

LA PRÉCISION DES ESTIMATIONS DE L'INÉGALITE DES REVENUS DANS LES ENQUÊTES AUPRÈS DES MÉNAGES

Jérôme Accardo, Madior Fall

Introduction

On illustre ici l'intérêt pratique de l'indicateur d'inégalité de Gini et du calcul de sa précision dans le cadre d'une analyse de l'évolution de la distribution des revenus des ménages, en France, sur une période qui couvre approximativement la seconde moitié des années 80.

L'approche suivie est fondée avant tout sur l'utilisation comparative de plusieurs enquêtes de l'Insee auprès des ménages réalisées durant cette période.

- une première partie descriptive expose les différentes évaluations de l'inégalité des revenus. Rétenant l'indice de Gini comme indicateur privilégié, elle montre la nécessité, pour établir valablement un diagnostic sur son évolution, de déterminer la précision avec laquelle cet indicateur est estimé.

- une deuxième partie met en oeuvre une formule de linéarisation proposée par Deville (1995) pour évaluer la variance de sondage du Gini du revenu dans les enquêtes-ménages de l'Insee.

- la troisième partie aborde le problème de l'absence de robustesse du Gini aux erreurs de mesure. Elle suggère que, malgré les apparences, ce problème est vraisemblablement de peu de poids pratique.

I. L'inégalité des revenus¹ dans les enquêtes ménages de l'Insee, 1984 -1993

On dispose de six familles d'enquête qui saisissent le revenu des ménages et, pour chacune, de deux réalisations, la première au milieu de la décennie 80 (1984, 1986), la seconde de 5 à 8 ans postérieure, soit :

- Budget de Famille (BDF) 1984 et 1989,
- Revenus Fiscaux (RF) 1984 et 1990,
- Logement (LOG) 1984 et 1992,
- Enquête Trimestrielle de Conjoncture (CJ) mars 1984 et mars 1990,
- Actifs Financiers (AF) 1986 et 1992,
- Situations Défavorisées (SD) 1986 et 1993

Le rapprochement de divers indicateurs d'inégalité calculés sur ces enquêtes met en évidence la difficulté à établir un diagnostic clair sur l'évolution de l'inégalité des revenus sur la période.

Comme le montre le tableau ci-après,

- d'une part les niveaux des indicateurs ne coïncident pas toujours, même pour des enquêtes de même millésime, et il paraît souhaitables de rationaliser ces écarts.
- d'autre part le sens d'évolution de l'inégalité dépend
 - i) du type d'indicateur choisi
 - ii) de l'enquête retenue

1. Il s'agit, dans tout ce qui suit et sauf mention explicite, du revenu moyen par unité de consommation d'Oxford.

Tableau 1 : Indicateurs d'inégalité, revenu/uc

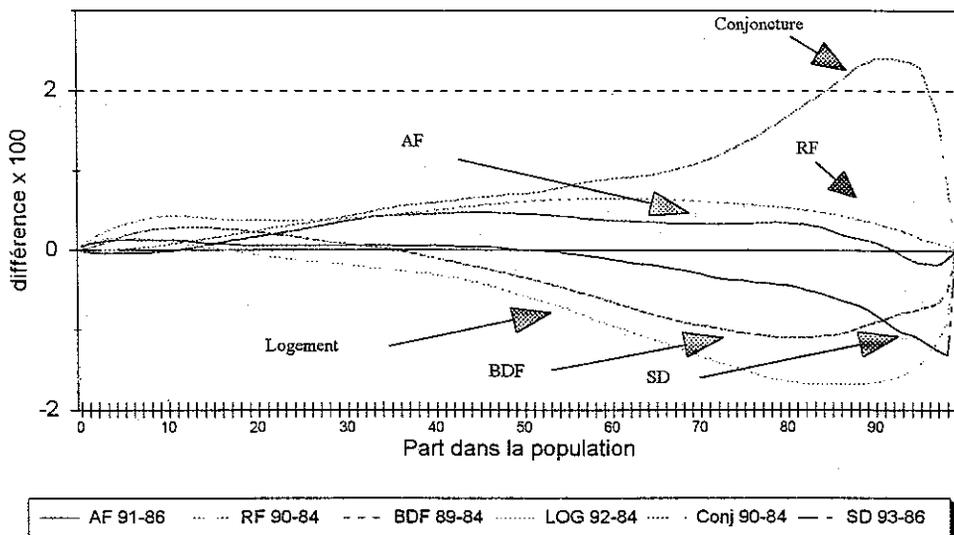
Enquêtes	D9/D1	Theil	Gini
AF 1986	4.91	0.218	0.352
AF 1992	4.78	0.214	0.347
BDF 1984	5.16	0.200	0.334
BDF 1989	4.84	0.214	0.341
RF 1984	4.15	0.235	0.341
RF 1990	3.85	0.219	0.329
LOG 1984	4.20	0.204	0.328
LOG 1992	4.39	0.227	0.342
SD 1986	4.34	0.189	0.328
SD 1993	4.08	0.213	0.332
CJ 1984	4.60	0.217	0.349
CJ 1990	4.56	0.186	0.332

L'analyse des courbes de Lorenz sur les deux vagues d'enquêtes successives confirme la complexité des évolutions (*graphique 1*).

Remarque : quelle que soit l'enquête considérée, la courbes de Lorenz de la première vague est indiscernable de la seconde (quand elles sont représentées aux formats usuels). Aussi on a choisi de représenter, pour chaque enquête, la différence des courbes de Lorenz de chacune des deux réalisations.

Sur ce graphique, on remarque que pour plusieurs enquêtes la différence des deux courbes change au moins une fois de signe (sauf pour RF où l'inégalité paraît baisser légèrement). Il paraît ainsi difficile de se prononcer, sur une base descriptive, sur le sens d'évolution de l'inégalité des revenus sur la période. Ce constat souligne l'intérêt d'une évaluation de la précision des estimateurs des indices d'inégalité fournis par les enquêtes auprès des ménages. On se borne, dans ce qui suit, au cas du Gini.

**Graphique 1 : Différence de courbes de Lorenz
revenu par uc**



II. Estimation de la précision de la mesure d'enquête du Gini.

On expose ici succinctement les procédures de calcul de la précision pour l'écart-type de sondage.

Le calcul de variance de sondage utilise la formule de linéarisation proposée par J.C. Deville avec les étapes suivantes (les calculs sont effectués sur les enquêtes LOG 1992 et SD 1993)

I. L'estimation pratique de la précision du coefficient de Gini dans les enquêtes "Logement 1992" et "Situations Défavorisées 1993".

1) Le calcul de la précision repose en premier lieu sur une linéarisation du coefficient de Gini due à J.C. Deville (Deville, 1995)².

² Deville, Jean-Claude (1995), "Estimation de la variance du coefficient de Gini mesuré par sondage", Document de Travail n° F9510, Insee.

On suppose ici qu'on dispose de n observations du revenu, (y_i) , $i=1, \dots, n$, tirées selon un certain plan de sondage. Soit (w_i) , $i=1, \dots, n$ les poids de sondage associés (i.e. l'inverse des probabilités de tirage). La procédure proposée par Deville consiste à :

- a) ordonner les observations dans le sens croissant. Soit (z_i) , $i=1, \dots, n$ la séquence ainsi obtenue.

- b) calculer les cumuls $W_k = \sum_{i=1}^k w_i$ et $Z_k = \sum_{i=1}^k w_i z_i$

- c) calculer les totaux N (= la population totale) et Z (la masse totale des revenus).

- d) calculer $R = S/Z$, où $S = \sum_{i=1}^n z_i w_i \frac{W_i}{N}$. On en déduit la valeur du coefficient de Gini par $G=2R-1$.

- e) calculer les n valeurs $u_k = F_k(z_k - Z_k/W_k) - Rz_k$, où $F_k = (W_k - 0.5w_k)/N$, autrement dit F_k est la fonction de répartition empirique des (z_k) , légèrement lissée.

Il suffit alors de calculer la variance V du total $\sum_{k=1}^n w_k u_k$. L'écart-type du coefficient de Gini est alors

$$\sigma = 2 \frac{\sqrt{V}}{Z}$$

2) On est ainsi ramené au calcul de la variance du total (pondéré) d'une variable artificielle u_k . Les enquêtes considérées sont tirées selon un plan de sondage complexe : stratification, tirage à plusieurs degrés, en plusieurs phases. La pondération de ces enquêtes est en outre redressée par Calmar. Pour pouvoir mener à bien le calcul de variance, on est ainsi amené à faire les simplifications suivantes :

A) Dans les strates de tirage appartenant aux strates de gestion 0, 1, 2 (i.e. respectivement, les communes rurales au RP 1990, les communes appartenant à des unités urbaines de moins de 20 000 habitants au RP 1990, les communes appartenant à des unités urbaines ayant entre 20 000 et 100 000 habitants au RP 1990).

Dans chacune de ces strates on suppose que le premier degré de tirage (tirage des Unités Primaires de l'échantillon-maître) s'est fait avec remise. On peut alors facilement calculer la variance (voir P. ARDILLY, "Les techniques de sondage",

ed. Technip, 1994, p 131 et sq.) , dans chaque strate d'un total . Pour chaque UP d'une part on estime le total de la variable revenu y dans l'UP n° j par

$$Y_j = \sum_{i \in UP_j} w_i y_i .$$

(Ce calcul suppose évidemment qu'on connaît, pour chaque observation, l'UP dont elle provient). D'autre part on calcule Q_j , la probabilité d'une UP (de la strate) d'être tirée (à chacun des tirages du tirage avec remise), soit $Q_j = (\text{Nombre de logements dans l'UP})/(\text{nombre de logements dans la strate})$.

Soit alors

$$\bar{Y} = \frac{1}{m} \sum_{j=1}^m \frac{Y_j}{Q_j} ,$$

où m est le nombre d'UP tirées (dans la strate considérée). La variance (intra-strate) de Y est alors donnée par

$$V(Y) = \frac{1}{m(m-1)} \sum_{j=1}^m \left(\frac{Y_j}{Q_j} - \bar{Y} \right)^2$$

Il suffit alors de sommer ces variances sur l'ensemble des strates considérées.

B) Dans les strates de gestion 3 et 4 (i.e. respectivement les communes appartenant à des unités urbaines de plus de 100 000 habitants au RP 1990, sauf l'unité urbaine de Paris, et l'unité urbaine de Paris) : on considère ici que, dans chacune de ces deux strates, le sondage est un sondage aléatoire simple, chaque observations étant tirée avec la probabilité : n/N , où n est le nombre d'observations dans la strate et N le nombre total de logements dans la strate.

Pour chacune des deux strates, la variance est alors classiquement donnée par

$$V(Y) \approx \frac{1-f}{f^2} \sum (y_k - \bar{y})^2$$

où f est le taux de sondage dans la strate.

On obtient ainsi (en francs courants) :

Logement 1992 : écart-type du revenu moyen=679 F , écart-type du Gini=0.0058

Situ.Défav1993 : écart-type du revenu moyen=772 F, écart-type du Gini=0.0074

3) Si on admet que les ordres de grandeur obtenus sont généralisables à toutes les enquêtes-ménages analysées dans ce travail, il peut être intéressant de déterminer l'analogie du design effect (ou effet de sondage). Posons

$$DEFF = \frac{\sigma_{\text{sondage}}}{\sigma_{\text{1 passe}}}$$

où le numérateur est l'écart-type calculé comme ci-dessus, et où le dénominateur désigne l'écart-type obtenu en supposant que les observations ont été tirées en une passe, selon l'inverse de leur pondération effective (i.e. celle, éventuellement issue de Calmar, qui apparaît dans le fichier d'enquête final). Cet écart-type est facilement calculable en utilisant une formule asymptotique fournie par Rosen (1991), qui évite le calcul des probabilités d'inclusion double. On obtient alors :

Logement 1992 : DEFF du revenu moyen = 0.77 , DEFF du Gini = 0.77

Situ.Défav1993 : DEFF du revenu moyen = 0.50, DEFF du Gini = 0.50

4) Conclusion : on retiendra 0.007 comme écart-type du Gini dans les différentes enquêtes ménages. Un rapide calcul de linéarisation à partir de l'écart-type du revenu moyen fournit, pour l'indice de Theil, un écart-type de 0.015 environ.

Remarque : On a procédé à une simulation numérique pour vérifier la solidité de l'approche par la linéarisation. Dans le cadre certes limité de cette simulation, les résultats apparaissent tout à fait satisfaisants. En outre l'approche par linéarisation présente des performances supérieures à une approche de type bootstrap.

1) Simulations

On a constitué une population artificielle de ménages en empilant les différents fichiers d'enquête analysés ici (RF, AF, LOG, ...) ; on obtient ainsi une population d'environ 200 000 ménages. Chaque ménage est affecté de la probabilité d'inclusion égale à l'inverse de son poids de sondage tel qu'il apparaît dans l'enquête. Les probabilités d'inclusion présentent donc une dispersion considérable (cf. Document de travail n° F 9602, Insee, section I).

On tire alors, d'abord en utilisant un tirage uniforme, pour différents taux de sondage, 50 échantillons. pour chacun de ces échantillons, $k=1, \dots, 50$ on calcule

a) le coefficient de Gini, $G(k)$

b) la variance $\text{varD}(k)$ de ce coefficient, telle qu'on la calcule par la formule de linéarisation de Deville. On pose alors $\text{VD} = \text{moyenne des } \text{varD}(k), k=1, \dots, 50$.
Soit $\sigma_d = \sqrt{\text{VD}}$ l'écart-type correspondant.

c) la variance empirique VE de l'échantillon des $G(k)$ et $\sigma_e = \sqrt{\text{VE}}$.

On effectue ensuite ces calculs pour des tirages selon les probabilités d'inclusion du fichier, en faisant varier le taux de sondage.

Tableau 2 : vérification du calcul de variance par linéarisation

Taille de l'échantillon	Type de pondération	Écart type empirique (σ_e)	Écart type Deville (σ_d)
102	uniforme	0.13	0.93
1 017	uniforme	0.04	0.05
4 067	uniforme	0.02	0.02
2 180	inégaie	0.09	0.08
4 356	inégaie	0.06	0.06
8 711	inégaie	0.04	0.04

2) Bootstrap

La lourdeur des calculs impliqués par le bootstrap impose de se borner à des sous-populations de taille réduite. On a ainsi retenu des tranches d'âge quinquennale. Les calculs ont été menés sur LOG 1992 et sur RF 1990. Bien entendu il n'est ici pas tenu compte de la stratification : on suppose ici que le tirage est "en une passe". On constate qu'une estimation de la variance par bootstrap fournit une évaluation un peu trop "optimiste" de la vraie variance, et biaise ainsi l'inférence en faisant trop souvent considérer comme significatives les évolutions des indicateurs d'inégalité.

Tableau 3 : Comparaison du calcul de variance par bootstrap et par linéarisation

âge	Rev Fiscx			Logement		
	Gini	σ_b	σ_d	Gini	σ_b	σ_d
25-30	0.249	0.018	0.008	0.287	0.013	0.008
30-35	0.271	0.014	0.006	0.325	0.013	0.006
40-45	0.309	0.012	0.008	0.341	0.013	0.007
45-50	0.322	0.015	0.011	0.361	0.013	0.009
65-70	0.282	0.018	0.012	0.339	0.019	0.008
70-75	0.262	0.024	0.010	0.332	0.022	0.010

Note : σ_b = écart-type calculé par bootstrap, σ_d = écart-type selon la formule de linéarisation de Deville

On déduit alors de ce calcul de précision précédent :

- i) les écarts de niveau du Gini entre les enquêtes n'apparaissent pas incompatibles avec la variance de sondage
- ii) on peut proposer un test de significativité des évolutions du Gini, dans chaque enquête.

Tableau 4 : p-values d'un test de significativité des évolutions de l'inégalité

	AF	RF	BDF	LOG	SD	CJ
Δ Gini	- 0.005	- 0.012	+ 0.007	+0.014	+ 0.004	- 0.017
p-value	30 %	11%	24 %	8 %	34%	4 %

Note : pour chaque enquête on calcule la variation du Gini entre ses deux réalisations. On peut alors calculer, sous l'hypothèse que cette variation suit une loi normale, de moyenne nulle et de variance $2\sigma^2$, la probabilité P d'une variation supérieure ou égale en valeur absolue à l'écart observé.

Au vu de ces résultats, il ne paraît pas possible de conclure à une variation significative de ces indicateurs : elles sont en effet contradictoires et seule CJ (dont on a pu noter précédemment le caractère atypique, cf. *Graphique 1*), atteint le seuil usuel de rejet (5%).

III. La question de la robustesse du Gini

1. La critique théorique

Le calcul de la variance de sondage du Gini clôt-il la discussion ? Malheureusement non. En effet, aux aléas de sondages viennent s'ajouter les erreurs de mesure. Or, le Gini, comme la plupart des indicateurs d'inégalité, se révèle peu robuste à certaines erreurs de mesure ; plus précisément, considérons les trois types suivants d'erreur sur la mesure du revenu : si on note x le vrai revenu du ménage et \hat{x} sa réponse telle qu'elle apparaît dans le fichier final,

(i) le ménage se trompe sur la période considérée : il déclare son revenu mensuel alors qu'on l'interroge sur son revenu annuel (i.e. $\hat{x} = x/12$) ou l'inverse (i.e. $\hat{x} = 12x$)

(ii) le ménage se trompe sur l'unité et répond en anciens francs : $\hat{x} = 100x$

(iii) lors du processus de saisie, des erreurs de report se produisent qui modifient la réponse du ménage de l'ordre d'une décimale : $\hat{x} \approx 10^n x$, avec $n = -2, -1, 1, 2, \dots$

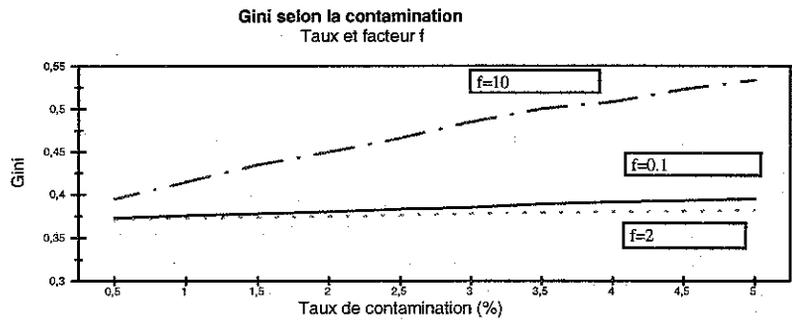
Dans ce cas, des taux d'erreur même faibles, inférieurs à 5 % par exemple, peuvent affecter très sensiblement la valeur du Gini. Soit ainsi la distribution des revenus de l'enquête « Actifs Financiers 1992 » portant sur 9530 ménages³. Le Gini est de 0.37. Si on la perturbe en multipliant par $f=10$ le revenu déclaré d'une proportion $\tau=5\%$ des ménages, le Gini passe à 0.55 (le Theil passe de 0.23 à 0.69).

Les graphiques suivants présentent l'effet de ce type de contamination sur le Gini (pour différents taux de contamination τ et trois facteurs d'erreur $f=0.1, f=2, f=10$) dans deux cas :

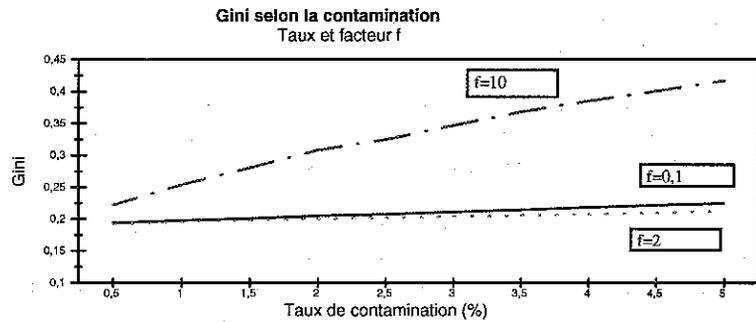
- pour la distribution observée, de niveau de concentration 0.37 (*graphique 2-1*),
- pour la même distribution, mais transformée par l'opérateur $\sqrt{\cdot}$, de façon à obtenir une distribution moins concentrée, de $\text{gini}=0.19$ (*graphique 2-2*).

³ Il s'agit, ici, du revenu global du ménage.

Graphique 2.1 : (distribution observée)



Graphique 2.2 (distribution modifiée)



On peut alors constater en particulier que :

α) si le facteur d'erreur est élevé ($f=10$), une contamination très faible (de l'ordre de 1 %) à déjà une incidence notable sur le Gini.

β) en revanche, quand le facteur d'erreur est petit, cette incidence reste modérée, même pour des contaminations notables.

γ) il n'y a pas de relation monotone entre la valeur du facteur d'erreur et l'effet de la contamination sur le Gini (ceci étant essentiellement dû au fait que le Gini est plus sensible à ce qui se passe dans le haut de la distribution).

δ) les effets de la contamination apparaissent dépendre peu de la concentration initiale de la distribution.

Pour plus de détail sur ce type d'approche, voir Cowell, Victoria-Feser (1996)⁴.

De façon très similaire, on peut arguer que les fractiles extrêmes (par exemple les 1 % des ménages les plus pauvres et les 1 % des ménages les plus riches) de la distribution des revenus sont très mal appréhendés : la mauvaise perception (et donc mauvaise déclaration) des revenus socialisés (y.c. les remboursements maladie) est particulièrement gênante chez les plus pauvres dans la mesure où, chez ces ménages, ces revenus constituent une part relativement plus importante des ressources, chez les plus riches les revenus du patrimoine sont mal mesurés, les bénéfices sont souvent difficilement isolables du chiffre d'affaires, quand il s'agit d'indépendants, etc.

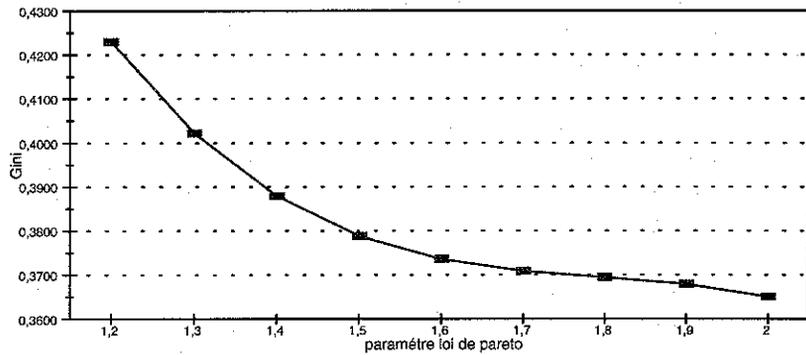
On peut ainsi, par exemple dans le cas du haut de la distribution, se demander quel est l'effet sur le Gini de l'incertitude sur la forme exacte de la distribution dans cette zone des plus hauts revenus. Par exemple en expérimentant l'incidence de différentes hypothèses quant à cette forme. On constate que le Gini, là encore, varie sensiblement selon la forme retenue.

Soit ainsi, dans l'enquête « Actifs Financiers 1992 », le dernier centile des revenus : il est composé d'environ 200 000 ménages qui disposent chacun de plus de 352 000 F (par uc). Le plus haut revenu (par uc) mesuré est : 1 177 000.

On peut se défier de la distribution empirique des revenus dans ce dernier centile, et préférer supposer d'autres distributions ; à titre d'exemple, on a supposé que cette distribution était en réalité une loi de Pareto, de seuil 352 000 F, et de paramètre α , α étant choisi pour que la moyenne des revenus dans le dernier centile soit compatible avec l'ordre de grandeur de la moyenne empiriquement observée dans le dernier centile. On a ainsi fait varier α entre 1.2 et 2. Pour un α donné, on a tiré 10 loi de Pareto, obtenant ainsi 10 distributions alternatives des revenus dans le dernier centile et donc, 10 distributions globales des revenus qui ne diffèrent que par leur queue. On a calculé alors la moyenne des 10 Gini correspondants. Le *graphique 3* montre que pour des valeurs de α proches de 1.2, les revenus apparaissent plus concentrés que ce que mesure l'enquête. L'effet est cependant sensiblement moins fort que celui, vu ci-dessus, des erreurs de mesure.

4 Cowell, Frank A. and Maria-Pia Victoria-Feser (1996), « Robustness Properties of Inequality Measures », *Econometrica*, Vol. 64, n° 1, 77-101

Graphique 3 : Simulation Gini
Dernier centile suit la loi de Pareto



Des analyses de ce genre paraissent remettre en cause l'utilisation habituelle du Gini, même suppléée par le calcul de la variance de sondage : elles impliquent en effet que l'estimation usuelle est extrêmement imprécise, ce qui en particulier interdit pratiquement tout diagnostic quant à l'accroissement ou la baisse des inégalités entre deux dates. Dans cette optique la variance de sondage ne représenterait qu'une borne (très) inférieure de l'erreur probable, sans intérêt pratique.

2. Le constat empirique

Pourtant, de façon surprenante étant donné ce qui précède, les estimations du Gini de la distribution des revenus effectuées dans les diverses enquêtes auprès des ménages, en France, ne présentent pas une dispersion particulièrement frappante.

Tableau 5 : Évolutions du Gini entre différentes enquêtes

	Actifs Fin.	Rev. Fiscx	Budg. Famil.	Logement	Sit. défav.	Enq. Conj.	Ecart Max.
Gini - 1	0.352	0.341	0.334	0.328	0.328	0.349	0.024
Gini - 2	0.347	0.329	0.341	0.342	0.332	0.332	0.018
Évolution	+ 0.005	- 0.012	+ 0.007	+ 0.014	+ 0.004	- 0.017	

La dernière colonne (Ecart Max) donne, pour chaque vague, la différence maximale observée entre les Gini d'une même période. Il est clair qu'on ne retrouve pas, ici, la très forte variabilité suggérée par l'analyse de (manque de) robustesse précédente. A chaque période, l'écart observé est au plus de 0.024 (c'est aussi l'écart maximum entre toutes les enquêtes du tableau).

A l'inverse, sous l'hypothèse de constance de la distribution réelle des revenus sur la période couverte par les enquêtes d'une même vague, l'écart observé est compatible avec l'intervalle de confiance à 95 % prédit par la variance de sondage, $[g-2\sigma, g+2\sigma]$, avec $\sigma=0.007$.

En toute rigueur, ce constat est insuffisant pour invalider les conclusions de l'analyse de robustesse : on peut en effet soutenir que l'apparente convergence des Gini résulte de ce que toutes les enquêtes sont affectées de façon comparable par les erreurs de mesure. Le niveau serait donc grossièrement erroné. En outre, les fortes variations dues, en principe, aux erreurs de mesure interdiraient de tabler sur une constance de leurs effets, donc interdiraient une utilisation du Gini (en tous cas de ses estimateurs empiriques usuels) dans une analyse en évolution. Voir Cowell, Victoria-Feser, op.cit.

Pourtant, encore une fois, le rapprochement de différentes enquêtes suggère des conclusions nettement plus optimistes. Les enquêtes ne sont en effet pas également exposées aux trois types (i) (ii), (iii) d'erreur de mesure mentionnés plus haut. Si, par exemple, une erreur de type (iii) (où une erreur de saisie a divisé ou au contraire multiplié par 10 le revenu déclaré par le ménage) est plausible dans une enquête comme Budget de Famille ou Revenus Fiscaux, elle est tout à fait improbable dans Actifs Financiers où le ménage est invité à déclarer un revenu en tranche. De même une erreur de type (i) (le ménage déclare un revenu annuel à une question sur son revenu mensuel ou vice et versa) est possible dans Actifs Financiers, déjà moins probable dans Budget de Famille, enquête dont le questionnaire sur le revenu est très détaillé et réduit très certainement ce genre de confusion. Elle est très peu vraisemblable dans le cas de Revenus Fiscaux, où l'on peut raisonnablement penser que les ménages font particulièrement attention à ne pas multiplier fortuitement par 10 leur revenu fiscal !⁵

On peut ainsi proposer le tableau suivant qui indique, de façon qualitative, la probabilité a priori des différents types d'erreurs sur les six familles d'enquêtes citées supra.

⁵ Il est possible certes qu'ils fraudent...mais, dans ce cas, on observe $\hat{x} \ll x$, et on a vu que l'effet de ce type d'erreur était relativement négligeable.

Tableau 6 : Types d'erreur plausibles dans les 6 enquêtes.

	Actifs Fin.	Rev. Fiscx	Budg. Famil.	Logement	Sit. Défav.	Enq. Conj.
erreur (i)	+	-	?	+	?	+
erreur (ii)	?	-	?	+	?	?
erreur (iii)	-	+	+	+	+	-

Légende :

- « + » indique que l'erreur est a priori plausible, dans une proportion non négligeable (de l'ordre de quelques %),
- « ? » indique qu'il est difficile de se prononcer a priori sur la plausibilité de l'occurrence de ce type d'erreur,
- « - » indique que ce type d'erreur est a priori très peu probable.

On voit ainsi apparaître deux cas polaires :

- l'enquête Revenus Fiscaux, la moins susceptible d'être entachée d'erreur de mesure
- l'enquête Logement, dans laquelle le statut relativement mineur de la partie « revenus » du questionnaire (elle n'a pour fonction essentielle que de fournir une indication de la capacité du ménage à investir dans un logement) suggère a priori une plus forte proportion d'erreur de différents types.

Si on admet alors que les 3 types d'erreur sont équiprobables, de taux α , Logement 1992 contient vraisemblablement une proportion 3α d'erreurs et Revenus Fiscaux seulement α . L'écart de Gini observé est : $\Delta=0.013$ qui doit ainsi provenir du différentiel de contamination 2α . D'après les graphiques 2-1 et 2-2, indiquant le niveau du Gini en fonction du taux de contamination α , on a au plus $\alpha \approx 0.3 \%$, autrement dit une contamination globale (de facteur 10 et plus) de très faible ampleur, dont l'effet sur le niveau du Gini est pratiquement négligeable.