

30 December (Atlantico)

[Thomas Piketty, une autorité incontestée ? : mais pourquoi personne ne s'intéresse au débunkage de ses données sur les inégalités ? | Atlantico.fr](#)

MANQUE DE CURIOSITÉ ?

## Thomas Piketty, une autorité incontestée ? : mais pourquoi personne ne s'intéresse au débunkage de ses données sur les inégalités ?

Outre-Atlantique, de nombreux économistes ont souligné que les données sur lesquelles Thomas Piketty a construit sa grande thèse sur le creusement des inégalités sociales ont mis en lumière les limites et inexactitudes des données statistiques.



L'économiste Thomas Piketty lors d'une séance photo pour l'AFP.

[Vincent Geloso](#)

**Atlantico Rédaction** : Dans sa recension de l'ouvrage de [Thomas Piketty et Julia Cagé, Le Monde](#) (où il est chroniqueur et elle présidente de la société des lecteurs) affirme qu'il est l'auteur « d'ouvrages faisant autorité ».

Manque de curiosité...?

Le Monde oublie en effet soigneusement de préciser que les données statistiques sur lesquelles Thomas Piketty a construit ses précédents livres ont largement été débunkées outre-Atlantique...

En résumé, Piketty s'appuie sur des données fiscales américaines dont il a été solidement établi qu'elles étaient incomplètes et donc absolument pas appropriées pour mesurer l'évolution des inégalités de revenus.

Au regard de la place qu'a pris le discours sur le creusement des inégalités dans le débat public (au mépris le plus souvent des faits en ce qui concerne la France), est-il anodin que le « quotidien de référence » n'ait aucun recul critique sur son économiste star... ?



## **Atlantico : Dans quelle mesure est-ce que les travaux de Zucman (et Piketty et Saez) sont-ils devenus une référence dans leur domaine (et la base de nombreux travaux) ?**

Vincent Geloso : Tout d'abord, il est clair que Zucman, Piketty et Saez sont devenus la première source de référence dans leur domaine. Ceci dit, ce n'est pas tant en raison de la qualité, c'est largement en raison de la politisation de leurs travaux. Devenus incontournables dans les débats publics, les économistes universitaires se retrouvent obligés de les citer. Cependant, je vous dis aussi que depuis 10 ans environ, ils ont graduellement perdu en faveur parmi les académiques. Plusieurs travaux ont remis en question la qualité de leurs estimations et plusieurs se sont mis à les critiquer féroce­ment (notamment Larry Summers qui est l'ancien président de l'Université Harvard et Wojtek Kopczuk qui est l'éditeur en chef de Journal of Public Economics). Ainsi, ils sont vus avec une certaine suspicion. Ce qui est particulièrement intéressant, c'est que les critiques couvrent la totalité de l'échiquier politique. Je suis moi-même un libéral de la vieille espèce alors que Summers est un démocrate américain qui est plus centriste alors que d'autres comme Kopczuk et James Galbraith se retrouvent plus à gauche. Avant son décès, il y avait aussi Richard Sutch de UC-Davis (qui avait féroce­ment critiqué Piketty et Saez) qui se situait plutôt à gauche.

La partie triste de cela c'est que la profession joue un double jeu -- celui de la petite politique et celui du débat académique. Les deux se mélangent fréquemment et c'est pour cela que je pense que Zucman, Piketty et Saez réussissent à s'en sortir avec ce qui aurait coûté la carrière d'un chercheur plus junior ou d'un chercheur "moins à la mode" (voir prochaine réponse). Finalement, il faut ajouter que la majorité des académiques ne parlent pas notre chère langue de Molière. Comme Zucman, Piketty et Saez se permettent souvent d'exprimer leurs opinions politiques plus extrêmes dans la presse française, les académiciens qui ne connaissent pas le français ne sont pas au courant de l'étendue des biais de Zucman, Piketty et Saez. En consé­quent, il est plus difficile d'appliquer un taux d'escompte à leurs arguments/résultats.

Dans quelle mesure cela pose-t-il problème que leur travaux doivent être pris avec autant de précaution que vous le démontrez ?

Vincent Geloso : Zucman, Piketty et Saez se pavanent fréquemment dans les médias. Ainsi, Zucman, Piketty et Saez trouvent souvent des résultats qui frappent l'imaginaire politique. Cependant, lorsqu'on creuse, on réalise qu'ils vont de pair avec l'idéologie qu'ils confessent. Dans mon cas particulier, je vous admet que mes résultats sont en défaut avec mon idéologie. Étant libéral, j'aimerais bien que les tendances d'inégalités n'aient point augmenté depuis 1980. Mais elles ont augmenté. Ceci dit, ce n'est pas vrai que le niveau est plus élevé qu'au sommet des années 1920. Ce n'est pas vrai non plus qu'il y a une explosion des inégalités (elles semblent avoir augmenté beaucoup entre 1970 et 1990 et s'être stabilisées depuis). Il n'est pas vrai que ce sont les vastes augmentations de taxe qui ont permis de réduire les inégalités. Plutôt, il semble que ce nivellement est le résultat de la Grande Dépression et il faut bien admettre que l'égalité grandissante dans la pauvreté grandissante (ce qu'était la Grande Dépression) ne constitue en rien un fait qui mérite d'être répliqué! Ajoutez les nuances que je fais et vous voyez que le résultat époustouflant qui permet à Zucman, Piketty et Saez de s'immiscer dans l'imaginaire collectif disparaît entièrement. Imaginez si ils avaient présenté un résultat plus nuancé (i.e., plus proche de la vérité). Est-ce-que Zucman, Piketty et Saez auraient été aussi présents dans la sphère médiatique? J'en doute fortement. Cependant, c'est pour cela qu'il faut les considérer avec précaution -- ils veulent avoir raison sur deux plans -- a) le plan académique et b) le plan politique. Lorsqu'il y a conflit entre les deux, je pense qu'ils donnent prédominance au second. Zucman, Piketty et Saez vivent bien trop dans la sphère entre la politique (et les politiques qu'ils cotôient) et l'académique. Si je suis partisan de parler à tout le monde (gauche et droite), il faut garder une muraille de Chine entre l'université et les politiques lorsqu'on fait nos recherches. La muraille de Zucman, Piketty et Saez est un fromage gruyère qui est à moitié mangé par les souris!

**Vous avez mené différents travaux universitaires critiques de ces analyses dont [How Pronounced is the U-Curve? Revisiting Income Inequality in the United States, 1917–60](#). Que peut-on exactement reprocher aux auteurs ?**

**Vincent Geloso** : Soyons clairs : les travaux de Gabriel Zucman sont à prendre avec des pincettes. En grande partie parce que lui et ses collègues ont été très négligents. Je ne vais pas mâcher mes mots et énumérer la litanie des négligences.

Tout d'abord, vous devez comprendre la chose suivante : le travail de Zucman avec Piketty et Saez s'appuie fortement sur un article publié en 2003 dans le Quarterly Journal of Economics. J'ai réexaminé les données, les hypothèses et les méthodes de cet article dans deux articles publiés - l'un à Economic Inquiry et l'autre à The Economic Journal.

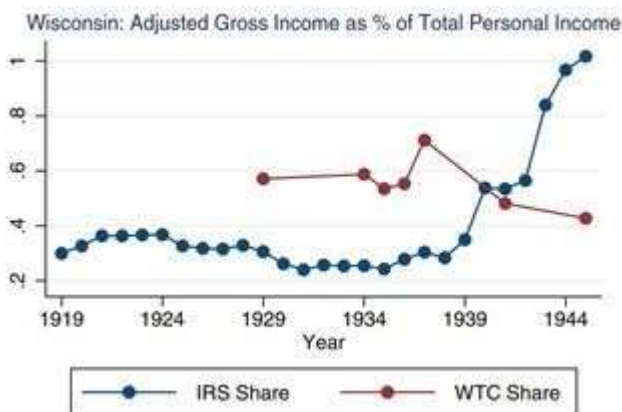
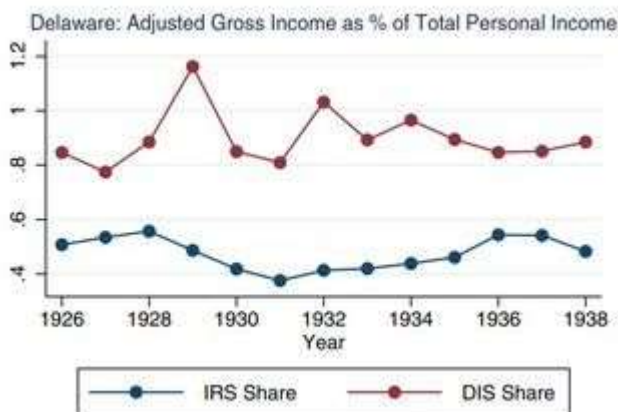
Sans cet article, les autres ne suivent pas. Il s'agit de l'élément de base essentiel et le fait d'y apporter des modifications modifie les autres articles (sur la fiscalité optimale, sur la progressivité, sur l'inégalité des richesses).

Permettez-moi également d'expliquer que lorsque j'ai commencé ces articles avec [Phil Magness](#) en 2016, je pensais que nous améliorions simplement une série de données. Cependant, plus je me plongeais dans les détails et les archives, plus je me rendais compte qu'il y avait un immense laisser-aller.

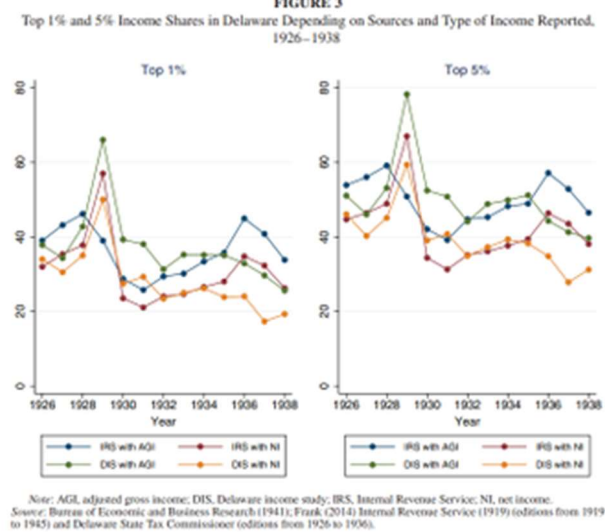
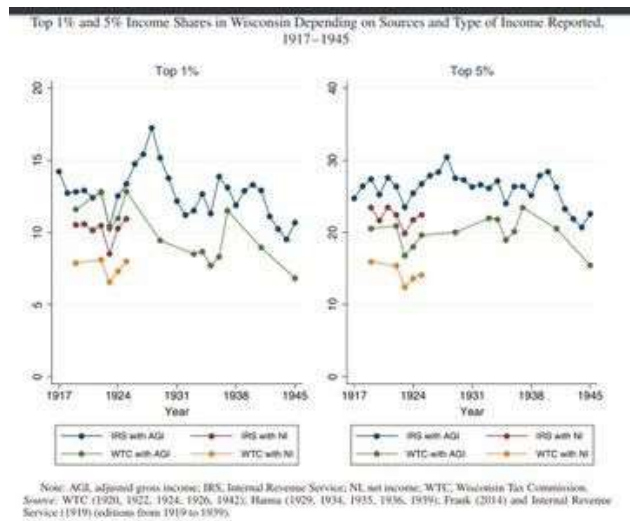
Et je vais me concentrer sur la période antérieure à 1960 parce que [David Splinter et Gerald Auten](#) ont montré des problèmes bien plus graves après 1960 que ceux que je montre avant 1960.

Commençons par la qualité des données elles-mêmes. Jusqu'à la Seconde Guerre mondiale, les revenus étaient déclarés au niveau fédéral aux États-Unis. Par conséquent, les revenus étaient sous-déclarés dans les déclarations fiscales. Cependant, ce qui est moins connu, c'est que la sous-déclaration était beaucoup plus importante dans les classes de revenus inférieures et j'ai remarqué qu'il y avait eu des tonnes d'articles dans les 5 meilleures revues au cours de la période 1940-1970 qui utilisaient les systèmes d'impôt sur le revenu de l'État pour créer des estimations au niveau de l'État. Nous avons donc vérifié si ces articles faisaient état de revenus similaires à ceux de l'IRS.

Ils ne l'ont pas fait (voir les deux graphiques du bas). Ils ont déclaré plus de revenus et proportionnellement plus dans les classes à faibles revenus.

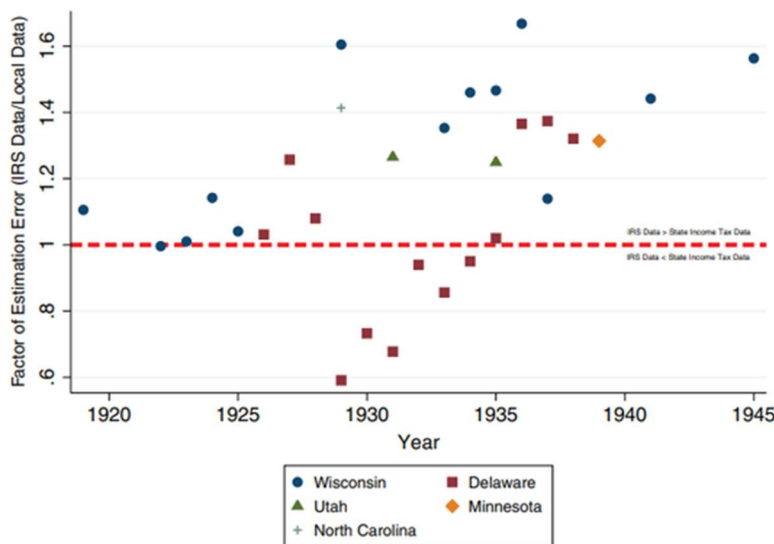


La raison en est que l'IRS contrôle davantage les hauts revenus que les bas revenus. Les rapports sont donc de meilleure qualité au sommet de l'échelle. Nous avons utilisé les données de deux États (le Wisconsin et le Delaware) qui disposaient de systèmes de contrôle très fermes sur l'ensemble de l'échelle des revenus. Et nous avons constaté que les niveaux d'inégalité étaient parfois inférieurs de plus de 5 points (pour les 1 % les plus aisés) à ceux de Piketty et Saez.



Mais comme nous ne sommes pas paresseux, nous nous sommes demandé si le Delaware et le Wisconsin n'étaient pas des bizarreries. Et si nous utilisions TOUS les États que nous pouvons trouver ? Encore une fois... même... schéma. Les données de l'IRS surestiment les inégalités. Globalement, en moyenne, les données de l'IRS sont supérieures de 1,18 aux données correctes et ce chiffre est faible parce que nous ne remettons même pas en question leurs hypothèses. Le graphique ci-dessous montre l'ampleur de la surestimation par l'IRS.

**FIGURE 5**  
Ratio of IRS Data Results to State Income Tax–Data Results (Top 1% Only), 1919–1945



À aucun moment, que ce soit dans l'article de 2003 ou depuis, Zucman avec Piketty et Saez n'ont envisagé que leurs données puissent être faibles. Aucun effort n'a été fait pour documenter les limites des données dues aux incitations.

À aucun moment -- dans le document de 2003 ou depuis -- Zucman, Piketty et Saez n'ont considéré que ses données pouvaient être insuffisantes. Aucun effort n'a été fait pour documenter les limites des données dues aux incitations.

Premièrement : Il est nécessaire d'ajuster les données manquantes car souvent moins de X% (X=fractile de choix, 10,5,1 etc.) des unités fiscales ont rempli une déclaration d'impôts. Les auteurs procèdent à un ajustement basé sur le marché du travail en 1942, en faisant valoir qu'il existe un échantillon de l'IRS qui leur permet de le faire.



Deux choses : 1942 est une année de guerre et l'ajustement sera faussé par la guerre. Deuxièmement, dans la même source qu'ils utilisent, il est fait mention d'un autre ensemble de données plus important, de meilleure qualité et portant sur l'année fiscale 1941. Il y a beaucoup moins de distorsions et c'est mieux.

Le simple fait de passer à 1941 change beaucoup de choses. Suffisamment pour enlever un point à la part des 10 % de revenus les plus élevés avant 1940.

Deuxièmement : avant 1943, l'IRS communique les données en termes de revenu net. Il y a donc un besoin évident d'ajuster le revenu brut. PS a simplement fait des hypothèses arbitraires qu'il a qualifiées de raisonnables et qu'il n'a pas détaillées (que vous ne pouvez trouver que dans leur fichier excel... en français (que je parle)).

>>> les taux de déductions/ajustements à prendre en compte:  
 (ces taux n'ont pas été obtenus par extrapolations: je me suis contenté de reprendre des valeurs raisonnables à partir des tabcomp reproduits sur la feuille Comp1665 et des fichiers Stata; les erreurs sont très faibles anyway) (j'ai retenu des taux de déductions pas trop élevés pour 1913-1943, afin de prendre en compte les pb de loss comptés en déduction, notamment dans les 30s; cf. feuille ResComp1698) (pour 1944-1998, je prend des taux d'ajustements uniformes égaux aux taux agrégés (cf. feuille AGI1944-1998), ce qui est conforme à ce que l'on observe dans les 1990s; cpt, lors des années de tx de déduc à 4-4,5% des mid-80s,

Cependant, les hypothèses sont beaucoup trop libérales. Même si l'on comptabilise TOUTES les déductions de l'IRS et qu'on les compare aux déductions implicites (sans hypothèses sur la répartition entre les différentes catégories de revenus), le revenu brut implicite de la FP est presque toujours supérieur au revenu brut total de l'IRS.

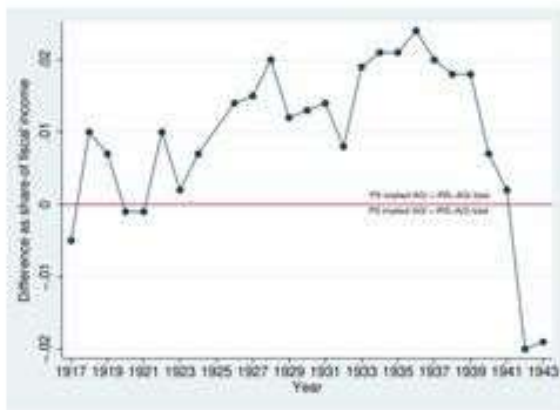
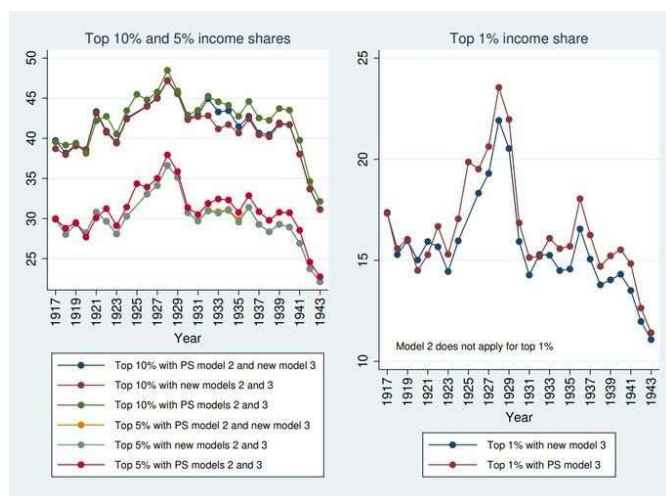
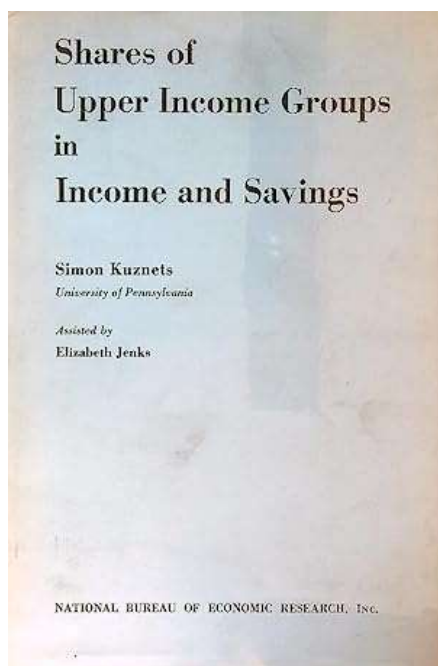


Figure 7. Difference between implied deductions from PS model 3 multipliers and AGI-eligible deduction IRS totals

De plus, dans les mêmes sources qu'ils utilisent, l'IRS fournit suffisamment de données pour créer des ajustements afin de convertir le net en brut au niveau de la classe de revenu. Cela permet d'enlever quelques points supplémentaires



Troisièmement, Piketty et Saez poursuivent en dénigrant le travail de Simon Kuznets sur le dénominateur des revenus pour estimer l'inégalité, mais il oublie de remarquer qu'il consacre beaucoup de temps, dans son livre, à expliquer comment les employés de l'État et des collectivités locales n'étaient pas obligés de déclarer leurs impôts avant 1939.



C'est important puisqu'ils passent d'environ 2 % de l'emploi à 5 % de l'emploi et que leur masse salariale semble augmenter encore plus rapidement. Cela réduit d'environ 1 à 2 points la part de revenu des 10 % les plus riches. Cela est dû au fait que je n'ai pas lu correctement les sources historiques.

Quatrièmement, ils doivent créer un dénominateur de revenu pour diviser les revenus les plus élevés. Pour ce faire, ils supposent qu'il s'agit de 80 % du revenu personnel (PNIA) moins les transferts. Cependant, ils commettent quatre grosses erreurs.

4a) Ils n'ont pas remarqué que Hawaï et l'Alaska sont inclus dans les données de l'IRS mais pas dans le PNIA. Ainsi, le numérateur et le dénominateur ont une couverture différente. Il s'agit d'une petite erreur, mais d'un manque de rigueur.

4b) Ils justifient la création de leurs dénominateurs au lieu de les relier à ceux de Simon Kuznets. Pour ce faire, ils ridiculisent Kuznets en soulignant que sa méthode signifie qu'en 1948, les non-déclarants avaient des revenus plus élevés que les déclarants. C'est invraisemblable, disent-ils.

Ils affirment que la méthode de Kuznets n'a été utilisée que pour les années 4b-cont), mais que cela n'est vrai que pour 1948. Pour toutes les autres années, Kuznets n'a pas ce problème.

PS attribute this discrepancy to insufficient adjustments by Kuznets, producing a denominator that is consistently too large. PS base this contention on Kuznets's calculations for 1948, which implausibly suggest that the average non-filer's income exceeded the average filer's income in a year where almost all income earners filed returns.<sup>20</sup> Given the distribution of earnings by tax filers, PS intuitively conclude that they should display larger total

earnings than non-filers. They therefore assume a similar problem affects the remainder of Kuznets's denominator, including before WWII.

The PS criticism of Kuznets initially appears reasonable when limited to the postwar period, however it is not apparent that this justification holds prior to the 1943-44 tax code and IRS accounting revisions. From 1917 to 1941, the Kuznets ratio of the implied income of non-filers to filers' income varies between 27% and 46% (with an average of 39%). It is only in 1942 that the ratio starts increasing while remaining below 100% (at 86%) and in 1943 that it surpasses 100%.<sup>21</sup> The pre-1944 proportions for Kuznets are close to the post-1944 assumptions of PS regarding non-filer income (imputed at 50% of the average filer's income in 1944-45 and 20% thereafter).<sup>22</sup> As such, the Kuznets series still yields very plausible results for years prior to 1944 based on PS's own benchmark for imputing the income of non-filers.

4c) Ils disent qu'ils utilisent 80 % du revenu personnel moins les transferts parce que cela correspond à peu près aux chiffres de l'après-guerre. Ce n'est pas le cas ! C'est loin d'être le cas.

As a solution, PS point out that since 1944, "the ratio between total tax return gross income reported on tax returns and total personal income minus transfers estimated in national accounts has been fairly stable since the late 1940s (around 75-80%)" (Piketty and Saez, 2007, p. 169). Based on this observation and recognizing a residual gap between the two terms that derives from the small number of persons falling below the postwar tax eligibility threshold, they assume that the total income series from 1913 to 1943 is equal to 80% of total personal income minus transfers.

4c-cont) En fait, il est presque toujours supérieur à 80 % - et cela provient de leurs propres fiches techniques ! La différence entre 80 % et 82,7 % peut sembler minime. Toutefois, elle représente 1,35 point de pourcentage supplémentaire pour les 10 % les plus performants.

themselves. For the period from 1944 to 1974, their ratio of fiscal income to personal income

<sup>21</sup>We subtracted reported income from tax returns *after* the model 2 adjustments to preserve comparability. We thank an anonymous referee for pointing this out. We can also compare this method to the PS denominator and our denominator (see more below). Between 1917 and 1943, the PS data suggest that the income of non-filers was equal to 25% of the income of filers as opposed to 34% with our denominator.

<sup>22</sup>In fact, the series are probably even closer together than above. PS offered their critique based on Kuznets' unadjusted personal income figures. Once we remove Kuznets' adjustments for imputed rent and tax exempt government employees, the ratio of the income of filers to non-filers moves closer to those of PS by 4 percentage points on average.

(excluding transfers) is equal to 82.7% on average.<sup>23</sup> After 1974, the ratio falls constantly to a low of just over 77% in the late 1990s. This is not what PS implied when they said that the ratio "fluctuates between 77% and 83%" (2007, p. 169). Rather, the data suggest

4d) Ils affirment ensuite qu'ils utilisent l'hypothèse selon laquelle les non-remplaçants disposaient de 20 % du revenu des remplaçants pour créer le dénominateur après 1943. Mais ce n'est pas ce qu'ils font. Ils utilisent des proportions différentes qu'ils n'expliquent jamais et qui ne correspondent pas à leurs descriptions en annexe.

basis to link their two separately-constructed denominators across the WWII era accounting changes. To construct their denominator series, PS begin with the period from 1944 to the present. They state that they "imputed to non-filers a fixed fraction of filers' average income (50% in 1944-45 and 20% thereafter)" by linking it to the personal income component of national accounts, giving them what they call "tax return gross income" (Piketty and Saez, 2007, p. 169). Already, we find discrepancies between PS's explanation and the adjustments that appear in their calculations. Their datafile shows an adjustment equal to 50% for 1944, 60% in 1945, 25% in 1946 and 20% thereafter.<sup>19</sup>

There is nonetheless a plausible reason to begin their denominator construction in 1944. As PS explain, their post-1944 method of computing total income could not be extended to earlier years because of the high tax eligibility threshold before the war and the use of

<sup>18</sup>Differences occur because of imputed rent, interest and dividends received from pension plans, life insurance carriers, non-profit institutions, non-taxable employer and employee contributions to non-monetary forms of compensation, capital and inventory adjustments. An additional residual difference stems from low filing rates.

<sup>19</sup>The PS code files (see "AGI vs Personal" in their ComptaNat.xls) show that they used different proportions from their descriptive appendix. It is only from 1947 onward that the denominator sets the income of non-filers at 20% of filers' income. PS offer no indication for how they decided upon these percentages.

Phil Magness et moi avons donc simplement décidé qu'il valait mieux construire notre dénominateur de revenu à partir de zéro en utilisant les lignes directrices du BEA pour créer un dénominateur de revenu qui soit, par définition, cohérent avec les chiffres du revenu brut de l'IRS. Nous l'avons fait ligne par ligne et année par année (un an de travail)

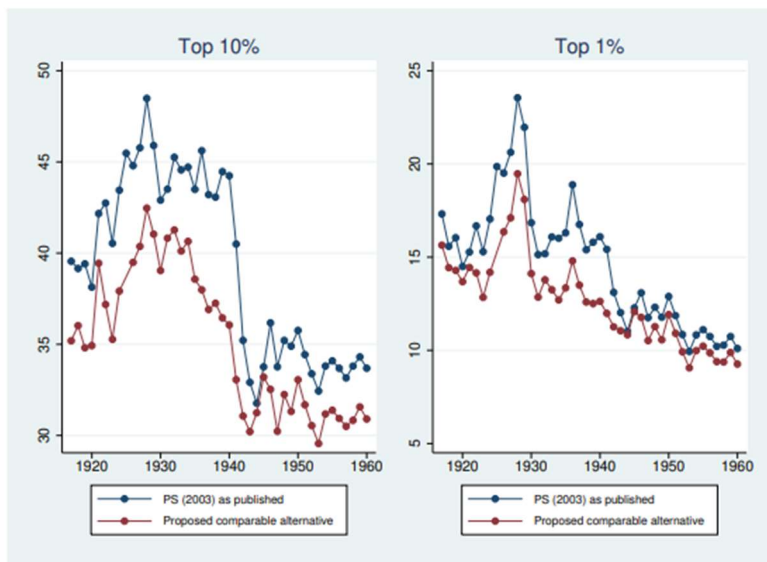
Ce dénominateur nous place entre les estimations de Kuznets et les estimations du PS. Mais il retranche entre 3 et 6 points de pourcentage de la part des 10 % de revenus les plus élevés et entre 0,5 et 2 points de la part des 1 % de revenus les plus élevés.

Table 1. *Effect of Different Data Adjustments on Original PS Estimates at Model 3 (Selected Years)*

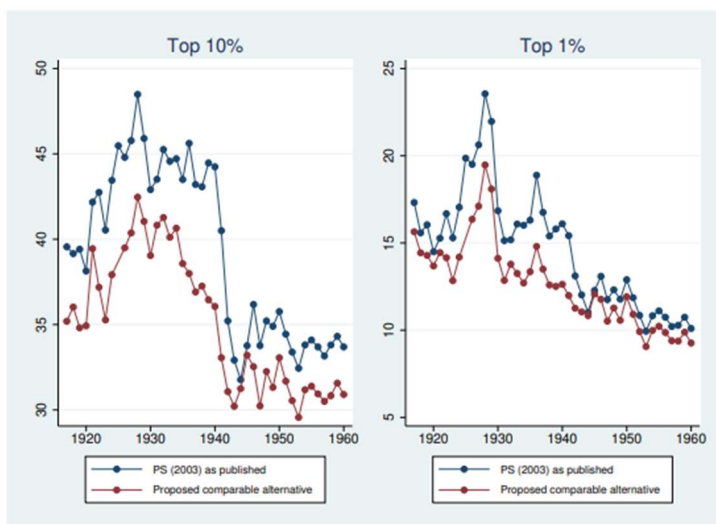
	1917	1929	1936	1939	1943
<i>Panel A: Top 10% Income Shares</i>					
Piketty-Saez Model 3	39.6	45.9	44.6	43.7	32.1
New Models 2 and 3	-0.9	-0.3	-2.2	-2.0	-1.0
New Models 2 and 3 + Denominator Effects	-4.4	-4.9	-7.5	-7.9	-2.7
<i>Panel B: Top 1% Income Shares</i>					
Piketty-Saez Model 3	17.3	22.0	18.0	15.2	11.4
New Model 3	0.0	-1.4	-1.5	-1.2	-0.3
New Model 3 + Denominator Effects	-1.7	-3.9	-3.9	-3.2	-0.9

Vous aboutissez également à une tendance radicalement différente où l'essentiel du nivellement de l'inégalité se produit pendant la dépression et non pendant la seconde guerre mondiale.





Mais nous savons que les erreurs méthodologiques qui ont conduit PS à surestimer les inégalités doivent être ajoutées aux problèmes de qualité des données de l'IRS. La ligne rouge est donc probablement encore plus basse que ce que nous montrons.



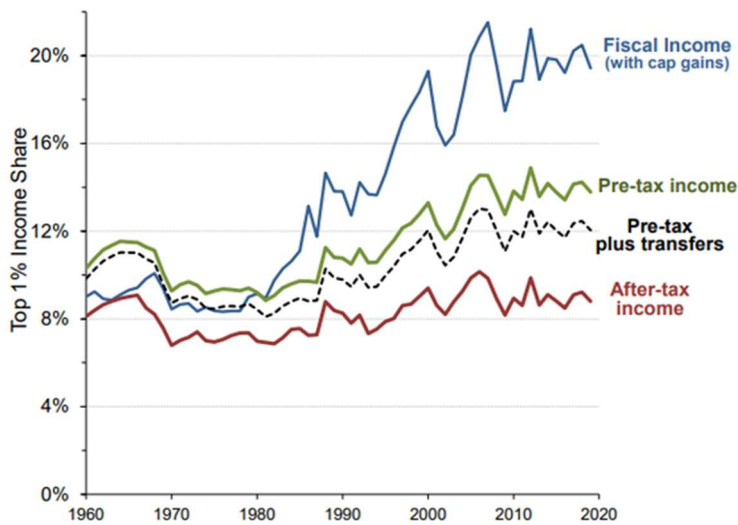
D'accord, vous pouvez vous moquer de moi maintenant et dire "on s'en fout, Vincent". Moi, je m'en soucie ! Je tiens à ce que les données soient traitées avec respect et à ce que les sources historiques ne soient pas malmenées.

Je pense que mes corrections suggèrent que l'inégalité a été traitée davantage en raison de la dépression qu'en raison des changements de politique fiscale dans les années 1940. Cela correspond à l'histoire de Walter Scheidel selon laquelle les catastrophes tendent à être les facteurs de nivellement les plus efficaces.

Mais même si vous n'êtes pas d'accord, vous ne pouvez pas négliger les données. Je passe des heures et des jours à documenter TOUS mes choix de données dans tous mes articles et je partage mes données pour que les gens puissent me critiquer.

Mais ce n'est pas tout, il faut faire son autocritique et se demander si l'on travaille de manière à voir ce que l'on veut voir. Est-ce que j'obtiens les résultats que je veux à cause de la façon dont je le fais ou est-ce que ce sont les vrais résultats ?

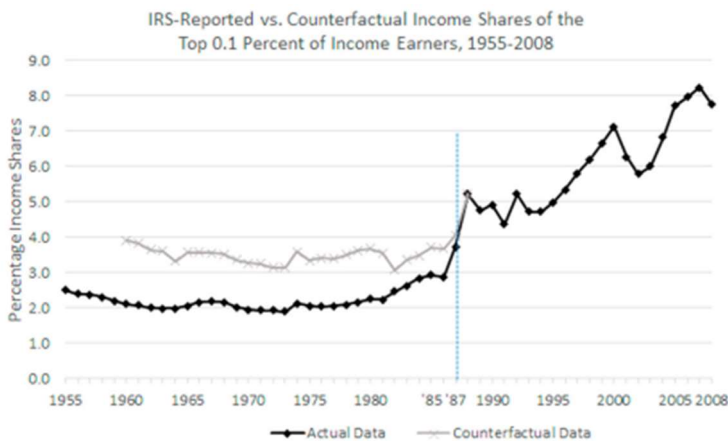
N'oubliez pas que je ne suis pas le seul à le souligner. [Gerald Auten et David Splinter](#) ont mis en évidence d'autres problèmes liés au traitement des données, au choix des unités pour les ménages/unités fiscales et aux définitions des revenus.



**Figure 1: Top 1% income shares**

Notes: Fiscal income includes capital gains with thresholds set without capital gains. Adjustments used to estimate pre-tax national income, pre-tax plus transfers, and after-tax income are listed in Tables 1 and 2 and described in detail in the online appendix. Annual values are shown in appendix Table A1.  
Sources: Authors' calculations and in Piketty and Saez (2003 and updates) for fiscal income.

[Mechling et al.](#) ont montré que leurs choix de données permettent aux réformes fiscales de 1986 d'influencer les résultats uniquement pour des raisons de comptabilité fiscale. Une fois la correction effectuée, on constate que le creux de l'inégalité est plus élevé (de sorte que l'augmentation après 1980 est plus faible)



**Figure 5.** IRS-reported vs. counterfactual income shares of the top 0.1 per cent of income earners, 1955–2008.

En fait, je suis tellement sceptique à l'égard de leur travail que s'ils me disaient qu'ils sont d'accord avec moi, je referais probablement tout mon travail une fois de plus pour être sûr de ne pas avoir oublié quelque chose d'important.