

# Le Panel 2008 de la Dares des bénéficiaires de contrats aidés et d'une population témoin

Un outil dédié à l'évaluation par *matching*

I. BENOITEAU, S. LE MINEZ, L. LEQUIEN et V. RÉMY

Département des Politiques de l'Emploi  
Dares, Ministère du Travail

JMS, 25/01/2012

DARES

# Introduction

Longue tradition de panels à la Dares depuis le début des années 1990, pour évaluer les effets « causaux » du passage en contrat aidé (CA)

Améliorations méthodologiques au fil des générations de panels :

- 1 Définition d'un groupe de témoins : individus éligibles mais non entrés en CA
- 2 Meilleure comparabilité des bénéficiaires et des témoins
- 3 Plus grande richesse du questionnaire :
  - amélioration de la qualité du *matching*
  - diversification des variables d'intérêt : emploi et qualité de l'emploi, état de santé, conditions de vie, etc.

Dernier panel en date : le Panel 2008

# Plan de la présentation

- 1 Présentation du Panel 2008
- 2 Méthodologie sur données d'enquête
- 3 Résultats illustratifs

# Plan

- 1 Présentation du Panel 2008
- 2 Méthodologie sur données d'enquête
- 3 Résultats illustratifs

# La réforme des contrats aidés de 2005

**Définition** : emplois subventionnés par l'État pour favoriser l'insertion professionnelle de publics cibles : chômeurs de longue durée, allocataires de minima sociaux, jeunes non qualifiés, etc.

**Loi de cohésion sociale de 2005** : réforme des contrats aidés

<b>public</b> <b>secteur</b>	Personnes sans emploi rencontrant des difficultés sociales et professionnelles d'accès à l'emploi	Allocataires de minima sociaux (RMI, ASS, API, AAH)
Marchand	<b>CIE</b> Contrat Initiative Emploi	<b>CI-RMA</b> Contrat Insertion - Revenu Minimum d'Activité
Non marchand	<b>CAE</b> Contrat d'Accompagnement dans l'Emploi	<b>CAV</b> Contrat d'Avenir

Remplacement de ces contrats par le contrat unique d'insertion en 2010.

# Le Panel 2008 : une enquête par et pour le *matching*

**Objectif** : corriger le biais d'hétérogénéité rencontré avec l'estimateur naïf (simple différence de moyenne entre traités et témoins)

**Principe** : appairer chaque bénéficiaire à un « témoin » ayant des caractéristiques observables proches mais non entré en traitement ( $T$ )

$$\begin{aligned}\Delta^{ATT} &= E(Y_1|T = 1, X) - E(Y_0|T = 0, X) \\ &= E(Y_1|T = 1, X) - E(Y_0|T = 1, X)\end{aligned}$$

$\Delta^{ATT}$  = moyenne des écarts entre l'*outcome* des traités et celui des témoins servant de contrefactuel au  $Y_0$  des traités en l'absence de traitement.

- **Hyp. 1 : indépendance conditionnelle (CIA) :**  $Y_0 \perp T|X$  ou  $Y_0 \perp T|p(X)$   
⇒ le  $Y_0$  des témoins est un bon contrefactuel du  $Y_0$  des traités à condition qu'il n'existe pas d'autres caractéristiques que  $X$  influant sur  $Y$  et sur l'entrée en traitement (hyp. forte)
- **Hyp. 2 : existence d'un support commun :**  $0 < P(T = 1|X) < 1$   
⇒ présence de témoins ayant les mêmes caractéristiques que les bénéficiaires.

# Échantillonnage

## Premier constat

Le mode d'échantillonnage des individus dont on étudie la trajectoire joue sur la capacité à identifier les effets de la politique, ainsi que sur la précision des estimateurs.

Une **population de référence unique** : les demandeurs d'emploi (DE) inscrits à l'ANPE au T2-2005 (1 100 000)

**Base utilisée : le fichier historique des DE de Pôle Emploi (FH).**

# Échantillonnage : sélection des bénéficiaires

... dont des bénéficiaires embauchés pour la 1<sup>re</sup> fois en CAE, CAV, CIE ou CI-RMA sur la période [T2-2005,T2-2007] (83 000)

- 1 échantillonnés dans le fichier des conventions individuelles d'embauches en CA (type de contrat, date d'entrée, etc.) apparié au FH
- 2 sur-représentés dans l'échantillon (*choice based sampling*)
- 3 sélectionnés selon un plan de sondage équilibré

⇒ **Échantillon retenu** : 10 000 bénéficiaires de CAE et de CIE et 6 700 bénéficiaires de CAV et de CI-RMA

# Échantillonnage : sélection des témoins et *matching*

... dont des témoins éligibles aux CA mais non entrés en CA sur la période de traitement

- 1 spécialisés par type de public (témoins 1 : CAE et CIE ; témoins 2 : CAV et CI-RMA)
- 2 sélectionnés pour garantir *ex ante* l'existence d'un support commun.
  - témoins 1 échantillonnés aléatoirement au sein de strates d'individus de mêmes caractéristiques que les bénéficiaires,
  - témoins 2 sélectionnés aléatoirement au sein de strates d'individus ayant la même probabilité instantanée d'entrer en contrat aidé que celle des bénéficiaires (modèle de Cox).

⇒ **Échantillon retenu** : 7 000 témoins 1 et 6 000 témoins 2.

## Collecte des données

Enquête réalisée en deux vagues, fin 2008 et fin 2009  
⇒ choix des dates pour garantir une fenêtre d'observation suffisante mais limiter les biais de mémoire.

### Deuxième constat

Les modes de collecte entre les différents types de bénéficiaires et de témoins doivent être homogènes pour éliminer toute corrélation entre le traitement et les erreurs de mesure (ou comportements de non-réponse).

## Richesse des informations recueillies

- trajectoire professionnelle avant entrée en traitement (panel + FH)  
⇒ parcours du demandeur d'emploi, type d'emploi recherché, motif d'inscription, etc.
- informations au T2-2005 sur l'hétérogénéité individuelle habituellement inobservée :
  - environnement familial, conditions de vie (revenu, logement), littératie,
  - conditions de vie (subjectives), état de santé, rapport aux études, freins à l'emploi, objectifs lors de l'inscription à l'ANPE, etc.

⇒ La richesse des données disponibles permet de garantir au mieux l'hypothèse *CIA*.

# Plan

- 1 Présentation du Panel 2008
- 2 Méthodologie sur données d'enquête
- 3 Résultats illustratifs

# Le calcul des pondérations

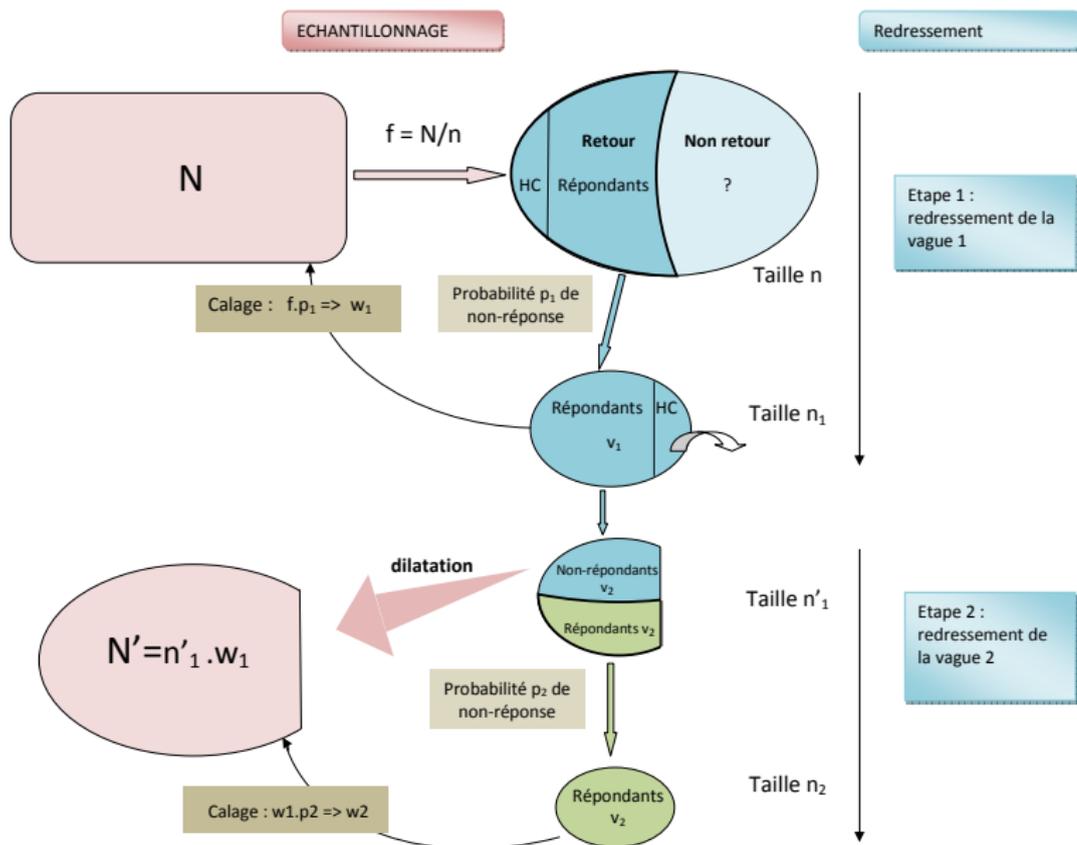
Correction de la **non-réponse** par la méthode des groupes de réponse homogènes (GRH). Réalisée en deux étapes :

- 1 non-réponse en vague 1 (54%)
- 2 attrition en vague 2 (83%)  
⇒ enrichissement par la prise en compte des variables d'enquête de la vague 1

Complétée par un **calage** sur marges à chaque étape (variables identiques, issues du FH)

Traitement des **hors champ** :

- assimilés aux répondants ⇒ calcul de leur poids calé et redressé
- supprimés en fin de première étape ⇒ détermination de la population dans le champ, sur laquelle caler la vague 2



# Matching et pondérations

Les pondérations  $w$  interviennent à plusieurs étapes (Frölich, 2003) :

- 1 estimation pondérée du **score de propension** :

$$\ln(L(\beta)) = \sum_{i=1}^n [w_i \cdot T_i \cdot \ln(F(X_i\beta)) + w_i \cdot (1 - T_i) \cdot \ln(1 - F(X_i\beta))]$$

- 2 calcul du **contrefactuel** des bénéficiaires par noyau pondéré :

$$\hat{Y}_{i0} = \frac{\sum_{E_0} w_k \cdot K\left(\frac{\|p_i - p_k\|}{h}\right) \cdot Y_k}{\sum_{E_0} w_k \cdot K\left(\frac{\|p_i - p_k\|}{h}\right)}$$

- 3 **ATT** calculé comme la moyenne pondérée sur l'ensemble des bénéficiaires des écarts entre leur *outcome* observé et leur *outcome* contrefactuel :

$$\hat{\Delta}^{ATT} = \frac{\sum_{E_1} w_i \cdot (Y_{i1} - \hat{Y}_{i0})}{\sum_{E_1} w_i}$$

Ecart-type de l'ATT calculé par *bootstrap* : les poids initiaux ne permettent plus de retrouver la structure de la population sur les différents échantillons répliqués.

Adaptation du *bootstrap* standard aux données pondérées (Davezies et D'Haultfœuille, 2009). Pour chaque réplication :

- 1 Tirage de  $n_{b1}$  ( $n_{b0}$ ) dans une loi binomiale  $B \sim (N_1, n_1/N_1)$  ( $B \sim (N_0, n_0/N_0)$ ),
- 2 Tirage aléatoire simple avec remise d'un échantillon de bénéficiaires (témoins) de taille  $n_{b1}$  ( $n_{b0}$ ),
- 3 Recalcul des nouvelles pondérations selon la méthode employée sur l'échantillon original,
- 4 Estimation de l'ATT (y.c. recalcul du score)

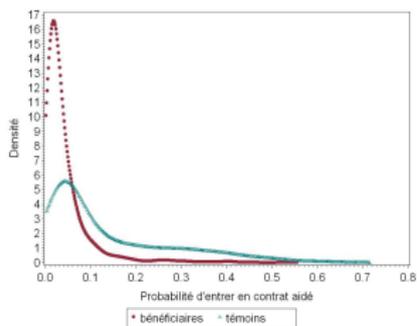
# Plan

- 1 Présentation du Panel 2008
- 2 Méthodologie sur données d'enquête
- 3 Résultats illustratifs

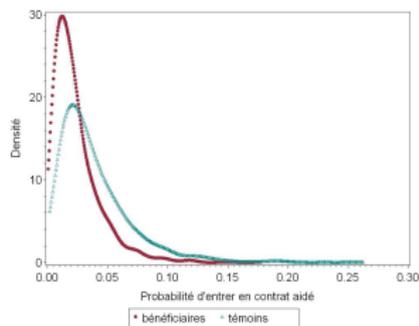
# Validité du score

## Validation du *matching* :

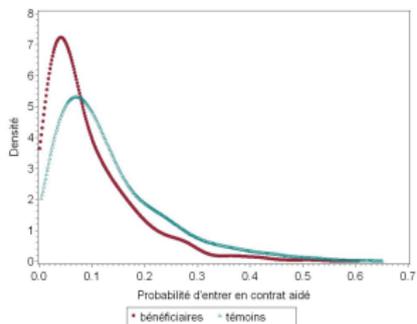
- nombreuses variables de conditionnement  $X$   
⇒ crédibilité de l'hypothèse CIA
- échantillons bénéficiaires et témoins équilibrés sur les  $X$   
conditionnellement au score  
⇒ validité du conditionnement par le score plutôt que par les  $X$
- large support commun (déterminé par le score max des témoins)  
⇒ utilisation optimale des observations de l'échantillon des bénéficiaires



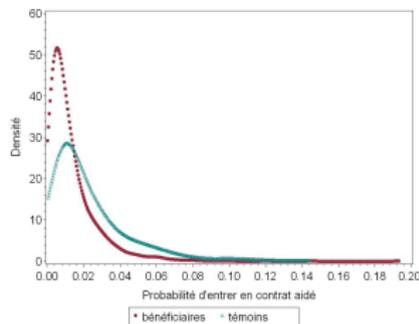
(a) CAE



(b) CIE



(c) CAV



(d) CI-RMA

FIGURE: Comparaisons des distributions des scores de propension

# Comparaison des résultats

**Tableau:** Estimation de l'ATT, non pondéré *vs* pondéré

$\hat{\Delta}^{ATT}$ non pondéré	CAE	CIE	CAV	CI-RMA
Être en CDI deux ans après l'entrée en contrat aidé	-0,173***	0,096***	-0,153***	0,131***
Durée cumulée en CDI sur les deux ans suivant l'entrée en contrat	-4,737***	-1,891***	-3,809***	-1,198***

Notes : écart-types calculés par *bootstrap* avec 500 réplifications.

$\hat{\Delta}^{ATT}$ pondéré	CAE	CIE	CAV	CI-RMA
Être en CDI deux ans après l'entrée en contrat aidé	-0,160***	0,086***	-0,148***	0,125***
Durée cumulée en CDI sur les deux ans suivant l'entrée en contrat	-4,537***	-1,785***	-3,483***	-1,042***

Notes : écart-types calculés par *bootstrap* avec 500 réplifications, sans recalcul des poids à chaque itération.

# Hétérogénéité du traitement

Deux niveaux d'hétérogénéité :

- 1 Hétérogénéité des publics visés  $\Rightarrow$  évaluation par contrat
- 2 Hétérogénéité du traitement :
  - *le délai avant l'entrée en traitement* : hétérogénéité du traitement en termes de durée passée au chômage avant l'entrée en contrat aidé (large fenêtre de référence).
  - *le contenu du traitement* : accompagnement (interne ou externe), formation, type d'employeur, temps de travail, métier exercé durant le contrat.
  - *la durée du traitement* : définie comme la durée totale du contrat, reconductions éventuelles comprises .

# Résultats

**Tableau:** Estimation de l'effet moyen du passage en contrat aidé

$\hat{\Delta}^{ATT}$	Non pondéré		Pondéré	
	CAE	CI-RMA	CAE	CI-RMA
<i>Durée du contrat inférieure à 6 mois</i>				
Être en CDI deux ans après l'entrée en contrat aidé	-0,109***	0,076***	-0,119***	0,073**
<i>Durée du contrat comprise entre 6 et 12 mois</i>				
Être en CDI deux ans après l'entrée en contrat aidé	-0,156***	0,132***	-0,162***	0,119**
<i>Durée du contrat supérieure à 12 mois</i>				
Être en CDI deux ans après l'entrée en contrat aidé	-0,230***	0,182***	-0,224***	0,148***
<i>Accompagnement</i>				
Être en CDI deux ans après l'entrée en contrat aidé	-0,156***	0,149***	-0,161***	0,165***
<i>Sans accompagnement</i>				
Être en CDI deux ans après l'entrée en contrat aidé	-0,197***	0,119***	-0,186***	0,110***

## Conclusion / discussion

Conception de l'enquête : des avancées méthodologiques

- échantillonnage pour garantir un support commun
- questionnaire pour capter l'hétérogénéité

Analyses causales : de nombreux enjeux

- définition du traitement : date d'entrée, durée, contenu
- *matching* avec pondération

Des prolongements pour affiner l'évaluation :

- croisement entre données d'enquête et administratives (ASP, FH)
- *matching dynamique* pour corriger le biais de sélection
- *modèles de durée* pour tenir compte des données censurées