

**Estimation communale de l'enquête  
Famille et Logements 2011  
et du recensement**

*Pascal ARDILLY*

INSEE  
Dmsci

## Le contexte

EFL : enquête associée à l'EAR 2011, auprès d'individus adultes en ménage ordinaire, en métropole, dans 1438 communes volontaires.

**Plan de sondage complexe**, s'appuyant sur l'intersection de deux échantillons : EAR 2011 + échantillon de logements spécifique (plan à deux degrés) :

Phase 1 : tirage stratifié de communes (recensées en 2011) puis de grappes de logements ;

Phase 2 : 1 grappe = 1 sexe (avec  $\frac{2}{3}$  'femmes') ;

**Echantillon** : 239 000 femmes répondantes et 121 000 hommes répondants .

L'enquête concerne le plus souvent **entre 200 et 400 logements dans la commune.**

## **Restitution aux communes** (« 4 pages ») :

1. nombre d'hommes, de femmes et de couples ayant souscrit un pacs ;
2. nombre d'hommes et de femmes vivant en couple mais non cohabitant ;
3. nombre d'enfants de moins de 4 ans gardés principalement par les parents, les grands-parents, une assistante maternelle, une crèche, ou autre ;
4. nombre de grands-pères et de grands-mères, respectivement de moins et de plus de 75 ans ;
5. nombre d'hommes et de femmes de plus de 75 ans vivant seuls et ayant au moins un enfant résidant dans la même commune ;
6. nombre d'hommes et de femmes de plus de 75 ans vivant seuls et ayant au moins un enfant résidant à moins de 30 minutes de leur domicile ;
7. nombre d'enfants de moins de 18 ans vivant dans une famille recomposée, traditionnelle, monoparentale, en distinguant trois tranches d'âge ;
8. nombre de familles recomposées et traditionnelles ;
9. nombre d'enfants de moins de 18 ans résidant habituellement dans deux logements pour cause de vie avec un autre parent;
10. nombre d'hommes et de femmes résidant habituellement dans deux logements.

# L'esprit de la méthode

Nécessité d'un modèle de comportement : **le comportement « moyen » de la commune est assimilé à celui d'autres communes de même type. Cela se conçoit au niveau de différentes catégories d'individus.**

**OBJECTIF : chercher une relation entre la probabilité d'appartenir à une sous-population d'intérêt et un ensemble de variables explicatives (= catégorie d'individus), la géographie n'ayant (presque) plus d'influence.**

Les variables explicatives sont tirées du **recensement** (appariement à EFL)

Par exemple : deux individus appartenant à deux communes différentes (de même type) mais ayant le même sexe, le même âge et le même niveau de diplôme ont la même probabilité d'être pacsés.

L'estimation communale s'appuiera donc sur un échantillon rassemblant de nombreuses communes.

**Finalement, on substitue du biais à de la variance.**

## Formalisation du modèle

$com$  = la commune

$ssp$  = une sous-population d'intérêt (ex : les 'pacsés' )

$ch$  = un champ (ex : l'ensemble des individus adultes vivant en couple et cohabitant).

**Un ensemble de facteurs explicatifs collectés au recensement prédit l'appartenance à  $ssp$  : ces facteurs forment des catégories de population  $h$**

(par exemple un croisement sexe X tranche d'âge X diplôme).

$$\forall h, \frac{N_{com}^{ch,h,ssp}}{N_{com}^{ch,h}} \text{ ne dépend que de } h$$

Modèle appliqué au sein de catégories de communes (redonne un peu d'influence à la géographie), soit :

$$\frac{N_{com}^{ch,h,ssp}}{N_{com}^{ch,h}} = \frac{\sum_{com} N_{com}^{ch,h,ssp}}{\sum_{com} N_{com}^{ch,h}}$$

## Définition de l'estimateur local

$$N_{com}^{ssp} = \sum_{h=1}^H N_{com}^{ch,h} \cdot \frac{N_{com}^{ch,h,ssp}}{N_{com}^{ch,h}}$$

Il reste à estimer les probabilités / proportions

$$\frac{\sum_{com} N_{com}^{ch,h,ssp}}{\sum_{com} N_{com}^{ch,h}}$$

Les poids « nationaux » d'EFL n'étaient pas disponibles.

La probabilité de réponse à EFL est *in fine* constante par grappe, estimée par le taux de réponse dans la grappe.

$$\hat{N}_{com,L}^{ssp} = \sum_{h=1}^H N_{com}^{ch,h} \cdot \frac{\sum_{i \in r \cap com \cap h \cap ssp} \frac{w_i}{\hat{R}_i}}{\sum_{i \in r \cap com \cap h} \frac{w_i}{\hat{R}_i}} \cdot \frac{1}{\sum_{com} N_{com}^{ch,h}}$$

$w_i$  = poids de l'individu  $i$  dans l'EAR 2011 et  $\hat{R}_i$  l'estimation de sa probabilité de réponse.

Pour certaines variables liées à la famille, on utilise la pondération de l'exploitation complémentaire « au quart » dans les petites communes.

**Au final : l'estimateur local a considérablement gagné en stabilité. Il peut se calculer en toutes circonstances, dans n'importe quelle commune.**

En particulier, on peut estimer un effectif portant sur un sexe NON enquêté dans la commune !!!

Noter le décalage temporel :

EFL → 2011

RP → 2009

## Constitution des catégories homogènes, selon la variable d'intérêt

\* \* \* \* \*

### Nombre d'hommes, de femmes, de couples ayant souscrit un pacte civil de responsabilité

Filtre : personnes en couple mais non mariées (EFL) +  
vivant dans le même logement (= cohabitant).

Mais situation matrimoniale incohérente entre EFL  
et EAR !

D'où :  $ch$  = personnes vivant en couple dans le même  
logement, **mariées ou non** au sens du recensement.  
Et les personnes mariées selon le RP constitueront une  
catégorie  $h$  spécifique !

Régression **logit** pour sélectionner les régresseurs  $X$  :

$$\text{Pr oba}(pacs) = \frac{e^{X.B}}{1 + e^{X.B}}$$

Convention : même catégories  $h$  pour hommes et  
femmes

Deux catégories de communes créées (selon la ZEAT).



## Les catégories formées :

- individus mariés selon le recensement (h=1) .

*Parmi les individus non-mariés selon le recensement :*

- diplôme de niveau inférieur au bac (h = 2) ;
- diplôme de niveau bac, naissance en 1969 ou avant (h = 3) ;
- (diplôme de niveau bac, né(e) entre 1970 et 1979) ou (diplôme de niveau supérieur au bac, né(e) avant 1970) (h = 4) ;
- diplôme de niveau bac, naissance en 1980 ou plus tard (h = 5) ;
- diplôme de niveau supérieur au bac, né(e) entre 1970 et 1979 (h = 6) ;
- diplôme de niveau supérieur au bac, né(e) après 1979 (h = 7).

### Probabilité de pacs des hommes

HOMMES		Catégorie de communes N°1		Catégorie de communes N°2	
Catégorie de population	Effectif répondant	Effectif répondant	Probabilités de pacs estimées (%)	Effectif répondant	Probabilités de pacs estimées (%)
1	63 709	21 623	<b>0.9</b>	42 086	<b>1.2</b>
2	7 693	2 393	<b>8.3</b>	5 300	<b>8.5</b>
3	1 841	529	<b>11.8</b>	1 312	<b>12.3</b>
4	2 643	1 012	<b>11.7</b>	1 631	<b>9.3</b>
5	3 962	1 220	<b>28.2</b>	2 742	<b>18.4</b>
6	2 446	675	<b>34.0</b>	1 771	<b>31.0</b>
7	2 334	694	<b>23.2</b>	1 640	<b>16.4</b>
Ensemble	84 628	28 146	<b>5.4</b>	56 482	<b>5.0</b>

- individus mariés selon le recensement (h=1) .

*Parmi les individus non-mariés selon le recensement :*

- diplôme de niveau inférieur au baccalauréat (h = 2) ;
- diplôme de niveau bac, naissance en 1969 ou avant (h = 3) ;
- (diplôme de niveau bac, né(e) entre 1970 et 1979) ou (diplôme de niveau supérieur au bac, né(e) avant 1970) (h = 4) ;
- diplôme de niveau bac, naissance en 1980 ou plus tard (h = 5) ;
- diplôme de niveau supérieur au bac, né(e) entre 1970 et 1979 (h = 6) ;
- diplôme de niveau supérieur au bac, né(e) après 1979 (h = 7).

### Probabilité de pacs des femmes

FEMMES		Catégorie de communes N°1		Catégorie de communes N°2	
Catégorie de population	Effectif répondant	Effectif répondant	Probabilités de pacs estimées (%)	Effectif répondant	Probabilités de pacs estimées (%)
1	112 757	38782	<b>0.9</b>	73975	<b>0.9</b>
2	10304	3434	<b>6.3</b>	6870	<b>6.4</b>
3	3183	910	<b>10.9</b>	2273	<b>10.4</b>
4	5931	2171	<b>10.9</b>	3760	<b>11.0</b>
5	6382	1874	<b>14.4</b>	4508	<b>12.5</b>
6	5020	1446	<b>31.0</b>	3574	<b>29.0</b>
7	6851	2231	<b>21.1</b>	4620	<b>17.2</b>
Ensemble	150 428	50 848	<b>4.5</b>	99 580	<b>4.4</b>

## Nombre d'enfants de moins de 4 ans, selon leur mode de garde principal

Questions portant cette fois sur le ménage.

5 modes de garde distingués = 5 probabilités à estimer

**Modèle** : pour un enfant donné dans une catégorie de communes donnée, sa probabilité d'être gardé selon un mode donné ne dépend pas de sa commune, mais « seulement » de la catégorie à laquelle appartient son ménage.

*\* Formation des catégories de ménages :*

1/ On retient pour chaque adulte du ménage 3 tranches d'activité selon l'EAR 2011 (question 10 du BI) :

- Emploi, apprentissage, études
- Chômage
- Retraite, personne au foyer, autre

2/ Pour les personnes qui travaillent, on prend également en compte le temps de travail déclaré (question 22 du BI) :

- Emploi occupé à temps complet
- Emploi occupé à temps partiel

- Ménage avec 1 adulte qui travaille à temps plein :  
catégorie  $h = 1$
- Ménage avec 1 adulte qui travaille à temps partiel OU qui  
est en stage / étude : catégorie  $h = 2$
- Ménage avec 1 adulte au chômage : catégorie  $h = 3$
- Ménage avec 1 adulte retraité ou "au foyer" : catégorie  
 $h = 4$
- Ménage avec au moins 2 adultes, et au moins un retraité  
ou une personne "au foyer" : catégorie  $h = 5$
- Ménage avec au moins 2 adultes, tous les adultes  
travaillent à temps plein : catégorie  $h = 6$
- Ménage avec au moins 2 adultes, tous les adultes  
travaillent ou sont en stage / étude mais il y a au moins un  
adulte qui ne travaille pas à temps plein (temps partiel ou  
alors études) : catégorie  $h = 7$
- Ménage avec au moins 2 adultes, aucun n'est retraité ni  
personne "au foyer" mais au moins l'un d'eux est au  
chômage : catégorie  $h = 8$

+ croisement avec la TU (en 4 tranches) dans les catégories 4, 5, 6 et 7  $\Rightarrow$  *in fine* 20 catégories de ménage, et 2 catégories de communes.

Prise de risque : l'activité est une information conjoncturelle !

**Catégorie de communes N°1, probabilité d'un enfant de moins de 4 ans d'être gardé à titre principal par un des modes de garde (en %).**

Sous-population		Effectif répondant	Probabilité par mode de garde principal (%)				
Catégorie de ménage	Tranche d'UU		Parents	Grands - parents, famille	Assistante maternelle	Crèche	Autre mode
1	toutes	383	<b>26.0</b>	<b>17.0</b>	<b>25.9</b>	<b>27.5</b>	3.6
2	toutes	160	<b>41.0</b>	<b>6.8</b>	<b>21.2</b>	<b>28.2</b>	2.7
3	toutes	210	<b>69.8</b>	<b>2.6</b>	<b>5.7</b>	<b>18.8</b>	3.1
4	1	62	<b>93.2</b>	<b>2.1</b>	<b>1.6</b>	<b>2.1</b>	1.0
4	2	65	<b>88.0</b>	<b>1.8</b>	<b>8.0</b>	<b>2.3</b>	0.0
4	3	168	<b>93.5</b>	<b>0.7</b>	<b>3.3</b>	<b>1.5</b>	1.0
4	4	157	<b>87.6</b>	<b>1.4</b>	<b>2.3</b>	<b>4.6</b>	4.1
5	1	1358	<b>89.1</b>	<b>2.2</b>	<b>5.6</b>	<b>2.4</b>	0.6
5	2	945	<b>88.7</b>	<b>2.3</b>	<b>5.1</b>	<b>3.7</b>	0.2
5	3	1533	<b>91.1</b>	<b>0.7</b>	<b>3.0</b>	<b>4.8</b>	0.4
5	4	1464	<b>91.0</b>	<b>1.4</b>	<b>2.9</b>	<b>4.0</b>	0.8
6	1	2553	<b>27.5</b>	<b>9.7</b>	<b>52.4</b>	<b>7.9</b>	2.4
6	2	1213	<b>27.3</b>	<b>9.4</b>	<b>44.4</b>	<b>17.1</b>	1.8
6	3	1376	<b>28.6</b>	<b>10.3</b>	<b>35.9</b>	<b>21.5</b>	3.7
6	4	1537	<b>28.7</b>	<b>9.8</b>	<b>33.9</b>	<b>24.8</b>	2.9
7	1	1466	<b>33.2</b>	<b>7.1</b>	<b>50.1</b>	<b>8.0</b>	1.6
7	2	717	<b>38.5</b>	<b>8.7</b>	<b>36.7</b>	<b>14.2</b>	1.9
7	3	822	<b>40.4</b>	<b>5.4</b>	<b>29.3</b>	<b>21.9</b>	3.0
7	4	979	<b>37.3</b>	<b>8.1</b>	<b>23.7</b>	<b>27.3</b>	3.6
8	toutes	2174	<b>67.5</b>	<b>3.3</b>	<b>11.7</b>	<b>16.2</b>	1.3
Ensemble		19 342	<b>62.1</b>	<b>5.0</b>	<b>16.8</b>	<b>14.2</b>	1.9

*Confrontation finale avec l'offre « d'équipement » :*

Utilisation de la 'Base Permanente des Equipements' (BPE) + distancier = calcul du temps pour atteindre la crèche la plus proche, en minutes (heures creuses) ;

- Si temps > 20 mn, alors absence de garde en crèche (hyp.) → concerne 15.6% des communes de métropole ;

Utilisation du fichier administratif 'Particuliers-employeurs' + distancier.

- Si temps > 20 mn, alors absence de garde par Ass. Mat. (hyp.) → concerne 0.4% des communes de métropole.

## Dernière étape : la production des effectifs communaux

Effectifs communaux estimés **pour l'ensemble des communes de métropole.**

Opération 1 : vieillissement des données du recensement de 2009 à 2011 (EFL : situation au 1<sup>er</sup> janvier 2011) : techniques de prolongement de tendance (distinguant petites/grandes communes).

Opération 2 : opération de calage sur les estimations nationales EFL - dite '*benchmarking*'.

Le benchmarking s'effectue selon 2 dimensions :

- l'effectif national EFL ;
- les effectifs RP du champ (*ch*) pour chaque commune.

On utilise un algorithme de type *raking-ratio*.

Il faut auparavant que, pour la population du champ :

estimation nationale EFL = estimation nationale RP.

Cette contrainte se satisfait par un coefficient correcteur appliqué au RP (référence = EFL).

## Appréciation du biais des modèles, avant benchmarking, par variable diffusée

Variable d'intérêt	Estimation nationale EFL	Somme des estimations communales	Erreur relative (%)	Coeff. correcteur
Nombre d'hommes pacsés	701 000	659 000	- 6,0	0.989
Nombre de femmes pacsées	659 000	583 000	<b>- 11,5</b>	0.990
Nombre d'hommes en couple, non cohabitants	602 000	689 000	<b>14,4</b>	0.999
Nombre de femmes en couple, non cohabitantes	714 000	788 000	<b>10,4</b>	0.999
Nombre de grands-pères, moins de 75 ans	4 552 000	4 268 000	- 6,2	1.004
Nombre de grands-mères, moins de 75 ans	6 035 000	5 810 000	- 3,7	1.002
Nombre de grands-pères, 75 ans ou plus	1 584 000	1 495 000	- 5,6	<b>1.033</b>
Nombre de grands-mères, 75 ans ou plus	2 547 000	2 427 000	- 4,7	<b>1.025</b>
Nombre d'hommes âgés, vivant seuls, au moins un enfant résidant dans la même commune	111 000	100 000	- 9,9	<b>1.062</b>
Nombre de femmes âgées, vivant seules, au moins un enfant résidant dans la même commune	625 000	671 000	7,4	<b>1.027</b>
Nombre d'hommes âgés, vivant seuls, au moins un enfant résidant à moins de 30 minutes	216 000	174 000	<b>- 19,4</b>	<b>1.062</b>
Nombre de femmes âgées, vivant seules, au moins un enfant résidant à moins de 30 minutes	1 103 000	1 025 000	- 7,1	<b>1.027</b>
Nombre d'enfants gardés par les parents	1 652 000	1 608 000	- 2,7	<b>1.027</b>



Nombre d'enfants gardés par les grands-parents	177 000	158 000	- 10,7	1.027
Nombre d'enfants gardés en crèche	397 000	430 000	8,3	1.027
Nombre d'enfants gardés par assistante maternelle	816 000	749 000	- 8,2	1.027
Nombre d'enfants gardés par un autre mode	97 000	106 000	9,3	1.027
Nombre d'enfants vivant en famille monoparentale	2 450 000	2 478 000	1,1	0.986
Nombre d'enfants vivant en famille recomposée	1 476 000	1 460 000	- 1,1	0.986
Nombre d'enfants vivant en famille traditionnelle	9 774 000	9 941 000	1,7	0.986
Nombre de familles traditionnelles	5 473 000	5 476 000	0,1	0.992
Nombre de familles recomposées	723 000	725 000	0,3	0.992
Nombre d'hommes résidant dans deux logements	2 701 000	2 821 000	4,4	0.999
Nombre de femmes résidant dans deux logements	2 767 000	2 809 000	1,5	0.999
Nombre d'enfants résidant dans deux logements pour vivre avec leur autre parent	905 000	835 000	- 7,7	0.998

Coefficient correcteur « éloigné » de 1 = défavorable à la qualité des estimations, dû à :

- processus de vieillissement du recensement,
- décalage des concepts définissant le champ entre EFL et RP (ex : vivre seul),
- erreur d'échantillonnage EFL.

Les erreurs relatives = indicateur de la qualité GLOBALE des estimations locales → **pas de sécurité au niveau communal**

Peut paraître parfois excessif (standard national de CV), mais **exercice extrêmement acrobatique qui a fait prendre tous les risques !!!** Quintuple peine :

- zones d'estimation (les communes) extrêmement petites ;
- échantillonnage à fort effet de grappe (par construction) ;
- certaines populations rares et/ou compliquées à isoler ;
- information disponible, pour expliquer des phénomènes qui pour la plupart relèvent de déterminants complexes, limitée au recensement - donc assez « pauvre » ;
- aucun effectif estimé sans erreur dans les grandes communes + dans les petites communes à chaque fois que l'unité 'famille' entre en jeu.

D'autres expériences conduisent à des erreurs comparables.

Normes possibles ???

- erreur inférieure à 5% : très satisfaisante
- erreur entre 5% et 10% : satisfaisante
- erreur entre 10% et 15% : acceptable
- erreur supérieure à 15% : dégradée

RAPPEL : erreurs du tableau = situation avant *benchmarking*.

Les estimations finales intègrent le *benchmarking*, qui par construction rend toutes ces erreurs relatives nulles !

**Cette correction devrait atténuer les biais « locaux ».**

Pas de calcul de variance d'échantillonnage car ce n'est pas essentiel : **le vrai problème vient du BIAIS** (estimateurs '*modèle dépendants*').

*Autres outils de mesure/contrôle de qualité :*

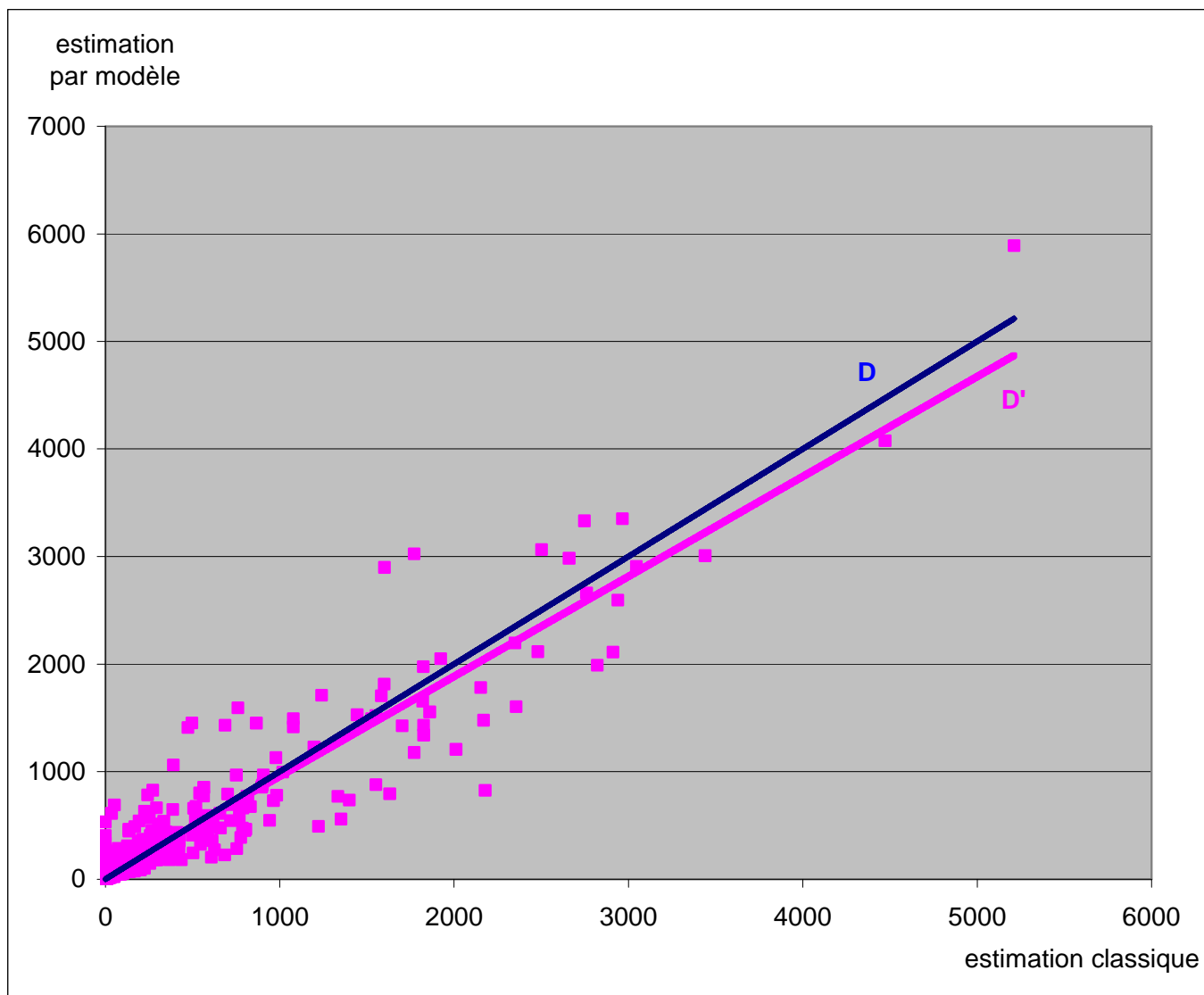
1/ Outil de régression ;

2/ Cartographie ;

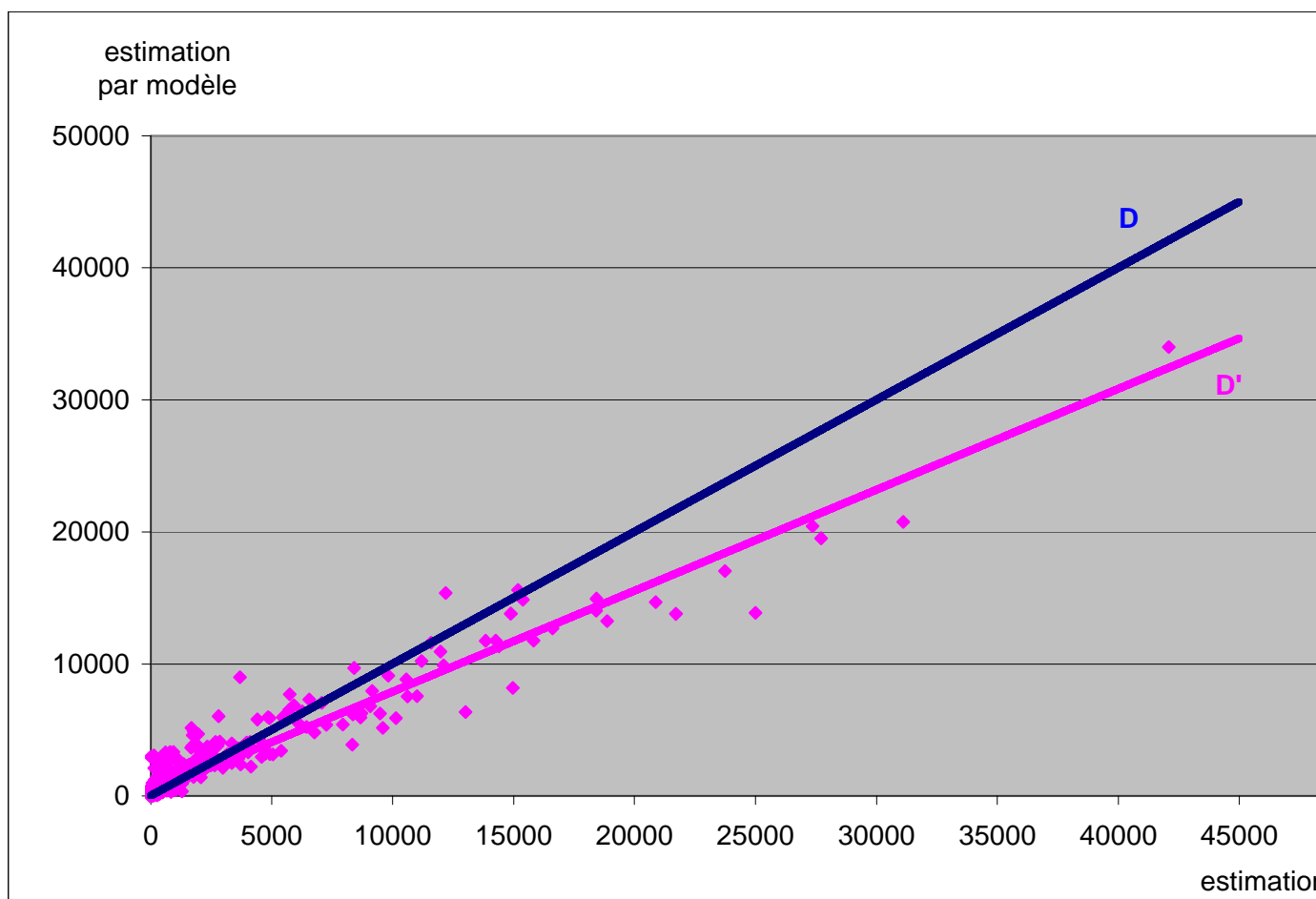
3/ Par sécurité : calcul de ratios communaux :  
estimation / population.

**Restitution finale** : politique d'arrondi de tous les effectifs : évite de donner aux utilisateurs une impression de précision illusoire + renforce l'anonymisation des données (communes de petite taille).

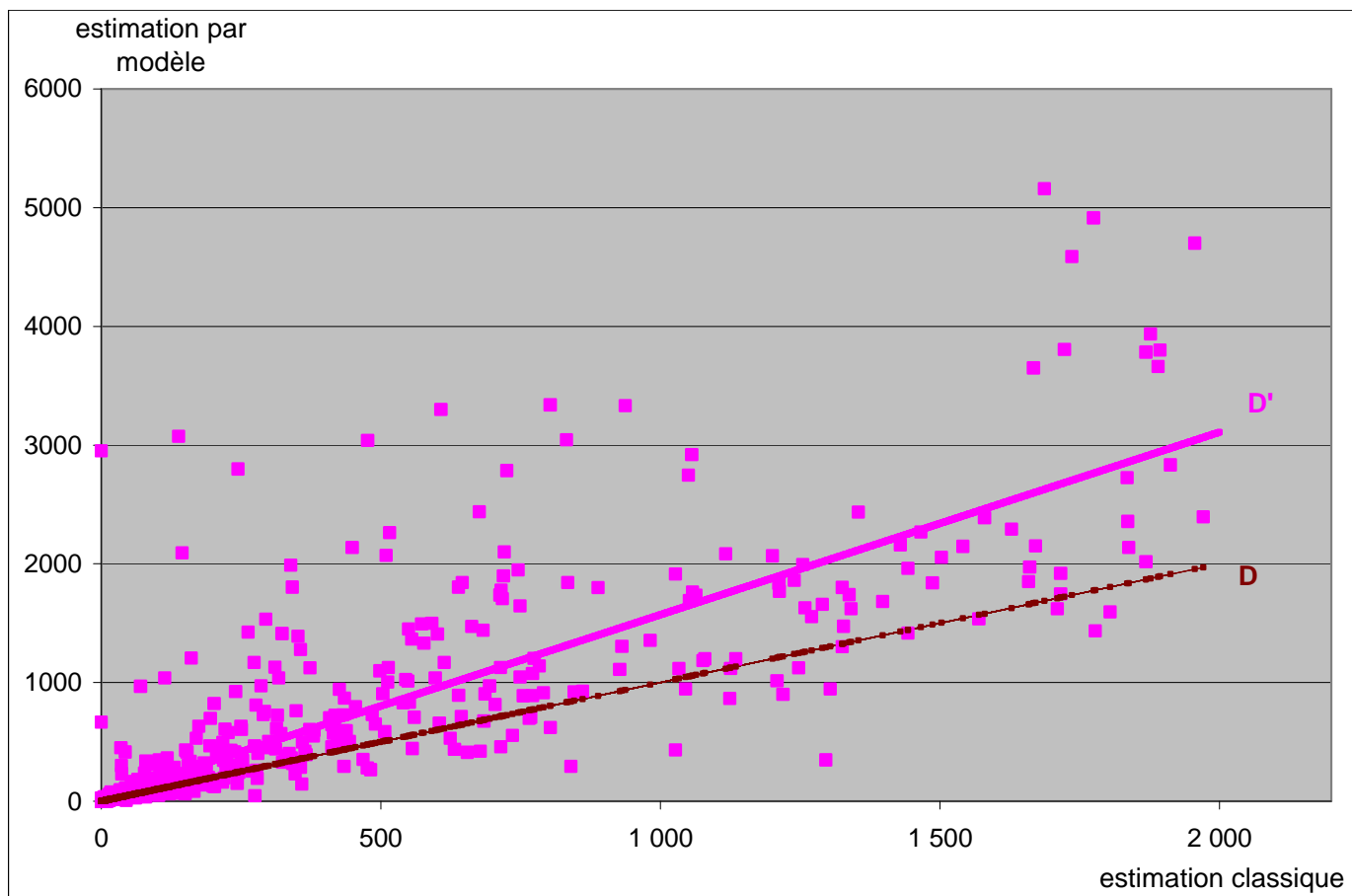
## Estimations classiques et par modèle du nombre d'hommes pacsés : détection du biais



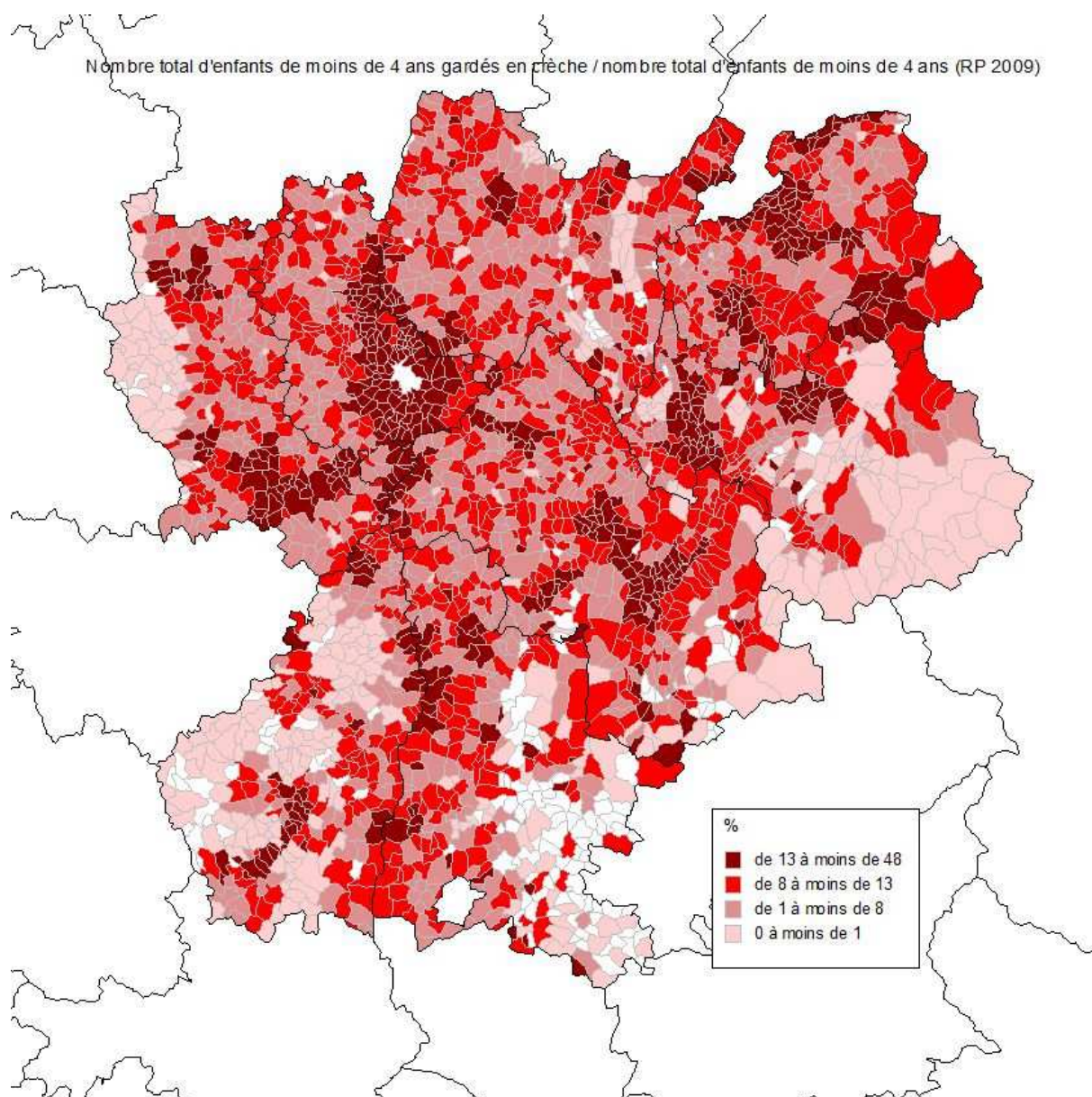
## Estimations classiques et par modèle du nombre de femmes en multi résidence : détection du biais



## Estimations classiques et par modèle du nombre de femmes en multi résidence, cas d'une sélection de communes : détection du biais

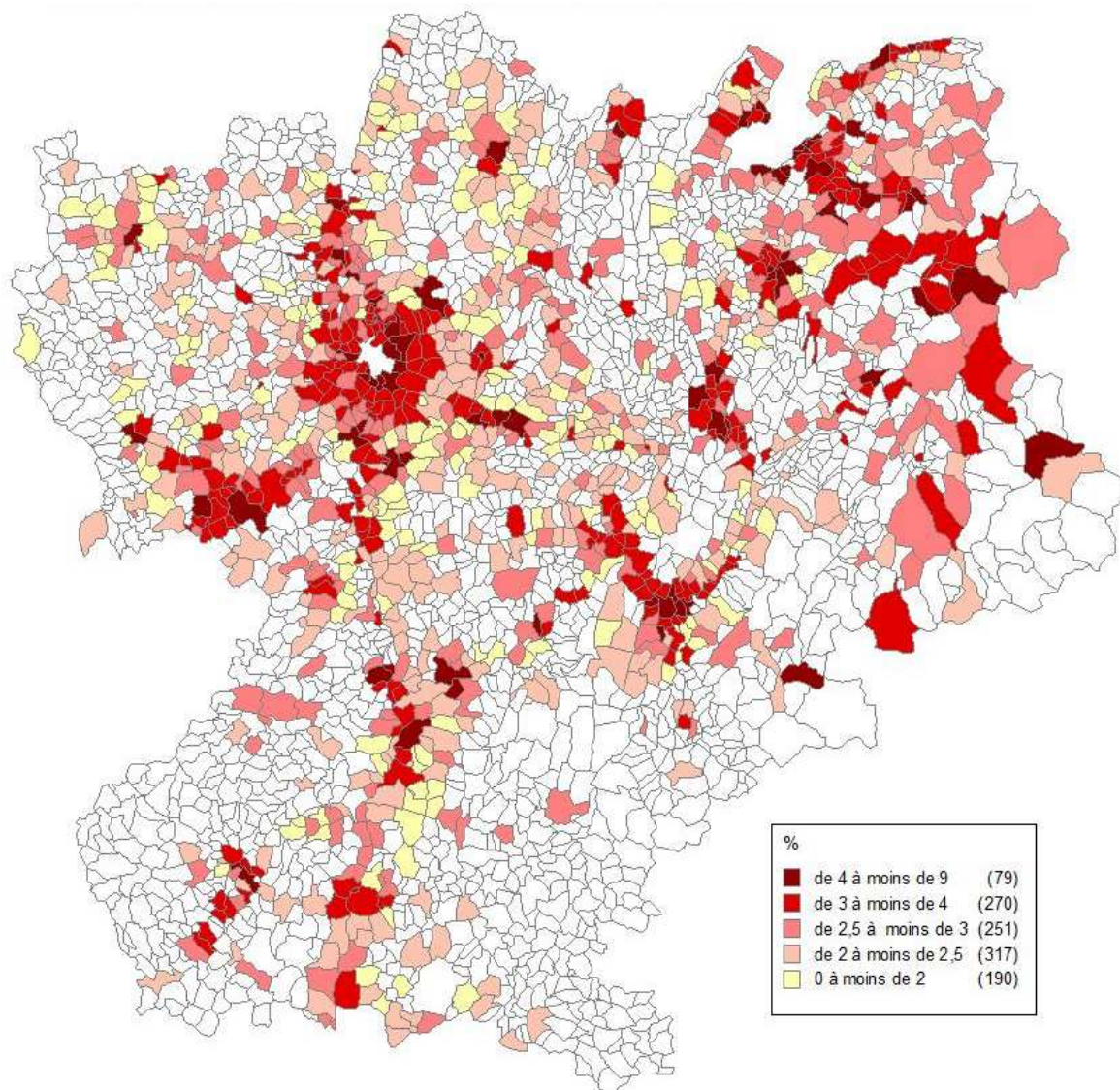


## Proportion d'enfants de moins de 4 ans gardés en crèche





## Proportion d'hommes vivant en couple non cohabitants



## Pourquoi pas une méthodologie plus « avancée » ?

L'estimation utilisée est « synthétique » - le modèle est **descriptif**, et **non** stochastique.

1/ Le nombre de domaines est énorme : réaction des logiciels ?

2/ *Dans les situations extrêmes*, je doute que l'on prenne moins de risque en utilisant un modèle sophistiqué

3/ Je ne crois pas (du tout) que l'on mesure mieux la qualité avec un modèle stochastique

4/ Aspect pédagogique : argument de communication

5/ Crainte d'**invraisemblances non maitrisables** : finalement, le *shrinkage* a un coté protecteur intéressant pour un INS...

## Les modalités de garde des enfants de moins de 4 ans

En France, selon les réponses des parents à l'enquête, 13 % des enfants de moins de 4 ans sont accueillis habituellement dans une crèche en semaine dans la journée, en dehors de l'école s'ils y vont. Pour 53 % des enfants, les parents déclarent les garder eux-mêmes. Enfin, 5 % sont confiés habituellement à un grand-parent ou à un membre de la famille et 29 % à une assistante maternelle ou un autre mode d'accueil (garde à domicile, baby-sitter...). Les places en crèches sont plus nombreuses dans les grandes agglomérations, alors que le recours aux assistantes maternelles est plus fréquent dans les zones rurales ou les petites unités urbaines.

### A Saint-Étienne, on peut estimer ainsi les modes de garde des jeunes enfants :

	Nombre d'enfants de moins de 4 ans
Gardés par leurs parents ou leur famille	5 400
Gardés par une assistante maternelle (garde à domicile, baby-sitter...)	1 800
Gardés dans une crèche collective (collective, familiale)	1 300

Source : Enquête « Famille et logements »

## Vivre seul et vivre en couple

En France, deux tiers des personnes majeures déclarent être en couple, qu'elles vivent ou non avec leur conjoint. Parmi elles, 73 % sont mariées. Jusqu'à 31 ans, moins d'une personne en couple sur deux est mariée et cohabite avec son conjoint. Les unions hors mariage se sont en effet largement développées depuis une quarantaine d'années. Parmi l'ensemble des personnes en couple, 4 % ont contractualisé leur union avec leur conjoint par un Pacte civil de solidarité (Pacs) et 23 % n'ont pas officialisé leur union, notamment les plus jeunes.

Dans votre commune, 56 % des adultes vivent en couple dans le même logement.

### Nombre de couples cohabitant à Saint-Étienne

Couples mariés	29 500
Couples unis par un PACS	1 500
Couples ni mariés ni pacsés	6 500
<b>Total des couples « cohabitants »</b>	<b>37 500</b>

Source : Enquête « Famille et logements » - Champ : couples cohabitant

## Conclusion

- Opération inédite et audacieuse (voire risquée !).
- Le recensement reste essentiel : dans l'attente du développement de répertoires inspirés principalement des fichiers fiscaux, quelle source pourrait s'y substituer ?
- Cumul d'erreurs de diverses natures qui rendent l'estimation délicate : se contenter de l' 'à peu près'.  
Mais l'ensemble garde bonne allure !

- 1438 restitutions personnalisées diffusées aux élus : aucune critique venant d'acteurs dont on redoutait l'aptitude à déceler des invraisemblances.
  
- L'esprit de la méthode, expliqué en toute transparence aux élus, n'a pas soulevé de désapprobation, ce qui est remarquable car il est bien connu qu'on ne ressemble jamais vraiment à son voisin ...
  
- Le principe de modélisation est critiquable mais quelle est l'alternative ?
  1. ne rien produire,
  2. s'en tenir aux seules données collectées dans la commune

→ l'estimation 'classique' comptabilise :

  - 0 femme pacsée à Saint-Etienne,
  - 0 grand-mère de moins de 75 ans à Aurillac,
  - 0 enfant de moins de 4 ans gardé par une assistante maternelle à Cergy,
  - 0 enfant gardé en crèche à Lorient !!!