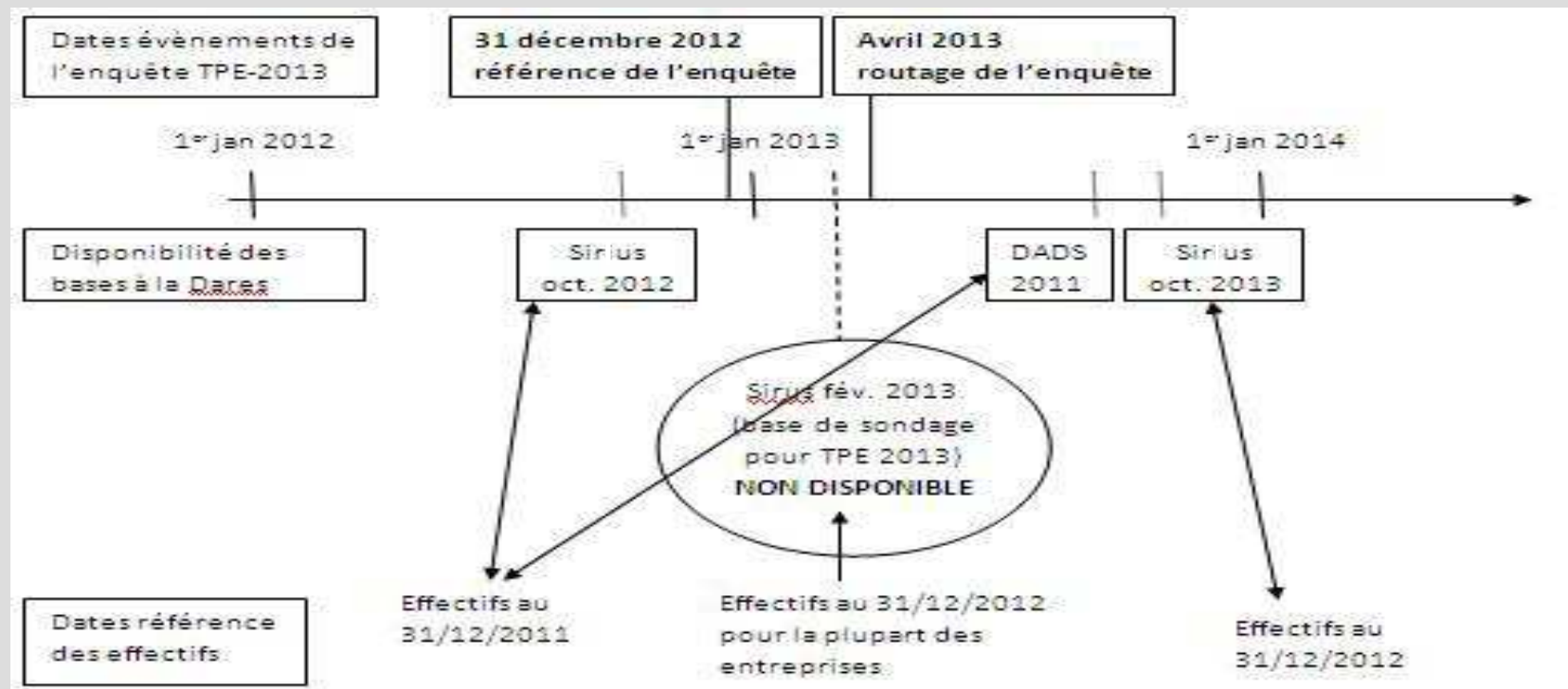
Non Rép. Dans le champ	Non Rép. Dans le champ	Non Rép. · HC	Non Rép. · HC
------------------------------	------------------------------	------------------------	------------------------

Application à TPE-2013 (1/6)

- **Présentation de l'enquête**
 - Une des 4 enquêtes Acemo (Activité et Condition d'Emploi de la Main d'Oeuvre) de la Dares
 - Entreprises de 1 à 9 salariés
 - Entreprises de Métropole
 - Entreprises du Secteur Marchand Non-Agricole*
- **Objectif de l'enquête**
 - Obtenir de l'information sur les salaires et les salariés des entreprises concernées
 - Permettre notamment de connaître chaque année les bénéficiaires de la revalorisation du Smic

Application à TPE-2013 (2/6)

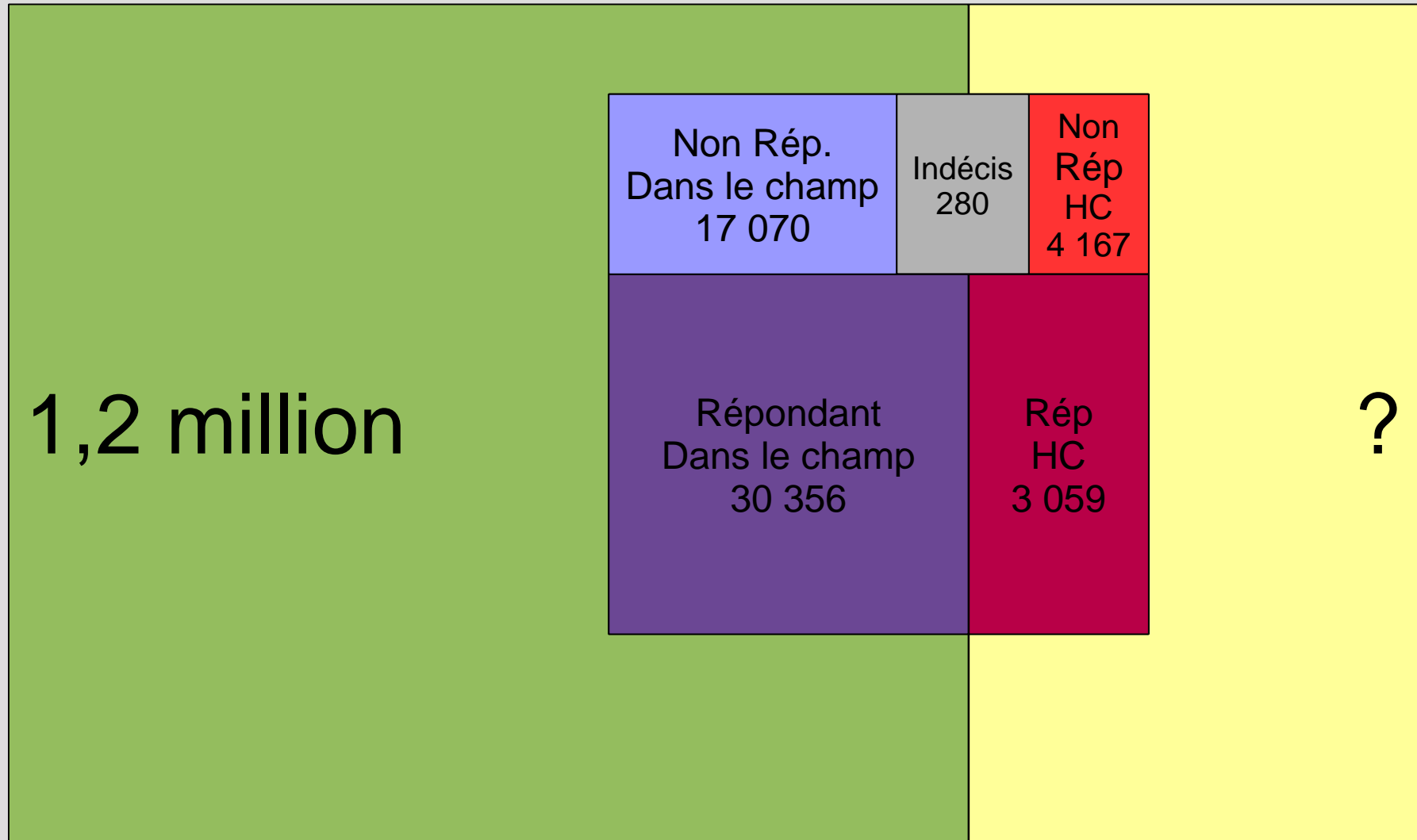
- Constituer précisément la base de sondage
 - Un champ défini précisément, et plutôt fiable :
 - *Unités légales de 1 à 9 salariés*
 - *Restriction sur quelques codes APE et catégories juridiques*
 - Une base datant de février 2013 pour des résultats en référence au 31 décembre 2012



Application à TPE-2013 (3/6)

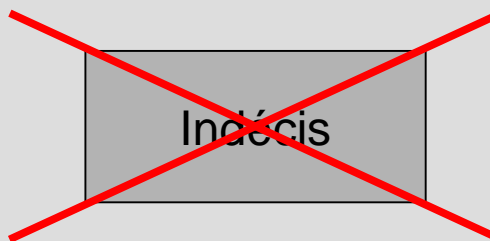
- Apurer les questionnaires retournés au regard du hors champ
 - du fait de cessations
 - à l'aide des réponses sur le nombre de salariés
 - à l'aide des modifications de code APE (marginal)
- Déterminer le statut des questionnaires non retournés
 - Sont considérées hors champ :
 - *Les unités légales présentes dans Sirius en 2012 et pas en 2013*
 - *Les unités légales ayant un effectif nul ou > 9 dans Sirius en 2013*

Application à TPE-2013 (4/6)



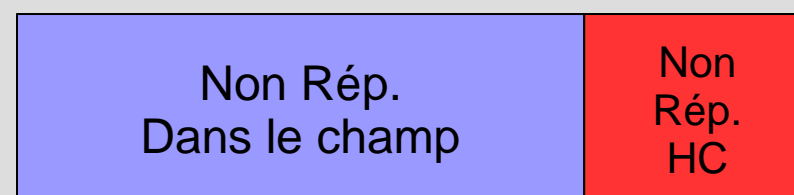
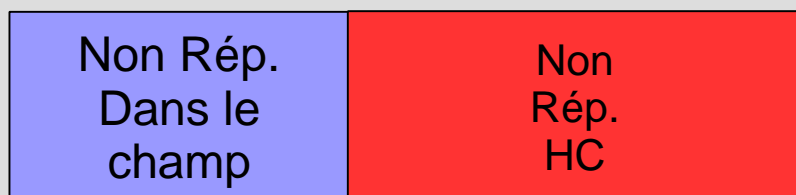
Application à TPE – 2013 (5/6)

- Une approche originale : **le traitement des indécis**
 - Utilisation de deux méthodes :
 - *L'imputation « radicale » des indécis en hors champ (méthode 1)*
 - *La repondération des unités au statut connu pour représenter les indécis (méthode 3).*



Méthode 1 : les indécis sont **hors champ**

Méthode 3 : On **repondère** les « non indécis »

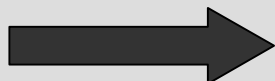


Application à TPE – 2013 (6/6)

Estimation du taux de temps partiel et du taux de bénéficiaires de la revalorisation du Smic au 1er janvier 2013 par taille d'entreprises en fonction de la méthode de repondération choisie.

		Résultats publiés	Résultat avec la pondération 1 pour le hors champ	Résultat avec la pondération 3 pour le hors champ
taux de temps partiel	1 salarié	41,2%	41,1%	41,1%
	2 salariés	36,8%	36,9%	36,9%
	3 à 5 salariés	30,2%	30,3%	30,3%
	6 à 9 salariés	23,9%	22,9%	22,9%
	Ensemble	30,26%	30,09%	30,10%
taux de bénéficiaires de la revalorisation du Smic au 1er janvier 2013	1 salarié	36,0%	35,9%	35,9%
	2 salariés	33,3%	33,1%	33,1%
	3 à 5 salariés	27,8%	27,8%	27,8%
	6 à 9 salariés	22,2%	21,7%	21,7%
	Ensemble	27,61%	27,53%	27,52%

source : Acemo-TPE2013



Dans le cas présent, peu de différences...

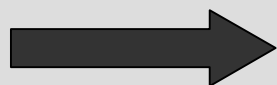
Et si tous les non retours étaient indécis ? (1/3)

- La méthode proposée :
 - Modèle de régression sur les répondants
 - Application à l'ensemble des non retours
- Le modèle :
 - Modèle de régression logistique
 - Variable expliquée : unité hors champ pour TPE - 2013
 - Variables explicatives :
 - *Tranches d'effectifs à partir des DADS 2011*
 - *Présence dans Sirius (2012 et 2013)*

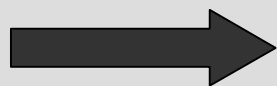
Et si tous les non retours étaient indécis ? (2/3)

- Les résultats du modèle :

Analysis of Maximum Likelihood Estimates					
Parameter	DDL	Estimate	Standard Error	Wald Chi-Square	Pr > Khi-2
Intercept	1	-0.7701	0.2953	6.7995	0.0091
0 salarié	1	1.0658	0.0906	138.4022	<.0001
10 salariés ou +	1	2.2125	0.1665	176.5889	<.0001
1 ou 2 salariés	1	0.5813	0.0446	169.9908	<.0001
8 ou 9 salariés	1	1.3594	0.0597	519.2388	<.0001
Sirus 2012	1	-1.1651	0.2658	19.2188	<.0001
Sirus 2013	1	-0.8649	0.1293	44.7320	<.0001



Si l'effectif DADS est proche des seuils (0 et 10 salariés), alors la probabilité d'être hors champ est plus importante

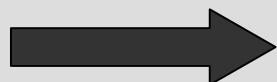


Si l'entreprise est dans SIRUS 2012 (respectivement 2013) alors la probabilité d'être hors champ est plus faible

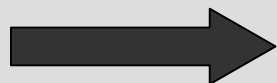
Et si tous les non retours étaient indécis ? (3/3)

- Les résultats de l'imputation sur le taux de réponse :

	Méthode 1	Méthode 2	Méthode 3	Méthode 4	Méthode 5
Répondants du champ	30356	30356	30356	30356	30356
Répondants hors champ	3059	3059	3059	3059	3059
Non répondants du champ estimé	17070	17350	17070	17278	19363
Non retour hors champ estimé	4447	4167	4167	4239	2154
Répondants du champ	30356	30356	30356	30356	30356
Echantillonnés du champ	47426	47706	47426	47634	49719
Taux de réponse estimé	64,0%	63,6%	64,0%	63,7%	61,1%



Des estimateurs peu différents par cette méthode (du fait du calage)



Moins de hors champ estimé → un moins bon taux de réponse

Conclusion

- Le hors champ, un écueil inévitable
- Un impact massif dans certaines enquêtes
- L'objectif de la méthode : le restreindre
- Vers une prédiction du hors champ, dès la base de sondage ?