

Economie Statistique **ET**

Economics **AND** Statistics

Varia

Incertitude des résultats et demande de football -
Discriminations dans l'accès au logement -
Comportement d'épargne des ménages - Inégalités
de qualité nutritionnelle de l'alimentation -
Élasticité de la demande d'électricité

Outcome Uncertainty and the Demand
for Football - Discriminations in Access
to Housing - Household Savings Behaviour -
Inequality in Food Nutritional Quality -
Elasticity of the Demand for Electricity

Economie Statistique ^{ET}

Economics AND Statistics

Conseil scientifique / Scientific Committee

Jacques LE CACHEUX, président (Université de Pau et des pays de l'Adour)
Jérôme BOURDIEU (École d'économie de Paris)
Pierre CAHUC (Sciences Po)
Gilbert CETTE (Banque de France et École d'économie d'Aix-Marseille)
Yannick L'HORTY (Université de Paris-Est - Marne la Vallée)
Daniel OESCH (Life Course and Inequality Research (LINES) et Institut des sciences sociales - Université de Lausanne)
Sophie PONTHEUX (Insee)
Katheline SCHUBERT (École d'économie de Paris, Université Paris I)
Claudia SENIK (Université Paris-Sorbonne et École d'économie de Paris)
Louis-André VALLET (Observatoire sociologique du changement-Sciences Po/CNRS)
François-Charles WOLFF (Université de Nantes)

Comité éditorial / Editorial Advisory Board

Luc ARRONDEL (École d'économie de Paris)
Lucio BACCARO (Max Planck Institute for the Study of Societies-Cologne et Département de Sociologie-Université de Genève)
Antoine BOZIO (Institut des politiques publiques/École d'économie de Paris)
Clément CARBONNIER (Théma/Université de Cergy-Pontoise et LIEPP-Sciences Po)
Erwan GAUTIER (Banque de France et Université de Nantes)
Pauline GIVORD (Ocde et Crest)
Florence JUSOT (Université Paris-Dauphine, Leda-Legos et Irdes)
François LEGENDRE (Erudite/Université Paris-Est)
Claire LELARGE (Université de Paris-Sud, Paris-Saclay et Crest)
Claire LOUPIAS (Direction générale du Trésor)
Pierre PORA (Insee)
Ariell RESHEF (École d'économie de Paris, Centre d'économie de la Sorbonne et CEPII)
Thepthida SOPRASEUTH (Théma/Université de Cergy-Pontoise)

Directeur de la publication / Director of Publication:

Jean-Luc TAVERNIER

Rédactrice en chef / Editor in Chief:

Sophie PONTHEUX

Responsable éditorial / Editorial Manager: Pascal GODEFROY

Assistant éditorial / Editorial Assistant: Étienne de LATUDE

Traductions / Translations: RWS Language Solutions
Chiltern Park, Chalfont St. Peter, Bucks, SL9 9FG Royaume-Uni

Maquette PAO et impression / CAP and printing: JOUVE
1, rue du Docteur-Sauvé, BP3, 53101 Mayenne

La revue est en accès libre sur le site www.insee.fr. Il est possible de s'abonner aux avis de parution sur le site. La revue peut être achetée sur le site www.insee.fr, rubrique « Services / Acheter nos publications ». La revue est également en vente dans 200 librairies à Paris et en province.

The journal is available in open access on the Insee website www.insee.fr. Publication alerts can be subscribed on-line. The printed version of the journal (in French) can be purchased on the Insee website www.insee.fr.

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Directeur Général : Jean-Luc TAVERNIER

Direction Générale : 88, avenue Verdier, CS 70058, 92541 MONTRouGE Cedex

Tél : +33 (0)1 87 69 50 00

Economie
Statistique **ET**

Economics
AND Statistics

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes,
et non les institutions auxquelles ils appartiennent, ni *a fortiori* l'Insee.

Economie et Statistique / Economics and Statistics

Numéro 513 – 2019

VARIA

- 5 Les supporters français de football sont-ils sensibles à l'incertitude du résultat ?**
Luc Arrondel et Richard Duhautois
- 29 Les discriminations dans l'accès au logement en France : un *testing* sur les aires urbaines métropolitaines**
Julie Le Gallo, Yannick L'Horty, Loïc du Parquet et Pascale Petit
- 49 Les liens entre taux d'épargne, revenu et incertitude : une analyse à partir de l'enquête *Budget de famille* 2011**
Céline Antonin
- 73 Quatre décennies d'achats alimentaires : évolutions des inégalités de qualité nutritionnelle en France, 1971-2010**
France Caillavet, Nicole Darmon, Flavie Létoile et Véronique Nichèle
- 95 L'élasticité-prix de la demande d'électricité en France**
Stéphane Auray, Vincenzo Caponi et Benoît Ravel

Les supporters français de football sont-ils sensibles à l'incertitude du résultat ?

Are French Football Fans Sensitive to Outcome Uncertainty?

Luc Arrondel* et Richard Duhautois**

Résumé – L'hypothèse que l'équilibre compétitif augmente l'utilité des supporters, soit leurs dépenses et donc les revenus des clubs professionnels, est au cœur de l'économie du sport en général, du football en particulier. Cette notion d'équilibre compétitif est mise en avant pour justifier les décisions des ligues professionnelles pour changer les règles des compétitions ou la distribution des droits TV. Mais comme le montre la littérature empirique, la relation entre équilibre compétitif et demande des supporters est loin d'être évidente. Dans cet article, nous nous intéressons à l'équilibre compétitif tel qu'il est perçu par les supporters. Pour la Ligue 1, il s'explique par une incertitude de moyen et de long terme alors que pour la Ligue des champions, il s'agit d'un suspense de long terme. L'incertitude du résultat est loin d'être le seul facteur explicatif de la demande de football puisque environ 30 % des supporters se déclarent prêts à suivre les matches même dans le cas où il n'y aurait plus aucun suspense.

Abstract – *The idea that competitive balance increases the utility of fans, and therefore their spending and the revenue of professional clubs, lies at the heart of sports economics in general and the economics of football in particular. This notion of competitive balance is often invoked to explain the decisions of professional leagues to change the rules of competitions or the distribution of TV rights. However, the empirical literature shows that the relationship between competitive balance and fan demand is far from obvious. In this paper, we examine the idea of competitive balance as perceived by football fans. In the case of Ligue 1, it is mainly explained by medium- and long-term uncertainty, while in the case of the Champions League it is more a matter of long-term suspense. But uncertainty over the outcome is far from being the only factor explaining the demand for football since around 30% of fans report that they would always be willing to attend or watch games even in the hypothetical case that there is no suspense left.*

Codes JEL / JEL Classification : D12, L83

Mots-clefs : équilibre compétitif, hypothèse d'incertitude du résultat, demande de football

Keywords: *competitive balance, uncertainty of outcome hypothesis, demand for football*

* CNRS, PSE (luc.arrondel@ens.fr) ; ** CNAM-Lirsa et CEET (richard.duhautois@ecnam.net)

Les auteurs remercient particulièrement J. Le Bail du quotidien L'Équipe qui leur a permis de réaliser l'enquête utilisée dans cet article. Les remarques de deux rapporteurs anonymes nous ont été particulièrement utiles.

Reçu le 22 juillet 2018, accepté après révisions le 12 juillet 2019.

Citation : Arrondel, L. & Duhautois, R. (2019). Are French Football Fans Sensitive to Outcome Uncertainty? *Economie et Statistique / Economics and Statistics*, 513, 5–26. <https://doi.org/10.24187/ecostat.2019.513.2001>

L'hypothèse d'incertitude du résultat (*uncertainty of outcome hypothesis* – UOH en anglais) est une hypothèse centrale de l'économie du sport depuis les années 1950 (Rottenberg, 1956) pour expliquer la « consommation » de sport des spectateurs. L'idée est que les championnats où les compétitions déséquilibrés – par exemple lorsqu'une équipe est plus forte que toutes les autres – ont un effet négatif sur la « demande » des supporters, entraînant une baisse de la fréquentation des stades et du nombre de téléspectateurs, et en conséquence, une baisse de revenu pour les clubs. Cette notion d'équilibre des compétitions est connue sous le nom « d'équilibre compétitif » (*competitive balance*). Neale (1964) distingue l'incertitude du résultat au niveau du match (le paradoxe Louis-Schmeling¹) et l'incertitude du résultat d'une compétition (par exemple le championnat de France de football de Ligue 1, qu'il appelle « *League Standing Effect* » qu'on peut traduire par « effet de classement ») : si les équipes dans un championnat sont de force égale, il y a régulièrement des changements dans le classement, augmentant l'intérêt des spectateurs et par conséquent les revenus.

La plupart des études empiriques concernant le football professionnel ne trouvent cependant aucun lien entre incertitude du résultat et demande des supporters (voir en annexe 1 un recensement annoté des études). Ainsi, par exemple, on observe que les stades sont bien remplis en Angleterre et en Allemagne, et relativement vides en France et en Italie sans qu'on puisse y trouver un lien avec l'équilibre compétitif des championnats : c'est en France où l'incertitude était la plus importante jusqu'au rachat du Paris Saint-Germain Football Club (PSG) par le Qatar (Andreff, 2014, 2018).

La première raison est que l'attractivité des compétitions dépend de leur fonctionnement (encadré 1). En effet, il existe des ligues ouvertes, avec système de promotions-relégations (comme les championnats de football européens) et des ligues fermées, sans un tel système (comme les

1. L'idée du paradoxe vient des faits suivants : lorsqu'après une série de victoires écrasantes le boxeur Joe Louis connaît face à Max Schmeling la première défaite de sa carrière professionnelle en 1936 (après 23 victoires consécutives), l'intérêt autour du boxeur américain va paradoxalement rapidement augmenter, et ses revenus également.

ENCADRÉ 1 – L'organisation du football en France et Europe

Les compétitions nationales de football en France – comme toutes celles en Europe – sont organisées de manière pyramidale. Une de leurs caractéristiques est le système de promotions (les montées d'équipes dans la division supérieure) et de relégations (les descentes d'équipes dans la division inférieure). On appelle ces compétitions des « ligues ouvertes » par opposition aux « ligues fermées » nord-américaines dans lesquelles le système de promotions-relégations n'existe pas. Il fait place aux systèmes de *salary cap* (plafond salarial) et à la *draft* (incorporer) qui permettent de rééquilibrer les équipes d'une saison sur l'autre. Même si le *salary cap* et la *draft* ont des règles spécifiques dans les différents sports, leur but est le suivant : d'une part, que la masse salariale de chaque franchise (équipe) soit équivalente et d'autre part choisir les meilleurs joueurs lorsqu'on a fini dernier de la compétition.

Le nombre de promus et de relégués dans les différents championnats européens n'est pas identique à chaque niveau de la hiérarchie et a changé de nombreuses fois au cours du temps. En France, dans les deux plus hautes divisions (Ligue 1 et Ligue 2), tous les clubs ont un statut professionnel obligatoire : la plupart sont des sociétés anonymes sportives professionnelles (SASP), statut créé à la fin des années 1990 qui permet notamment de distribuer des dividendes. Dans les grands championnats européens, les premières et deuxième divisions sont le plus souvent les divisions professionnelles. En Angleterre, les quatre premières divisions sont professionnelles et la plupart des clubs sont des sociétés depuis le début du professionnalisme en 1888.

Viennent ensuite les divisions du football amateur avec des clubs à statut associatif et des joueurs non professionnels. En France, à partir de la saison 2017-2018, les divisions nationales amateurs s'appellent dans l'ordre le championnat National 1, le championnat National 2 (ancien championnat de France amateur, CFA), et le championnat National 3 (ancien CFA 2). La plupart des clubs de ces championnats ont le statut amateur. Seuls les clubs qui descendent de Ligue 2 à National 1 peuvent garder leur statut professionnel pendant deux ans. Ils redeviennent des clubs amateurs s'ils ne réussissent pas à remonter en Ligue 2 au bout de deux saisons. En deçà du niveau national, les compétitions s'effectuent au niveau régional, gérées par les ligues de chaque région, et au niveau départemental, gérées par les districts départementaux. Une réforme de l'organisation s'est achevée lors de la saison 2018-2019.

Les compétitions internationales européennes – la Ligue des champions et la Ligue Europa – sont réservées aux équipes des différentes compétitions nationales ayant remplies certains critères sportifs : vainqueurs du championnat et des coupes nationales, ainsi que les mieux classés du championnat (le nombre varie en fonction des pays). Étant donné le nombre important d'équipes, il y a des tours préliminaires avant les phases finales. Les phases finales s'organisent en deux temps : d'abord un mini championnat dans un groupe de quatre équipes (chaque équipe joue six matchs en aller-retour) et ensuite pour les deux premiers de chaque groupe, des matchs à élimination directe (en aller-retour), à partir des 8^e ou des 16^e de finales.

championnats nord-américains). Dans les ligues fermées, la proximité d'équipes d'une même ville dans des ligues concurrentes permet aux consommateurs de changer de stade si l'incertitude du résultat diminue. Dans le football européen, la concurrence entre ligues est moins prégnante car elles sont hiérarchisées (Ligue 1, Ligue 2, National, etc.). Surtout, les équipes dans les mêmes villes en Europe se sont souvent construites sur des oppositions de classe ou religieuses, ce qui, « moralement », interdit aux supporters de « changer » de club (Kuper & Szymanski, 2018). La seconde raison est que la demande des fans de football ne se résume pas à un comportement de pur consommateur : ils préfèrent voir leur équipe favorite gagner (Buraimo & Simmons, 2008) et voir une équipe de renom même si elle est beaucoup plus forte (Pawlowski & Anders, 2012).

D'une manière générale, en dehors de l'incertitude, la littérature identifie cinq facteurs potentiels qui influencent la demande de football (Caruso *et al.*, 2019).

(1) La performance sportive : plus l'équipe est performante, plus il y a de monde au stade.

(2) La qualité des matchs : meilleures sont les deux équipes, plus il y a de spectateurs. Ainsi, plus on descend dans la hiérarchie sportive et moins il y a de monde au stade.

(3) Le confort : plus les stades sont neufs, plus les supporters se déplacent ; ils sont également sensibles aux conditions climatiques et à l'heure programmée des rencontres.

(4) Le prix, même si la sensibilité des fans aux prix des places dépend des équipes qui jouent.

(5) La diffusion des matchs à la télévision : si la diffusion peut faire concurrence à la fréquentation des stades, la demande globale a fortement augmenté depuis la massification des retransmissions des matchs dans les années 1990.

Une des caractéristiques de l'ensemble des études sur l'effet de l'incertitude sur la demande de football est que l'incertitude est mesurée par des données *a priori* sur les matchs et les équipes et non sur les perceptions des supporters (encadré 2). Pour cela, il faut des enquêtes et les informations concernant la « consommation » de football sont longues et difficiles à administrer. Ainsi, Pawlowski (2013) a fait passer environ 1 000 entretiens pendant deux mois dans six stades et quelques bars en Allemagne avec un nombre de questions limité. L'incertitude du résultat est alors mesurée subjectivement, telle qu'elle est perçue par les fans. Ses résultats montrent que les supporters sont sensibles à l'incertitude et qu'ils pensent qu'en Allemagne le championnat de football de première division est assez équilibré pour continuer d'aller au stade ou de regarder les matchs à la télévision.

Pour mieux comprendre les comportements des supporters de football en France, nous avons réalisé une enquête originale à la fin de la saison 2015-2016. Le questionnaire a été mis en ligne fin mai sur le site du quotidien sportif *L'Équipe* (lequipe.fr) et nous avons enregistré plus de 22 000 réponses. À partir des informations collectées dans cette enquête, nous analysons – pour la première fois avec des données françaises – ce qui explique l'équilibre compétitif perçu par les supporters. Cet équilibre compétitif peut

ENCADRÉ 2 – Les différentes mesures de l'incertitude

L'équilibre compétitif (EC) et l'hypothèse d'incertitude du résultat (UOH) sont étroitement liés et sont souvent traités comme interchangeable, à quelques différences près. L'UOH est traitée comme un concept *ex ante* tourné vers l'avenir, défini en termes de distribution de probabilité des résultats possibles. En revanche, l'évaluation de l'EC repose généralement sur des mesures rétrospectives. Mais la principale différence entre les deux notions se situe dans le court terme : il y a une dimension plus marquée sur l'incertitude des matchs ainsi que sur l'incertitude au cours de la saison.

Comme nous le soulignons dans le texte, différentes dimensions temporelles sont pertinentes pour mesurer l'UOH : le court terme, le moyen terme et le long terme.

L'incertitude de court terme est la dimension la plus fréquemment examinée. Sa mesure s'appuie sur deux sources d'information principales : les statistiques sur les performances relatives des deux équipes avant un match ou les cotes des paris sportifs (voir annexe 1).

L'incertitude de moyen terme est le degré d'incertitude au sein du championnat : la victoire en championnat, les qualifications pour les compétitions européennes et la relégation. Sa mesure s'appuie sur des indicateurs de concentration comme l'indice de Gini.

L'incertitude de long terme fait référence à la domination à long terme d'une équipe ou d'un petit nombre d'équipes. Sa mesure est basée sur des indicateurs comme l'indice de Herfindahl-Hirschman, la variation des indicateurs de Gini, etc.

concerner le court terme (l'issue d'un match), le moyen terme (l'issue d'un championnat lors d'une saison) ou le long terme (la récurrence des vainqueurs des compétitions).

Nos résultats montrent que :

- les supporters estiment que le suspense des championnats est différent selon les compétitions ;
- sans aucune incertitude, environ 30 % des personnes interrogées continueraient d'aller au stade ou de regarder les matchs à la télévision ;
- l'équilibre compétitif perçu de la Ligue 1 s'explique autant par l'incertitude de moyen et long terme et que celui de la Ligue des champions s'explique essentiellement par l'incertitude de long terme, c'est-à-dire sur l'identité du vainqueur chaque année.

1. La relation entre incertitude et demande de football : revue de la littérature

Pour estimer la relation entre la demande des supporters et l'UOH, de nombreuses études mettent en relation la fréquentation des stades et des mesures de l'équilibre compétitif basées sur des indicateurs d'inégalités appliqués aux classements des championnats. Les premières analyses ont été faites sur les championnats anglais, écossais et allemand alors que les plus récentes portent sur d'autres championnats (brésiliens, portugais, etc.). Falter & Pérignon (2000) et Falter *et al.* (2008) s'intéressent au championnat français de première division, même si le cœur du sujet n'est pas toujours l'analyse de l'effet de l'incertitude mais plutôt les déterminants de la demande de football. Par exemple, le contexte économique régional en est un élément moteur : plus les salaires sont faibles et plus le taux de chômage est élevé, plus les individus vont au stade, traduisant l'engouement populaire pour le football.

Au départ, les études se concentrent sur l'affluence au stade et ultérieurement, avec l'augmentation des retransmissions, sur l'audience télévisée². Parallèlement, d'autres études essaient d'analyser la possible substitution entre fréquentation du stade et retransmission télévisée.

1.1. Les effets de l'incertitude sur l'affluence au stade

Hart *et al.* (1975) analysent l'effet de la différence de classement sur l'affluence au stade

pour quatre équipes du championnat anglais entre 1969 et 1972. Les auteurs de cette étude montrent qu'il n'y a pas d'effet significatif, ce qui sera par la suite souvent le cas. Deux études sur le championnat écossais de première division (Jennett, 1984 ; Cairns, 1987) concluent, cette fois, que l'incertitude du résultat influence positivement l'affluence au stade, dans la lignée des résultats de Rottenberg (1956) et des sports américains. Les mesures de l'incertitude utilisées sont la différence de classement des équipes et le nombre de points nécessaires pour gagner le championnat, c'est-à-dire une incertitude de court et moyen terme. En particulier, Jennett (1984) montre que si une équipe a encore une chance de gagner le championnat, elle attire plus de monde. Wilson & Sim (1995) trouvent le même résultat à partir de données du championnat de première division de Malaisie. Baimbridge *et al.* (1996), pour l'Angleterre, ne trouvent aucun lien entre les variables d'incertitude et la fréquentation du stade. Simplement, ils montrent que l'affluence décline progressivement jusqu'au milieu de la saison et remonte ensuite.

Peel & Thomas (1988, 1992) étudient l'effet de l'incertitude sur l'affluence au stade dans les quatre divisions professionnelles anglaises en utilisant pour la première fois la probabilité de gagner *a priori* de l'équipe qui reçoit, basée sur les paris sportifs. Leurs résultats montrent un lien positif entre la probabilité de gagner et l'affluence au stade, ce qui signifie que ce n'est pas l'incertitude mais le fait de voir gagner son équipe qui stimule les supporters. Ce sont surtout des buts que les supporters veulent voir. Peel & Thomas (1996) continuent à creuser l'effet de l'incertitude avec les trois premières divisions écossaises. Un des intérêts de ces championnats est que, au sein d'une même saison, les équipes se rencontrent plusieurs fois à domicile. Cela permet de rajouter un contrôle sur les caractéristiques des matchs. Leurs résultats confirment les résultats de Peel & Thomas (1992) : la relation entre l'affluence au stade et la probabilité que l'équipe receveuse gagne est une courbe en forme de U, les supporters étant surtout motivés par une victoire facile ou un exploit. Buraimo & Simmons (2008, 2009) analysent respectivement le championnat anglais et espagnol de première division et retombent sur cette relation en forme de U.

2. Jusqu'en 1983 en Angleterre et en 1984 en France, les matchs de championnat n'étaient pas diffusés en direct à la télévision. Les acteurs du football avaient peur de la concurrence que la diffusion pourrait engendrer sur la fréquentation des stades. Certains matchs de coupes (coupes d'Europe, coupe du Monde, Euros, coupes nationales, etc.) étaient déjà depuis 1937, date du premier match diffusé en direct (finale de la coupe d'Angleterre).

Finalement, il est très difficile de connaître la vraie demande de football en raison des contraintes de capacité des stades. Czarnitzki & Stadtmann (2002) et Benz *et al.* (2009), analysent le championnat allemand de première division en tentant de prendre en compte cette contrainte mais aboutissent au même résultat : pas d'effet de l'incertitude sur l'affluence au stade. Les études les plus récentes (Anders & Pavlowski, 2012 ; Cox, 2015 ; Bedina & Pershakov, 2017) vont également dans ce sens. Notons toutefois l'étude de Jang & Lee (2015), qui conclut à l'existence d'un effet de l'incertitude dans le championnat sud-coréen.

1.2. Les retransmissions télévisées ont-elles eu un effet sur l'affluence au stade ?

Aujourd'hui, la demande de football ne peut plus se résumer à la fréquentation des stades. Depuis la fin des années 1980, le nombre de matchs de football – mais aussi de nombreuses autres compétitions – diffusés à la télévision a explosé. Pour cette raison, les droits de diffusion sont devenus la principale source de financement du football professionnel, ce qui permet notamment pour les grands clubs européens d'acheter à prix d'or les meilleurs joueurs et de remplir les stades.

Le fait que tous les matchs de championnat soient diffusés peut avoir un effet négatif sur la fréquentation des stades mais a un effet positif sur la demande globale. Garcia & Rodriguez (2002) s'intéressent aux variables qui expliquent l'affluence au stade dans le championnat espagnol de première division entre 1993 et 1996. Leurs résultats montrent que les matchs diffusés à la télévision et ceux programmés en dehors du week-end sont ceux qui génèrent significativement moins de public. L'effet de la diffusion est d'autant plus important que les matchs sont diffusés sur une chaîne gratuite. Buraimo & Simmons (2009) trouvent les mêmes résultats pour le championnat de première division en Espagne. Forrest *et al.* (2004) estiment une relation similaire pour la première division anglaise entre 1992 et 2001 et montrent que les matchs diffusés le week-end n'engendrent pas mécaniquement de baisse de la fréquentation des stades ; cela dépend plutôt encore une fois de la chaîne qui diffuse le match, payante ou non. Buraimo & Simmons (2008) trouvent l'effet contraire : pour le même championnat, entre 2001 et 2006, les matchs diffusés le week-end engendrent une légère baisse de la fréquentation des stades, alors que ceux diffusés la semaine

sont sans effet. Buraimo *et al.* (2006) analysent les déterminants de l'affluence au stade de la deuxième division anglaise entre 1998 et 2004. Les auteurs montrent que la diffusion d'un match sur une chaîne gratuite réduit de plus de 20 % la fréquentation au stade (5 % sur une chaîne payante). Ils observent également que lorsque des matchs plus « haut de gamme » – soit des matchs des meilleures équipes de première division ou des matchs internationaux – ont lieu en même temps, les spectateurs se déplacent moins. Allan & Roy (2008) décomposent les spectateurs en trois catégories : les abonnés, les supporters occasionnels de l'équipe qui reçoit et les supporters de l'équipe visiteuse (qui peuvent être bien entendu des abonnés au stade de l'équipe visiteuse). Sans surprise, les abonnés ne changent pas leurs habitudes et continuent d'aller au stade même en cas de retransmission, tout comme les fans de l'équipe visiteuse, très motivés, alors que la fréquentation du stade des spectateurs occasionnels baisse d'environ 30 %.

1.3. L'incertitude a un effet plus fort sur l'audience télévisée

Si l'effet de l'incertitude sur l'affluence au stade est rarement significatif, son effet sur l'audience télévisée semble légèrement plus important. Forrest *et al.* (2005) montrent que l'incertitude du résultat a un effet positif sur les audiences télévisées des matchs de première division anglaise. Buraimo (2008) ne trouve pas d'effet significatif de l'incertitude mais un effet positif de la qualité des joueurs (les stars) et du nombre de spectateurs au stade. Buraimo & Simmons (2015) analysent les saisons de la décennie 2000 du championnat anglais et montrent que l'incertitude a un effet positif seulement pour les deux premières saisons, confirmant l'idée que c'est les meilleurs joueurs du monde qui attirent les téléspectateurs. En revanche, Buraimo & Simmons (2009), pour le championnat espagnol, estiment une relation significative entre audience télévisée et incertitude du résultat. Alavy *et al.* (2010) analysent minute par minute le comportement des téléspectateurs des matchs du championnat anglais. Ces derniers ont tendance à changer de chaîne lorsque la probabilité d'un match nul se dégage. Schreyer *et al.* (2016, 2017) montrent que l'incertitude a un effet positif sur l'audience télévisée lors des matchs à enjeu dans les grandes compétitions internationales, lors des matchs de championnat, mais n'a pas d'effet significatif lors des matchs de coupe nationale. Enfin, Caruso *et al.* (2019) s'intéressent au championnat italien et concluent que ce qui attirent les supporters, ce sont les stars

et les équipes qui gagnent : une augmentation de 1 % de la masse salariale des équipes qui se rencontrent augmente le nombre de téléspectateurs d'environ 0.75 % et une augmentation de 1 % de la moyenne des points augmente le nombre de téléspectateurs d'environ 0.7 %.

2. L'enquête et la mesure du « suspense » des compétitions

Pour analyser les effets de l'incertitude sur la consommation de football en France, nous utilisons un questionnaire original incluant une centaine de questions (y compris les caractéristiques socio-démographiques des répondants). L'enquête *Quel supporter êtes-vous ?* a été mise en ligne fin mai sur le site du quotidien sportif *L'Équipe* (lequipe.fr), à la fin de la saison 2015-2016, pendant trois jours. La base finale contient 21 716 observations, certaines avec des valeurs manquantes. La durée de remplissage du questionnaire était d'environ vingt minutes.

Comme il s'agit de la première enquête de ce type sur les supporters, il est difficile d'avoir une idée de sa représentativité. Selon *Nielsen Sport* (2018), 50 % des Français de plus de seize ans se déclarent « intéressés » ou « très intéressés » par le football. Parmi cette moitié de la population ayant un intérêt pour le football, environ 85 % seraient « supporter » d'au moins un club de Ligue 1. Cette population des supporters serait, selon *Statista* (2018), en majorité des hommes (63 %) et nettement plus jeunes que la population totale (les plus de 65 ans seraient ainsi deux fois moins représentés, 8 % vs 16 %). Il est possible de restreindre encore plus finement la définition d'un « supporter » en s'intéressant aux « inconditionnels », aux « passionnés », aux « fans », c'est-à-dire à ceux qui suivent « systématiquement ou presque » les compétitions et l'actualité du football.

C'est vraisemblablement ce segment de population « passionné de foot » qui a répondu à notre enquête sur le site de *L'Équipe* (tableau 1) : ceux-ci sont nettement plus jeunes que la population des supporters en général (moins de 2 % de plus de 60 ans) et les femmes y seraient largement sous-représentées (moins de 5 %). Par rapport aux visiteurs habituels du site du quotidien sportif, les biais de représentativité vont dans le même sens. Pour tenir compte de la spécificité de notre échantillon, nous avons procédé dans notre étude à des traitements spécifiques en fonction de l'âge (18-26 ans vs plus de

26 ans) afin de vérifier nos principaux résultats (voir annexe).

Le tableau 1 synthétise la distribution des variables reflétant l'intérêt et la consommation de football et les caractéristiques socio-démographiques des répondants. Sans surprise, pratiquement tous les répondants sont très intéressés par les différentes compétitions européennes – excepté la Ligue 2 qui ne passionne qu'une minorité – et sont des consommateurs assidus de football, tant à la télévision qu'au stade. La plupart sont des hommes (96.5 %), jeunes (près de 85 % ont moins de 40 ans) et vivant seul (55.2 %), n'habitant pas à proximité des stades (la distance médiane entre le lieu d'habitation et le stade est de 75 kilomètres). Une majorité est sur le marché du travail (63.1 %) et près de 20 % ont des revenus supérieurs à 30 000 euros annuels. Enfin, 25 % habitent en région parisienne, 97 % sont français et les deux-tiers n'ont pas d'enfants.

Dans un premier temps, nous nous sommes concentrés sur le niveau d'incertitude perçu par les fans dans les principales compétitions de football professionnel en France (Ligue 1 et Ligue 2) et en Europe (Ligue des champions et Ligue Europa). Pour chaque compétition, nous avons posé la question suivante : « *Sur les dernières saisons, comment évaluez-vous le « suspense » de [la compétition en question] (sur une échelle de 0 : pas de suspense du tout à 10 : beaucoup de suspense) ?* ». Les figures I et II donnent la distribution des réponses à ces questions. Pour la Ligue 2, la Ligue des champions et la Ligue Europa, les résultats sont sensiblement identiques : le mode est à 8 (relativement beaucoup de suspense), la médiane à 7 et la moyenne est respectivement de 6.64, 7.02 et 6.96. Les résultats de Pawlawski (2013) sont comparables pour le championnat allemand : le mode est également à 8 mais avec une moyenne plus élevée, légèrement supérieure à 8, la distribution étant très concentrée sur les grandes valeurs. Ce résultat peut surprendre pour un championnat dominé depuis les années 1970 par le Bayern Munich. En revanche, pour la Ligue 1, les supporters ont intégré dans leur perception du suspense, la domination du PSG depuis 2011 et son rachat par les Qataris : le mode est à 0 (pas de suspense du tout), la médiane à 3 et la moyenne à 3.36.

Nous avons reporté en annexe 2 quelques traitements complémentaires en fonction de l'âge et des clubs supportés. On constate que même s'il existe quelques écarts de pourcentage entre

Tableau 1 – Caractéristiques des répondants (N= 21 716) et variables d'intérêt (en %)

Intérêt pour le football (7-10 ; 4-6 ; 0-3)	(93.1 ; 5.8 ; 1.1)
Intérêt pour la Ligue 1 (intéressé ; pas intéressé)	(95.2 ; 4.8)
Intérêt pour la Ligue 2 (intéressé ; pas intéressé)	(44.7 ; 55.3)
Intérêt pour la Coupe de France (intéressé ; pas intéressé)	(74.8 ; 25.2)
Intérêt pour la Ligue des champions (intéressé ; pas intéressé)	(96.3 ; 3.7)
Intérêt pour la Ligue Europa (intéressé ; pas intéressé)	(81.8 ; 18.2)
Intérêt pour la <i>Premier League</i> anglaise (intéressé ; pas intéressé)	(84.0 ; 16.)
Intérêt pour la <i>Liga</i> espagnole (intéressé ; pas intéressé)	(79.3 ; 20.7)
Intérêt pour la <i>Bundesliga</i> allemande (intéressé ; pas intéressé)	(59.0 ; 41.0)
Intérêt pour la <i>Serie A</i> italienne (intéressé ; pas intéressé)	(51.8 ; 48.2)
Intérêt pour d'autres championnats (intéressé ; pas intéressé)	(19.8 ; 80.2)
Club de supporter (non ; oui)	(82.5 ; 17.5)
Abonnement au stade (non ; oui)	(83.8 ; 16.2)
Nombre de clubs supportés (aucun ou un club ; plusieurs clubs)	(79.4 ; 20.6)
Degré d'attachement à un club (7-10 ; 4-6 ; 0-3)	(93 ; 6 ; 1)
Distance (en km) du stade de son équipe favorite (Q1 ; médiane ; Q3)	(19 ; 75 ; 450)
Consultation sites sportifs sur internet (non ; oui)	(12.9 ; 87.1)
Lecture presse papier sportive (jamais ; 1 fois par semaine ; plus d'1 fois par semaine)	(16.7 ; 42.9 ; 40.4)
Abonnement chaînes sportives (non ; dans le passé ; oui)	(26.1 ; 12.1 ; 61.8)
Pratique du football (non ; dans le passé ; oui)	(15.8 ; 51.8 ; 32.5)
Culture footballistique (inexistante ; faible ; moyenne ; élevée ; très élevée)	(0.5 ; 0.9 ; 16.3 ; 54.8 ; 27.5)
Père et / ou mère qui jouent / ont joué au foot (ne sait pas ; non ; oui)	(2.5 ; 58.3 ; 39.2)
Sexe (homme ; femme)	(96.5 ; 3.5)
Âge (<26 ; 26-40 ; 41-60 ; >60)	(44.8 ; 38.7 ; 14.6 ; 1.9)
Nationalité (française ; autre)	(97.1 ; 2.9)
Diplôme (supérieur ; secondaire)	(58.8 ; 41.2)
Statut familial (seul ; en couple)	(55.2 ; 44.8)
Salaire annuel net (pas de revenu ; <30 000 € ; >30 000 €)	(31.7 ; 48.6 ; 19.8)
Marché du travail (en emploi ; étudiant ou sans emploi)	(63.1 ; 36.9)
Nombre d'enfants (0 ; 1 ; 2 ; >2)	(64,5 ; 11,6 ; 16,4 ; 7,5)
Région (région parisienne ; province)	(25.0 ; 75.0)

Note de lecture : 95.2 % des personnes interrogées sont intéressées par la Ligue 1.
Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

« jeunes » et « vieux », les caractéristiques des distributions sont les mêmes (voir figure A2-I en annexe). Tout au plus notons-nous un léger effet positif des jeunes sur l'équilibre compétitif perçu de la Ligue des champions.

La figure A2-II indique le suspense perçu en fonction du club soutenu. Quatre catégories ont été retenues (voir tableau A2-1 du classement de la saison en annexe) : le champion de la saison (le PSG qui avait 31 points d'avance sur son dauphin), les clubs qualifiés pour les

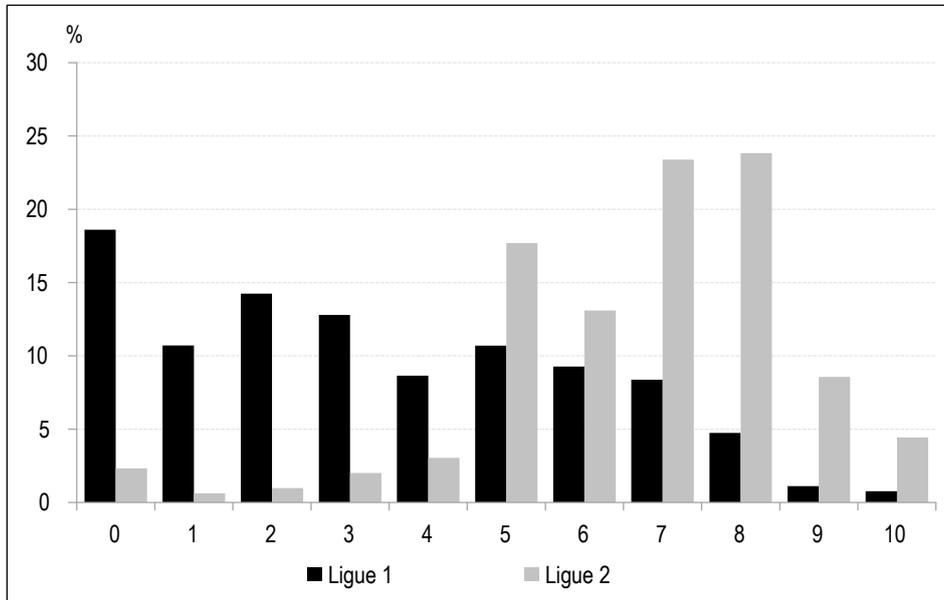
compétitions européennes (de la 2^e à la 6^e place), ceux du « ventre mou » du championnat (de la 7^e à la 16^e place) et enfin ceux ayant joué la relégation jusqu'à la dernière journée (de la 17^e à la 20^e place). On constate que ce sont bien les supporteurs des équipes qui jouent l'Europe ou le maintien en première division qui perçoivent le plus de suspense, ce qui n'est pas le cas des clubs du « ventre mou », sans défi ; les fans du PSG évaluent le suspense à un niveau intermédiaire. L'équilibre compétitif de la Ligue 1 perçu par les supporteurs est donc une notion très

subjective qui semble dépendre fortement du niveau du club que l'on soutient. Cette perception individuelle de l'intensité de la compétition est néanmoins un facteur à prendre en compte pour expliquer la « demande » de football.

3. Suspense et intérêt des supporters pour les compétitions de football

Pour évaluer l'effet du suspense sur les intentions de consommer du football, nous avons

Figure I – L'équilibre compétitif perçu pour la Ligue 1 et la Ligue 2

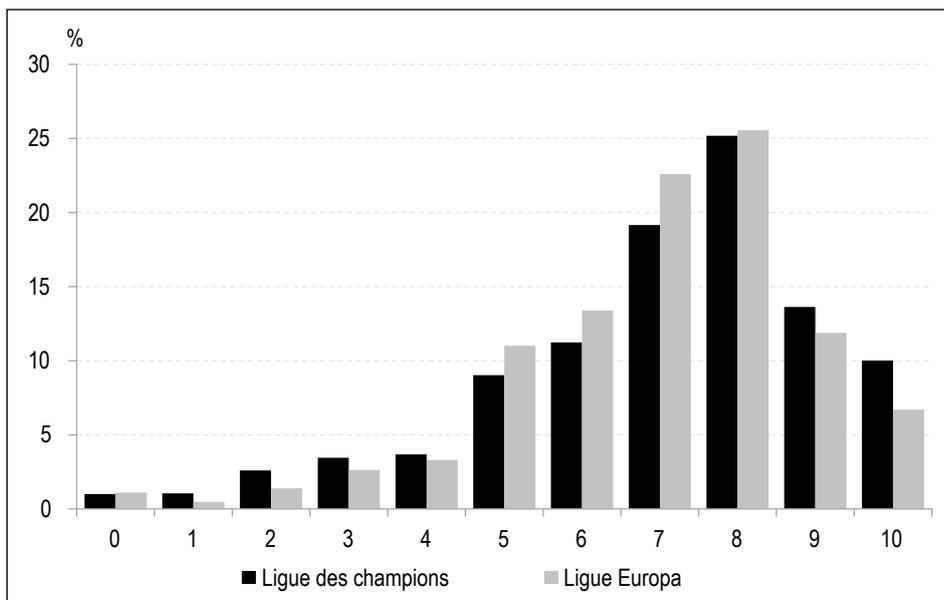


Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ». Ligue 1 : 21 283 réponses, 433 valeurs manquantes ; Ligue 2 : 21 045 réponses, 671 valeurs manquantes.

Note de lecture : 18 % des supporters estiment qu'il n'y a aucun suspense dans la Ligue 1.

Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Figure II – L'équilibre compétitif perçu pour la Ligue des champions et la Ligue Europa



Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ». Ligue des champions : 21 604 réponses, 112 valeurs manquantes ; Ligue Europa : 21 513 réponses, 203 valeurs manquantes.

Note de lecture : 25 % des supporters estiment qu'il n'y a beaucoup de suspense (8 sur 10) dans la Ligue des champions.

Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

posé les questions suivantes pour la Ligue 1 et la Ligue des champions, les deux compétitions phares que suivent les supporteurs vivant en France :

(1) « À partir de quel niveau minimum de suspense (sur une échelle de 0 : pas de suspense du tout à 10 : beaucoup de suspense) seriez-vous susceptible de perdre de l'intérêt pour [la compétition en question] ? » ;

(2) « À partir de quel niveau minimum de suspense (sur une échelle de 0 : pas de suspense du tout à 10 : beaucoup de suspense) seriez-vous susceptible de ne plus aller au stade ? de [la compétition en question] » ;

(3) « À partir de quel niveau minimum de suspense (sur une échelle de 0 : pas de suspense du tout à 10 : beaucoup de suspense) seriez-vous susceptible de ne plus regarder les matchs de [la compétition en question] à la TV ? ».

La première question est relative à la « demande globale », la deuxième question à la demande de « stade » et la troisième question à la demande de « télévision ». Les figures III-A à III-C indiquent les réponses à ces trois questions. Les courbes sont construites comme une fonction de survie à l'aide de l'estimateur Kaplan-Meier. On classe les réponses par ordre décroissant et on cumule le nombre de réponses. Ensuite, on calcule la proportion d'individus qui « survivent » après chaque niveau de « suspense ». Ainsi, si l'incertitude était maximale (valeur 10), aucune des personnes interrogées ne renonceraient à aller au stade ou à regarder le match à la télévision (figures III-B et III-C). Autrement dit, pour 100 % des enquêtés, le niveau minimum de suspense qui les ferait renoncer est inférieur à 10. Au contraire, s'il n'y avait aucune incertitude (valeur 0 du niveau de suspense), seuls environ 30 % des supporteurs ne renonceraient pas à aller au stade (figure III-B). Nous avons tracé en annexe (voir figures A2-III à A2-V en annexe) les mêmes courbes en différenciant les « jeunes » des « vieux » supporteurs. Par rapport aux commentaires précédents, on constate que les plus jeunes apparaissent moins sensibles à l'incertitude des résultats que leurs homologues plus âgés : de manière générale, 30 à 40 % des 26 ans et moins continueraient à s'intéresser aux compétitions même si le suspense en était totalement absent, alors que cette proportion n'est que de 20 % chez les plus de 26 ans.

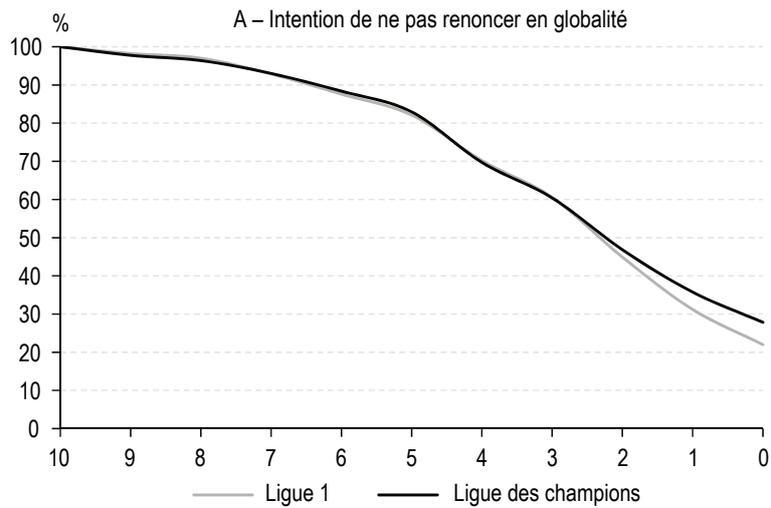
Deux résultats importants se dégagent de toutes ces figures. Le premier est évidemment

la similitude des courbes pour les différentes demandes et les différentes compétitions. Finalement, les intentions de renoncer au football ne dépendent ni des compétitions ni de la manière de consommer. Le second résultat intéressant concerne le manque de lien entre l'incertitude perçue et l'intention de consommer. Alors que le niveau déclaré d'incertitude est totalement différent entre la Ligue 1 et la Ligue des champions (figures I et II), les intentions de renoncer ont quasiment le même profil. Cet effet s'inscrit dans la littérature évoquée plus haut concernant le football professionnel, dont les résultats sont peu probants sur le lien entre incertitude du résultat et demande des supporteurs. La demande des fans de football ne se résume pas à un comportement de pur consommateur : ils préfèrent voir leur équipe favorite gagner (Buraimo & Simmons, 2008), comme c'est le cas en Ligue 1, et voir une équipe de renom même si elle est beaucoup plus forte (Pawlowski & Anders, 2012), comme c'est le cas en Ligue des champions. C'est d'ailleurs ce que déclarent spontanément les supporteurs qui vont au stade quand on leur demande à quoi ils sont le plus sensibles (figure IV) : l'ambiance (84 %), la victoire (78 %) et la combativité des joueurs (72 %) sont citées le plus souvent, le suspense arrivant très loin derrière (21 %).

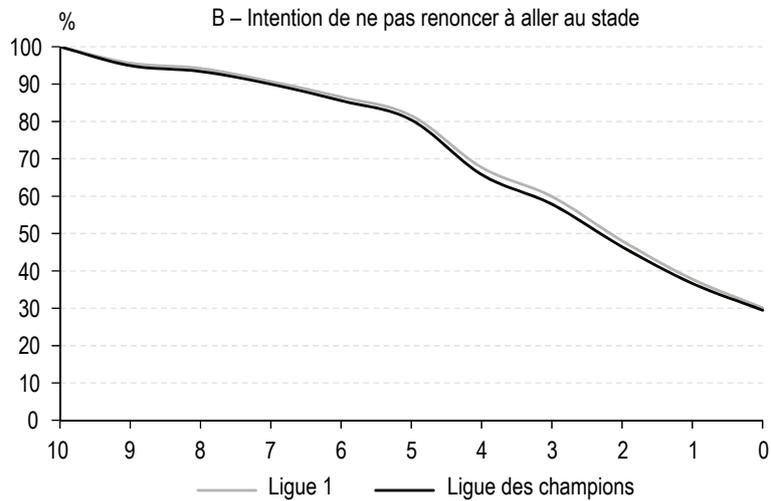
Nous avons également demandé aux enquêtés de se prononcer sur 17 items relatifs à plusieurs types d'incertitude en Ligue 1 et en Ligue des champions : l'incertitude de court terme, l'incertitude de moyen terme et l'incertitude de long terme. Quatre réponses sont possibles pour chaque item : « pas du tout d'accord », « plutôt pas d'accord », « plutôt d'accord » et « tout à fait d'accord ». Le tableau 2 présente la distribution des réponses. La plupart se concentrent sur les deux réponses médianes sauf pour les quelques items suivants. Pour la Ligue 1, 45 % des personnes interrogées sont « tout à fait d'accord » sur le fait que la lutte pour le maintien en Ligue 1 est intéressante, et près des trois-quarts des personnes interrogées (72.1 %) ne sont « pas du tout d'accord » sur le fait que de nombreuses équipes différentes gagnent le championnat de France de Ligue 1. Cela renvoie d'une part à la domination du PSG depuis 2011 et d'autre part au fait que le championnat de France comptait trois équipes reléguées jusqu'à la saison 2015-2016.

À la fin de la saison 2016-2017, au lieu de trois relégués de Ligue 1 vers la Ligue 2, le 18^e a fait un match de barrage contre le 3^e de Ligue 2, donnant une chance supplémentaire de rester au

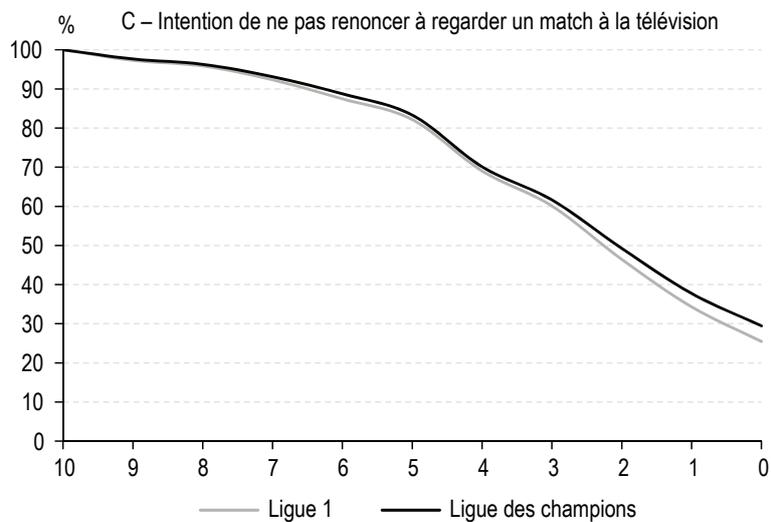
Figure III – Intention de consommer du football



Note de lecture : 70 % des supporters continueront à s'intéresser à la Ligue 1 et à la Ligue des champions si le niveau de suspense est à 4.



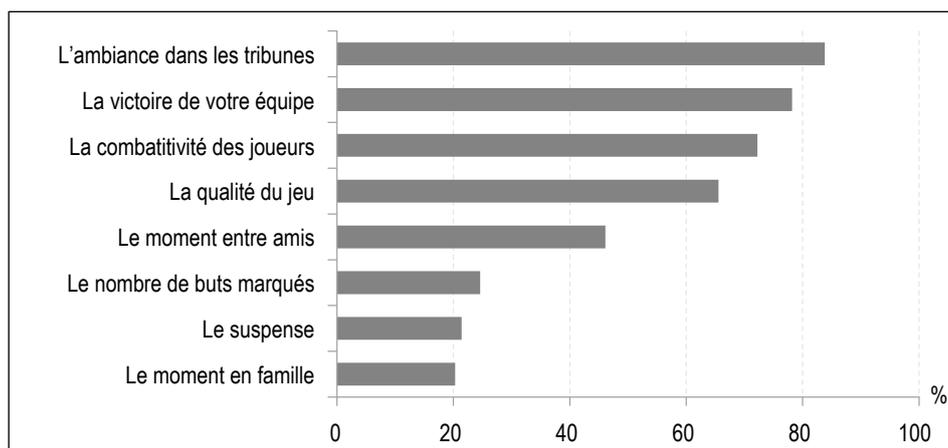
Note de lecture : 66 % des supporters continueront à aller au stade pour la Ligue des champions si le niveau de suspense est à 4.



Note de lecture : 70 % des supporters continueront à regarder un match de la Ligue des champions à la télévision si le niveau de suspense est à 4.

Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ».
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Figure IV – Au stade, à quoi êtes-vous le plus sensible... (plusieurs réponses possibles)



Note de lecture : 21.4 % des supporters qui vont au stade sont sensibles au « suspense » du match.
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Tableau 2 – Distribution des réponses aux questions sur l'incertitude perçue par les fans

	Pas du tout d'accord	Plutôt pas d'accord	Plutôt d'accord	Tout à fait d'accord	Valeurs manquantes	Pas d'accord	D'accord
Court terme (CT)							
<i>Avant un match, il est impossible de savoir quelle équipe va gagner</i>							
Ligue 1	9.8	42.0	39.8	4.9	3.4	51.8	44.7
Ligue des champions	9.8	39.2	38.1	8.9	4.1	49.0	47.0
<i>Les équipes qui reçoivent ont autant de chances de gagner que les équipes qui se déplacent</i>							
Ligue 1	9.7	39.8	40.7	6.2	3.6	49.5	46.9
Ligue des champions	13.8	46.3	29.8	5.6	4.5	60.1	35.4
<i>Il n'y a pas de véritable favori à chaque match</i>							
Ligue 1	19.6	49.3	24.9	2.5	3.8	68.9	27.4
Ligue des champions	17.7	51.9	22.0	3.9	4.5	69.6	25.9
Moyen terme (MT)							
Ligue 1							
<i>La lutte pour les places qualificatives en Ligue des champions reste un enjeu intéressant jusqu'à la fin de la saison</i>	5.8	12.1	46.0	32.7	3.4	17.9	78.7
<i>La lutte pour les places qualificatives en Ligue Europa reste un enjeu intéressant jusqu'à la fin de la saison</i>	5.0	13.2	51	27.3	3.5	18.2	78.3
<i>La lutte pour le maintien reste un enjeu intéressant jusqu'à la fin de la saison</i>	3.7	5.9	41.9	45	3.5	9.6	86.9
Ligue des champions							
<i>La lutte pour le titre reste un enjeu intéressant jusqu'à la fin de saison</i>	3.7	7.3	38	46.5	4.6	11.0	84.5
<i>Les phases de poule en Ligue des champions restent un enjeu intéressant</i>	6.8	19.7	49.6	19.7	4.6	26.5	69.3
<i>Les phases finales en Ligue des champions restent un enjeu intéressant</i>	2.6	4.2	36.5	52.2	4.5	6.8	88.7

→

Tableau 2 (suite)

	Pas du tout d'accord	Plutôt pas d'accord	Plutôt d'accord	Tout à fait d'accord	Valeurs manquantes	Pas d'accord	D'accord
Long terme (LT)							
Ligue 1							
<i>De nombreuses équipes différentes gagnent le championnat</i>	72.1	17.9	3.7	1.7	4.6	90.0	5.4
<i>De nombreuses équipes différentes se qualifient pour la Ligue des champions</i>	33.7	47.3	12.3	2.2	4.5	81.0	14.5
<i>De nombreuses équipes différentes se qualifient pour la Ligue Europa</i>	12.0	35.4	42.5	5.9	4.3	47.4	48.4
<i>De nombreuses équipes différentes ont été reléguées</i>	7.3	23.1	49.3	16.3	4.0	30.4	65.6
Ligue des champions							
<i>De nombreuses équipes différentes gagnent la Ligue des champions</i>	16.3	43.3	25.8	10.0	4.7	59.6	35.8

Note : 21 716 observations.

Note de lecture : 5.0 % des supporters ne sont pas du tout d'accord avec l'affirmation : « La lutte pour les places qualificatives en Ligue Europa reste un enjeu intéressant jusqu'à la fin de la saison ».

Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

plus haut niveau et donc limitant l'incertitude sur le maintien en première division. C'est pourquoi la Ligue de football professionnel (LFP) a introduit un système de pré-barrages similaire à des *play-offs* à partir de la saison 2017-2018 pour les clubs qui finissent la saison 3^e, 4^e et 5^e de Ligue 2. Selon le format retenu, le cinquième affronte d'abord le quatrième chez celui-ci, puis le vainqueur de cette confrontation joue contre le troisième (sur le terrain de ce dernier) ; enfin, le vainqueur de cette dernière rencontre est opposé au 18^e de Ligue 1 en barrage de promotion/relégation³.

Pour la Ligue des champions, environ la moitié des personnes interrogées sont « tout à fait d'accord » sur le fait que l'enjeu pour le titre (46.5 %) et les phases finales à élimination directe (52.2 %) restent intéressants. L'incertitude de la Ligue des champions est relativement importante car aucune équipe entre 1992 et 2016 n'a gagné deux fois de suite⁴ dans la version actuelle. Dans la phase de poules, les équipes à gros budgets se retrouvent avec des équipes à « petits » budgets et n'ont aucune difficulté à se qualifier du fait de la grande variabilité des budgets. Ensuite, dans la phase finale, lorsqu'elles se rencontrent entre elles, la différence de budget se réduisant, les qualités sportives et collectives reprennent le dessus, l'incertitude également.

4. Les déterminants de l'équilibre compétitif perçu en Ligue 1 et en Ligue des champions

Pour tester l'effet des différentes incertitudes sur l'équilibre compétitif perçu, nous avons construit dans un premier temps un « score » regroupant les items dans leur incertitude respective grâce aux réponses des personnes interrogées : « pas du tout d'accord » vaut 1, « plutôt pas d'accord » vaut 2, « plutôt d'accord » vaut 3 et « tout à fait d'accord » vaut 4. Les non-réponses valent 0. Pour la Ligue 1, l'incertitude de court terme (CT) est composée des trois items suivants : « *Avant un match, il est impossible de savoir quelle équipe va gagner* », « *Les équipes qui reçoivent ont autant de chances de gagner que les équipes qui se déplacent* » et « *Il n'y a pas de véritable favori à chaque match* ». Ainsi, le score de l'incertitude de CT est compris entre 0 et 12. Tout comme l'incertitude de moyen terme (MT) qui est composée des trois items suivants : « *La lutte pour les places qualificatives en Ligue des champions reste un enjeu intéressant jusqu'à la fin de la saison* », « *La lutte pour les places qualificatives en Ligue Europa reste un enjeu intéressant jusqu'à la fin de la saison* » et « *La*

3. Ces matchs supplémentaires ont pour objectif d'éviter le désintérêt des supporters.

4. Ce n'est arrivé qu'en 2017-2018, saison où le Real Madrid a gagné sa troisième Ligue des champions consécutive.

lutte pour le maintien reste un enjeu intéressant jusqu'à la fin de la saison ». L'incertitude de long terme (LT) est composée des quatre items suivants : « *De nombreuses équipes différentes gagnent le championnat* », « *De nombreuses équipes différentes se qualifient pour la Ligue des champions* », « *De nombreuses équipes différentes se qualifient pour la Ligue Europa* » et « *De nombreuses équipes différentes ont été reléguées* ». Dans ce cas, le score de l'incertitude de LT est compris entre 0 et 16. Pour la Ligue des champions, l'incertitude de CT est composée des mêmes trois items que la Ligue 1. L'incertitude de MT est composée des trois items suivants : « *La lutte pour le titre reste un enjeu intéressant jusqu'à la fin de saison* », « *Les phases de poule en Ligue des champions restent un enjeu intéressant* » et « *Les phases finales en Ligue des champions restent un enjeu intéressant* ».

L'incertitude de LT est composée du seul item « *De nombreuses équipes différentes gagnent la Ligue des champions* ». Le score est compris entre 0 et 4. Pour mesurer la cohérence globale de l'agrégation des items, nous utilisons la mesure classique de l'alpha de Cronbach. Les valeurs calculées sont comprises entre 0.69 pour l'incertitude de CT de la Ligue 1 et 0.86 pour l'incertitude de MT de la Ligue des champions. En dehors de la première, toutes sont supérieures à 0.7. On peut donc considérer que les indicateurs construits représentent bien ces différentes notions d'incertitude.

Nous estimons l'équilibre compétitif perçu par les supporters en Ligue 1 et en Ligue des champions en fonction des indicateurs d'incertitude de CT, MT et LT construits et en contrôlant des caractéristiques individuelles (tableau 3). Les

Tableau 3 – Les variables qui influencent la perception de l'équilibre compétitif (EC)

Variable dépendante : niveau de suspense perçu de 0 à 10	MCO		Logit ordonné	
	Ligue 1	Ligue des champions	Ligue 1	Ligue des champions
Incertitude de court terme (CT)	0.018 (0.011)	0.027*** (0.008)	0.018** (0.008)	0.037*** (0.007)
Incertitude de moyen terme (MT)	0.112*** (0.010)	0.064*** (0.007)	0.086*** (0.007)	0.051*** (0.007)
Incertitude de long terme (LT)	0.110*** (0.009)	0.555*** (0.017)	0.083*** (0.006)	0.531*** (0.016)
<i>Intéressé par le foot</i> (0 à 10)	0.146*** (0.016)	0.221*** (0.011)	0.109*** (0.011)	0.200*** (0.011)
<i>Degré d'attachement à un club de foot</i> (0 à 10)	0.097*** (0.015)	0.011 (0.011)	0.055*** (0.010)	0.025** (0.010)
<i>Supporter de plusieurs clubs</i> (réf. Aucun ou 1 club)	0.241*** (0.048)	0.117*** (0.036)	0.163*** (0.034)	0.123*** (0.034)
<i>Abonné</i> (réf. Pas d'abonnement)	0.142*** (0.055)	-0.205*** (0.041)	0.091** (0.038)	-0.217*** (0.038)
<i>Club de supporters</i> (réf. Aucun club de supporters)	-0.286*** (0.054)	-0.219*** (0.040)	-0.228*** (0.037)	-0.192*** (0.037)
<i>Lecture de la presse papier</i> (réf. Jamais)				
1 fois par semaine	0.177*** (0.053)	0.060 (0.040)	0.162*** (0.037)	0.006 (0.037)
Abonnés à une revue ou tous les jours	0.167*** (0.054)	0.218*** (0.041)	0.148*** (0.038)	0.166*** (0.038)
<i>Lecture d'articles Internet</i> (réf. Pas de lecture)	0.088 (0.056)	0.044 (0.042)	0.074* (0.039)	0.046 (0.039)
<i>Abonnement à une chaîne payante</i> (réf. Jamais)				
Dans le passé	-0.242*** (0.063)	0.061 (0.047)	-0.171*** (0.044)	0.051 (0.044)
Oui	0.016 (0.044)	0.377*** (0.033)	0.010 (0.031)	0.332*** (0.031)

→

Tableau 3 (suite)

Variable dépendante : niveau de suspense perçu de 0 à 10	MCO		Logit ordonné	
	Ligue 1	Ligue des champions	Ligue 1	Ligue des champions
<i>Pratique du football</i> (réf. Non)				
Dans le passé	0.046 (0.055)	-0.053 (0.041)	0.035 (0.038)	-0.055 (0.038)
Oui	0.077 (0.061)	0.158*** (0.046)	0.046 (0.042)	0.149*** (0.042)
<i>Culture footballistique</i> (réf. Inexistante ou faible)				
Moyenne	-0.103 (0.168)	-0.238* (0.126)	-0.069 (0.118)	-0.294** (0.117)
Élevée	-0.274 (0.167)	-0.269** (0.125)	-0.186 (0.117)	-0.321*** (0.116)
Très élevée	-0.486*** (0.171)	-0.255** (0.128)	-0.359*** (0.120)	-0.224* (0.119)
<i>Au moins un des parents pratiquant le football</i> (réf. Non ou ne sait pas)				
Oui	0.002 (0.038)	0.001 (0.029)	0.007 (0.027)	0.012 (0.027)
<i>Homme</i> (réf. Femme)	-0.447*** (0.102)	-0.050 (0.077)	-0.306*** (0.071)	-0.025 (0.071)
<i>Âge</i> (réf. Inférieur ou égal à 25 ans)				
De 26 à 40 ans	-0.110* (0.057)	-0.354*** (0.043)	-0.083** (0.040)	-0.338*** (0.040)
de 41 à 60 ans	-0.072 (0.078)	-0.711*** (0.058)	-0.065 (0.054)	-0.664*** (0.054)
61 ans et plus	-0.392*** (0.145)	-0.526*** (0.108)	-0.322*** (0.101)	-0.548*** (0.100)
<i>Vit en Île-de-France</i> (réf. Non)	-0.146*** (0.053)	-0.067* (0.040)	-0.087** (0.037)	-0.064* (0.037)
<i>Études supérieures</i> (réf. Pas d'études supérieures)	0.035 (0.040)	-0.243*** (0.030)	0.038 (0.028)	-0.242*** (0.028)
<i>Nombre d'enfants</i> (réf. Aucun)				
Un seul enfant	0.016 (0.072)	-0.051 (0.054)	0.003 (0.050)	-0.068 (0.050)
Deux enfants ou plus	0.219*** (0.063)	0.010 (0.047)	0.148*** (0.044)	0.001 (0.043)
<i>En emploi</i> (réf. Sans activité)	0.116* (0.065)	0.020 (0.049)	0.085* (0.045)	0.029 (0.045)
<i>Revenu</i> (réf. Pas de revenus)				
Inférieur à 30 000 €	-0.147** (0.062)	-0.131*** (0.047)	-0.102** (0.043)	-0.141*** (0.043)
Supérieur ou égal à 30 000 €	-0.185** (0.078)	-0.108* (0.059)	-0.131** (0.054)	-0.109** (0.054)
<i>Personne vivant seule</i> (réf. Ménage d'au moins deux personnes)	0.117*** (0.047)	-0.069* (0.035)	0.077** (0.033)	-0.072** (0.033)
N	19 386	19 609	19 386	19 609
R ²	0.06	0.21		

Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

estimations sont obtenues par des MCO (deux premières colonnes) et par un Logit ordonné (deux dernières colonnes). Étant donné le nombre élevé de modalités de la variable dépendante (entre 0 et 10), les résultats sont très proches. Avant d'analyser quel type d'incertitude reflète l'équilibre compétitif perçu par les supporters, intéressons-nous aux caractéristiques individuelles qui influencent le suspense perçu (les variables socio-démographiques ont été introduites à titre de contrôle).

Quelle que soit la spécification, un des éléments qui conditionne la valorisation du suspense est le fort intérêt du football en général, et pour la Ligue 1 et pour la Ligue des champions en particulier. Toutes choses égales par ailleurs, les variables reflétant ce fort intérêt sont positives et significatives (« *Intéressé par le foot* », « *Intéressé par la Ligue 1* », « *Intéressé par la Ligue des champions* », « *Supporter de plusieurs clubs* »). Ce qui signifie que les personnes intéressées par les compétitions sont conscientes que, même s'il existe un fort déséquilibre entre les équipes, le football reste le sport collectif le plus aléatoire et que rien n'est perdu d'avance. En d'autres termes, l'intérêt du football gomme la baisse de l'incertitude perçue, même en cas de domination forte d'une équipe comme c'est le cas en Ligue 1 depuis 2011.

Intéressons-nous maintenant à deux variables : « *Abonné* » et « *Appartenance à un club de supporters* ». Les abonnés sont les personnes qui en début de saison réservent une place pour les dix-neuf matchs de championnat à domicile – souvent prioritaires pour les autres matchs – et qui ont donc un lien fort avec leur club. Les coefficients associés aux estimations sont positifs et significatifs pour la Ligue 1 et négatifs et significatifs pour la Ligue des champions. Les abonnés de l'enquête sont tous abonnés à des équipes françaises. Ainsi, être abonné c'est être fortement intéressé par son club et son histoire et on retrouve un effet qui gomme la diminution de l'incertitude perçue lorsqu'on estime les effets sur la Ligue 1. En revanche, ces mêmes abonnés sont plus critiques vis-à-vis des « grosses » équipes européennes – notamment des sommes d'argent en jeu – et pensent que l'incertitude est faible car ce sont toujours les mêmes équipes qui gagnent. Ce qui est à la fois vrai et faux (*supra*). L'effet de la variable « *Appartenance à un club de supporters* » est négatif et significatif pour les deux compétitions, comme si les supporters de ce type étaient conscients que leur équipe n'avait aucune ou inversement beaucoup de chance de gagner.

Les supporters utilisateurs de média valorisent différemment l'incertitude perçue. Si le fait de consulter ou non des sites internet sur le football ne modifie pas la perception de l'incertitude, les supporters qui lisent la presse tous les jours ou sont abonnés à une revue spécialisée ont tendance également à penser qu'il existe une forte incertitude tant en Ligue 1 qu'en Ligue des champions. Les résultats pour la variable « *Abonnement à une chaîne payante sportive* » révèlent deux comportements distincts. Pour la Ligue 1, il y a les supporters qui pensent que l'incertitude est faible et qui ont résilié leur abonnement (l'effet négatif et significatif devant la modalité « dans le passé ») et pour la Ligue des champions, il y a les supporters qui pensent que l'incertitude est forte et sont abonnés à une chaîne payante (l'effet positif et significatif devant la modalité « oui »).

La pratique du football, passée ou présente, et un des parents pratiquant ou ayant pratiqué le football n'a aucune influence sur la perception de l'incertitude en Ligue 1 et en Ligue des champions. En revanche, les individus déclarant avoir une culture footballistique élevée ou très élevée ont le sentiment qu'il y a moins de suspense dans les compétitions (Ligue 1 et Ligue des champions).

À quel type d'incertitude les supporters sont-ils le plus sensibles lorsqu'on leur demande de mesurer le suspense des deux compétitions « phares » : le championnat de Ligue 1 et la Ligue des champions. Concernant la Ligue 1, c'est l'incertitude de MT et de LT qui semble déterminer leur jugement, à savoir l'incertitude sur le champion et sur la récurrence des titres : il n'est alors pas étonnant qu'ils placent le degré de suspense assez bas car depuis 2011, le championnat est dominé par le PSG. Le degré de suspense perçu de la Ligue des champions est déterminé surtout par l'incertitude de LT liée à la récurrence des trophées : il est vrai que passé les phases de poule, l'incertitude des matchs entre les grosses « écuries » européennes reste importante.

* *
*

L'idée que l'équilibre compétitif augmente l'utilité des supporters, donc leurs dépenses et donc les revenus des clubs professionnels, est au cœur de l'économie du sport en général, et de l'économie du football en particulier. Budzinski & Pawloski (2017) soulignent que la notion

d'équilibre compétitif est souvent mise en avant pour justifier les décisions des ligues professionnelles pour changer les règles des compétitions ou la distribution des droits TV. Comme le montre la littérature, la relation entre équilibre compétitif et demande des supporters est loin d'être évidente. Leur perception de l'équilibre des compétitions est bien déterminée et relativement cohérente mais d'autres facteurs apparaissent plus importants en termes de « consommation » de football : attachement au club, probabilité de victoire, rencontre contre une grosse équipe, présence des « stars » sur la feuille de match, etc.

Dans cet article, nous nous sommes intéressés à l'équilibre compétitif tel qu'il est perçu par les supporters de football ayant répondu à l'enquête *Quel supporter êtes-vous ?* sur le site lequipe.fr. Cet équilibre compétitif subjectif revêt cependant plusieurs dimensions qui dépendent des compétitions : pour le championnat de l'élite française il s'explique surtout par une incertitude de moyen et de long terme (le nom du champion et la récurrence des titres) alors que pour la Ligue

des champions, il s'agit plutôt d'un suspense de long terme (vainqueurs successifs). De ce fait, la Ligue 1 sera perçue comme moins incertaine que la Ligue des champions. Enfin, l'incertitude du résultat est loin d'être le seul facteur explicatif de la demande de football puisqu'environ 30 % des supporters se déclarent prêts à toujours regarder les matchs ou à aller au stade dans le cas hypothétique où il n'y aurait plus aucun suspense...

Malgré sa forte croissance liée principalement à celle des droits TV ces vingt dernières années (environ 10 % par an en France) et contrairement à une idée reçue liée à sa très forte popularité, le football reste aujourd'hui plutôt un « petit business » comparé à d'autres secteurs d'activité (Arrondel & Duhautois, 2018) : environ 0.2 % du PIB Français. L'évolution de la structure du budget des clubs (plus de droits audiovisuels, moins de billetterie) montre que la recherche sur la demande de football doit viser aujourd'hui à en expliquer les déterminants aussi bien sur les abonnements média que sur la fréquentation des stades. □

BIBLIOGRAPHIE

- Alavy, K., Gaskell, A., Leach, S. & Szymanski, S. (2010).** On the Edge of Your Seat: Demand for Football on Television and the Uncertainty of Outcome Hypothesis. *International Journal of Sport Finance*, 5(2), 75–95. <https://econpapers.repec.org/RePEc:jsf:intjsf:v:5:y:2010:i:2:p:75-95>
- Andreff, W. (2014).** French professional football: how much different? In: Goddard, J. & Sloane, P., *Handbook on the Economics of Professional Football*, pp. 298–321. Cheltenham, UK: Edward Elgar.
- Andreff, W. (2018).** Financial and Sporting Performance in French Football Ligue 1: Influence on the Player Market. *International Journal of Financial Studies*, 6(4), 1–17. <https://doi.org/10.3390/ijfs6040091>
- Arrondel, L. & Duhautois, R. (2018).** *L'argent du football*. Paris : Cepremap.
- Baidina, K. & Parshakov, P. (2017).** Uncertainty of Outcome and Attendance: Evidence from Russian Football. Mimeo.
- Benz, M.-A., Brandes, L. & Franck, E. (2009).** Do Soccer Associations Really Spend on a Good Thing? Empirical Evidence on Heterogeneity in the Consumer Response to Match Uncertainty of Outcome. *Contemporary Economic Policy*, 27(2), 216–235. <https://dx.doi.org/10.1111/j.1465-7287.2008.00127.x>
- Budzinski, O. & Pawlowski, T. (2017).** The Behavioral Economics of Competitive Balance: Theories, Findings, and Implications. *International Journal of Sport Finance*, 12(2), 109–122.
- Buraimo, B. (2008).** Stadium attendance and television audience demand in English league football. *Managerial and Decision Economics*, 29(6), 513–523. <https://doi.org/10.1002/mde.1421>
- Buraimo, B. & Simmons, R. (2008).** Do sports fans really value uncertainty of outcome? Evidence from the English Premier League. *International Journal of Sport Finance*, 3(3), 146–155.

- Buraimo, B. & Simmons, R. (2009).** A tale of two audiences: Spectators, television viewers and outcome uncertainty in Spanish football. *Journal of Economics and Business*, 61(4), 326–338.
<https://doi.org/10.1016/j.jeconbus.2008.10.002>
- Buraimo, B. & Simmons, R. (2015).** Uncertainty of outcome or star quality? Television audience demand for English Premier League football. *International Journal of the Economics of Business*, 22(3), 449–469.
<https://doi.org/10.1080/13571516.2015.1010282>
- Cairns, J. A. (1987).** Evaluating changes in league structure: The reorganization of the Scottish Football League. *Applied Economics*, 19(2), 259–275. <https://doi.org/10.1080/00036848700000101>
- Caruso, R., Addesa, F. & Di Domizio, M. (2019).** The Determinants of the TV Demand of Soccer: Empirical Evidence On Italian Serie A for The Period 2008-2015. *Journal of Sports Economics*, 20(1), 25–49.
<https://doi.org/10.1177%2F1527002517717298>
- Czarnitzki, D. & Stadtmann, G. (2002).** Uncertainty of outcome versus reputation: Empirical evidence for the first German football division. *Empirical Economics*, 27(1), 101–112. <https://doi.org/10.1007/s181-002-8361-7>
- Falter, J.-M. & Perignon, C. (2000).** Demand for Football and intramatch winning probability: An essay on the glorious uncertainty of sports. *Applied Economics*, 32(13), 1757–1765. <https://doi.org/10.1080/000368400421101>
- Falter, J.-M., Pérignon, C. & Vercruyse, O. (2008).** Impact of Overwhelming Joy on Consumer Demand: The Case of a Soccer World Cup Victory. *Journal of Sports Economics*, 9(1), 20–42.
<https://doi.org/10.1177%2F1527002506296548>
- Forrest, D. & Simmons, R. (2002).** Outcome uncertainty and attendance demand: The case of English soccer. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 51(2), 229–241.
<https://doi.org/10.1111/1467-9884.00314>
- Forrest, D. & Simmons, R. (2006).** New Issues in Attendance demand: The Case of the English Football League. *Journal of Sports Economics*, 7(3), 247–266. <https://doi.org/10.1177%2F1527002504273392>
- Forrest, D., Beaumont, J., Goddard, J. & Simmons, R. (2005).** Home advantage and the debate about competitive balance in professional sports leagues. *Journal of Sports Sciences*, 23(4), 439–445.
<https://doi.org/10.1080/02640410400021641>
- Forrest, D., Simmons, R. & Buraimo, B. (2005).** Outcome Uncertainty and the Couch Potato Audience. *Scottish Journal of Political Economy*, 52(4), 641–661. <https://dx.doi.org/10.1111/j.1467-9485.2005.00360.x>
- Garcia, J. & Rodriguez, P. (2002).** The determinants of football match attendance revisited: Empirical evidence from the Spanish Football League. *Journal of Sports Economics*, 3(1), 18–38.
<https://doi.org/10.1177%2F152700250200300103>
- Hart, R. A., Hutton, J. & Sharot, T. (1975).** A Statistical Analysis of Association Football Attendances. *Journal of the Royal Statistical Society, Series C Applied Statistics*, 24(1), 17–27. <https://doi.org/10.2307/2346700>
- Jang, H. & Lee, Y. H. (2015).** Outcome uncertainty, governance structure, and attendance: A study of the Korean professional football league. In: Lee, Y. H. & Fort, R. (Eds.), *The sports business in the Pacific rim*, pp. 59–81. New York: Springer.
- Jennett, N. (1984).** Attendances, Uncertainty of Outcome and Policy in Scottish League Football. *Scottish Journal of Political Economy*, 31(2), 176–198. <https://doi.org/10.1111/j.1467-9485.1984.tb00472.x>
- Kuper, S. & Szymanski, S. (2018).** *Soccernomics: Why England lose: and other curious phenomena explained*. London: HarperSport.
- Neale, W. C. (1964).** The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. *Quarterly Journal of Economics*, 78(1), 1–14.
<https://doi.org/10.2307/1880543>
- Nielsen Sports (2018).** *World Football Report*. <https://niensports.com/reports/world-football-2018/>
- Pawlowski, T. & Anders, C. (2012).** Stadium attendance in German professional football – The (un)importance of uncertainty of outcome reconsidered. *Applied Economics Letters*, 19(16), 1553–1556.
<https://doi.org/10.1080/13504851.2011.639725>

- Peel, D. A. & Thomas, D. A. (1988).** Outcome uncertainty and the demand for football: An analysis of match attendances in the English Football League. *Scottish Journal of Political Economy*, 35(3), 242–249. <https://doi.org/10.1111/j.1467-9485.1988.tb01049.x>
- Peel, D. A. & Thomas, D. A. (1992).** The demand for football: Some evidence on outcome uncertainty. *Empirical Economics*, 17(2), 323–331. <https://econpapers.repec.org/RePEc:spr:empeco:v:17:y:1992:i:2:p:323-31>
- Peel, D. A. & Thomas, D. A. (1996).** Attendance demand: An investigation of repeat fixtures. *Applied Economics Letters*, 3(6), 391–394. <https://doi.org/10.1080/135048596356294>
- Rottenberg, S. (1956).** The baseball players' labor market. *Journal of Political Economy*, 64(3), 242–258. <https://doi.org/10.1086/257790>
- Scelles, N. (2017).** Star quality and competitive balance? Television audience demand for English Premier League football reconsidered. *Applied Economics Letters*, 24(19), 1399–1402. <https://doi.org/10.1080/13504851.2017.1282125>
- Schreyer, D., Schmidt, S. L. & Torgler, B. (2017).** Game outcome uncertainty and the demand for international football games: Evidence from the German TV market. *Journal of Media Economics*, 30(1), 31–45. <https://econpapers.repec.org/RePEc:taf:jmedec:v:30:y:2017:i:1:p:31-45>
- Schreyer, D., Schmidt, S. L. & Torgler, B. (2018).** Game Outcome Uncertainty and Television Audience Demand: New Evidence from German Football. *German Economic Review*, 19(2), 140–161. <https://doi.org/10.1111/geer.12120>
- Schreyer, D., Schmidt, S. L., Sascha, L. & Torgler, B. (2016).** Against all odds? Exploring the role of game outcome uncertainty in season ticket holders' stadium attendance demand. *Journal of Economic Psychology*, 56, 192–217. <https://doi.org/10.1016/j.joep.2016.07.006>
- Simmons, R. (1996).** The demand for English league football: A club-level analysis. *Applied Economics*, 28(2), 139–155. <https://doi.org/10.1080/000368496328777>
- Statista (2018).** *European Football Benchmark: France*.
- Wilson, P. & Sim, B. (1995).** The demand for semi-pro league football in Malaysia 1989–91: A panel data approach. *Applied Economics*, 27(1), 131–138. <https://doi.org/10.1080/00036849500000015>
-

ANNEXE 1

RECENSEMENT ANNOTÉ DES ÉTUDES SUR L'EFFET DE L'INCERTITUDE DANS LE FOOTBALL

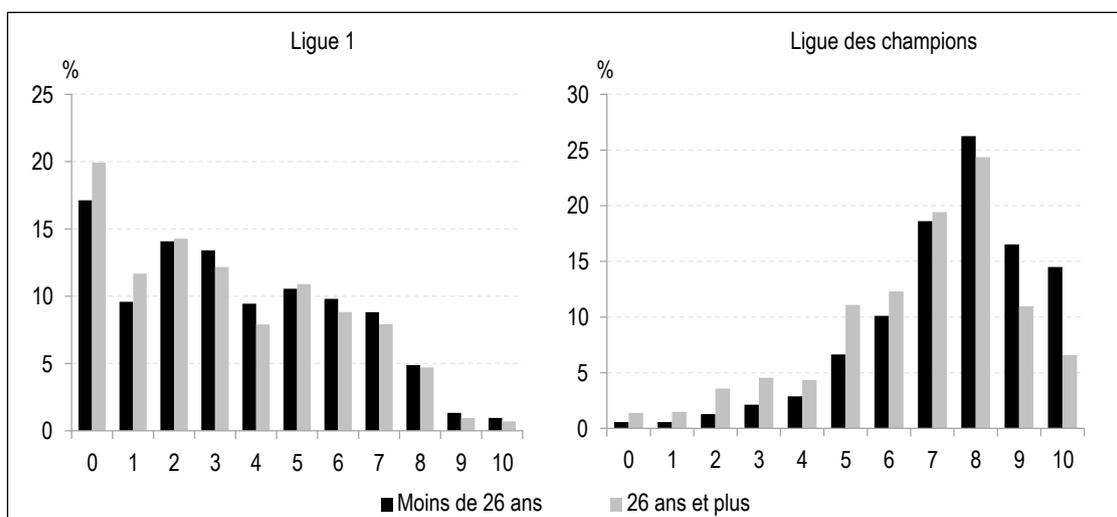
Auteurs (ordre chronologique)	Championnat	Saisons analysées	Variable dépendante	Mesure de l'incertitude	Effet de l'incertitude (résultat principal)
Hart, Hutton & Sharott (1975)	4 équipes anglaises	1969-1972	Affluence au stade	Différence de classement des équipes	Non
Jennett (1984)	Écosse Division 1	1975-1981	Affluence au stade	Nombre de points nécessaires pour gagner le championnat	Oui
Cairns (1987)	Écosse Division 1	1979-1980	Affluence au stade	Classement des deux équipes	Oui
Peel & Tomas (1988)	Angleterre Divisions 1-4	1981-1982	Affluence au stade	Paris sportifs (probabilité de gagner à domicile)	Non
Peel & Tomas (1992)	Angleterre Divisions 1-4	1986-1987	Affluence au stade	Paris sportifs	Non
Wilson & Sim (1995)	Malaisie (Ligue avec deux divisions)	1989-1992	Affluence au stade	Différence de points entre les équipes	Non
Bairbridge, Cameron & Dawson (1996)	Angleterre Division 1	1993-1994	Affluence au stade	Différence de classement	Non
Peel & Tomas (1996)	Écosse Divisions 1-3	1991-1992	Affluence au stade	Paris sportifs	Non
Falter & Pérignon (2000)	France Division 1	1997-1998	Affluence au stade	Différence de classement	Oui
Czarnitzki & Stadtmann (2002)	Allemagne Division 1	1996-1998	Affluence au stade	Paris sportifs	Non
Forrest & Simmons (2002)	Angleterre Divisions 1-4	1997-1998	Affluence au stade	Paris sportifs	Non
Garcia & Rodriguez (2002)	Espagne Division 1	1992-1996	Affluence au stade	Différence de classement	Oui (les auteurs ne croient pas à leurs résultats)
Forrest, Buraimo & Simmons (2005)	Angleterre Division 1	1993-2002	Audience télévisée	Différence de points entre les équipes	Oui
Forrest, Beaumont, Goddard & Simmons (2005)	Angleterre Division 1	1997-1998	Affluence au stade	Paris sportifs	Oui
Forrest & Simmons (2006)	Angleterre Divisions 2-4	1999-2002	Affluence au stade	Différence de points entre les équipes	Non
Johnsen & Solvoll (2007)	Norvège et Danemark	1998-2007	Audience télévisée	Différence de points entre les équipes	Non
Buraimo & Simmons (2008)	Angleterre Division 1	2000-2006	Affluence au stade	Paris sportifs	Non
Buraimo (2008)	Angleterre Division 1	1997-2004	Affluence au stade / Audience télévisée	Différence de points entre les équipes	Non
Falter, Pérignon & Vercoyuisse (2008)	France Division 1	1996-2000	Affluence au stade	Différence de points entre les équipes	Non
Benz, Brandes & Franck (2009)	Allemagne Division 1	1999-2004	Affluence au stade	Différence de classement, différence de points entre les équipes, paris sportifs	Non (seulement pour les équipes avec une forte affluence)



Auteurs (ordre chronologique)	Championnat	Saisons analysées	Variable dépendante	Mesure de l'incertitude	Effet de l'incertitude (résultat principal)
Buraimo & Simmons (2009)	Espagne Division 1	2003-2007	Affluence au stade / Audience télévisée	Paris sportifs	Non (affluence) / Oui (audience)
Madalozzo & Berber Villar (2009)	Bésil Division 1	2003-2006	Affluence au stade	Différence de classement	Non
Alavy, Gaskell, Leach & Szymanski (2010)	Angleterre	2002-2006	Audience télévisée	Paris sportifs	Oui
Anders & Pavlowski (2012)	Allemagne Division 1	2005-2006	Affluence au stade	Paris sportifs	Non
Pawlowski (2013)	Allemagne Division 1 (6 matchs)	Sept.-Oct. 2011	Intention de consommer (enquêtes stades et bars)	Questions sur différentes mesures de l'incertitude	Oui
Jang & Lee (2015)	Corée du sud K-League (ligue fermée)	1987-2011	Affluence au stade	Paris des victoires et points moyens avant le match	Oui
Cox (2015)	Angleterre Division 1	2004-2012	Affluence au stade / Audience télévisée	Paris sportifs	Non (affluence) / Oui (audience)
Buraimo & Simmons (2015)	Angleterre Division 1	2001-2008	Audience télévisée	Paris sportifs	Non
Schreyer, Schmidt & Torgler (2018)	Allemagne Division 1	2012-2013	Affluence au stade des abonnés	Paris sportifs	Oui (les abonnées arrivent plus tôt dans le stade lorsque l'incertitude augmente)
Schreyer, Schmidt & Torgler (2018)	Allemagne Division 1 et Coupe	2011-2015	Audience télévisée	Paris sportifs	Oui (championnat) / Non (coupe)
Schreyer, Schmidt & Torgler (2017)	Coupes du monde et championnats d'Europe (Euro), Équipe nationale Allemande	6 coupes du monde, 5 Euro et matchs amicaux	Audience télévisée	Différence de classement FIFA	Oui sur les matchs avec intérêt
Scelles (2017)	Angleterre Division 1	2013-2014	Audience télévisée	Paris sportifs	Non
Caruso, Addesa & Di Domizio (2019)	Italie Division 1	2008-2015	Audience télévisée	Différence de classement, différence de points entre les équipes, paris sportifs	Non
Bedina & Pershakov (2017)	Russie Division 1	2012-2014	Affluence au stade	Paris sportifs	Non

ANNEXE 2

Figure A2-I – L'équilibre compétitif perçu pour la Ligue 1 et la Ligue des champions

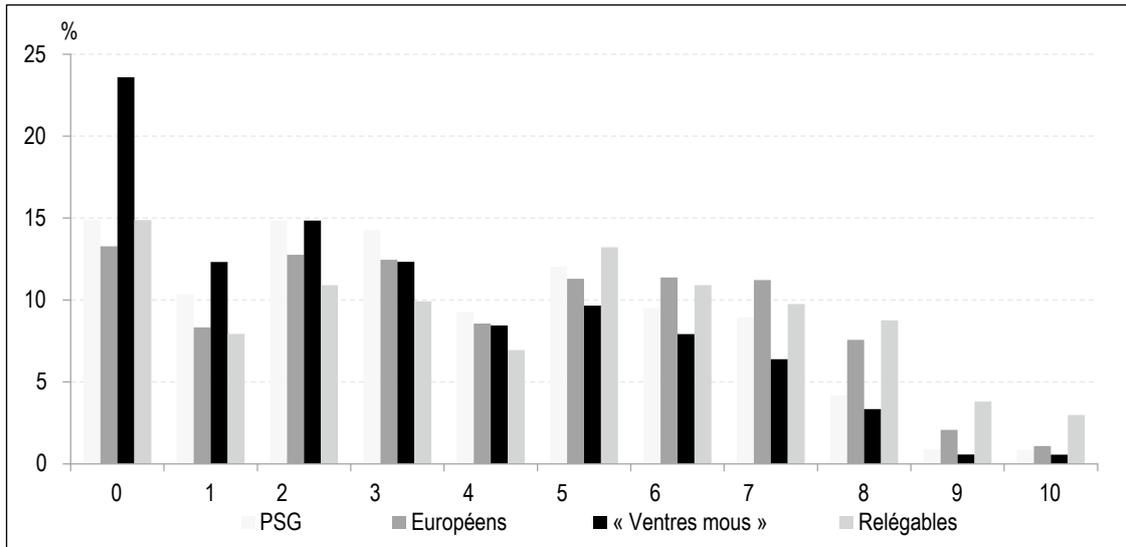


Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ».
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Tableau A2-1 – Classement de la Ligue 1 saison 2015-2016

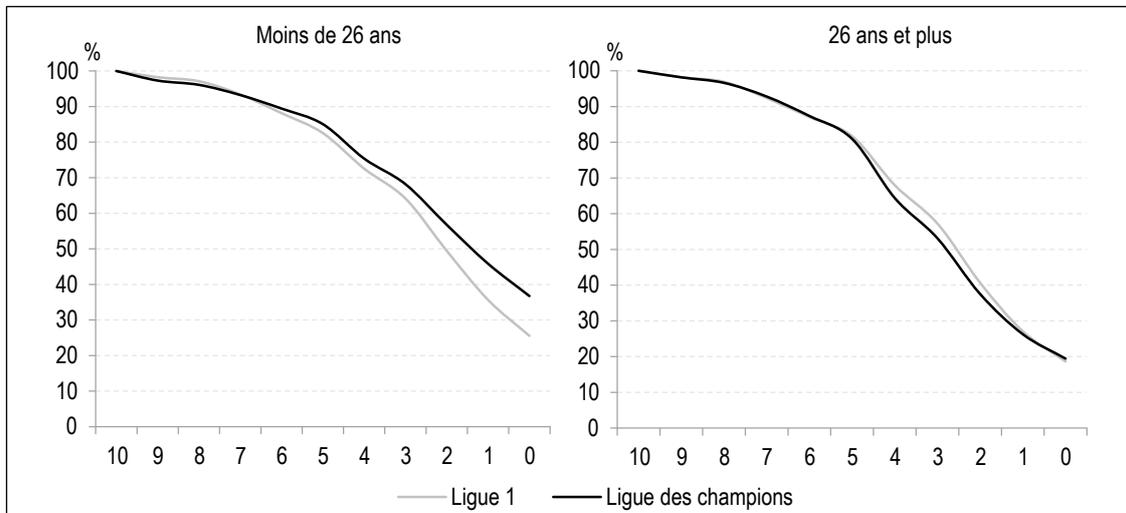
	Club	Points
1	Paris Saint-Germain	96
2	Olympique lyonnais	65
3	AS Monaco	65
4	OGC Nice	63
5	Lille OSC	60
6	AS Saint-Etienne	58
7	SM Caen	54
8	Stade Rennais FC	52
9	Angers SCO	50
10	SC Bastia	50
11	Girondins de Bordeaux	50
12	Montpellier HSC	49
13	Olympique de Marseille	48
14	FC Nantes	48
15	FC Lorient	46
16	EA Guingamp	44
17	Toulouse FC	40
18	Stade de Reims	39
19	GFC Ajaccio	37
20	ES Troyes AC	18

Figure A2-II – L'équilibre compétitif perçu pour la Ligue 1 selon les clubs



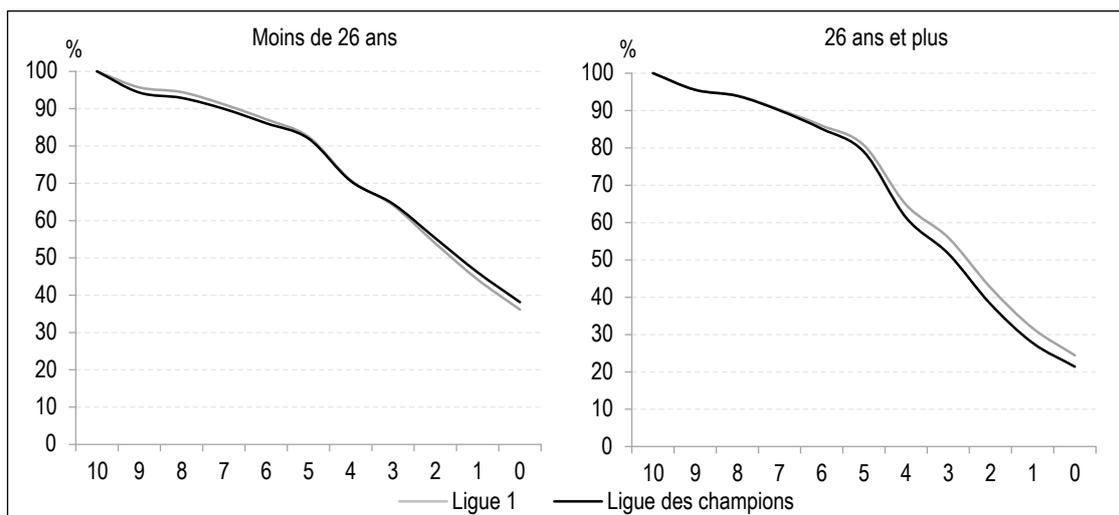
Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ».
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Figure A2-III – Intention de ne pas renoncer en globalité à la Ligue 1 ou la Ligue des champions



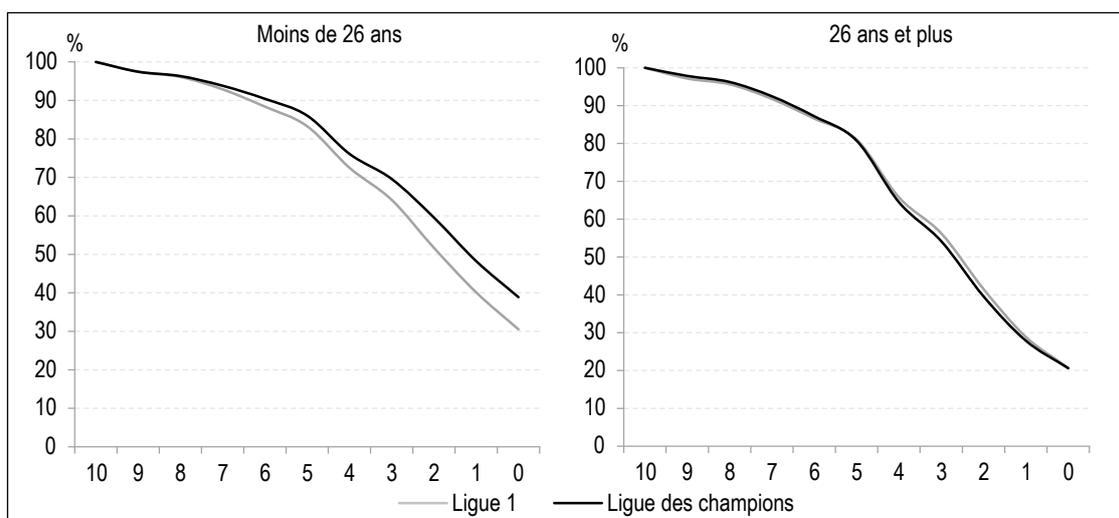
Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ».
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Figure A2-IV – Intention de ne pas renoncer à aller au stade pour la Ligue 1 ou la Ligue des champions



Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ».
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Figure A2-V – Intention de ne pas renoncer à regarder un match à la télévision de Ligue 1 ou de Ligue des champions



Note : 0 = « pas de suspense » / 10 = « beaucoup de suspense ».
 Source : PSE-L'Équipe, enquête *Quel supporter êtes-vous ?* - 2016.

Les discriminations dans l'accès au logement en France : un *testing* sur les aires urbaines métropolitaines

Discrimination in Access to Housing: A Test on Urban Areas in Metropolitan France

Julie Le Gallo*, Yannick L'Horty**, Loïc du Parquet***
et Pascale Petit**

Résumé – Nous mesurons l'ampleur des discriminations dans l'accès au logement locatif du parc privé à l'aide d'un *testing*, sur les 50 plus grandes aires urbaines de France métropolitaine, couvrant plusieurs motifs de discriminations : l'âge, l'origine, le lieu de résidence et leurs combinaisons. Le protocole a consisté à envoyer, entre juin et décembre 2016, cinq candidatures fictives en réponse à une sélection de 5 000 annonces de locations dans le parc privé réparties sur l'ensemble du territoire métropolitain. Nous vérifions si les discriminations dans l'accès au logement dépendent des caractéristiques des candidats, de celles de l'annonceur et de celles du contexte local. Nous ne mettons pas en évidence de discriminations en raison de l'âge du candidat à la location. Nous trouvons un effet positif du fait de signaler habiter en HLM ou dans une cité. Nous trouvons en outre d'importantes discriminations selon l'origine, pénalisant les candidats dont les noms et prénoms signalent une origine maghrébine ou une origine africaine.

Abstract – *We measure the extent of discrimination in access to rental housing in the private sector using a test in the 50 largest urban areas in Metropolitan France, covering several grounds of discrimination: age, origin, place of residence and combinations thereof. The protocol consisted of sending, between June and December 2016, five fictitious applications in response to a selection of 5,000 advertisements for private rental housing spread throughout Metropolitan France. We check whether discrimination in access to housing depends on the characteristics of the applicants, those of the advertiser and those of the local context. We do not highlight any discrimination based on the applicant's age. We do find a positive effect of reporting living in low-rent housing or in a housing estate. We also find significant discrimination according to origin, which penalises applicants whose surnames and first names indicate a North African or African origin.*

Codes JEL / JEL Classification : J14, R31

Mots-clés : discrimination, logement, *testing*, expérience contrôlée

Keywords: *discrimination, housing, testing, controlled experiment*

* CESAER UMR1041, AgroSup Dijon, INRAE, Université de Bourgogne Franche-Comté (julie.le-gallo@agrosupdijon.fr) ; ** ERUDITE, Université Gustave Eiffel, UPEC, TEPP-CNRS (yannick.lhorty@u-pem.fr; pascal.petit@u-pem.fr) ; *** Université du Mans, GAINS et TEPP-CNRS (loic.duparquet@univ-lemans.fr)

Cette étude a bénéficié du soutien de l'Agence Nationale de la Recherche dans le cadre du projet « Discrimination dans l'Accès au Logement : un *testing* de couverture nationale » (projet DALTON ANR-15-CE28-0004). Nous remercions Abdel Diakité, Victor Simo et Ludovic Ebene Moussole pour leur assistance dans la construction de la base de données, ainsi que Stéphanie Abrial, Sarah Bénichou, François Bonnet, François Cochard, Quentin David et Patrick Simon pour leurs commentaires ainsi que l'ensemble des participants de la journée « Discrimination dans l'accès au logement » le 15 décembre 2017 à l'UPEM. Nous remercions enfin les rapporteurs anonymes pour leurs remarques constructives qui ont conduit à des améliorations substantielles de l'article.

Reçu le 13 mars 2018, accepté après révisions le 29 avril 2019.

Citation : Le Gallo, J., L'Horty, Y., du Parquet, L. & Petit, P. (2019). Discrimination in Access to Housing: A Test on Urban Areas in Metropolitan France. *Economie et Statistique / Economics and Statistics*, 513, 27–45. <https://doi.org/10.24187/ecostat.2019.513.2004>

Un propriétaire ne peut pas refuser un locataire en raison de son origine ethnique, de son âge, de son sexe, de son orientation sexuelle ou de l'un des autres critères prohibés par le droit. C'est contraire au principe d'égalité et constitue une atteinte à la dignité des personnes. Pour cette raison, les discriminations sont punies par la loi et exposent leur auteur à trois ans d'emprisonnement et 45 000 euros d'amende aux termes de l'article 225-2 du code pénal. Malgré ce cadre juridique répressif, le logement est perçu par près d'un français sur deux comme l'un des principaux domaines susceptibles de produire des discriminations, suivant en cela l'emploi ou les contrôles de police (Défenseur des Droits, 2017). Les discriminations dans l'accès au logement impliquent en outre un coût économique pour la collectivité. Les personnes discriminées subissent des coûts de recherche élevés et elles sont limitées dans leur choix de localisation, ce qui réduit leur mobilité résidentielle. Cela peut alimenter les processus de ségrégation urbaine et générer une mauvaise allocation spatiale qui pénalise en retour l'accès à l'emploi et à la formation. Certes, la ségrégation urbaine peut être observée en l'absence de toute discrimination sur le marché du logement, mais elle peut être encore plus critique en présence de ce type de discrimination. La mise en évidence de telles situations est donc utile à la fois pour analyser les déterminants des fractures spatiales qui organisent la séparation territoriale des groupes sociaux, mais aussi pour définir les politiques publiques les plus pertinentes. Les actions classiques de lutte contre les ségrégations urbaines, qui visent à organiser la mixité sociale des quartiers ou leur désenclavement par des opérations de rénovations urbaines, ne ciblent pas explicitement la question de la discrimination dans l'accès au logement. Cette question recouvre donc à la fois des enjeux de connaissance pour la recherche et des enjeux essentiels pour la conception même des politiques publiques.

Afin de mesurer de façon objective les discriminations dans l'accès au logement, à l'emploi ou à d'autres marchés, la méthode la plus utilisée est celle du *testing* qui consiste à comparer les retours obtenus en réponse à une annonce (ici, une annonce immobilière) par deux candidats semblables en tous points sauf sur la caractéristique dont on veut tester l'effet. Cette méthode a été appliquée aux États-Unis depuis les années 1980 et a apporté de multiples preuves expérimentales de discriminations dans l'accès au logement, en particulier pour les discriminations

ethno-raciales qui sont les plus étudiées (Yinger, 1986 ; Page, 1995 ; Choi *et al.*, 2005 ; Hanson & Hawley, 2011). Elle a aussi été appliquée dans de nombreux pays d'Europe¹. Avec la généralisation d'internet et des sites d'annonces immobilières, le test de correspondance, qui consiste à envoyer des courriels de demandes fictives, s'est imposé comme le moyen le plus efficace de réaliser des *testing* sur le marché du logement. Dans un survol récent de la littérature, Flage (2018) identifie 29 études scientifiques ayant appliqué cette méthode dans 15 pays différents. Il conclut que les candidats dont le nom de famille signale une origine étrangère ont deux fois moins de chance en moyenne d'être en location que les autres candidats.

Des *testings* ont été fréquemment utilisés en France dans d'autres domaines, principalement sur le marché du travail, et pour étudier divers motifs de discrimination : le sexe (Duguet & Petit, 2005 ; Petit, 2007), l'origine apparente (Berson, 2011), la réputation du lieu de résidence (Bunel *et al.*, 2016), la religion (Adida *et al.*, 2010 ; Pierné, 2013), et les effets croisés de plusieurs motifs (Duguet *et al.*, 2010 ; L'Horty *et al.*, 2011 ; Petit *et al.*, 2014). Cependant, cette méthode a encore été peu utilisée pour le marché du logement, à l'exception d'un petit nombre de travaux réalisés sur des échantillons restreints et/ou ciblés sur des territoires particuliers (Bonnet *et al.*, 2015 ; Acolin *et al.*, 2016 ; Bunel *et al.*, 2017).

Notre objectif est de mesurer et d'interpréter les discriminations dans l'accès au logement locatif du parc privé à l'aide d'un *testing* sur les grandes aires urbaines de la métropole permettant de mesurer trois motifs de discriminations : l'âge (discrimination à l'encontre des jeunes), l'origine (en distinguant Maghreb et Afrique de l'Ouest), le type de résidence (HLM ou cité) au moment de la prise de contact. Notre protocole a consisté à envoyer cinq candidatures fictives en réponse à une sélection de 5 000 annonces de locations dans le parc privé, réparties sur les 50 plus grandes aires urbaines du territoire métropolitain. Nous avons ainsi constitué une base de données expérimentales de 25 000 observations dont nous effectuons une exploitation statistique.

1. Pour la Suède : Ahmed *et al.*, 2008 et 2010 ; Bengtsson *et al.*, 2012 ; Carlsson & Eriksson, 2014. Pour l'Espagne : Bosch *et al.*, 2010. Pour l'Italie : Baldini & Federici, 2011. Pour la Belgique : Heylen *et al.*, 2015. Pour la Grèce : Drydakis, 2011.

L'article s'organise de la façon suivante. Dans la première section, nous proposons un survol rapide de la littérature sur les discriminations dans l'accès au logement. La deuxième section est consacrée à la présentation du protocole de collecte du *testing*. Les résultats obtenus sont présentés dans la section 3, par type de discriminations. Une dernière section résume les principaux résultats et discute leurs implications en termes de politique publique.

1. Discrimination dans l'accès au logement : un survol de la littérature

Les rares études qui ont mesuré les discriminations dans l'accès au logement en France reposent sur de petits échantillons et/ou couvrent des territoires particuliers. Ainsi, Bonnet *et al.* (2015) exploitent une campagne réalisée au printemps 2009 en réponse à 250 annonces de biens immobiliers à louer en Île-de-France, pendant laquelle des figurants ont téléphoné aux loueurs en révélant des informations personnelles selon un protocole préétabli. Acolin *et al.* (2016) ont envoyé des courriers électroniques pour six candidats dont cinq avec un nom de famille évoquant une origine étrangère, en réponse à 300 annonces publiées sur internet pendant deux mois, entre avril et mai 2014, jusqu'à atteindre un total de 1 800 réponses sur la France entière. Ces deux études concluent à l'existence de fortes discriminations dans l'accès au logement selon l'origine ethnique, sans pouvoir véritablement ni les spatialiser, ni en interpréter les causes.

Bunel *et al.* (2017) ont étudié l'ampleur de la discrimination dans l'accès au logement de candidats d'origine maghrébine à Paris à l'aide d'une campagne de *testing* menée entre début avril et fin mai 2016. Les chercheurs ont envoyé quatre messages sollicitant la visite d'un logement en réponse à 504 annonces immobilières émanant de particuliers ou d'agences immobilières, soit un envoi de 2 016 réponses. Ils concluent que les discriminations à l'encontre des individus d'origine maghrébine sont fortes dans l'accès au logement parisien et qu'elles sont peu liées à la fragilité financière supposée de ces individus. L'individu de référence d'origine française a un taux de réponse de 18.7 % à ses demandes de visite de logement. Pour l'individu d'origine maghrébine, ce taux est de 12.9 %, soit un tiers de chances en moins de recevoir une réponse favorable à sa demande de visite. Si le candidat maghrébin précise qu'il est

fonctionnaire, son taux de réponse est de 15.5 %, ce qui reste inférieur au candidat d'origine française qui ne fait pas mention de sa situation. Lorsque c'est le candidat d'origine française qui envoie le même signal de stabilité, son taux de réponse atteint 42.9 %. Un signal de stabilité professionnelle et financière n'augmente fortement les chances d'accès au logement que pour les candidats d'origine française, ce qui suggère une forte discrimination à la Becker à l'encontre des candidats maghrébins. Ce résultat est vérifié, que l'annonce émane d'un particulier ou d'une agence immobilière.

Le contraste est net avec la littérature qui s'est développée dans d'autres pays, en particulier aux États-Unis où les *testings* sont appliqués à l'accès au logement depuis les années 1970 et où les preuves expérimentales de discriminations sont multiples même si les discriminations ethno-raciales sont les plus étudiées (Yinger, 1986 ; Page, 1995 ; Ondrich *et al.*, 2000 et 2003 ; Zhao, 2005 ; Zhao *et al.* 2006 ; Ahmed & Hammarstedt, 2008 ; Ahmed *et al.*, 2010 ; Bosch *et al.*, 2010 ; Hanson & Hawley, 2011). Tous ces travaux mettent en évidence que les minorités sont victimes d'un traitement différencié sur le marché de l'immobilier. Globalement, ces candidats sont moins souvent contactés et le nombre et la qualité des biens qui leur sont présentés sont plus faibles. Flage (2018) à partir d'un *survey* de près de 30 études qui utilisent la méthode des tests de correspondance, conclut à une différence du simple au double entre les candidats au profil ethnique majoritaire et ceux appartenant à la minorité.

Les études montrent de façon générale que la discrimination sur le marché du logement peut varier selon les caractéristiques du loueur, le type de bien loué, l'environnement de la localisation du bien, le signal d'intégration et/ou de stabilité économique transmis par le candidat. Aux États-Unis, les premières études prenant en compte à un niveau désagrégé l'environnement ethnique sont celles de Yinger (1986), Page (1995) et Roychoudhury & Goodman (1996). Depuis, Ondrich *et al.* (2003) puis Hanson & Santas (2014), à l'aide d'échantillons plus importants, ont précisé la relation entre la proportion de blancs dans un quartier et l'ampleur de la discrimination. Les résultats obtenus par ces travaux indiquent l'existence d'un point critique de retournement (*neighborhood tipping*) dans la répartition de la discrimination. Ondrich *et al.* (2003) montrent ainsi que la discrimination est plus faible dans les quartiers où sont sur-représentées les

minorités et Hanson & Santas (2014) montrent que la discrimination est la plus forte dans les quartiers relativement mixtes.

Afin d'identifier le rôle joué par la discrimination statistique (à la Arrow) plusieurs auteurs ont croisé le signal ethnique ou religieux avec un signal suggérant, soit un niveau social élevé (maîtrise d'un langage soutenu), soit un niveau d'intégration (maîtrise de la langue), soit encore une plus grande stabilité professionnelle (bonne situation professionnelle). Globalement, quelle que soit sa forme, un signal de qualité réduit l'ampleur de la discrimination sans la faire disparaître complètement (Massey & Lundy, 2001 ; Ahmed *et al.*, 2010 ; Bosch *et al.*, 2010 ; Baldini & Federici, 2011 ; Drydakis, 2011).

Le rôle joué par les agences immobilières est plus ambigu. Dans leur survol de la littérature, Oh & Yinger (2015) soulignent que les agents immobiliers jouent un rôle central dans le processus d'orientation (*racial and ethnic steering*) des candidats vers les quartiers où leur ethnie est la plus représentée. Selon un test réalisé en Belgique par Heylen *et al.* (2015) afin de mesurer la proportion d'agences acceptant d'écarter des candidats à la location d'origine étrangère, les agences immobilières semblent être un canal massif de transmission de discrimination par la clientèle. Les résultats antérieurs d'Ahmed & Hammarstedt (2008) confirment cette conclusion pour la Suède. Toutefois, ceux de Carpusor & Lodge (2006) pour la ville de Los Angeles indiquent que les agences ne discriminent pas plus que les propriétaires privés et ceux de Bosch *et al.* (2010) pour différentes villes espagnoles identifient même un rôle modérateur des agences.

2. Le protocole de collecte des données

Le protocole consiste à effectuer 100 tests dans chacune des 50 plus grandes aires urbaines, soit 5 000 tests. Dans chaque territoire, et pour chaque offre de location dans le parc privé, le test consiste à envoyer de courts messages émanant de cinq individus fictifs en réponse à des offres de logement pour demander aux offreurs des informations complémentaires en vue d'une visite et relever les suites que ces derniers leur réservent. Les 50 plus grandes aires urbaines regroupent au total 36.6 millions d'habitants, soit 57.1 % de la population française.

2.1. Profil des candidats

Parmi les cinq candidats fictifs, deux figurent dans tous les envois. Les trois autres sont tirés au sort pour chaque annonce testée dans un ensemble de six. Nous détaillons dans le tableau 1 les profils des individus fictifs, dont les autres caractéristiques sont par ailleurs similaires.

Les comparaisons deux à deux des réponses données à ces différents candidats permettent de mesurer des discriminations selon plusieurs critères (l'origine, le lieu de résidence au moment de la réponse à l'annonce et l'âge du candidat) et selon leur croisement. Pour mesurer l'effet de l'origine, signalée par le nom de famille du candidat, il faut neutraliser celui de la nationalité : une nationalité étrangère peut en effet être perçue par un bailleur comme un signal d'instabilité dans le logement. Pour capter l'effet de l'origine de demandeurs français, tous les messages des candidats dont le nom signale une origine étrangère font mention de

Tableau 1 – Profils des huit candidats fictifs

	Individu Prénom NOM	Âge	Consonance du prénom et du nom	Autres caractéristiques
Pour toutes les annonces	Sébastien PETIT	41	Française	Neutre
	Mohamed CHETTOUH	41	Maghrébine	Neutre
En alternance, pour une annonce sur deux	Mounir MEHDAOUI	41	Maghrébine	Fonctionnaire
	Kévin DURAND	22	Française	Neutre
	Frédéric ROUSSEAU	41	Française	Habite dans une cité ou un HLM
	Désiré SAMBOU	41	Africaine (subsaharienne)	Prénom à consonance non-musulmane
	Nordine M'BAREK	22	Maghrébine	Neutre
	Karim BENCHARGUI	41	Maghrébine	Habite dans une cité ou un HLM

Source : *testing* DALTON – TEPP CNRS.

leur nationalité française, soit directement, soit indirectement comme dans le cas d'un candidat (Mounir Mehdaoui) qui fait état de son statut de fonctionnaire (voir encadré).

Tous nos candidats sont des hommes. Nous avons fait le choix de ne pas explorer l'effet du sexe du candidat qui est une question à part entière. Le survol extensif de littérature réalisé par Flage (2018) indique qu'il existe des pénalités à l'encontre des hommes sur le marché du logement et qu'elles sont plus marquées pour les personnes d'origine étrangère. En moyenne, dans les 14 études qui ont exploré ce phénomène, une femme a 30 % de chances de plus d'être conviée à visiter un appartement qu'un homme et cet écart est de 50 % entre candidats d'origine étrangère.

Les éventuelles différences de traitement des candidats peuvent provenir d'une discrimination statistique ou d'une discrimination par les préférences. Afin d'identifier ces deux sources de discrimination à l'instar de Massey *et al.* (2001), Ahmed *et al.* (2010), Bosch *et al.* (2010), Baldini *et al.* (2011) et Drydakis (2011), nous ajoutons un signal de qualité du candidat. Nous avons fait le choix de la mention du statut de fonctionnaire dans le message de prise de contact, comme l'avaient fait Bunel *et al.* (2017). Cette mention, qui signale la stabilité financière d'un candidat, va permettre de distinguer l'effet des deux formes de discrimination mentionnées plus haut. La comparaison deux à deux de ces profils nous permet d'identifier les effets respectifs de chaque caractéristique sur l'accès au logement.

2.2. Sélection des annonces

Notre expérimentation porte sur les locations de biens immobiliers dans le parc privé, à l'exclusion des locations saisonnières de vacances. Les annonces testées sont publiées soit par des particuliers, soit par des agences immobilières. Le fait de tester à la fois des annonces de particuliers et des annonces d'agences immobilières nous permet de déterminer si les éventuelles discriminations sont liées aux préférences personnelles des conseillers des agences et des loueurs et/ou aux caractéristiques des quartiers voire si les agences jouent un rôle actif dans l'orientation des candidats, ce qui constitue des hypothèses largement discutées dans la littérature (Choi *et al.*, 2005 ; Oh & Yinger, 2015).

Nous avons choisi de nous focaliser sur un type de bien intermédiaire, les F2, car il s'agit du bien le plus demandé et le plus offert. Le fait de se limiter aux F2 ne nous empêche pas de considérer une grande variété de qualité de biens, de surfaces, de localisations et de loyers. Nous répondons à des annonces récentes, publiées depuis moins de trois jours, en visitant quotidiennement les sites internet d'annonces de locations les plus utilisés : Le Bon Coin, seloger.com, logic-immo, etc. Ces annonces sont sélectionnées au hasard chaque jour dans celles qui répondent à nos critères de recherche : annonces publiées depuis moins de trois jours, pour un appartement de type F2, dans l'une des communes de l'aire urbaine, en dehors des locations saisonnières. Nous n'avons retenu que les annonces pour lesquelles on pouvait identifier à la fois la nature de l'offreur (agences ou particulier) et la localisation du bien (nom du quartier dans la commune). Nous n'avons jamais testé deux fois le même offreur.

Les caractéristiques des annonces auxquelles nous avons répondu sont les suivantes. Les loyers médians et moyens pour ces F2 sont aux environs de 500 €. Un cinquième des annonces correspond à des meublés. Ces annonces sont issues pour plus de 70 % d'entre elles du site Le Bon Coin, qui est, selon les données de Médiamétrie, le site le plus consulté et le plus utilisé en France. Le processus de réponse ainsi que le contenu des messages envoyés en réponse aux offres de location sont décrits en encadré.

Au total, sur l'ensemble des 50 principales aires urbaines de France métropolitaine, 5 008 annonces ont été testées, ce qui correspond à l'envoi de 25 040 messages personnalisés de demande d'information pour une visite. Parmi ces annonces, près des deux tiers émanent de particuliers (3 235) et plus d'un tiers émanent d'agences (1 773). Nous avons enrichi cette base en prenant en compte des variables complémentaires qui décrivent les caractéristiques de l'annonce, sa date de parution, le montant du loyer et des charges, la durée du bail, la surface du logement, l'étage, et la localisation du bien.

2.3. La sélectivité du marché immobilier

Sur les 5 008 annonces testées (5 messages envoyés pour chaque annonce), 1 228 ont reçu au moins une réponse, soit un taux de réponse de 24.5 %. Les trois quarts des annonces auxquelles nous avons répondu n'ont donc fait l'objet d'aucune suite. Le taux de réponse est particulièrement faible pour les annonces publiées sur le site

Le Bon Coin, avec un taux de réponse de 11.6 %. Le site Le Bon Coin est ainsi majoritaire dans les annonces auxquelles nous avons répondu (70.5 %) mais il est minoritaire dans les annonces ayant fait l'objet d'une réponse (33.3 %). La fréquence des non-réponses est très variable selon les aires urbaines. Le nombre maximum d'annonces ayant fait l'objet d'au moins un retour est de 45 à Perpignan et le minimum est de 13 à Annecy.

Nous nous focalisons sur les réponses non négatives reçues par les candidats. Sur l'ensemble des annonces testées, 20.9 % (soit 1 140 annonces) ont obtenu au moins une réponse non négative et 79.0 % (soit 3 868 annonces) n'ont fait l'objet d'aucune réponse non négative (tableau 2). Parmi les annonces ayant débouché sur au moins une réponse non négative, les annonceurs n'ont donné des réponses non négatives

ENCADRÉ – Contenu des messages envoyés en réponse aux offres de location

L'ordre d'envoi des demandes des cinq candidats à une même annonce a été déterminé par tirage au sort, de façon à ce que sur l'ensemble de l'échantillon le message de chacun ait été envoyé en premier un même nombre de fois. Parallèlement, nous avons permuté les messages entre les candidats au cours de la campagne de *testing*, de façon à ce qu'une inégalité de traitement ne puisse être imputable à la qualité différenciée des messages (nous alternons deux jeux de message, notés J1 et J2, tout au long du *testing*).

Les cinq candidats fictifs envoient le même jour et à quelques heures d'intervalle de courts messages par

Bonjour
Cette annonce correspond assez bien à ce que je recherche actuellement dans cette zone. Comment faire pour visiter cet appartement ? Quels documents sont nécessaires pour le louer ?
En vous remerciant
Sébastien PETIT

Madame, Monsieur,
L'appartement que vous proposez dans cette annonce correspond à ce que je cherche. Serait-il possible de le visiter ? Je voudrais préparer le dossier de location, pouvez-vous me donner la liste des pièces demandées ?
Merci par avance,
Mohamed Chettouh
Téléphone : XXX
Mail : XXX
Date de naissance : 13/03/1975
Nationalité Française

Bonjour,
Je suis fonctionnaire depuis 15 ans et je viens d'être muté. Je cherche un appartement dans ce quartier et votre annonce m'intéresse. Est-ce que je peux faire une visite ? Pouvez-vous également me dire quels seront les papiers à vous fournir ?
Cordialement,
Mounir MEHDAOUI

Bonjour, je suis intéressé par cet appartement à louer. Pourriez-vous me dire comment le visiter et quelles sont les formalités à remplir ? Je suis de nationalité Française et je vais avoir 41 ans.
Je vous remercie par avance.
Désiré SAMBOU

courriel en réponse aux annonces sélectionnées. Leurs caractéristiques distinctives apparaissent explicitement dans les champs obligatoires à remplir pour envoyer le message (nom/prénom) ou à défaut dans le contenu du message (âge, lieu de résidence actuel, statut professionnel).

Nous reportons ci-dessous les messages des candidats fictifs correspondant au jeu J1. Notons que l'identité, le numéro de téléphone et l'adresse de courrier électronique du candidat figurent dans les champs à remplir obligatoirement pour toute demande :

Bonjour, est-il encore possible de visiter cet appartement car il m'intéresse. Comme je vis en ce moment en HLM je ne sais pas quels documents vous voulez que je fournisse.
D'avance merci,
Frédéric Rousseau

Madame, Monsieur,
Je cherche un logement comme celui que vous louez. Quand pourrait-on se rencontrer pour le visiter et quels sont les papiers que je dois préparer (je vous précise que je suis de nationalité Française). Par contre, je ne suis pas libre demain après-midi car je passe mon permis.
Merci de votre réponse
Nordine M'BAREK

Bonjour, J'ai 22 ans et je cherche un appartement pour m'installer. Celui de cette annonce m'intéresse, est-ce que je pourrai le visiter, peut-on fixer un rendez-vous ? Quels documents je dois apporter ?
Merci
Kévin Durand

Bonjour,
J'habite en ce moment dans une cité et je veux déménager dans ce quartier. Est-ce que je peux visiter cet appartement ? Je vais préparer mon dossier de location, pouvez-vous me dire quelles sont les pièces administratives à prévoir ?
Bien cordialement,
Karim Benchargui – né le 17/02/1976 à Paris

Tableau 2 – Répartition du nombre de réponses non négatives par annonce testée

Nombre de réponses non négatives par annonce	Ensemble des annonces N = 5 008		Annonces émanant des agences immobilières N = 1 773		Annonces émanant des particuliers N = 3 235	
	Nombre	%	Nombre	%	Nombre	%
Aucune	3 868	79.1	1 090	61.8	2 778	85.9
1 et plus	1 140	20.9	683	38.5	457	14.1
1	367	32.2	148	21.7	219	47.9
2	245	21.5	136	19.9	109	23.8
3	172	15.1	110	16.1	62	13.6
4	162	14.2	130	19.0	32	7.0
5	194	17.0	159	23.3	35	7.7

Note de lecture : Parmi l'ensemble des annonces émanant d'agences immobilières testées, 683 annonceurs ont donné au moins une réponse non négative : 148 à un seul candidat (soit 21.7 %) et 159 aux cinq (soit 23.3 %).

Source : *testing* DALTON – TEPP CNRS.

à l'ensemble des cinq candidats que dans 17 % des cas. En d'autres termes, pour près de 83 % des 1 140 réponses non négatives de l'échantillon, les annonceurs ont été sélectifs et n'ont pas répondu à tous les candidats. Les différences de traitement entre les candidats – envoi ou pas d'une réponse – sont donc manifestes dans les données de ce *testing*.

Ces chiffres diffèrent selon que l'on considère les annonces publiées par les particuliers ou par des professionnels de l'immobilier. D'une part, les chances d'obtenir une réponse sont plus élevées lorsque l'annonceur est un professionnel : parmi l'ensemble des annonces testées, 14.1 % ont obtenu au moins une réponse non négative pour les 3 235 annonces publiées par les particuliers, contre 38.5 % pour les 1 773 annonces publiées par des agences immobilières. D'autre part, la sélectivité est marquée, mais les professionnels apparaissent moins sélectifs que les particuliers : si l'on se restreint aux annonces pour lesquelles au moins un candidat a obtenu une réponse non négative, 23.3 % des professionnels ont donné une réponse non négative aux cinq candidats contre seulement 7.7 % des offreurs particuliers.

3. Résultats du *testing*

3.1. De fortes discriminations selon l'origine des candidats à la location, pour les annonces publiées par les agences comme pour celles des particuliers

Les premiers résultats du *testing* sont présentés dans le tableau 3. Le taux de succès de notre

candidat de référence, Sébastien Petit, est de 13.9 % (il a reçu au total 698 réponses non négatives suite à ses 5 008 prises de contact avec les annonceurs). En revanche, le taux de succès du candidat qui signale une origine maghrébine par son patronyme, Mohamed Chettouh, est de 10.1 % (507 réponses non négatives suite à ses 5 008 prises de contact). La différence est donc de 3.8 points de pourcentage, soit, en termes relatifs 27.4 % de chances en moins pour le candidat d'origine maghrébine. Le taux de succès de Désiré Sambou, dont le patronyme signale une origine africaine, est quant à lui de 9.4 %. Il n'est comparable avec celui de Sébastien Petit, le candidat « référence », que pour les 2 776 réponses aux mêmes annonces ; sur ce champ, Petit a un taux de succès de 13.6 % (non présenté dans le tableau 3), soit 4.2 points de pourcentage de plus que Sambou. Tous ces écarts sont statistiquement significatifs au seuil de 1 %.

L'inférence statistique est réalisée par une procédure de type bloc-*bootstrap* dans laquelle le rééchantillonnage est effectué au sein de chaque aire urbaine, ce qui permet de prendre en compte la dimension « aire urbaine ». Nous devons corriger le problème de comparaison multiple lié au fait que nous effectuons plusieurs tests avec le même échantillon. Pour limiter le nombre de faux positifs lié aux comparaisons multiples et à l'instar de Carlsson & Eriksson (2014) dans leur étude de la discrimination ethnique sur le marché locatif en Suède, nous appliquons la procédure de correction proposée par Benjamini & Hochberg (1995). Contrairement à la correction de Bonferroni, cette procédure n'est pas trop conservatrice et elle est adaptée dans les cas où les conclusions globales ne dépendent pas

Tableau 3 – Tests de discrimination selon l'origine

	Nombre d'envois	Nombre de réponses non négatives	Taux de succès (%)	Écart (en points de pct) et test
Référence (PETIT)	5 008	698	13.9	
Candidat français d'origine maghrébine (CHETTOUH)	5 008	507	10.1	-3.8*** (<0.001) [<0.001]
Candidat français originaire d'Afrique subsaharienne (SAMBOU)	2 776	262	9.4	-4.2*** (<0.001) [<0.001]

Note : *** significatif au seuil de 1 %, ** au seuil de 5 %, * au seuil de 10 % ; probabilité critique non ajustée entre parenthèses ; probabilité critique ajustée par la méthode de Benjamini & Hochberg (1995) entre crochets ; inférence statistique par bloc-*bootstrap* (10 000 réplifications).
Source : *testing* DALTON – TEPP CNRS.

que d'un seul test². En effet, dans ce cas, il est raisonnable de tolérer quelques erreurs de type I (des faux positifs) pour augmenter la puissance statistique des tests. Nous indiquons dans le tableau 3 la probabilité critique obtenue sans ajustement et celle obtenue avec ajustement. Quel que soit l'indicateur considéré et même avec ajustement, les tests concluent à l'existence d'une discrimination entre les deux candidats, avec un risque d'erreur de 1 %.

Nous avons vérifié par ailleurs que ces écarts de taux de succès n'étaient pas dus à des différences dans l'ordre des envois des candidatures. À cette fin, nous avons recalculé les taux de réponses non négatives obtenues par les candidats, selon que leur demande a été formulée en premier, en deuxième, en troisième, en quatrième ou en cinquième position dans l'ordre des cinq demandes analogues qui ont été effectuées auprès de l'annonceur. C'est aussi une manière de vérifier que la permutation aléatoire des ordres d'envoi a été convenablement réalisée. On constate effectivement que les taux de succès sont plus élevés lorsque la demande est formulée en premier, avant les autres candidats, et que le profil d'ensemble des taux de succès est décroissant avec l'ordre d'envoi. On constate également que Sébastien Petit bénéficie plus fréquemment de réponses non négatives à ses demandes que Mohamed Chettouh ou Désiré Sambou, en fixant l'ordre d'envoi.

Les données indiquent donc qu'il existe des discriminations dans l'accès au logement selon l'origine des candidats à la location.

Nous vérifions aussi si la nature et l'intensité des discriminations diffèrent selon que l'annonce émane d'un particulier ou d'une agence

immobilière. Plusieurs études ont d'ores et déjà montré qu'il existait des différences de comportement entre les propriétaires privés et les agences : Choi *et al.* (2005), Ahmed *et al.* (2008), Bosch *et al.* (2010), et Heylen *et al.* (2015) trouvent que les professionnels discriminent globalement moins que les particuliers, tandis que Carpusor *et al.* (2006) aboutissent au résultat inverse. Il s'agit de savoir si les agents immobiliers qui sont partie prenante du processus de discrimination dans l'accès au logement, agissent selon leurs propres préférences ou s'ils répondent aux demandes exprimées de façon plus ou moins explicites par leurs clients propriétaires.

Nous avons vérifié tout d'abord que les résultats étaient similaires selon différentes définitions de l'indicateur de succès considéré. Nous prenons en compte trois types d'indicateurs : (i) le taux de réponses non négatives indique le nombre de réponses non négatives reçues par un candidat, par téléphone ou par courriel, rapporté au nombre de demandes envoyées ; (ii) le taux de « Rendez-vous de principe » indique le nombre de fois où un loueur a proposé le principe d'une visite au candidat, voire une date de visite, rapporté au nombre de demandes envoyées ; et (iii) le nombre de contacts est le nombre de fois où un loueur a contacté le candidat (par téléphone avec ou sans message, ou par courrier électronique).

2. La procédure de Benjamini & Hochberg est mise en œuvre en définissant $k = \max\{i : p_i \leq (i/m)q\}$, où i le test i , m est le nombre de tests effectués, q est le seuil de significativité. On rejette les hypothèses nulles H_{0i} , pour $j = 1 \dots k$. Cette procédure, initialement développée pour des statistiques de tests indépendantes, est également valable lorsque les statistiques de tests sont corrélées positivement, comme dans les comparaisons de traitement multiples pour une variable, comme c'est le cas ici. q doit alors être remplacé par $q \sum (1/i)$ dans la formule précédente. Voir également Bender & Lange (2001) pour une présentation non-technique des différentes procédures de correction.

Ensuite, en ventilant les résultats selon que les annonces sont publiées par des particuliers ou des agences, nous montrons que la hiérarchie des taux de succès est globalement la même selon l'origine des candidats³. Les niveaux des taux de succès sont nettement plus élevés lorsque les annonces émanent d'agences immobilières, mais le classement des candidats est similaire. Pour le

candidat au nom d'origine maghrébine relativement au candidat de référence, les écarts de taux de succès selon l'origine sont statistiquement différents de zéro (tableau 4-A). La conclusion est donc la même si l'on considère uniquement

3. Les résultats détaillés sont disponibles sur demande auprès des auteurs.

Tableau 4 – Tests de discrimination selon l'origine selon trois indicateurs de succès

A – PETIT / CHETTOUH

	Écart de taux de réponses non négatives (en points de pct)	Écart de taux de rendez-vous de principe (en points de pct)	Écart de nombre de contacts (%)
Ensemble des annonces	+3.8*** (<0.001) [<0.001]	+2.2*** (<0.001) [<0.001]	+0.06*** (<0.001) [<0.001]
Agences	+4.5*** (<0.001) [0.005]	+1.9* (0.071) [0.149]	+0.07*** (<0.001) [0.001]
Particuliers	+3.4*** (<0.001) [<0.001]	+2.4*** (<0.001) [<0.001]	+0.05*** (<0.001) [<0.001]

B – PETIT / SAMBOU

	Écart de taux de réponses non négatives (en points de pct)	Écart de taux de rendez-vous de principe (en points de pct)	Écart de nombre de contacts (%)
Ensemble des annonces	+4.2*** (< 0.001) [<0.001]	+2.3*** (< 0.001) [<0.001]	+0.06*** (< 0.001) [<0.001]
Agences	+3.7** (0.047) [0.116]	+0.771 (0.603) [0.698]	+0.05** (0.035) [0.081]
Particuliers	+4.6*** (< 0.001) [<0.001]	+3.2*** (< 0.001) [<0.001]	+0.07*** (< 0.001) [<0.001]

C – CHETTOUH / SAMBOU

	Écart de taux de réponses non négatives (en points de pct)	Écart de taux de rendez-vous de principe (en points de pct)	Écart de nombre de contacts (%)
Ensemble des annonces	+1.1 (0.145) [0.269]	+0.4 (0.480) [0.588]	+0.01 (0.240) [0.387]
Agences	+0.7 (0.692) [0.775]	-0.2 (0.846) [0.883]	+0.001 (0.982) [0.984]
Particuliers	+1.3** (0.020) [0.056]	0.7 (0.126) [0.230]	+0.02** (0.011) [0.037]

Note : *** significatif au seuil de 1 %, ** au seuil de 5 %, * au seuil de 10 % ; probabilité critique non ajustée entre parenthèses ; probabilité critique ajustée par la méthode de Benjamini & Hochberg (1995) entre crochets ; inférence statistique par bloc-*bootstrap* (10 000 réplications).
Source : *testing* DALTON – TEPP CNRS.

les agences ou les particuliers, à l'exception de l'écart du taux de rendez-vous de principe dans le cas des agences, significatif à 10 % sans la correction et qui n'est plus significatif avec la correction. Les agences ne paraissent donc guère avoir un rôle atténuateur des discriminations lorsque l'on considère les résultats de ce test.

Dans le cas du candidat français d'origine africaine, l'écart de taux de réponses non négatives n'est plus significatif avec la correction et l'écart de fréquence des rendez-vous de principe est significatif à 10 % (tableau 4-B). Sans correction, deux tests sur trois concluent donc que les agences immobilières discriminent les candidats d'origine africaine, comme le font les particuliers. Avec la correction, seul l'écart du nombre de contacts reste significativement différent de zéro, le rôle atténuateur des agences apparaît ici plus nettement.

Le troisième test compare les deux candidats dont le patronyme évoque une origine extra-européenne (tableau 4-C). On se restreint aux 2 776 annonces pour lesquelles Désiré Sambou a envoyé un courriel. Les résultats dépendent du statut de l'annonceur et des indicateurs considérés. Les particuliers sont les seuls à discriminer entre ces deux profils, en préférant le candidat d'origine maghrébine à celui d'origine africaine, si l'on observe les écarts dans les réponses non négatives ou le nombre de contacts, tandis que l'écart entre le nombre de rendez-vous de principe n'est pas significatif. On ne trouve en

revanche aucune différence de traitement entre ces deux candidats par les professionnels de l'immobilier (ils sont l'un et l'autre discriminés vis-à-vis du candidat dont le patronyme signale une origine française)⁴.

3.2. Pas de discrimination selon l'adresse et selon l'âge

Nous testons l'influence de l'âge du candidat en comparant les résultats obtenus par Sébastien Petit, qui précise dans sa prise de contact avoir 41 ans, et ceux de Kevin Durand qui indique avoir 22 ans. On se restreint alors aux 2 465 annonces pour lesquelles Kevin Durand a envoyé une réponse, dont 919 annonces publiées par une agence et 1 546 par un particulier. Le tableau 5 indique qu'il n'y a pas de différence significative entre les deux candidats d'origine française lorsque l'on retient comme indicateur le taux de réponses non négatives. C'est également le cas pour les deux autres indicateurs, sauf pour les « rendez-vous de principe » dans le cas des annonces publiées par les particuliers et sans correction. Le candidat âgé d'une quarantaine d'années dispose alors d'un avantage sur le candidat plus jeune, qui est pénalisé pour accéder à la visite d'un logement en location. Cependant, ce résultat n'est plus significatif dès lors que l'on corrige des comparaisons

4. Dans la suite de l'article, nous réalisons ces tests pour chaque aire urbaine et montrons que les discriminations selon l'origine ne sont significatives que dans un petit nombre d'aires urbaines.

Tableau 5 – Tests de discrimination selon l'adresse et l'âge

Candidat	Nombre d'envois	Nombre de réponses non négatives	Taux de succès (%)	Écart (en points de pct) et test
origine française, jeune (DURAND)	2 462	365	14.8 %	+0.89 (0.361) [0.495]
origine française habitant en HLM (ROUSSEAU)	2 465	366	14.8 %	-1.3 (0.184) [0.322]
origine maghrébine, jeune (MBAREK)	2 439	262	10.7 %	-0.6 (0.290) [0.418]
origine maghrébine, habitant en HLM (BENCHARGUI)	2 424	271	11.0 %	-0.9 (0.302) [0.418]

Note : *** significatif au seuil de 1 %, ** au seuil de 5 %, * au seuil de 10 % ; probabilité critique non ajustée entre parenthèses ; probabilité critique ajustée par la méthode de Benjamini & Hochberg (1995) entre crochets ; inférence statistique par bloc-*bootstrap* (10 000 répliquions).
Source : *testing* DALTON – TEPP CNRS.

multiples. Après correction, l'âge n'apparaît ainsi jamais comme un facteur discriminant pour les patronymes français, quel que soit l'indicateur utilisé.

Le test des effets de l'âge peut également être effectué pour les candidats d'origine maghrébine (cf. tableau 5). On se restreint alors aux 2 439 annonces pour lesquelles le candidat Mbarek a envoyé un mail. Ces tests montrent que l'âge n'a d'effet significatif pour les trois indicateurs que pour les annonces publiées par des particuliers. Le signe de l'effet est l'inverse du précédent : le candidat le plus âgé est désavantagé par rapport au plus jeune. Pour les réponses à ces annonces des particuliers, le sens de la discrimination en raison de l'âge n'est pas le même selon l'origine : avec un patronyme signalant une origine française, le candidat jeune est parfois pénalisé par rapport au plus âgé. Mais il est toujours avantagé si son nom signale une origine maghrébine.

Nous testons également l'effet du lieu de résidence, en signalant explicitement une adresse dans une cité ou dans un logement HLM au moment de la prise de contact avec l'offreur. On se restreint alors aux 2 462 annonces pour lesquelles le candidat Rousseau a envoyé une demande. Nous ne trouvons pas ce type d'effet sauf dans le cas des particuliers pour l'écart de taux de réponse et le nombre de contacts. L'effet est négatif, ce qui signifie que Frédéric Rousseau, le candidat dont le message de prise de contact indique qu'il réside dans un HLM ou dans une cité, est avantagé relativement au candidat de référence. Une interprétation est que le fait de déclarer quitter un HLM ou une cité pour accéder à un appartement du parc locatif privé dans un quartier neutre ou favorisé est un marqueur d'ascension sociale et de hausse du niveau de vie, ce qui envoie un signal financier positif pour les bailleurs sur la capacité du futur locataire à s'acquitter de ses loyers.

Un test analogue est décliné pour les candidats qui signalent une origine maghrébine par leur patronyme. On se restreint aux 2 464 annonces pour lesquelles le candidat Benchargui a envoyé une réponse. À nouveau, nous ne trouvons pas d'effet de l'adresse, sauf s'agissant des particuliers pour le nombre de contacts dans le cas non ajusté. L'effet est négatif ce qui signifie que Benchargui, le candidat qui indique résider dans une cité ou dans un HLM dans son message de prise de contact, est avantagé lui aussi relativement au candidat de référence. Toutefois, si

l'on corrige des comparaisons multiples, cette différence n'est plus significative.

3.3. Les discriminations selon l'origine sont réduites en présence d'un signal de stabilité financière

À l'issue de ces tests, seule l'origine semble le motif de discrimination confirmé de façon claire et robuste. Pour aller plus loin, il est utile d'identifier la nature des discriminations en œuvre. Nous mobilisons ici les deux principaux fondements des discriminations dans la littérature économique. D'une part, la discrimination peut résulter de préférences exogènes ou d'aversions individuelles pour telle ou telle caractéristique démographique des candidats, dans la logique de Becker (1957). D'autre part, il peut s'agir d'une discrimination statistique qui met en jeu les présupposés des offreurs sur les caractéristiques du groupe démographique du candidat en lien avec leur qualité de locataire, en particulier le risque de non-paiement des loyers, dans la lignée d'Arrow (1973). Pour faire la part de ces deux formes de discrimination, les chercheurs ajoutent un signal de « qualité » du candidat à la location, à l'instar de Massey & Lundy (2001), Ahmed *et al.* (2010), Bosch *et al.* (2010), Baldini & Federici (2011), Drydakis (2011) et Bunel *et al.* (2017). Dans ces travaux, ce type de signal réduit fortement l'ampleur de la discrimination sans la faire disparaître complètement, suggérant la coexistence d'une discrimination par l'information et d'une discrimination par les préférences.

Le signal que nous avons utilisé est la mention d'un statut de fonctionnaire dans le message envoyé par le candidat Mehdaoui. On se restreint aux 2 424 annonces pour lesquelles ce candidat a envoyé un courriel, publiées par des agences (818 annonces) ou des particuliers (1 606 annonces). Les résultats donnés dans le tableau 6 indiquent que ce signal de stabilité est bien perçu par les offreurs. Cela suggère qu'une partie des discriminations est liée à l'information. Pour déterminer de façon plus précise le poids de ce type de discrimination, il aurait fallu effectuer une comparaison avec une candidature française ayant le statut de fonctionnaire, ce qui n'a pas été prévu dans notre protocole de collecte de données, afin de limiter le nombre de profils de candidats. Cependant, nous avons effectué cette comparaison sur Paris dans un test préalable qui a fait l'objet d'une étude à part entière (Bunel *et al.*, 2017). Ce test indique que le signal de stabilité est nettement mieux perçu par les

Tableau 6 – Effet d'un signal de stabilité

CHETTOUH / MEHDAOUI (fonctionnaire)	Écart de taux de réponses non négatives (en points de pct)	Écart de taux de rendez-vous de principe (en points de pct)	Écart de nombre de contacts (%)
Ensemble des annonces	-1.2** (0.031) [0.068]	-2.1*** (0.001) [0.004]	-0.03*** (0.005) [0.012]
Agences	-3.3* (0.083) [0.186]	-3.7*** (0.009) [0.032]	-0.04* (0.093) [0.196]
Particuliers	-1.1 (0.118) [0.230]	-1.3** (0.022) [0.063]	-0.03*** (0.006) [0.017]

Note : *** significatif au seuil de 1 %, ** au seuil de 5 %, * au seuil de 10 % ; probabilité critique non ajustée entre parenthèses ; probabilité critique ajustée par la méthode de Benjamini & Hochberg (1995) entre crochets ; inférence statistique par bloc-*bootstrap* (10 000 réplifications).
Source : *testing* DALTON – TEPP CNRS.

offreurs lorsqu'il émane d'un candidat signalant par son patronyme une origine française.

4. Une confirmation économétrique

Dans cette section, nous réalisons une analyse visant à estimer, pour chaque courriel envoyé, la probabilité de recevoir une réponse non-négative ou un rendez-vous de principe. L'échantillon se compose de l'ensemble des 25 040 courriels envoyés. Les variables expliquées sont binaires : 1) recevoir une réponse non-négative et 2) recevoir une proposition de rendez-vous de principe. Les variables explicatives sont les suivantes :

- le type loueur : variable binaire prenant deux modalités : « agence » (référence) ou « particulier » ;
- la source de l'annonce : variable binaire prenant deux modalités : « Le Bon Coin » ou « autre site » (référence) ;
- le loyer exprimé en logarithme ;
- une variable binaire indiquant si le logement est meublé ou non (référence) ;
- l'origine du nom : une variable binaire indiquant si le patronyme évoque une origine étrangère (référence) ou française ;
- l'âge du candidat : une variable binaire indiquant si le candidat est âgé de 22 ans (référence) ou de 41 ans ;
- le statut de fonctionnaire : une variable binaire indiquant si le candidat est fonctionnaire ou non (référence) ;

- la localisation : une variable binaire indiquant si le candidat indique habiter dans une cité ou un HLM ou non (référence) ;

- le rang de l'envoi : une variable qualitative à cinq modalités, indiquant le rang de l'envoi, la référence étant l'ordre 1.

Pour chaque variable expliquée, nous avons estimé un modèle probit incluant, outre les variables explicatives précédentes, des effets fixes par aire urbaine et des effets fixes pour le mois de candidature. Les premiers captent l'hétérogénéité inobservable des aires urbaines, invariante dans le temps. Les seconds captent les effets de conjoncture s'appliquant à toutes les annonces. En outre, dans la mesure où notre unité d'observation est le courriel envoyé, les variables relatives au logement (type du loueur, source de l'annonce, loyer, meublé) sont répliquées pour chacun des cinq courriels envoyés à la même annonce. Puisque les annonces sont réparties dans 50 aires urbaines, cette structure particulière de l'échantillon peut générer des corrélations intra-aire urbaine et intra-annonces. Nous menons par conséquent une inférence statistique robuste par groupe (*cluster*) avec une double clusterisation des écarts-types, par aire urbaine et par annonce. Les résultats d'estimation sont reportés dans le tableau 7.

Pour chaque variable expliquée, la première colonne donne les résultats de l'estimation lorsque les variables caractérisant les logements sont incluses, la seconde colonne donne les résultats lorsque les variables caractérisant les candidats sont introduites et la 3^e colonne inclut des interactions entre la variable origine et les

variables binaires, type de loueur et source de l'annonce.

Les résultats indiquent que certaines caractéristiques des annonces ont un effet significatif sur les chances des cinq individus fictifs de recevoir une réponse non négative ou un rendez-vous de principe de l'offreur, toutes choses égales par ailleurs. Ainsi, quel que soit le profil, utiliser le site internet Le Bon Coin réduit fortement les chances de succès, relativement aux autres sites, sans doute parce que la concurrence entre les demandeurs est forte sur ce site populaire. Il apparaît également que les candidats à la

location reçoivent moins souvent une réponse non négative de la part des annonceurs particuliers. En revanche, il n'y a pas de différence entre particuliers et annonceurs professionnels pour un rendez-vous de principe. Les autres caractéristiques du logement (montant du loyer, logement meublé ou non) n'ont pas d'effet significatif sur la probabilité de recevoir une réponse non négative ou un rendez-vous de principe.

Concernant les caractéristiques des candidats à la location, la discrimination selon l'origine supposée à partir du patronyme apparaît clairement, avec un effet positif de la variable origine pour le taux

Tableau 7 – Taux de réponses non négatives, modèle probit

	Réponse non négative			Rendez-vous de principe		
	(1)	(2)	(3)	(4)	(5)	(6)
Particulier	-0.143*** (0.055)	-0.150*** (0.055)	-0.284*** (0.060)	0.024 (0.062)	0.021 (0.062)	-0.122* (0.062)
Le Bon Coin	-1.324*** (0.068)	-1.330*** (0.069)	-1.269*** (0.074)	-1.125*** (0.086)	-1.126*** (0.087)	-1.056*** (0.089)
Log(Loyer)	0.094 (0.123)	0.100 (0.124)	0.100 (0.125)	0.117 (0.140)	0.121 (0.141)	0.119 (0.141)
Meublé	0.047 (0.040)	0.048 (0.040)	0.049 (0.040)	0.055 (0.043)	0.054 (0.042)	0.054 (0.043)
Jeu B	0.149*** (0.045)	0.158*** (0.046)	0.157*** (0.046)	0.086 (0.058)	0.093 (0.058)	0.091 (0.059)
Origine française		0.261*** (0.021)	0.188*** (0.035)		0.186*** (0.027)	0.103*** (0.034)
41 ans		-0.056 (0.051)	-0.056 (0.051)		-0.046 (0.053)	-0.046 (0.053)
Fonctionnaire		0.064* (0.034)	0.069** (0.034)		0.114*** (0.036)	0.121*** (0.037)
Cité ou HLM		0.114* (0.062)	0.113* (0.063)		0.049 (0.058)	0.049 (0.059)
Rang d'envoi 2		-0.062* (0.033)	-0.060* (0.033)		-0.057* (0.033)	-0.056* (0.033)
Rang d'envoi 3		-0.073*** (0.027)	-0.070** (0.028)		-0.096*** (0.040)	-0.090** (0.040)
Rang d'envoi 4		-0.104*** (0.027)	-0.099*** (0.028)		-0.128*** (0.042)	-0.123*** (0.041)
Rang d'envoi 5		-0.076*** (0.029)	-0.075*** (0.029)		-0.130*** (0.032)	-0.126*** (0.032)
Origine française*Particulier			0.304*** (0.052)			0.328*** (0.058)
Origine française*Le Bon Coin			-0.146*** (0.053)			-0.164*** (0.066)
Nombre d'observations	24 885	24 885	24 885	24 885	24 885	24 885
Log-vraisemblance	-6698	-6631	-6618	-4509	-4483	-4473

Note : *** significatif au seuil de 1 %, ** au seuil de 5 %, * au seuil de 10 %. Les écarts-types figurent entre parenthèses.
Source : *testing* DALTON – TEPP CNRS.

de réponses non négatives et pour l'obtention d'un rendez-vous de principe (colonnes 3 et 6). Les discriminations par les particuliers sont plus fréquentes comme le montre le signe positif de l'interaction entre la modalité « origine française » et la modalité « particulier ». Le statut de fonctionnaire a également un effet « facilitant » dans les deux cas alors que l'âge ne semble pas avoir d'impact. Le fait d'habiter en HLM augmente les chances de recevoir une réponse non négative. Enfin, le rang d'envoi de la candidature n'est pas neutre : par rapport à une candidature reçue en premier, les candidatures reçues en deuxième, troisième, quatrième et cinquième position ont moins de chances d'aboutir positivement. Les résultats économétriques confirment donc globalement les résultats obtenus par les tests bilatéraux.

Afin de faire la synthèse de ces différents résultats, nous avons calculé la probabilité de recevoir une réponse non-négative (resp. rendez-vous de principe) pour un homme de 41 ans non fonctionnaire habitant à Paris en zone neutre au mois de juillet, pour une annonce émanant du site Le Bon Coin avec un loyer médian, le rang d'envoi étant fixé à 1. Nous confirmons le résultat selon lequel la discrimination est moins importante pour des annonces émanant des agences, sans pour autant être complètement neutralisée.

4.1. Prise en compte des non réponses au travers d'un deuxième type de test de discrimination

Nos résultats sont établis sur la base d'envoi de courriels en réaction à des annonces immobilières pour lesquels nous avons obtenu une proportion importante de non réponses. Dans le tableau 2 (*supra*), nous avons constaté que près de 80 % de nos envois de courriels n'ont obtenu aucune réponse. Cette proportion élevée nous amène à considérer un deuxième type d'indice de discrimination. Dans un test de correspondance, la discrimination est mesurée par la différence des taux de succès entre le candidat de référence et le candidat potentiellement discriminé. Deux modes de calcul du taux de succès (*call back rate*) sont envisageables : le taux calculé sur l'ensemble des envois et celui calculé sur l'ensemble des offres ayant fait l'objet d'au moins une réponse. Les deux calculs diffèrent uniquement par leur dénominateur. Ils sont évoqués l'un et l'autre dans le survol de littérature sur les discriminations sur le marché du travail réalisé par Neumark (2018). Le calcul sur le nombre des envois est le plus usuel dans la littérature internationale (voir par exemple Bertrand & Mullainathan,

2004, tableau 1, p. 997). On se place du point de vue du demandeur qui est intéressé par le volume de démarches nécessaires pour obtenir un rendez-vous pour une location d'appartement. Le deuxième ratio ne prend pas en compte les offres pour lesquelles aucune réponse n'a été donnée. Cette approche est utilisée par Riach & Rich (2002) et elle est préconisée par le Bureau International du Travail (Bovenkerk, 1992). La mesure de la discrimination, donnée par la différence des taux de succès, est alors appelée « taux net de discrimination ».

La distinction entre les deux modes de calcul est pertinente dans le cadre de ce *testing* parce que les taux de réponse sont faibles. Le traitement réservé aux offres pour lesquelles aucune réponse n'a été donnée est donc un point important. Si les deux calculs sont équivalents en l'absence de non réponse, ils diffèrent sensiblement lorsque les non réponses sont nombreuses. Comme les taux de succès sont plus élevés avec le calcul préconisé par le BIT, il va être statistiquement plus facile de détecter une discrimination. Le calcul du taux de réponse sur l'ensemble des envois est quant à lui plus conventionnel. Dans son survol, Neumark (2018) indique que la pratique standard est d'estimer des effets marginaux issus de modèles de probabilité linéaire ou probit sur la base de l'ensemble des observations, ce qui revient à privilégier la première approche.

La différence entre les deux approches réside dans l'interprétation qui est retenue pour les offres sans aucune réponse. Dans le calcul des indices de discriminations sur l'ensemble des envois, les non réponses sont assimilées à des réponses négatives. Dans celui effectué sur l'ensemble des réponses exprimées, les non réponses sont assimilées à des non envois. Notre point de vue est que les deux hypothèses sont extrêmes et c'est pourquoi nous retenons les deux modes de calcul. Il est vraisemblable qu'une partie des demandes par courrier électronique n'a pas été reçue par le destinataire, mais une partie seulement. Comme il est impossible de déterminer quelle proportion précise, nous encadrons la mesure des discriminations en combinant les deux mesures.

Cette nouvelle approche ne modifie pas nos résultats qui demeurent qualitativement inchangés⁵. Nous constatons un écart significatif dans les chances de succès des démarches effectuées par

5. Les tableaux détaillés de résultats sont disponibles auprès des auteurs.

le candidat d'origine française, d'une part, et par les candidats français d'origine maghrébine (Chettouh) ou africaine (Sambou). Cet écart significatif est confirmé pour nos trois indicateurs (taux de réponses non négatives, taux de rendez-vous de principe, nombre de contacts), que les annonces émanent de particuliers ou d'agences. De même, nous ne sommes pas en mesure de mettre en évidence des discriminations robustes liées à l'âge du demandeur ou à sa localisation dans un quartier défavorisé.

Nos résultats apparaissent globalement robustes à une grande variété d'indicateurs et de méthodes de calculs des indices de discrimination. Pour autant, plusieurs limites de ces tests doivent être soulignées. Tout d'abord, dans tous les calculs de nos différents indices de discrimination, on ne considère pas la non réponse comme un comportement de discrimination spécifique. On fait l'hypothèse implicite que le sous-échantillon des répondants est représentatif de l'ensemble des offreurs de logement, ce qui peut être discuté. En outre, en dehors des messages que nous avons envoyés, nous ne sommes pas en mesure d'observer les autres réponses aux annonces immobilières. Nous ne connaissons ni la nature ni le volume de ces autres candidatures que nous supposons uniformément réparties sur l'ensemble des annonces. Cependant, en matière de discriminations dans l'accès au logement, le nombre

et la qualité des candidats exercent une influence déterminante sur les réponses des offreurs. De surcroît, les annonceurs peuvent privilégier d'autres canaux de recherche au-delà de la réponse aux offres par voie de courrier électronique, ce qui peut limiter notre capacité à détecter les discriminations effectives sur l'ensemble du marché. Certes, ces limites existent pour tous les tests de discrimination, mais elles s'appliquent pleinement ici et méritent d'être rappelées.

4.2. Des différences locales assez marquées selon les aires urbaines

Pour étudier les différences locales dans l'exposition aux discriminations, nous nous concentrons sur le critère de l'origine, qui est le seul à produire des résultats significatifs, et nous considérons l'écart relatif de taux de réponse non négatives entre Sébastien Petit, notre candidat de référence, et Mohamed Chettouh. Cet écart est calculé de deux façons, sur l'ensemble des annonces d'une part, et uniquement sur les annonces pour lesquelles nous avons obtenu au moins une réponse d'autre part. Le résultat figure dans le tableau 8. On constate que les deux indicateurs sont très dispersés selon les aires urbaines, ce qui suggère une forte composante locale dans les déterminants des discriminations.

Tableau 8 – Proportion de fois où PETIT a reçu une réponse non négative et pas CHETTOUH

Aire urbaine	Différence de taux de réponses non négatives (en points de pct) calculée sur le nombre total d'annonces	Rang (du moins au plus discriminant)	Différence de taux de réponses non négatives (en points de pct) calculée sur les annonces avec au moins une réponse	Rang (du moins au plus discriminant)
Amiens	11.8**	45	30.8***	39
Angers	1.0	16	5.9	21
Angoulême	1.0	15	7.7	17
Annecy	2.9	24	27.3	33
Annemasse	2.9	23	14.3	23.5
Avignon	12.0*	47	31.6***	50
Bayonne	2.9	24	12.5	14
Besançon	2.0	19	13.3	16
Béthune	11.7**	44	33.3***	40
Bordeaux	0.0	7	0.0	18
Brest	3.0	26	15.8	31.5
Caen	6.9	41	20.0*	41
Chambéry	3.0	26	18.8	22
Clermont-Ferrand	6.0	39	25.0*	44
Dijon	2.0	20	13.3	15
Douai - Lens	3.0	29	14.3	46
Dunkerque	1.0	13	4.8	5

→

Tableau 8 (suite)

Aire urbaine	Différence de taux de réponses non négatives (en points de pct) calculée sur le nombre total d'annonces	Rang (du moins au plus discriminant)	Différence de taux de réponses non négatives (en points de pct) calculée sur les annonces avec au moins une réponse	Rang (du moins au plus discriminant)
Grenoble	1.9	17	7.7	11
La Rochelle	1.0	11	5.9	5
Le Havre	11.0*	43	27.5***	37
Le Mans	3.0	28	15.8	31.5
Lille	2.0	18	11.8	35
Limoges	12.0*	48	29.3***	43
Lorient	0.0	7	0.0	5
Lyon	3.9	33	19.1	23.5
Marseille - Aix-en-Provence	5.8	37	27.3*	47
Metz	4.9	36	13.9	36
Montpellier	0.9	10	7.1	9
Mulhouse	4.0	34	17.4	20
Nancy	11.9**	46	30.8***	34
Nantes	1.0	13	5.0	12
Nice	-4.8	1	-17.2	2
Nîmes	0.0	7	6.3	8
Orléans	7.8	42	24.2**	38
Paris	-2.9	2	-17.7	1
Pau	0.0	7	0.0	13
Perpignan	14.8***	50	34.9***	45
Poitiers	-2.0	3	-13.3	3
Reims	0.0	7	0.0	26
Rennes	4.0	35	28.6	42
Rouen	-1.1	4	-5.9	7
Saint-Étienne	2.0	21	10.0	27
Saint-Nazaire	3.1	31	17.7	28.5
Strasbourg	1.0	12	4.8	30
Toulon	6.4	40	26.1*	48
Toulouse	3.9	32	15.4	25
Tours	5.9	38	17.1	19
Troyes	3.0	30	21.4	10
Valence	2.0	22	11.8	28.5
Valenciennes	12.0**	48	30.8***	49

Note : *** significatif au seuil de 1 %, ** au seuil de 5 %, * au seuil de 10 %.
Source : *testing* DALTON – TEPP CNRS.

Parmi les 50 aires urbaines, 8 affichent des écarts de taux de succès statistiquement significatifs au niveau de l'aire urbaine avec le premier indicateur et 11 avec le deuxième. Les discriminations paraissent très concentrées dans un petit nombre d'aires urbaines. Avignon, Valenciennes et Perpignan apparaissent comme les villes où les discriminations dans l'accès au logement sont les plus fortes. Nancy, Limoges, Le Havre, Amiens et Béthune sont également des villes caractérisées par des discriminations significatives pour les deux indicateurs.

Il apparaît qu'aucune de ces aires urbaines ne correspond à une grande capitale régionale. En outre, ces aires urbaines les plus discriminantes ne sont pas les plus grandes. La plus grande de ces aires urbaines est celle d'Avignon, qui figure à la 16^e place du classement des aires urbaines par leur taille. Les aires les plus discriminantes ne sont pas non plus les plus petites. Aucune des 15 plus petites aires urbaines, au sein des 50 plus grandes, ne figure dans la liste des aires les plus discriminantes. La plus petite est Limoges. Les dix aires urbaines les plus

discriminantes sont des préfectures ou des sous-préfectures. Leur taille moyenne est proche de la médiane. Elle est de 365 000 habitants pour l'aire urbaine, de 266 000 habitants pour le pôle urbain et de près de 100 000 habitants pour la couronne urbaine.

* *
*

Dans cette étude, notre objectif était de mesurer les discriminations dans l'accès au logement du parc privé à l'aide d'un *testing* sur les grandes aires urbaines de la France métropolitaine, et en couvrant plusieurs motifs de discriminations : l'âge (discrimination à l'encontre des jeunes), l'origine (en distinguant Maghreb et Afrique subsaharienne), le lieu de résidence au moment de la réponse à une annonce (habiter un quartier inscrit dans la géographie prioritaire de la politique de la ville). Les tests ont été réalisés entre juin et décembre 2016 sur les 50 plus grandes aires urbaines de France métropolitaine. Nous avons répondu à 5 008 annonces au nom de cinq candidats à la location, soit 25 040 réponses à des annonces immobilières. Nous vérifions si les discriminations dans l'accès au logement dépendent des caractéristiques du candidat (en ajoutant à son profil un signal de stabilité financière avec un statut de fonctionnaire), de celles de l'annonceur (un particulier ou une agence), et de celles du contexte local. Nous tenons compte de plusieurs indicateurs complémentaires que nous calculons selon que l'annonce est publiée par un particulier ou un professionnel de l'immobilier, afin de prendre la mesure du rôle éventuellement atténuateur des agences immobilières dans les discriminations au logement.

La première exploitation de cette base de données conduit à plusieurs conclusions intéressantes. Nous ne mettons pas en évidence de discriminations en raison de l'âge du candidat. Nous trouvons un effet positif du fait de déclarer résider dans un HLM ou une cité lors de la recherche d'un appartement dans le parc locatif privé, qui signale une augmentation de niveau de vie et une capacité du futur locataire à s'acquitter de ses loyers. Nous trouvons en outre d'importantes discriminations selon l'origine qui pénalisent les candidats dont le nom ou prénom signale une origine maghrébine ou une origine africaine. Relativement au candidat de référence d'origine présumée française, Sébastien Petit, le candidat d'origine maghrébine Mohamed

Chettouh a 26.7 % de chances en moins de voir ses démarches d'accès au logement aboutir. Ces discriminations sont très marquées pour les offres qui émanent de particuliers et elles le sont aussi pour les offres diffusées par les agences immobilières. Elles ne sont que faiblement atténuées lorsque le candidat à la location ajoute un signal de qualité en précisant qu'il est fonctionnaire.

Nous trouvons que ces discriminations sont très différentes selon les territoires. Elles sont patentes dans un petit nombre d'aires urbaines dont nous établissons la liste. Perpignan, Limoges, Avignon et Nancy sont en tête des classements établis à partir d'indicateurs différents. Les dix aires urbaines où l'intensité des discriminations est la plus forte ne sont ni les plus grandes, ni les plus petites. Aucune n'est une capitale régionale. Toutes sont des préfectures ou des sous-préfectures. Leur taille est proche de la médiane des grandes aires urbaines et elles sont dispersées dans l'espace métropolitain, au centre (Limoges), au nord (Amiens, Béthune, Le Havre, Valenciennes,) au sud (Avignon, Perpignan) et à l'est (Nancy).

Soulignons qu'une limite de cette étude est de ne tester que la première marche de l'accès au logement : la prise de rendez-vous avec un loueur. En outre, nous nous sommes focalisés sur le logement locatif dans le parc privé en nous concentrant sur des biens immobiliers intermédiaires, les F2, qui sont les biens les plus demandés et les plus offerts. Il serait intéressant de prolonger l'étude en examinant d'autres types de biens. Néanmoins, malgré ces limites, nos résultats nous paraissent suffisamment robustes pour appeler des réactions de politique publique. Alors que la mixité sociale des quartiers est un objectif affiché de l'action publique en France et que les discriminations selon l'origine sont rigoureusement prohibées, le fait de constater de fortes discriminations dans l'accès au logement selon ce critère recouvre des enjeux de politique publique importants. Les résultats de cette étude invitent à une réflexion approfondie sur la régulation du marché immobilier et sur les différents outils qui peuvent être déployés pour lutter contre les discriminations dans l'accès au logement. Ces outils de politique publique vont du rappel de la règle de droit à des mesures concrètes visant à rendre le droit plus effectif. Parmi ces mesures, des actions de lutte contre la précarité sociale, qu'il s'agisse de l'accès au logement social ou de mécanismes de garanties financières permettant de sécuriser les loueurs du parc privé, ont sans doute une place importante à occuper. □

BIBLIOGRAPHIE

- Acolin, A., Bostic, R. & Painter, G. (2016).** A field study of rental market discrimination across origins in France. *Journal of Urban Economics*, 95, 49–63. <https://doi.org/10.1016/j.jue.2016.07.003>
- Adida, C. L., Laitin, D. D. & Valfort, M.-A. (2010).** Identifying barriers to Muslim integration in France. *PNAS, Proceedings of the National Academy of Sciences of the United States of America*, 107(52), 384–390. <https://doi.org/10.1073/pnas.1015550107>
- Ahmed, A. M., Andersson, L. & Hammarstedt, M. (2010).** Can Discrimination in the Housing Market be Reduced by Increasing the Information about the Applicants? *Land Economics*, 86(1), 79–90. <https://www.jstor.org/stable/27759719>
- Ahmed, A. M. & Hammarstedt, M. (2008).** Discrimination in the rental housing market: A field experiment on the Internet. *Journal of Urban Economics*, 64(2), 362–72. <https://doi.org/10.1016/j.jue.2008.02.004>
- Arrow, K. (1973).** The Theory of Discrimination. In: Ashenfelter, O. A. & Rees, A. (Eds.), *Discrimination in Labor Markets*. Princeton: Princeton University Press.
- Baldini, M. & Federici, M. (2011).** Ethnic discrimination in the Italian rental housing market. *Journal of Housing Economics*, 20(1), 1–14. <https://doi.org/10.1016/j.jhe.2011.02.003>
- Becker, G. S. (1957).** *The Economics of Discrimination*. Chicago: The University of Chicago Press.
- Bender, L. & Lange, S. (2001).** Adjusting for multiple testing – when and how? *Journal of Clinical Epidemiology*, 54, 343–349. [https://doi.org/10.1016/s0895-4356\(00\)00314-0](https://doi.org/10.1016/s0895-4356(00)00314-0)
- Bengtsson, R., Iverman, E. & Hinnerich, B. T. (2012).** Gender and ethnic discrimination in the rental housing market. *Applied Economics Letters*, 19(1), 1–5. <https://doi.org/10.1080/13504851.2011.564125>
- Benjamini, Y. & Hochberg, Y. (1995).** Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society B*, 57, 289–300. <https://www.jstor.org/stable/2346101>
- Benjamini, Y. & Yekutieli, D. (2001).** The Control of the False Discovery Rate in Multiple Testing Under Dependency. *The Annals of Statistics*, 29(4), 1165–1188. <https://www.jstor.org/stable/2674075>
- Berson, C. (2011).** Testing sur les jeunes français issus de l’immigration en fonction du degré de concurrence. *Rapport final d’évaluation*, AP2 091. http://www.experimentation.jeunes.gouv.fr/IMG/pdf/Rapport_Final_EVA_AP2_091-Testing.pdf
- Bertrand, M. & Mullainathan, S. (2004).** Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94(4), 991–1013. <https://doi.org/10.1257/0002828042002561>
- Bonnet, F., Lalé, E., Safi, M. & Wasmer, E. (2015).** Better residential than ethnic discrimination! Reconciling audit and interview findings in the Parisian housing market. *Urban Studies*, 53(13), 2815–2833. <https://doi.org/10.1177%2F0042098015596107>
- Bosch, M., Carnero, A. & Farre, L. (2010).** Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional Science and Urban Economics*, 40(1), 11–19. <https://doi.org/10.1016/j.regsciurbeco.2009.11.001>
- Bovenkerk, F. (1992).** *Testing discrimination in natural experiments: A Manual for International Comparative Research on Discrimination on the Grounds of Race and Ethnic Origin*. Geneva: International Labour Office.
- Bunel, M., L’Horty, Y. & Petit, P. (2016).** Discrimination based on place of residence and access to employment. *Urban Studies*, 53(2), 267–286. <https://doi.org/10.1177%2F0042098014563470>
- Bunel, M., L’Horty, Y. & Petit, P. (2017).** Les discriminations dans l’accès au logement à Paris : une expérience contrôlée. *Rapport de recherche de TEPP N° 17-01*. <https://halshs.archives-ouvertes.fr/halshs-01521995>
- Carlsson, M. & Eriksson, S. (2014).** Discrimination in the rental market for apartments. *Journal of Housing Economics*, 23, 41–54. <https://EconPapers.repec.org/RePEc:eee:jhouse:v:23:y:2014:i:c:p:41-54>

- Carpusor, A. & Loges, W. (2006).** Rental discrimination and ethnicity in names. *Journal of Applied Social Psychology*, 36(4), 934–952. <https://doi.org/10.1111/j.0021-9029.2006.00050.x>
- Choi, S. J., Ondrich, J. & Yinger, J. (2005).** Do rental agents discriminate against minority customers? Evidence from the 2000 Housing Discrimination Study. *Journal of Housing Economics*, 14, 1–26. <https://doi.org/10.1016/j.jhe.2005.02.001>
- Défenseur des Droits (2017).** Les discriminations dans l'accès au logement. *Enquête sur l'accès aux droits* N° 5. https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=16974
- Drydakis, N. (2011).** Ethnic discrimination in the Greek housing market. *Journal of Population Economics*, 24(4), 1235–1255. <https://EconPapers.repec.org/RePEc:spr:jopoe:v:24:y:2011:i:4:p:1235-1255>
- Duguet, E. & Petit, P. (2005).** Hiring discrimination in the French financial sector: an econometric analysis on field experiment data. *Annals of Economics and Statistics*, 78, 79–102. <https://EconPapers.repec.org/RePEc:adr:anecst:y:2005:i:78:p:79-102>
- Duguet, E., Leandri, N., L'Horty, Y. & Petit, P. (2010).** Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled Experiment in the Paris area. *Annals of Economics and Statistics*, 99-100, 187–215. <https://EconPapers.repec.org/RePEc:adr:anecst:y:2010:i:99-100:p:187-215>
- Flage, A. (2018).** Ethnic and Gender Discrimination in the Rental Housing Market: Evidence from a Meta-Analysis of Correspondence Tests, 2006-2017. *Journal of Housing Economics*, 41, 251–273. <https://doi.org/10.1016/j.jhe.2018.07.003>
- Gobillon, L., Selod, H. & Zenou, Y. (2007).** The Mechanisms of Spatial Mismatch. *Urban Studies*, 44(12), 2401–2427. <https://halshs.archives-ouvertes.fr/halshs-00826212>
- Hanson, A. & Hawley, Z. (2011).** Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics*, 70(2-3), 99–114. <https://EconPapers.repec.org/RePEc:eee:juecon:v:70:y:2011:i:2-3:p:99-114>
- Hanson, A. & Santas, M. (2014).** Field experiment tests for discrimination against hispanics in the U.S. rental housing market. *Southern Economic Journal*, 81(1), 135–167. <https://doi.org/10.4284/0038-4038-2012.231>
- Heylen, K. & Van den Broeck, K. (2015).** Differential Treatment of Rental Home Seekers According to their Sociodemographic and Economic Status by Real Estate Agencies in Belgium. *European Journal of Homelessness*, 9(2), 39–62. https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1898823&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1
- L'Horty, Y., Duguet, E., du Parquet, L., Petit, P. & Sari, F. (2011).** Les effets du lieu de résidence sur l'accès à l'emploi : une expérience contrôlée sur des jeunes qualifiés. *Économie et Statistique*, 447, 71–95. <https://www.insee.fr/fr/statistiques/1377348?sommaire=1377352>
- Massey, D. S. & Lundy, G. (2001).** Use of Black English and racial discrimination in urban housing markets: new methods and findings. *Urban Affairs Review*, 36(4), 452–469. <https://doi.org/10.1177%2F10780870122184957>
- Neumark, D. (2018).** Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, 56(3), 799–866. <https://doi.org/10.1257/jel.20161309>
- Oh, S. J. & Yinger, J. (2015).** What have we learned from paired testing in housing markets? In: Yinger, J. (Ed.), *Why Housing Agents Discriminate: The Measurement, Causes, and Consequences of Housing Discrimination*. E-book.
- Ondrich, J., Ross, S. L. & Yinger, J. (2000).** How common is housing discrimination? Improving on traditional measures. *Journal of Urban Economics*, 47(3), 470–500. <https://doi.org/10.1006/juec.1999.2148>
- Ondrich, J., Ross, S. L. & Yinger, J. (2003).** Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses from Black Customers? *Review of Economics and Statistics*, 85(4), 854–73. <https://www.jstor.org/stable/3211810>
- Page, M. (1995).** Racial and Ethnic Discrimination in Urban Housing Markets: Evidence from a Recent Audit Study. *Journal of Urban Economics*, 38(2), 183–206. <https://doi.org/10.1006/juec.1995.1028>

- Petit, P. (2007).** The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector. *Labour Economics*, 14(3), 371–392. <https://doi.org/10.1016/j.labeco.2006.01.006>
- Petit, P., Duguet, E., L'Horty, Y., du Parquet, L. & Sari, F. (2014).** Discriminations à l'embauche des jeunes franciliens et effets croisés du sexe et de l'origine : les résultats d'un testing. *Économie et Statistique*, 464-465-466, 141–153. <https://www.insee.fr/fr/statistiques/1378021?sommaire=1378033>
- Pierné, G. (2013).** Hiring discrimination based on national origin and religious closeness: results from a field experiment in the Paris area. *IZA Journal of Labor Economics*, 2(4). <https://doi.org/10.1186/2193-8997-2-4>
- Riach, P. A., & Rich. J. (2002).** Field Experiments of Discrimination in the Market Place. *Economic Journal*, 112 (483), 480–518. <https://doi.org/10.1111/1468-0297.00080>
- Roychoudhury, C. & Goodman, A. C. (1996).** Evidence of racial discrimination in different dimension of owner-occupied housing search. *Real estate Economics*, 24(2), 161–178. <https://doi.org/10.1111/1540-6229.00685>
- Yinger, J. (1986).** Measuring racial discrimination with fair housing audits: caught in the act. *American Economic Review*, 76(5), 881–893. <https://www.jstor.org/stable/1816458>
- Zhao, B. (2005).** Does the number of houses a broker shows depend on a homeseeker's race? *Journal of Urban Economics*, 57(1), 128–47. <https://doi.org/10.1016/j.jue.2004.10.001>
- Zhao, B., Ondric, J. & Yinger, J. (2006).** Why do real estate brokers continue to discriminate? Evidence from the 2000 Housing Discrimination Study. *Journal of Urban Economics*, 59(3), 394–419. <https://doi.org/10.1016/j.jue.2005.12.001>
-

Les liens entre taux d'épargne, revenu et incertitude : une analyse à partir de l'enquête *Budget de famille 2011*

The Links between Saving Rates, Income and Uncertainty: An Analysis based on the 2011 Household Budget Survey

Céline Antonin*

Résumé – L'article étudie, à partir des données de l'enquête *Budget de famille* 2010-2011 de l'Insee, les liens entre le taux d'épargne des ménages français et leur revenu. On montre que le taux d'épargne des ménages les plus riches croît avec le revenu permanent : ces derniers épargnent davantage sur le cycle de vie. En outre, l'analyse empirique permet de mettre en évidence et de quantifier un motif de précaution lié au risque de chômage : le motif de précaution entraîne un surplus de flux d'épargne d'environ 6.4 % pour les ménages actifs. Par ailleurs, la part du patrimoine de précaution liée à l'incertitude sur le revenu futur est comparable, autour de 6.3 % de la richesse globale. Enfin, l'importance du motif de précaution dépend du niveau de revenu selon une courbe en U inversé : il est près de deux fois plus fort pour les ménages des troisième et quatrième quintiles que pour les quintiles extrêmes de revenu.

Abstract – Using data from the 2010-2011 Insee Household Budget Survey (enquête Budget de famille), the article examines the links between the saving rate of French households and their income. It shows that the saving rate of the richest households increases with permanent income: they save more over their lifecycle. In addition, the empirical analysis makes it possible to identify and quantify a precautionary motive linked to the risk of unemployment: the precautionary motive leads to a surplus of savings flows of approximately 6.4% for working households. Furthermore, the proportion of precautionary wealth linked to uncertainty over future income is comparable, around 6.3% of overall wealth. Finally, the significance of the precautionary motive depends on the level of income, following an inverted U-shaped curve: it is almost twice as strong for households in the third and fourth quintiles than for the extreme income quintiles.

Codes JEL / JEL Classification : C81, D12, D14, D31

Mots-clés : taux d'épargne, consommation, revenu permanent, épargne de précaution, chômage

Keywords: saving rate, consumption, permanent income, precautionary savings, unemployment

* OFCE (celine.antonin@sciencespo.fr)

L'auteure tient à remercier Jérôme Bourdieu et Muriel Roger pour leurs conseils et leurs relectures attentives, ainsi que Philippe Aghion, Didier Blanchet, Anne Lavigne et Xavier Ragot pour leurs commentaires et discussions. Elle remercie également deux rapporteurs anonymes pour leurs remarques et suggestions.

Reçu le 14 mai 2018, accepté après révisions le 21 mars 2019.

Citation: Antonin, C. (2019), The Links between Saving Rates, Income and Uncertainty: An Analysis based on the 2011 Household Budget Survey. *Economie et Statistique / Economics and Statistics*, 513, 47–68. <https://doi.org/10.24187/ecostat.2019.513.2000>

En allouant une partie de leur revenu disponible à l'épargne et l'autre à la consommation, les ménages effectuent un arbitrage ayant des implications directes au niveau macroéconomique. Si la propension à épargner est plus forte chez les ménages riches, une politique de relance par la consommation sera plus efficace en ciblant les bas revenus. D'autres questions comme celle des dispositifs fiscaux visant à favoriser l'épargne (assurance-vie, livret A) ou celle de l'assiette pertinente en matière de fiscalité (travail vs consommation, revenu vs patrimoine) dépendent de l'arbitrage consommation/épargne. Par ailleurs, la mesure de l'épargne de précaution est cruciale, notamment pour comprendre les implications de la hausse du chômage lors d'un choc comme la crise de 2008. Si la hausse du chômage touche indifféremment tous les ménages, et si les ménages riches ont un motif de précaution plus fort que les autres, alors la récession sera plus violente. La chute de la demande agrégée accroît le taux de chômage, ce qui augmente l'épargne de précaution, enclenchant un cercle vicieux.

Historiquement, les modèles de cycle de vie et de revenu permanent, dus à Modigliani & Brumberg (1954) et Friedman (1957), ont fourni l'un des premiers cadres théoriques pour penser les comportements d'épargne. Ces modèles ont été enrichis de la théorie de l'épargne de précaution – déjà évoquée par Keynes (1936), puis modélisée par Leland (1968), Sandmo (1970), Drèze & Modigliani (1972) – qui montre que l'épargne joue également un rôle d'assurance contre les aléas affectant le ménage. Ainsi, les ménages n'épargnent-ils pas seulement pour compenser une baisse des revenus futurs, par exemple au moment de la retraite, mais également pour s'assurer contre d'autres sortes de risques liés au revenu, notamment celui d'une baisse non anticipée des revenus.

La principale difficulté, lorsque l'on cherche à évaluer ce comportement de précaution, est de trouver une mesure correcte du risque lié au revenu. Certains auteurs (Skinner, 1988) utilisent la catégorie socioprofessionnelle (CSP) comme *proxy* de ce risque. L'approche peut alors conduire à sous-estimer la part de l'épargne de précaution, si les personnes les plus averses au risque choisissent leur métier en fonction de cette aversion au risque. Une autre approche consiste à utiliser, sur données de panel, la variance liée au revenu passé comme mesure du risque (Carroll & Samwick, 1997). Cependant, il est possible que cette mesure « objective » du risque ne corresponde pas au ressenti du ménage. L'approche la

plus convaincante est celle qui consiste à utiliser les données subjectives recueillies par enquête, sur l'évolution du revenu ou de la probabilité de chômage (Guiso *et al.*, 1992 ; Lusardi, 1997 ; Lusardi, 1998 ; Arrondel, 2002 ; Carroll *et al.*, 2003 ; Arrondel, 2008).

Dans cet article, on cherche d'abord à tester empiriquement l'homogénéité des taux d'épargne en fonction du niveau de revenu, puis à démontrer l'existence d'un comportement d'épargne de précaution lié au revenu et à tenter de le quantifier, à partir de l'enquête *Budget de famille* de 2010-2011 de l'Insee. Le motif de précaution est appréhendé à travers une mesure subjective : la probabilité de chômage pour l'année à venir, déclarée lors de l'enquête par la personne de référence du ménage (pour elle-même et son conjoint). L'originalité de cet article tient à la quantification du motif de précaution pour analyser le taux d'épargne annuel (en flux), quand la plupart des travaux existants analysent l'accumulation patrimoniale (en stock).

Il ressort que les comportements d'épargne paraissent assez homogènes sur le cycle de vie, sauf pour les 20 % de ménages ayant les revenus les plus élevés, qui épargnent davantage en proportion de leur revenu permanent. Par ailleurs, le motif d'épargne de précaution existe chez tous les ménages français : le surplus d'épargne lié au risque de revenu est évalué à 6.4 %, et la part du patrimoine de précaution due au risque de revenu représente 6.3 % de la richesse globale. Le motif de précaution apparaît différencié, et ce sont les ménages des troisième et quatrième quintiles¹ de revenu qui accumulent le plus d'épargne de précaution.

1. Épargne, revenu et incertitude dans la littérature

Le modèle originel de revenu permanent de Friedman (1957) supposait une parfaite anticipation du revenu futur. Cette hypothèse s'est révélée trop restrictive et le cadre d'analyse a été progressivement enrichi. L'introduction de l'incertitude a notamment permis de mettre en évidence un comportement de précaution : le revenu du travail futur étant aléatoire, la consommation (donc l'épargne) dépend non seulement de l'espérance, mais également de la variance du revenu anticipé.

1. Afin de ne pas alourdir la rédaction, on désignera par « les ménages du n^e quintile de revenu » les ménages dont le revenu est compris entre les (n-1)^e et le n^e quintiles de revenu.

1.1. Épargne, revenu permanent et motif de précaution

Le modèle de Friedman (1957) prévoit que le taux d'épargne à chaque période ne dépend pas du revenu permanent (pour la formalisation du calcul voir Garbinti & Lamarche, 2014), mais uniquement du taux d'intérêt r , de la préférence pour le présent δ et du taux de remplacement au passage à la retraite k . Notons que ce constat n'est valide que si k , r et δ ne dépendent pas du revenu permanent. Il y a cependant plusieurs arguments qui plaident pour un relâchement de cette hypothèse (taux d'intérêt différencié selon le niveau de revenu permanent, taux de remplacement différencié entre ménages aisés et modestes, etc.). En outre, ce modèle fait l'hypothèse d'une anticipation parfaite du revenu futur.

Or, l'incertitude sur le revenu va également avoir une incidence sur le taux d'épargne, et l'importance de cet effet dépend de la courbure de la fonction d'utilité du consommateur. Lorsque l'utilité marginale n'est pas linéaire, i.e. lorsque les consommateurs sont prudents², la consommation est décalée dans le futur et l'épargne aujourd'hui augmente. Un cas particulier est celui d'une fonction d'utilité CARA (*constant absolute risk aversion* ou aversion absolue pour le risque constante), développée par Caballero (1990), qui établit une relation simple entre épargne et incertitude sur le revenu³. Dans ce modèle, le consommateur choisit, à chaque période, la consommation qui maximise sa fonction d'utilité $U(C) = -\frac{1}{\theta} \exp(-\theta C)$ avec θ le coefficient d'aversion au risque absolu $\left(\theta = -\frac{U''}{U'}\right)$. On

fait l'hypothèse que les consommateurs vivent T années, que le taux d'intérêt est égal au taux de préférence pour le présent et que le revenu suit une marche aléatoire, soumis à des chocs de variance σ^2 . La résolution du programme du consommateur montre que la consommation, donc l'épargne, est fonction de l'incertitude sur le revenu σ^2 , de l'âge et de l'aversion au risque de l'individu θ . Dans cette représentation simplifiée, l'incertitude portant sur le revenu réduit la consommation courante et augmente l'épargne.

Notons qu'une autre formalisation du comportement d'épargne de précaution, avec ou sans contraintes de liquidité, a été proposée par Deaton (1991) et Carroll (1992), à partir d'une fonction d'utilité isoélastique CRRA (*constant relative risk aversion* ou aversion relative pour le risque constante). Le motif de précaution apparaît également comme déterminant de la consommation et

de l'épargne, mais on ne peut dériver de forme linéaire simple.

1.2. L'importance de l'épargne de précaution dans la littérature : une absence de consensus

L'étude du motif de précaution dans les comportements d'épargne occupe une place prépondérante dans la littérature récente, et de nombreux travaux ont tenté de quantifier son importance (Browning & Lusardi, 1996, pour une recension).

Du côté « théorique », on trouve des calibrations de modèles de cycle de vie avec prise en compte de l'incertitude sur les revenus. Ces modèles, qui intègrent les effets des taux d'intérêt, des marchés de capitaux imparfaits, dérivent de la maximisation intertemporelle de l'utilité une forme de la consommation. Ils calibrent ensuite cette fonction de consommation à partir des données (Skinner, 1988 ; Caballero, 1991 ; Hubbard *et al.*, 1994 ; Krusell & Smith, 1994 ; Gourinchas & Parker, 2002 ; Cagetti, 2003). Les estimations de l'accumulation de richesse liée au risque de revenu varient largement, entre 0.7 % (Krusell & Smith, 1994) et 50 % (Skinner, 1988 ; Hubbard *et al.*, 1994) du patrimoine total.

Du côté « empirique », les travaux d'estimation sur micro-données proposent une accumulation de précaution comprise entre 1 % et 20 % du patrimoine total, des estimations qui peuvent sembler plus « raisonnables » (Guiso *et al.*, 1992 ; Lusardi, 1997 ; Arrondel & Calvo Pardo, 2008).

Il est difficile de comparer les résultats entre eux en raison de différences de concepts et de champs d'étude : pays, définition de la richesse (financière ou totale), définition de l'épargne (avec ou sans biens durables), méthode de construction des variables de risque (évaluation subjective ou objective du revenu futur, probabilité objective ou subjective de survenance du chômage), populations étudiées (totale, actifs, salariés ou indépendants, etc.). Arrondel & Calvo Pardo (2008) recensent ainsi 21 mesures de l'épargne de précaution dans la littérature récente.

2. Contrairement à l'intuition, l'aversion pour le risque ($U'' < 0$) n'est pas suffisante pour expliquer un comportement de protection vis-à-vis du risque, à savoir l'épargne de précaution. Pour observer un comportement de précaution, il faut également faire l'hypothèse de la prudence ($U''' > 0$) des consommateurs (Kimball, 1990).

3. Ce cas particulier conduit à séparer l'épargne de cycle de vie de l'épargne de précaution. Le coefficient de prudence absolu ne dépend pas de la richesse globale. L'intérêt de cette fonction CARA est néanmoins de permettre une expression « simple » et facilement interprétable de la fonction de consommation.

1.3. Estimer le risque spécifique portant sur le revenu : des approches diverses

Plusieurs approches ont été développées pour tenter de quantifier le risque spécifique portant sur le revenu. L'approche traditionnelle consiste à utiliser des indicatrices de la CSP comme *proxy* de la variance du revenu (Skinner, 1988 ; Fuchs-Schündeln & Schündeln, 2005). Le modèle de précaution est ainsi testé en comparant les taux d'épargne en fonction de la CSP, et l'on s'attend à observer un comportement de précaution plus fort pour les CSP plus risquées. Cela étant, cette mesure peut conduire à sous-estimer l'épargne de précaution, si les ménages plus prudents choisissent leur CSP en fonction de leur aversion au risque. Ainsi, on observera que des CSP plus sûres ont une épargne de précaution aussi importante que des CSP plus risquées.

Une deuxième méthode consiste à utiliser des données de panel (Carroll & Samwick, 1997 ; Kazarosian, 1997 ; Hurst *et al.*, 2005) pour estimer la variance du revenu à partir des revenus passés. Cela étant, cette approche se heurte aux problèmes d'erreur de mesure du revenu et de disponibilité des données.

Une troisième approche consiste à utiliser les mesures subjectives de la variance du revenu (Guiso *et al.*, 1992 ; Lusardi, 1997) et/ou de la probabilité de chômage (Lusardi, 1998 ; Arrondel, 2002 ; Carroll *et al.*, 2003 ; Arrondel & Calvo Pardo, 2008), à partir de questionnaires. Cette méthode présente l'avantage de prendre en compte la perception subjective de la personne de référence du ménage, et donc de s'approcher au mieux du coefficient de prudence. En effet, le ménage va agir en fonction de sa perception du risque, même si objectivement, le risque de chômage ou de variabilité du revenu est faible.

Dans le cas de la France, les estimations empiriques reposent essentiellement sur les travaux de Arrondel (2002), Arrondel & Calvo Pardo (2008), à partir des enquêtes Insee *Patrimoine*. L'effet de l'épargne de précaution, mesurée à partir de la variance du revenu, se révèle positif et statistiquement significatif, mais faible. Dans Arrondel (2002), la part de l'épargne de précaution⁴ représenterait 5 % de l'accumulation totale de patrimoine, à partir de l'enquête *Patrimoine* de 1997. Dans Arrondel & Calvo Pardo (2008), les données de l'enquête *Patrimoine* de 2004 donnent des estimations également faibles : lorsque l'on considère la probabilité de chômage, sur la population active, le motif de précaution

représente 2 à 3 % de l'accumulation patrimoniale. Pour les actifs salariés, les avoirs de précaution sont de l'ordre de 6 à 7 %, supérieurs à ceux des indépendants ou des agriculteurs.

Pour étudier les liens entre épargne et variabilité du revenu, on adopte une approche différente de l'approche purement patrimoniale, et on privilégie l'approche de l'épargne en flux. Après avoir étudié les liens entre épargne et revenu (courant et permanent), on quantifie l'impact du motif de précaution sur le taux d'épargne des ménages, et on en donne une estimation. On procède de même avec le patrimoine, ce qui permet de comparer les résultats à la littérature existante.

2. L'enquête *Budget de famille* et la construction du taux d'épargne

2.1. L'enquête *Budget de famille* de 2011

L'enquête *Budget de famille* 2010-2011 vise à reconstituer toute la comptabilité du ménage : dépenses, y compris celles qui ne relèvent pas de la consommation des biens et services (impôts, taxes, primes d'assurance, transferts inter-ménages, etc.), consommations ne donnant pas lieu à dépense (autoconsommation alimentaire, avantages en nature fournis par l'employeur) et enfin, ressources exceptionnelles (revenus, prestations sociales, sommes provenant d'autres ménages, héritage, primes de licenciement, loto, etc.). Quelques questions sur le patrimoine et l'épargne, la situation financière et son évolution, l'achat ou la vente de logement et de biens durables pendant l'année en cours figurent également dans l'enquête. On dispose de 15 797 observations dans l'échantillon initial, représentant 28.5 millions de ménages.

2.2. Construction du taux d'épargne

Pour définir l'épargne, on part de la contrainte budgétaire des ménages :

$$A_{t+1} = (1 + r)A_t + Y_t - C_t$$

avec A , r , Y et C respectivement le patrimoine, le taux d'intérêt réel, le revenu disponible hors

4. L'épargne de précaution est mesurée selon une méthode inspirée de Guiso *et al.* (1992), qui consiste à distribuer 100 points entre différents scénarios d'évolution du revenu dans les cinq prochaines années. Le montant de la richesse (divisée par le revenu permanent) est ensuite régressé sur la variance du revenu et d'autres variables exogènes.

revenu du capital et la consommation. L'épargne peut se définir en stock ou en flux. En stock, elle correspond à la variation du patrimoine entre t et $t+1$: $S_t = A_{t+1} - A_t$; en flux, elle correspond à la partie non consommée du revenu courant : $S_t = (Y_t + rA_t) - C_t$ avec rA_t le revenu du capital et $Y_t + rA_t$ le revenu disponible total.

On retient la définition en flux. L'épargne est ainsi construite comme la différence entre revenu disponible et dépense de consommation des ménages, et le taux d'épargne est simplement le rapport de l'épargne au revenu disponible. On définit le revenu disponible d'un ménage comme la somme des revenus d'activité (nets des cotisations sociales), des revenus du patrimoine, des transferts en provenance d'autres ménages et des prestations sociales (y compris les pensions de retraite et les indemnités chômage), et des impôts directs (impôt sur le revenu, taxe d'habitation, taxe foncière).

En faisant abstraction des entrepreneurs individuels, l'épargne des ménages a trois composantes : l'acquisition de biens durables, l'épargne logement et l'épargne financière, c'est-à-dire l'acquisition de monnaie et de titres (actions et obligations). La mesure de l'épargne dépend fortement de la définition du patrimoine. Il faut en effet distinguer du revenu les ressources tirées de la désaccumulation du patrimoine, et distinguer de la consommation les dépenses liées à l'acquisition de patrimoine. On retient une définition large du concept de patrimoine : actifs financiers, biens immobiliers (logements, terrains, etc.) et principaux biens durables (automobiles, grosses machines, etc.). On fait le choix d'exclure du revenu disponible les revenus tirés de la vente d'actifs financiers, de patrimoine immobilier et de biens durables⁵.

Les données portant sur les revenus d'activité, les revenus sociaux et les impôts de l'enquête de 2011 ont bénéficié d'appariements avec les fichiers administratifs et sociaux et sont de bonne qualité, même si la sous-déclaration des revenus n'a pas été totalement éliminée. En revanche, les revenus du patrimoine continuent à être collectés sur une base purement déclarative, d'où une très forte sous-déclaration. Cette sous-déclaration implique que, pour les ménages disposant de revenus du patrimoine (les ménages les plus riches), le taux d'épargne calculé sera sous-estimé. En outre, contrairement à la comptabilité nationale, on ne prend pas en compte les intérêts capitalisés.

Quant à la consommation finale des ménages, elle est plus faible dans l'enquête que dans la

comptabilité nationale. Cette différence est d'abord conceptuelle. Dans la comptabilité nationale, un bien ou un service mis gratuitement à la disposition d'un ménage peut entrer dans le champ de la consommation finale, alors qu'il n'est pas dans le champ de l'enquête. En outre, la consommation dans les enquêtes *Budget de famille* est mesurée à partir de carnets, sur lesquels sont reportées toutes les dépenses effectuées durant une ou deux semaines, ainsi que les dépenses importantes de biens durables de l'année. Les difficultés d'extrapolation sur une année de données hebdomadaires ou bimensuelles, combinées à de potentiels oublis, peuvent ainsi expliquer la sous-estimation de la consommation des ménages dans l'enquête. Malgré ces différences, on peut comparer la décomposition des montants de revenus, de consommation et d'épargne, obtenus dans l'enquête et dans les comptes nationaux (voir tableau A1-1 en annexe 1). Bien que les taux d'épargne obtenus, biens durables inclus, soient très proches, les montants de consommation et de revenu disponible sont largement sous-évalués dans l'enquête par rapport à la comptabilité nationale⁶. Malgré tout, le choix a été fait de conserver les données d'enquête déclaratives sans effectuer de calage, car l'ampleur de la sous-déclaration n'est pas forcément homogène selon les ménages et selon les niveaux de revenu. Par ailleurs, on choisit de s'écarter du concept de consommation adopté par la comptabilité nationale : les achats de biens durables ne sont pas considérés comme une consommation, mais comme de l'épargne⁷. L'échantillon d'étude est constitué de 13 393 ménages⁸.

On obtient ainsi le profil de taux d'épargne (tableau 1), et la distribution des taux d'épargne (voir figure A1-I en annexe 1). La moyenne élevée (29.3 %) est liée au choix de considérer

5. La plupart des études empiriques s'intéressent au taux d'épargne hors consommation de biens durables. En effet, les achats de biens durables ne se renouvellent pas tous les ans et il est difficile de décider de la durée d'amortissement d'un bien.

6. La comparaison de l'enquête Budget de famille avec les comptes par catégorie de ménages confirme une forte sous-estimation des montants de revenu disponible et de consommation pour chaque catégorie. Cela étant, le coefficient de corrélation entre dépenses de consommation par catégorie socio-professionnelle de la personne de référence dans la comptabilité nationale et dans l'enquête Budget de famille est de 0.98 (et de 0.94 pour le revenu disponible), ce qui montre que l'on a une hiérarchie du revenu et de la consommation comparable entre CSP dans les deux sources.

7. Cf. note 5.

8. On exclut de l'échantillon les ménages « atypiques » susceptibles de brouiller l'analyse : les ménages dont la personne de référence est membre du clergé, étudiant, chômeur n'ayant jamais travaillé, inactif autre que retraité (on écarte 1 615 ménages sur les 15 797 ménages initiaux). Par ailleurs, on exclut les ménages ayant un revenu disponible après impôt négatif (174 ménages). On tronque la distribution des 1 % de ménages affichant les taux d'épargne les plus faibles (inférieurs à -115 % du revenu disponible) et des 1 % de ménages ayant les taux d'épargne les plus élevés (supérieurs à 88 % du revenu disponible).

Tableau 1 – Taux d'épargne dans l'enquête *Budget de famille*

(En milliards d'euros courants)

Revenu disponible, hors loyers imputés (a)	846.7
Consommation finale totale, hors loyers imputés	695.2
dont : biens non et semi-durables (b)	598.8
investissement en biens durables (c)	96.4
Épargne, biens durables exclus (a-b-c)	151.5
Taux d'épargne, biens durables exclus ((a-b-c)/a)	17.9 %
Épargne, biens durables inclus (a-b)	247.9
Taux d'épargne, biens durables inclus ((a-b)/a)	29.3 %

Source et champ : Insee, enquête *Budget de famille* 2011. Ensemble des ménages, à l'exclusion des 1 % de taux d'épargne extrêmes et des ménages dont la personne de référence est membre du clergé, étudiant, chômeur n'ayant jamais travaillé ou inactif autre que retraité.

les biens durables comme de l'épargne et non comme une consommation. On vérifie également la cohérence des taux d'épargne avec l'aisance financière des ménages, et on retrouve bien l'effet selon lequel le taux d'épargne croît avec l'aisance financière déclarée des ménages (voir tableau A1-2 en annexe 1). Le taux d'épargne est toujours positif, même pour les catégories qui déclarent s'endetter ; cela vient du fait que l'on inclut tous les biens durables dans l'épargne, donc la variable d'épargne dans cette approche est beaucoup plus large que la variable d'épargne hors biens durables.

L'analyse porte d'abord sur les liens entre revenu (courant et permanent) et taux d'épargne ; on introduit ensuite l'incertitude portant sur le revenu afin de mesurer l'ampleur de l'épargne de précaution, puis celle du patrimoine de précaution.

3. Résultats

3.1. Épargne, revenu courant et revenu permanent

3.1.1 Épargne et revenu courant

On commence par vérifier le fait bien établi dans la littérature selon lequel le taux d'épargne est croissant avec le niveau du revenu courant. On estime des régressions « moindres carrés ordinaires » (MCO) et des régressions à la médiane. En effet, étant en présence d'une distribution dispersée des taux d'épargne, la médiane est robuste aux valeurs extrêmes, contrairement à la moyenne. Le modèle est de la forme :

$$s_c = \frac{S}{Y^C} = f(Y^C) + X\beta + \epsilon$$

avec s_c le taux d'épargne courant, S l'épargne en montant, Y^C le revenu courant, X les variables explicatives (âge moyen du ménage en tranches⁹, type de ménage détaillé, sexe de la personne de référence, ménage urbain ou rural, période de maladie, perception d'un héritage) et ϵ le résidu. La fonction f est soit la décomposition en quintiles de Y^C soit la fonction identité (dans ce cas on régresse le taux d'épargne sur le revenu courant). On estime à la fois des régressions de la moyenne et de la médiane (tableau 2). Le ménage de référence est un ménage formé d'un couple de deux actifs avec enfant(s), d'âge moyen compris entre 40 à 49 ans, urbain, dans lequel aucun des conjoints n'a connu de période de maladie ou perçu d'héritage.

Considérons d'abord les régressions (ii) et (iv). On observe que le taux d'épargne (moyen et médian) croît avec le quintile de revenu courant, un résultat similaire à ce que prédit la littérature. Ainsi, le taux d'épargne moyen des ménages de référence des deuxième, troisième, quatrième et cinquième quintiles de revenu est respectivement de 3.3 %, 15.4 %, 24.2 % et 36.7 %. Pour un quintile donné, le taux d'épargne du quintile supérieur est systématiquement plus élevé. Le résultat est similaire lorsque l'on considère les régressions médianes : toutes choses égales par ailleurs, le taux d'épargne médian d'un quintile donné est toujours plus élevé que le taux d'épargne médian du quintile inférieur. On procède également à une régression du taux d'épargne sur le revenu courant en continu (régressions (i) et (iii)), ce qui revient à tester les non linéarités entre épargne et revenu courant. Le coefficient portant sur le

9. L'âge moyen du ménage est défini comme l'âge de la personne de référence pour les personnes seules, et comme la moyenne de l'âge de la personne de référence et de l'âge du conjoint pour les couples.

revenu est positif et significatif dans les deux régressions : ainsi, le taux d'épargne moyen et médian croît avec le revenu courant. Pour un ménage ayant un revenu disponible médian et un taux d'épargne médian, l'augmentation du revenu disponible de 1 % augmente l'épargne du ménage de 1.5 %¹⁰. L'épargne augmente donc plus que proportionnellement au revenu, elle est assimilable à un « bien de luxe ».

La hiérarchie des taux d'épargne est en phase avec les résultats de Boissinot (2003) sur l'enquête *Budget de famille* de 2000-2001 ou d'Antonin (2009). Le différentiel de 42 points

entre taux d'épargne médians extrêmes est cependant moins fort que le différentiel de 60 points obtenu par Garbinti & Lamarche (2014) sur l'enquête *Patrimoine* 2010, mais cette différence peut s'expliquer en partie par le choix d'exclure de la base les centiles extrêmes de taux d'épargne. La spécification

10. Le revenu médian étant de 25 800 euros sur l'échantillon, une augmentation de 1 % augmente le revenu médian de 258 euros supplémentaires. À partir de la régression (iii), cela augmente le taux d'épargne de 0.14 point (5.4×0.0258). Ainsi, le taux d'épargne médian passe de 29.7 % à 29.84 %, alors que le revenu passe de 25 800 à 26 058 euros. L'épargne initiale passe de 7662.60 euros à 7775.40 euros, soit une augmentation de 1.5 %.

Tableau 2 – Régressions moyenne et médiane du taux d'épargne sur le revenu courant

	Régression moyenne		Régression médiane	
	(i)	(ii)	(iii)	(iv)
Constante	1.9* (1.1)	-9.4*** (1.2)	7.6*** (1.6)	-1.2 (1.7)
Revenu courant (10 ⁴)	5.4*** (0.2)		5.4*** (0.3)	
Revenu courant en quintiles				
Q1		réf.		réf.
Q2		12.7*** (0.9)		10.2*** (1.5)
Q3		24.8*** (1.0)		22.6*** (1.4)
Q4		33.6*** (1.1)		30.7*** (1.4)
Q5		46.1*** (1.1)		41.6*** (1.4)
Âge moyen du ménage ^(a)				
Moins de 30 ans	-3.2*** (1.2)	2.3** (1.2)	-3.7** (1.8)	-3.5** (1.6)
30 à 39 ans	0.8 (1.0)	0.6 (1.0)	-0.8 (1.2)	-0.3 (1.2)
40 à 49 ans	réf.	réf.	réf.	réf.
50 à 59 ans	0.5 (1.0)	0.9 (1.0)	-0.6 (1.3)	-0.5 (1.3)
60 à 69 ans	-4.9*** (1.5)	-5.1*** (1.5)	-6.7*** (2.3)	-7.7*** (2.1)
70 ans et plus	4.8*** (1.7)	6.5*** (1.7)	1.8 (2.4)	2.8* (2.3)
Nombre d'observations	13 393	13 393	13 393	13 393

(a) Âge de la personne de référence pour les personnes seules, moyenne de l'âge de la personne de référence et de l'âge du conjoint pour les couples.

Note : ***, **, * significatif aux seuils respectifs de 1 %, 5 % et 10 %. Les écarts-types figurent entre parenthèses. Les variables de contrôles suivantes ont été introduites dans les régressions : type de ménage, sexe de la personne de référence, urbain/rural, maladie, héritage. Le ménage de référence est un couple de deux actifs avec enfant(s), d'âge moyen compris entre 40 et 49 ans, urbain, n'ayant pas connu de période de maladie et n'ayant pas perçu d'héritage.

Source et champ : Insee, enquête *Budget de famille* 2011. Ensemble des ménages, à l'exclusion des 1 % de taux d'épargne extrêmes et des ménages dont la personne de référence est membre du clergé, étudiant, chômeur n'ayant jamais travaillé ou inactif autre que retraité.

la plus semblable à celle utilisée ici est celle de Bozio *et al.* (2013) lesquels, sur données anglaises 2007-2009, trouvent un différentiel de 50 points entre médianes des quintiles extrêmes, en contrôlant l'âge et la structure familiale.

Les modèles de cycle de vie prédisent une augmentation du taux d'épargne jusqu'à l'âge de la retraite, puis une désaccumulation ensuite. Comme le notent Dynan *et al.* (2004), le comportement d'épargne est difficilement interprétable en fonction de l'âge, d'autant que les données étant en coupe, les estimations mélangent les effets d'âge et de génération. L'effet de l'âge n'apparaît pas clairement, sauf pour la tranche des 60-69 ans qui affiche un taux d'épargne significativement plus faible que les autres tranches d'âge dans toutes les régressions, et les plus de 70 ans qui épargnent davantage. Ce dernier résultat peut s'expliquer par une mortalité différentielle, c'est-à-dire une plus grande probabilité de survie des ménages les plus riches, qui conservent des taux d'épargne élevés en fin de vie (Bommier *et al.*, 2005). Par ailleurs, le motif de transmission intergénérationnelle peut également expliquer cette persistance d'un taux d'épargne élevé (voir Kotlikoff & Summers¹¹, 1981). Les 20-29 ans affichent un taux d'épargne significativement plus faible que les autres, peut-être lié à la faiblesse de leurs revenus en début de carrière et à une insertion professionnelle plus chaotique. Le profil atypique des taux d'épargne n'est pas très surprenant lorsque l'on sait combien la littérature empirique est contrastée à ce sujet. Garbinti & Lamarche (2014) montrent que les plus jeunes et les plus âgés épargnent significativement davantage, mais sans contrôler la structure familiale. Bozio *et al.* (2013), qui contrôlent la structure familiale, trouvent que les ménages plus âgés affichent un taux d'épargne plus élevé, y compris à des âges avancés.

3.1.2 Épargne et revenu permanent

Le revenu permanent est, selon Friedman (1957), le flux de revenu constant qui, actualisé sur un horizon intertemporel infini, est égal à la somme des ressources anticipées actualisées. De façon plus intelligible, Carroll (1992) le décrit comme le niveau de revenu du travail que le ménage recevrait en l'absence de tout choc transitoire sur le revenu. Empiriquement, cette description est assez fruste : elle consisterait à exclure les ménages qui font face à une variation transitoire de leur revenu. En ne retenant dans les échantillons utilisés que les ménages déclarant

une évolution stable (« normale ») de leur revenu pour l'année courante, on pourrait alors assimiler revenu courant et revenu permanent. Cette méthode, appliquée aux données *Budget de famille*, ne donne pas de résultats concluants. On peut distinguer deux autres grandes méthodes d'estimation. La première utilise un instrument pour approximer le revenu permanent : niveau d'éducation, historique de la situation d'emploi du ménage, catégorie socio-professionnelle, consommation de biens non durables, etc. La seconde identifie le revenu permanent à la somme actualisée des revenus perçus sur l'ensemble du cycle de vie, selon la méthode utilisée par Dicks-Mireaux & King (1982), améliorée par Lollivier & Verger (1999).

Garbinti & Lamarche (2014) montrent que l'utilisation de la méthode consistant à reconstruire un revenu permanent sur toute la durée de vie ne change que marginalement les résultats et ne modifie pas la hiérarchie des taux d'épargne entre quintiles, par rapport à la méthode instrumentale. Par conséquent, c'est la première approche qui a été retenue. L'instrument doit être corrélé au revenu permanent, n'avoir d'effet sur l'épargne qu'à travers le revenu permanent et ne pas avoir de lien avec les chocs transitoires affectant le revenu. À l'instar de Dynan *et al.* (2004) et de Bozio *et al.* (2013), on choisit ici de retenir le plus haut diplôme obtenu¹², qui apparaît comme une composante stable du capital humain d'un individu et qui n'est pas lié aux chocs transitoires sur le revenu. Cet instrument n'est pas parfait : en effet, le niveau de diplôme peut être corrélé positivement avec un goût pour l'épargne (Mayer, 1972), une moindre préférence pour le présent, ou avoir un effet propre sur la capacité à anticiper sa retraite. Par ailleurs, il est corrélé avec une meilleure culture financière (Lusardi, 1999). Pour autant, les conclusions demeurent valides, même si l'on met en évidence un lien plus général entre épargne et niveau de diplôme, qui n'est pas totalement imputable à l'effet du revenu. On estime également une spécification alternative en utilisant la catégorie socio-professionnelle comme instrument du revenu permanent.

11. Pour Kotlikoff & Summers (1981), les transferts intergénérationnels constituent l'explication la plus importante de l'épargne. Alors que le ménage de cycle de vie cherche à consommer, à terme, la totalité des ressources perçues, le ménage dynastique transmet des ressources à ses enfants, car ces transferts lui apportent une certaine utilité (voir les travaux de Barro & Becker, 1988, sur l'altruisme dynastique).

12. Une autre possibilité est d'utiliser la catégorie socio-professionnelle, mais c'est une composante moins stable en raison de la mobilité professionnelle.

On souhaite estimer la relation suivante :

$$s_c = \frac{S}{Y^c} = f(Y^p) + X\beta + \epsilon \quad (1)$$

où s_c est le taux d'épargne courant, S et Y^c représentent l'épargne courante et le revenu courant, Y^p le revenu permanent du ménage, et X sont des variables de contrôle affectant directement le taux d'épargne (âge moyen, type de ménage, sexe de la personne de référence, urbain ou rural, période de maladie, perception d'un héritage).

Comme Dynan *et al.* (2004), on utilise une procédure en deux étapes. Dans une première étape, on régresse le revenu courant du ménage sur un instrument, le diplôme le plus élevé de la personne de référence et de son éventuel conjoint (regroupés en huit catégories), et sur des variables de contrôle, ce qui permet d'obtenir une valeur prédite pour le revenu permanent Y^p (voir tableau A2-3 en annexe 2). Dans une deuxième étape, les valeurs prédites du revenu permanent (variable continue ou classée en quintiles) servent de régresseur dans l'équation (1), que l'on estime par régression moyenne et régression à la médiane. Dans le cas où Y^p est une variable continue, on estime la relation par la méthode des doubles moindres carrés ; lorsque l'on introduit les quintiles pour tester la non-linéarité de la relation, les erreurs moyennes sont calculées par *bootstrap*.

Au préalable, afin d'écarter de l'échantillon les ménages ayant de fortes fluctuations de revenu, on exclut 1 658 ménages ayant eu une forte hausse ou une forte baisse de leurs revenus depuis un an, ce qui ramène le nombre d'observations à 11 735 ménages.

Les coefficients estimés pour la variable continue de revenu permanent du tableau 3 sont faibles mais positifs, donc les taux d'épargne croissent lorsque le revenu permanent croît. Par rapport au revenu courant, le fait d'utiliser un instrument a pour effet de lisser les taux d'épargne : les coefficients de la variable de revenu permanent sont moins forts que ceux de la variable de revenu courant. Ainsi, pour un revenu permanent médian et un taux d'épargne médian, l'augmentation du revenu permanent de 1 % augmente l'épargne du ménage de 1.2 % (contre 1.7 % dans le cas du revenu courant). Concernant l'âge, on observe bien une baisse du taux d'épargne significative pour les 60-69 ans, et une hausse du taux d'épargne entre 30 et 60 ans, sauf dans la tranche 40-49 ans.

Dans la littérature, c'est principalement l'hypothèse que les plus aisés épargnent une part plus importante de leur revenu permanent qui semble se dégager des travaux empiriques, même si certaines études soutiennent le point de vue de Friedman (Gustman & Steinmeier, 1998 ; Venti & Wise, 2000). Sur les données *Budget de famille*, il n'y a que pour le quintile de revenu permanent le plus élevé qu'on observe un taux d'épargne significativement plus fort (et encore, la significativité est au seuil de 10 %). Ce résultat est corroboré par l'étude directe de l'impact du diplôme sur le taux d'épargne (cf. la régression directe du taux d'épargne sur le diplôme dans le tableau A2-2 en annexe 2) : seuls les diplômés du supérieur (au-dessus de Bac+2) ont un taux d'épargne significativement plus élevé que les non diplômés. Il est difficile de conclure dans la mesure où l'on observe un taux d'épargne significativement plus élevé seulement pour les 20 % de plus hauts revenus. C'est également ce qu'observent Bozio *et al.* (2013) sur données britanniques et américaines. La différence de taux d'épargne, en contrôlant l'âge et le type de ménage, serait, sur les données *Budget de famille*, assez faible, de 4 à 7 points de pourcentage selon la spécification (4 points pour Bozio *et al.*, 2013). Pour Garbinti & Lamarche (2014), ou Dynan *et al.* (2004), l'effet est plus prononcé, avec une hiérarchie croissante des taux d'épargne en fonction des quintiles de revenu, alors qu'on ne l'observe, sur ces données, que pour les revenus les plus élevés.

On constate que le revenu permanent n'explique qu'une faible partie du comportement d'épargne. D'autres facteurs influencent les comportements des ménages, notamment l'aléa portant sur le revenu ; on se propose de compléter le modèle en l'intégrant. Pour cela, on se concentre sur les ménages actifs : la notion d'épargne de précaution s'analyse en effet à travers le risque de chômage, donc pour des ménages actifs (8 082 ménages, contre 11 735 pour l'ensemble des ménages).

3.2. Le motif de précaution et l'épargne

Les ménages n'épargnent pas seulement pour compenser la baisse des revenus futurs, mais aussi pour s'assurer contre l'incertitude. Pour mesurer cette incertitude, on régresse à nouveau le taux d'épargne courant sur le revenu permanent (instrumenté par le diplôme), l'âge, le nombre d'enfants, la situation maritale, la nationalité et le sexe de la personne de référence, le

lieu de résidence (urbain ou rural), l'existence de périodes de maladie et la perception d'un héritage, ainsi que sur deux variables issues des enquêtes : (1) une variable décrivant la perception de la personne de référence du ménage sur l'évolution de son niveau de vie dans les douze prochains mois ; (2) la variance du revenu futur calculée à partir de la probabilité de chômage (tableau 4). On dispose dans l'enquête d'une

variable indiquant le risque de chômage perçu par la personne de référence et son conjoint dans les douze mois à venir. On reprend la méthode de Lusardi (1998) et on affecte une valeur quantitative à la probabilité de chômage de la personne de référence : $p_{PR} = 0$ pour un risque nul, $p_{PR} = 0.3$ pour un risque faible, $p_{PR} = 0.5$ pour un risque moyen, $p_{PR} = 0.7$ pour un risque élevé et $p_{PR} = 0.9$ pour une

Tableau 3 – Régression du taux d'épargne moyen (en %) sur le revenu permanent (instrument : diplôme)

	Ensemble des ménages (hors indépendants)		Ménages (hors indépendants) dont la personne de référence est active	
	(i)	(ii)	(iii)	(iv)
Constante	20.9*** (2.2)	24.6*** (2.1)	16.5*** (2.6)	21.6*** (2.4)
Revenu permanent (10 ⁴)	1.2*** (0.4)		1.3*** (0.4)	
Revenu permanent				
Q1		réf.		réf.
Q2		-1.9 (1.2)		-1.0 (1.4)
Q3		0.0 (1.4)		-0.2 (1.8)
Q4		0.4 (1.6)		-0.1 (1.9)
Q5		3.2** (1.7)		2.2* (1.4)
Top 5 %		6.8*** (2.0)		6.8*** (2.0)
Âge moyen du ménage ^(a)				
Moins de 30 ans	-5.2*** (1.3)	-5.6*** (1.4)	-5.9*** (1.3)	-6.7*** (1.3)
30 à 39 ans	-0.1 (1.11)	-0.3 (1.13)	-0.3 (1.06)	-0.5 (1.09)
40 à 49 ans	réf.	réf.	réf.	réf.
50 à 59 ans	2.3** (1.1)	2.1* (1.2)	2.0* (1.1)	2.2* (1.2)
60 à 69 ans	-1.3 (1.3)	-1.6 (1.3)	-6.3*** (2.2)	-5.9*** (2.3)
70 ans et plus	6.2*** (1.3)	5.6*** (1.3)	5.5 (9.9)	7.7 (10.1)
Nombre d'observations	10 840	10 840	7 205	7 205

(a) Âge de la personne de référence pour les personnes seules, moyenne de l'âge de la personne de référence et de l'âge du conjoint pour les couples.

Note : ***, **, * significatif aux seuils respectifs de 1 %, 5 % et 10 %. Les écarts-types figurent entre parenthèses. Les variables de contrôles suivantes ont été introduites dans les régressions : type de ménage, sexe de la personne de référence, urbain/rural, maladie, héritage. Le ménage de référence est un couple de deux actifs avec enfant(s), d'âge moyen compris entre 40 et 49 ans, urbain, n'ayant pas connu de période de maladie et n'ayant pas perçu d'héritage.

Source et champ : Insee, enquête *Budget de famille* 2011. Ensemble des ménages (hors indépendants), à l'exclusion des 1 % de taux d'épargne extrêmes et des ménages dont la personne de référence est membre du clergé, étudiant, chômeur n'ayant jamais travaillé ou inactif autre que retraité.

quasi-certitude¹³. Pour les ménages mono-actifs, la variance du revenu futur, y_{t+1} , se calcule à partir de la formule : $Var_{a,PR}(y_{t+1}) = p_{PR}(1-p_{PR})(1-a)^2 y_t^2$, où a désigne le taux de remplacement du revenu en cas de chômage. Pour les couples dans lesquels les deux conjoints travaillent, on raffine la formule pour tenir compte de la probabilité de chômage du conjoint p_{CJ} . x désignant la part dans les revenus du ménage du revenu de la personne de référence, on définit : $Var_{a,couple}(y_{t+1}) = [x \times p_{PR}(1-p_{PR}) + (1-x)p_{CJ}(1-p_{CJ})](1-a)^2 y_t^2$. En France, le taux de remplacement moyen en début de période d'indemnisation est de 67 % et la dispersion du taux de remplacement est particulièrement faible sur différents cas-types (Dhont-Peltrault, 2017). Cela permet de faire l'approximation d'attribuer le même taux de remplacement à tous les ménages. Dans ce cas, $(1-a)^2$ est une constante, et les estimations économétriques ne sont pas sensibles à la valeur de a (Lusardi, 1998) ; on fixe $a=0$ par souci de simplicité.

On retrouve le fait que le taux d'épargne croît avec le revenu permanent (tableau 5). Le taux d'épargne est significativement plus faible dans le cas des personnes seules et des familles monoparentales. Les couples sans enfant ont en revanche un taux d'épargne plus fort. Les habitants de zones rurales épargnent également davantage que les citadins. Ceux qui reçoivent un héritage épargnent moins ; alors qu'ils accumulent paradoxalement davantage de patrimoine (voir tableau 6) : en effet, le taux d'épargne est mesuré en excluant les dons ou héritages reçus, alors que la mesure du patrimoine en tient compte. Par conséquent, la consommation est susceptible d'augmenter en raison de ces entrées d'argent exceptionnelle (et donc l'épargne baisse). Les ménages d'âge moyen 60-69 ans sont les seuls qui affichent un taux d'épargne statistiquement plus faible que les autres dans les deux régressions, alors que l'effet de l'âge

13. Pour tester la robustesse de l'estimation, on considère des quantifications alternatives de la probabilité de chômage. Le tableau A2-4 en annexe 2 montre que le surplus d'épargne est quasiment inchangé dans les différentes variantes testées.

Tableau 4 – Les mesures du risque de revenu

<i>Risque de chômage pour la personne de réf. dans les 12 mois (en % du total)</i>	
Risque nul	45.6
Risque faible	36.0
Risque moyen	12.6
Risque élevé	4.0
Quasi-certitude	1.8
Nombre d'observations	5 613
<i>Risque de chômage du conjoint éventuel dans les 12 mois (en % du total)</i>	
Risque nul	50.8
Risque faible	29.8
Risque moyen	12.1
Risque élevé	5.0
Quasi-certitude	2.3
Nombre d'observations	2 426
<i>Évolution du niveau de vie du ménage d'ici cinq ans (en % du total)</i>	
Va beaucoup s'améliorer	6.1
Va un peu s'améliorer	27.1
Va se maintenir	36.5
Va un peu se détériorer	22.9
Va beaucoup se détériorer	7.5
Nombre d'observations	5 408

Source et champ : Insee, enquête Budget de famille 2011. Ménages dont la personne de référence est active.

est non significatif pour les plus de 70 ans. Ce résultat est néanmoins difficile à interpréter dans la mesure où il ne concerne que les ménages encore actifs, très peu nombreux après 70 ans.

Quant au risque portant sur le revenu, le fait d'anticiper une amélioration de son revenu courant est corrélé à un taux d'épargne significativement plus faible (13.4 points en moyenne), mais l'inverse n'est pas vérifié. En outre, plus la variance du revenu (calculée à partir de la probabilité de chômage) est élevée, plus le taux d'épargne augmente : même si le coefficient est faible, il

est significatif au seuil de 1 %. Si l'on compare le surplus d'épargne lorsque le revenu est incertain (par rapport à un revenu certain), la régression 5 implique que, à la moyenne de l'échantillon, ce surplus d'épargne serait de 6.4 %¹⁴.

14. Soit s_i le taux d'épargne en situation d'incertitude et s_0 le taux d'épargne si la variance du revenu est nulle. En notant b le coefficient estimé de la variable variance du revenu, on a :

$$\frac{s_i - s_0}{s_i} = \frac{b \frac{\sigma^2}{Y^P}}{\frac{1}{n} \sum (s_i)}$$

Tableau 5 – Épargne de précaution (mesure du risque : variance du revenu)

	Coefficient	Écart-type	Moyenne
Constante	13.9***	(2.9)	1.00
Revenu permanent (10 ⁴)	1.3***	(0.5)	36 447
Variance du revenu (10 ⁴)	3.2***	(0.4)	4 876
Niveau de vie dans 5 ans			
Va beaucoup s'améliorer	-13.4***	(2.0)	0.05
Va un peu s'améliorer	-1.4	(1.2)	0.26
Va rester constant	réf.	réf.	réf.
Va un peu se détériorer	-0.3	(1.2)	0.22
Va beaucoup se détériorer	-2.4	(1.8)	0.07
Ne sait pas	-3.7*	(2.0)	0.06
La personne de référence du ménage est une femme (référence homme)	-0.4	(1.0)	0.34
Âge moyen du ménage			
20-29 ans	-4.1***	(1.5)	0.18
30-39 ans	0.6	(1.2)	0.28
40-49 ans	réf.	réf.	réf.
50-59 ans	2.0	(1.3)	0.22
60-69 ans	-6.6**	(2.7)	0.03
Plus de 70 ans	-1.8	(14.6)	0.00
Type de ménage			
Personne seule	-4.1***	(1.6)	0.25
Famille monoparentale	-11.1***	(1.9)	0.10
Couple sans enfant	5.0***	(1.3)	0.20
Couple avec enfant(s)	réf.	réf.	réf.
Autre type de ménage	-1.4	(3.0)	0.02
Vit en zone rurale	6.1***	(1.1)	0.20
Héritages et donations reçus	-6.1***	(2.4)	0.04
Maladies ou invalidité	6.9***	(1.9)	0.94
R ²	0.07		
Nombre d'observations	5 613		

Note : ***, **, * : significatif aux seuils respectifs de 1 %, 5 % et 10 %. On utilise une régression en deux étapes (DMC) pour estimer le revenu permanent, puis le taux d'épargne. La moyenne du taux d'épargne est de 24.0 %.

Source : Insee, enquête *Budget de famille* 2011.

Étant donné que le taux d'épargne moyen hors biens durables est de 24.3 % sur l'échantillon, l'épargne de précaution représente 1.6 % du revenu disponible, soit une part assez marginale. Autrement dit, pour un revenu courant moyen¹⁵, le surplus d'épargne lié à l'incertitude sur le revenu atteint 560 euros par an. Notons qu'un biais d'endogénéité n'est pas à exclure, si l'on a des ménages averse au risque qui valorisent à la fois le fait d'avoir un revenu stable et une épargne élevée, et *a contrario* des ménages « risquophiles » qui ont un goût pour les revenus risqués et épargnent peu. Dans ce cas, le coefficient de la variance du revenu sera sous-estimé, même si le fait d'avoir exclu les indépendants de l'échantillon permet une plus grande homogénéité de l'aversion au risque au sein de la population étudiée.

Par ailleurs, à partir de la régression 5, on peut calculer l'ampleur « maximale » de l'épargne de précaution, dans le cas hypothétique où tous les ménages auraient une incertitude maximale pesant sur leur revenu¹⁶. Dans cette situation extrême, l'épargne de précaution représente 2.9 % du revenu disponible, contre 1.6 % dans la situation d'incertitude moyenne observée sur l'enquête. D'après la méthodologie adoptée, cela signifie que quelle que soit la probabilité de chômage, la part de l'épargne de précaution se situe entre 0 et 2.9 % du revenu disponible.

Ces résultats empiriques sont intéressants mais il est difficile de les comparer à la littérature. En effet, depuis une vingtaine d'années, la plupart des articles portant sur l'estimation du motif de précaution s'attachent à mesurer la part du patrimoine lié au motif de précaution. Selon Deaton (1991) et Carroll (1992), la caractéristique des épargnants dans le modèle d'épargne tampon (*buffer stock*) est de vouloir atteindre une cible de richesse rapportée au revenu, la taille de cette cible étant, au moins en partie, fonction de l'incertitude liée au revenu.

3.3. Patrimoine et motif de précaution

Afin de pouvoir comparer les résultats à la littérature empirique existante, on évalue maintenant la part du patrimoine accumulé par motif de précaution à partir des données de *Budget de famille*.

Si l'incertitude liée au revenu accroît le taux d'épargne, elle accroît également le patrimoine accumulé. En principe, l'épargne et le patrimoine sont reliés par la contrainte de

budget intertemporelle, et calculer l'impact de l'incertitude sur l'épargne et l'accumulation de patrimoine devrait être équivalent. Cependant, étant données les différences dans le calcul de l'épargne (différence entre revenu disponible et consommation) et le patrimoine (somme des actifs détenus), les impacts peuvent être différenciés (Guiso *et al.*, 1992).

Pour mesurer la part du patrimoine liée au motif de précaution, on estime l'équation inspirée de la théorie du cycle de vie (Dicks-Mireaux & King, 1982) :

$$\ln \frac{A}{Y^P} = f \left(\text{age}, \frac{\sigma^2}{Y^P}, X \right) + \epsilon \quad (2)$$

où A/Y^P est le ratio patrimoine/revenu permanent du ménage, $\frac{\sigma^2}{Y^P}$ la variance subjective du revenu futur rapportée au revenu permanent, X d'autres caractéristiques du ménage et ϵ le terme d'erreur. Le revenu permanent peut figurer parmi les variables X si les préférences ne sont pas homothétiques (Masson & Arrondel, 1989)¹⁷.

L'équation de patrimoine est estimée à partir du patrimoine des ménages de l'enquête *Budget de famille*. Au niveau du montant du patrimoine, l'enquête est assez fruste. Le patrimoine brut est de type déclaratif, renseigné en 15 tranches. On rend cette variable continue en simulant, pour chaque ménage, un résidu issu d'une loi uniforme, que l'on ajoute à la borne inférieure du patrimoine déclaré¹⁸. La distribution comparée de la variable ainsi obtenue et du montant de patrimoine issu de l'enquête Insee *Patrimoine* 2010 montre de faibles différences (figure A1-II en annexe 1). Le patrimoine moyen par ménage issu de l'enquête *Budget de famille* est de 253 000 euros, contre 259 000 euros avec l'enquête *Patrimoine*. Le patrimoine représente en moyenne 4.6 fois le revenu permanent pour les 20 % de ménages ayant les revenus les plus faibles, contre 8 fois le revenu permanent pour les 20 % ayant les revenus les plus élevés.

On estime l'équation de patrimoine sur la population de ménages dont la personne de référence

15. Le revenu courant moyen est, sur l'échantillon, de 36 121 euros.

16. Cette incertitude maximale est obtenue en fixant la probabilité de chômage $p = 0.5$. La moyenne de la variance du revenu passe alors de 4 876 à 8 950, et le taux d'épargne augmente de 2.5 points d'après la régression.

17. Dans ce cas, la proportionnalité entre patrimoine et revenu permanent n'est plus assurée, contrairement à la théorie standard du cycle de vie.

18. A l'exception du plus haut centile de patrimoine, simulé à partir d'une loi exponentielle.

est active, en dehors des indépendants. On régresse le ratio richesse/revenu permanent sur le logarithme du revenu permanent instrumenté par le diplôme, les principales variables explicatives (type de ménage, tranche d'âge moyen du ménage, nationalité et sexe de la personne de référence, perception d'allocations maladie ou invalidité, perception d'héritages ou de donations, zone de résidence, catégorie sociale), la probabilité de chômage et la variance du revenu (tableau 6). Le coefficient positif sur la variable de revenu permanent montre que le patrimoine augmente plus que proportionnellement aux

ressources de cycle de vie. On constate que le ratio de richesse/revenu permanent augmente avec l'âge, et culmine pour les plus de 70 ans : on n'observe donc pas de désaccumulation aux âges avancés¹⁹. L'une des explications tient au fait que l'on ne conserve dans l'échantillon que les ménages dont la personne de référence est active. Ces ménages âgés qui poursuivent

19. Notons que la régression identique dans laquelle les tranches d'âge sont remplacées par les variables d'âge et d'âge au carré, montre que l'âge a un impact positif sur l'accumulation de patrimoine, mais que l'âge au carré a un impact faible et significatif sur l'accumulation de patrimoine.

Tableau 6 – Patrimoine de précaution (mesure du risque : variance du revenu)

	Coefficient	Écart-type	Moyenne
Constante	-15.4***	(1.0)	1.0
log(Y ^P)	1.6***	(0.1)	10.4
Variance du revenu (10 ⁵)	1.3***	(0.2)	4 876
Niveau de vie dans 5 ans			
Va beaucoup s'améliorer	-0.4***	(0.1)	0.1
Va un peu s'améliorer	-0.3***	(0.1)	0.3
Constant	réf.	réf.	réf.
Va un peu se détériorer	-0.1	(0.1)	0.2
Va beaucoup se détériorer	-0.1	(0.1)	0.1
Ne sait pas	-0.5	(0.1)	0.1
La personne de référence du ménage est une femme (référence homme)	0.1*	(0.0)	0.3
Âge moyen du ménage			
20-29 ans	-0.8***	(0.1)	0.2
30-39 ans	-0.3***	(0.1)	0.3
40-49 ans	réf.	réf.	réf.
50-59 ans	0.2**	(0.1)	0.2
60-69 ans	0.3**	(0.1)	0.0
Plus de 70 ans	1.6**	(0.7)	0.0
Type de ménage			
Personne seule	0.7***	(0.1)	0.3
Famille monoparentale	-0.2*	(0.1)	0.1
Couple sans enfant	0.2**	(0.1)	0.2
Couple avec enfant(s)	réf.	réf.	réf.
Autre type de ménage	-0.4***	(0.1)	0.0
Vit en zone rurale	0.5***	(0.1)	0.2
Héritages et donations reçus	0.3**	(0.1)	0.0
Maladies ou invalidité	-0.1	(0.1)	0.9
R ²	0.24		
Nombre d'observations	5 613		

Note : ***, **, * : significatif aux seuils respectifs de 1 %, 5 % et 10 %. La moyenne de la variable dépendante $\ln \frac{A}{Y}$ est de 0.9.
Source : Insee, enquête Budget de famille 2011.

une activité professionnelle ont des caractéristiques particulières qui pourraient expliquer une moindre désaccumulation de patrimoine que pour le reste de la population. Par ailleurs, les ménages ayant reçu un héritage ou une donation accumulent davantage de patrimoine. Les ménages habitant en zone rurale sont également plus riches en patrimoine (avec un patrimoine moyen de 244 000 euros, contre 212 000 euros pour les ménages urbains), ce qui peut s'expliquer par le nombre plus important de propriétaires fonciers en zone rurale que dans les grandes villes : seuls 22.0 % des ménages ruraux sont locataires, contre 50.4 % des ménages urbains.

Par ailleurs, les ménages dont le niveau de vie va s'améliorer épargnent significativement moins que les autres, mais l'inverse n'est pas vérifié. L'effet de l'incertitude du revenu sur le patrimoine est statistiquement significatif. On compare le patrimoine accumulé en situation d'incertitude W_i (avec $\frac{\sigma^2}{Y^P} > 0$) et le patrimoine accumulé lorsque le revenu W_0 est certain (avec $\sigma^2 = 0$) : il représente 6.3 % de la richesse totale accumulée²⁰, b étant le coefficient estimé de la variance du revenu dans la régression. Ces pourcentages sont comparables aux estimations que l'on trouve dans la littérature (Guiso *et al.*, 1992 ; Lusardi, 1997 ; Lusardi, 1998 ; Arrondel, 2002 ; Arrondel & Calvo Pardo, 2008). Sur l'enquête *Patrimoine* de 2010, Arrondel & Calvo Pardo (2008) obtiennent un coefficient de 2.4 % sur les ménages non retraités et de 7.5 % en tronquant la population non retraitée aux deux extrêmes des montants de patrimoine, sans exclure les indépendants²¹.

Pour savoir si les ménages ayant les plus faibles revenus accumulent une part de patrimoine de

précaution plus forte que les autres, on effectue cinq régressions distinctes pour chacun des cinq quintiles de revenu permanent, avec les mêmes variables explicatives que dans la régression 6, en s'intéressant aux coefficients affectant la variance du revenu (tableau 7). On constate que ce sont les ménages des quintiles intermédiaire et supérieur (troisième et quatrième quintiles) qui affichent le patrimoine de précaution le plus fort. Tous les quintiles accumulent un patrimoine de précaution, représentant entre 6.2 % et 16.2 % de leur patrimoine total. On observe une courbe en U inversé : les 40 % de ménages les plus pauvres et les 20 % de ménages les plus riches accumuleraient moins d'épargne de précaution, alors que les ménages appartenant aux quintiles de revenus intermédiaires auraient un patrimoine de précaution supérieur à 12 % de leur patrimoine total. On peut supposer que les raisons de la faible épargne de précaution est différente entre les deux quintiles inférieurs et le quintile supérieur de revenu permanent. Pour les quintiles inférieurs, on peut penser à un comportement « au jour le jour », avec une difficulté pour les ménages à accumuler du patrimoine, alors que pour les 20 % de ménages ayant les revenus les plus élevés, la facilité plus grande pour retrouver un emploi permet une vision plus « sereine » du risque de chômage. Ce résultat diffère de celui obtenu par Carroll *et al.* (2003) qui, sur données américaines, mettaient en évidence un surplus d'épargne de précaution pour les ménages à revenus moyens ou élevés.

$$20. \text{ D'après l'équation (2), } \frac{W_0}{W_i} = \frac{1}{e^{\frac{b\sigma^2}{Y^P}}}, \text{ donc } \frac{W_i - W_0}{W_i} = 1 - \frac{1}{e^{\frac{b\sigma^2}{Y^P}}}$$

21. La première différence vient d'une différence dans la population étudiée ; une autre différence vient de ce qu'Arrondel & Calvo Pardo choisissent de ne considérer que la probabilité de chômage de la personne de référence dans leur mesure de l'épargne de précaution, y compris pour les couples.

Tableau 7 - Quintiles de revenu permanent et patrimoine de précaution (mesure du risque : variance du revenu)

	Coefficient	Écart-type	Moyenne	Effet en %
Q1	2.60E-05	(1.20E-05)	3196	7.9
Q2	1.70E-05	(4.10E-06)	3974	6.6
Q3	4.60E-05	(7.30E-06)	3823	16.2
Q4	2.40E-05	(4.80E-06)	5236	12.0
Q5	8.00E-06	(1.70E-06)	8053	6.2

Note : estimation du coefficient b dans l'équation (2), pour chaque quintile de revenu permanent.
Source : Insee, enquête *Budget de famille* 2011.

* *
*

Deux résultats importants se dégagent de cet article. D'abord, le taux d'épargne des ménages les plus riches croît avec le revenu permanent, ce qui montre que ceux-ci épargnent davantage que les autres sur le cycle de vie. Ensuite, cet article permet de confirmer l'existence d'un motif de précaution lié au risque portant sur le revenu, et de le quantifier. Ainsi, pour l'année 2010 en France, le surplus d'épargne annuel engendré par l'incertitude sur le revenu serait de 6.4 %, soit 1.6 % du revenu disponible brut – environ 560 euros par an. En cas d'incertitude maximale sur le revenu, les calculs montrent que l'épargne de précaution représenterait 2.9 % du revenu disponible, soit 9 % de l'épargne totale. En termes de stock, le patrimoine accumulé par précaution

serait d'ampleur faible et représenterait 6.3 % du patrimoine total des ménages, avec un effet plus marqué pour les ménages qui se situent dans les troisième et quatrième quartiles de la distribution des revenus.

Ces résultats demandent à être confirmés et approfondis. Ainsi, il faudrait idéalement pouvoir instrumenter la variance du revenu, qui peut en partie capturer l'aversion au risque. Une idée consisterait à tester un indicateur non pas subjectif, mais objectif de la probabilité de chômage en disposant de données plus détaillées sur le type de contrat de travail et d'emploi occupé. Enfin, il serait intéressant de prolonger cette étude sur d'autres pays, pour comparer l'ampleur de l'épargne de précaution, et déterminer dans quelle mesure la flexibilité du marché du travail influence ce comportement de précaution. □

BIBLIOGRAPHIE

- Antonin, C. (2009).** Âge, revenu et comportements d'épargne des ménages : une analyse théorique et empirique sur la période 1978-2006. *Mémoire de Master*, EHESS/École d'Économie de Paris.
- Arrondel, L. (2002).** Risk management and wealth accumulation behavior in France. *Economics Letters*, 74(2), 187–194. [https://doi.org/10.1016/S0165-1765\(01\)00528-6](https://doi.org/10.1016/S0165-1765(01)00528-6)
- Arrondel, L. & Calvo Pardo, H. (2008).** Les Français sont-ils prudents ? Patrimoine et risque sur le marché du travail. *Économie et Statistique*, 417-418, 27–53. <https://www.insee.fr/fr/statistiques/1376746?sommaire=1376762>
- Barro, R. J. & Becker, G. S. (1988).** A Reformulation of the Economic Theory of Fertility. *Quarterly Journal of Economics*, 103(1), 1–25. <https://doi.org/10.2307/1882640>
- Boissinot, J. (2003).** L'épargne des hauts revenus. *Mémoire de Master*, EHESS/École d'Économie de Paris.
- Bommier, A., Magnac, T., Rapoport, B. & Roger, M. (2005).** Droits à la retraite et mortalité différentielle. *Économie et Prévision*, 168(2), 1–16. <https://doi.org/10.3406/ecop.2005.7417>
- Bozio, A., Emmerson, C., O'Dea, C. & Tetlow, G. (2013).** Savings and wealth of the lifetime rich: Evidence from the UK and US. *IFS Working Paper* N° W13/30. <https://doi.org/10.1920/wp.ifs.2013.1330>
- Browning, M. & Lusardi, A. (1996).** Household Saving: Micro Theories and Micro Facts. *Journal of Economic Literature*, 34(4), 1797–1855. <https://www.jstor.org/stable/2729595>
- Caballero, R. J. (1990).** Consumption puzzles and precautionary savings. *Journal of Monetary Economics*, 25(1), 113–136. [https://doi.org/10.1016/0304-3932\(90\)90048-9](https://doi.org/10.1016/0304-3932(90)90048-9)
- Caballero, R. J. (1991).** Earnings Uncertainty and Aggregate Wealth Accumulation. *American Economic Review*, 859–871. <https://www.jstor.org/stable/2117784>
- Cagetti, M. (2003).** Wealth Accumulation Over the Life Cycle and Precautionary Savings. *Journal of Business and Economic Statistics*, 21(3), 339–353. <https://doi.org/10.1198/073500103288619007>

- Carroll, C. D. (1992).** The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence. *Brookings Papers on Economic Activity*, 23(2), 61–156.
<https://econpapers.repec.org/RePEc:bin:bpeajo:v:23:y:1992:i:1992-2:p:61-156>
- Carroll, C. D., Dynan, K. E. & Krane, S. D. (2003).** Unemployment Risk and Precautionary Wealth: Evidence from Households' Balance Sheets. *Review of Economics and Statistics*, 85(3), 586–604.
<https://doi.org/10.1162/003465303322369740>
- Carroll, C. D. & Samwick, A. A. (1997).** The nature of precautionary wealth. *Journal of Monetary Economics*, 40(1), 41–71. [https://doi.org/10.1016/S0304-3932\(97\)00036-6](https://doi.org/10.1016/S0304-3932(97)00036-6)
- Deaton, A. (1991).** Saving and Liquidity Constraints. *Econometrica*, 59(5), 221–248.
<https://econpapers.repec.org/RePEc:em:emetrp:v:59:y:1991:i:5:p:1221-48>
- Dhont-Peltrault, E. (2017).** L'indemnisation du chômage en France au regard des pratiques européennes. *Économie & prévision*, (1), 137–146. <https://doi.org/10.3917/ecop.210.0137>
- Dicks-Mireaux, L. & King, M. A. (1982).** Pension Wealth and Household Savings: Tests of Robustness. NBER Working Paper N° 962. <https://doi.org/10.3386/w0962>
- Drèze, J. & Modigliani, F. (1972).** Consumption Decisions under Uncertainty. *Journal of Economic Theory*, 5(3), 308–335. [https://doi.org/10.1016/0022-0531\(72\)90044-0](https://doi.org/10.1016/0022-0531(72)90044-0)
- Dynan, K. E., Skinner, J. & Zeldes, S. P. (2004).** Do the rich save more? *Journal of Political Economy*, 112(2), 397–444. <https://doi.org/10.1086/381475>
- Friedman, M. (1957).** *Introduction to "A Theory of the Consumption Function"*. Princeton: Princeton University Press.
- Fuchs-Schündeln, N. & Schündeln, M. (2005).** Precautionary Savings and Self-Selection: Evidence from the German Reunification "Experiment". *Quarterly Journal of Economics*, 120 (3), 1085–1120.
<https://doi.org/10.1093/qje/120.3.1085>
- Garbinti, B. & Lamarche, P. (2014).** Les hauts revenus épargnent-ils davantage ? *Économie et Statistique*, 472-473, 49–64. <https://www.insee.fr/fr/statistiques/1377765?sommaire=1377781>
- Gourinchas, P. O. & Parker, J. A. (2002).** Consumption over the life cycle. *Econometrica*, 70(1), 47–89.
<https://doi.org/10.1111/1468-0262.00269>
- Guiso, L., Jappelli, T. & Terlizzese, D. (1992).** Earnings uncertainty and precautionary saving. *Journal of Monetary Economics*, 30(2), 307–337. [https://doi.org/10.1016/0304-3932\(92\)90064-9](https://doi.org/10.1016/0304-3932(92)90064-9)
- Gustman, A. L. & Steinmeier, T. L. (1998).** Effects of Pensions on Saving: Analysis with Data from the Health and Retirement Study. NBER Working Paper N° 6681. <https://doi.org/10.3386/w6681>
- Hubbard, R. G., Skinner, J. & Zeldes, S. P. (1994).** Expanding the Life-Cycle Model: Precautionary Saving and Public Policy. *American Economic Review*, 84(2), 174–179. <https://www.jstor.org/stable/2117824>
- Hurst, E., Kennickell, A., Lusardi, A. & Torralba, F. (2005).** Precautionary Savings and the Importance of Business Owners. NBER Working Paper N° 11731. <https://doi.org/10.3386/w11731>
- Kazarosian, M. (1997).** Precautionary Savings – A Panel Study. *Review of Economics and Statistics*, 79(2), 241–247. <https://www.jstor.org/stable/2951457>
- Kimball, M. S. (1990).** Precautionary Saving in the Small and in the Large. *Econometrica*, 58(1), 53–73.
<https://doi.org/10.2307/2938334>
- Kotlikoff, L. & Summers, L. (1981).** The Role of Intergenerational Transfers in Aggregate Capital Accumulation. *Journal of Political Economy*, 89(4), 706–732. <https://www.jstor.org/stable/1833031>
- Krusell, P. & Smith, A. A. (1994).** Income and wealth heterogeneity, aggregate fluctuations, and the representative agent. *Manuscript (University of Rochester, Rochester, NY)*.
- Leland, H. E. (1968).** Saving and Uncertainty: The Precautionary Demand for Saving. *Quarterly Journal of Economics*, 82(3), 65–473. <https://doi.org/10.2307/1879518>

- Lollivier, S. & Verger, D. (1999).** Inégalités et cycle de vie : les liens entre consommation, patrimoine et revenu permanent. *Annales d'Économie et de Statistique*, 54, 203–246. <https://doi.org/10.2307/20076184>
- Lusardi, A. (1997).** Precautionary saving and subjective earnings variance. *Economics Letters*, 57(3), 319–326. [https://doi.org/10.1016/S0165-1765\(97\)00239-5](https://doi.org/10.1016/S0165-1765(97)00239-5)
- Lusardi, A. (1998).** On the Importance of the Precautionary Saving Motive. *American Economic Review*, 88(2), 449–453. <https://econpapers.repec.org/RePEc:aea:aecrev:v:88:y:1998:i:2:p:449-53>
- Masson, A. & Arrondel, L. (1989).** Hypothèse du cycle de vie et accumulation du patrimoine : France 1986. *Économie & prévision*, 90(4), 11–30. <https://doi.org/10.3406/ecop.1989.5120>
- Modigliani, F. & Brumberg, R. (1954).** Utility analysis and the consumption function: An interpretation of cross-section data. In: Kurihara, K. K. (Ed.). *Post Keynesian Economics*, pp. 388–436. London : Allen and Unwin.
- Sandmo, A. (1970).** The Effect of Uncertainty on Savings Decisions. *Review of Economic Studies*, 37(3), 353–360. <https://doi.org/10.2307/2296725>
- Skinner, J. (1988).** Risky income, life cycle consumption, and precautionary savings. *Journal of Monetary Economics*, 22(2), 237–255. [https://doi.org/10.1016/0304-3932\(88\)90021-9](https://doi.org/10.1016/0304-3932(88)90021-9)
- Venti, S. F. & Wise, D. A. (2000).** Choice, Chance, and Wealth Dispersion at Retirement. NBER *Working Paper* N° 7521. <https://doi.org/10.3386/w7521>
-

ANNEXES

ANNEXE 1

CONSTRUCTION DES VARIABLES

Tableau A1-1 – Comparaison entre comptabilité nationale et enquête *Budget de famille*

(En milliards d'euros)

Agrégats	Comptabilité nationale	Budget de famille
<i>Revenus du travail</i>		
Salaires et traitements bruts	1 068.0	
Impôts sur les salaires et traitements	-82.8	
Cotisations sociales sur les salaires et traitements	-400.4	
Salaires et traitements, nets de prélèvements	584.8	552.6
Revenu mixte des indépendants	120.4	
Impôts sur revenu des indépendants	-9.2	
Cotisations sociales des non-salariés	-13.2	
Revenu mixte des indépendants, net de prélèvements	98.0	54.8
Solde revenus du travail	682.8	607.4
<i>Revenus immobiliers</i>		
Excédent d'exploitation	166.2	133.3
Solde des revenus des terrains et gisements	0.6	0.6
Solde revenus immobiliers	166.8	133.9
<i>Revenus financiers</i>		
Revenus financiers	134.4	18.1
Intérêts versés	-22.7	0.0
Solde revenus financiers	111.7	18.1
<i>Revenus sociaux et de transfert</i>		
Prestations autres que transferts sociaux en nature	424.2	333.4
Autres transferts courants reçus	61.9	27.6
Autres transferts courants versés	-60.0	-78.6
Solde revenus sociaux	426.1	282.4
<i>Impôts (à déduire)</i>		
Impôt sur le revenu	-65.9	-49.4
Autres impôts courants	-20.3	-29.7
Total impôts	-86.3	-79.1
<i>Revenu disponible brut</i>		
Dépense de consommation finale individuelle ⁽¹⁾	1 094.6	811.2
Épargne brute	206.5	151.5
Taux d'épargne brute en %	15.9	15.7

(1) À des fins de comparaison, la dépense de consommation inclut la consommation de biens durables.
Sources : Comptes de la Nation, 2010 et 2011 ; Insee, enquête *Budget de famille* 2011.

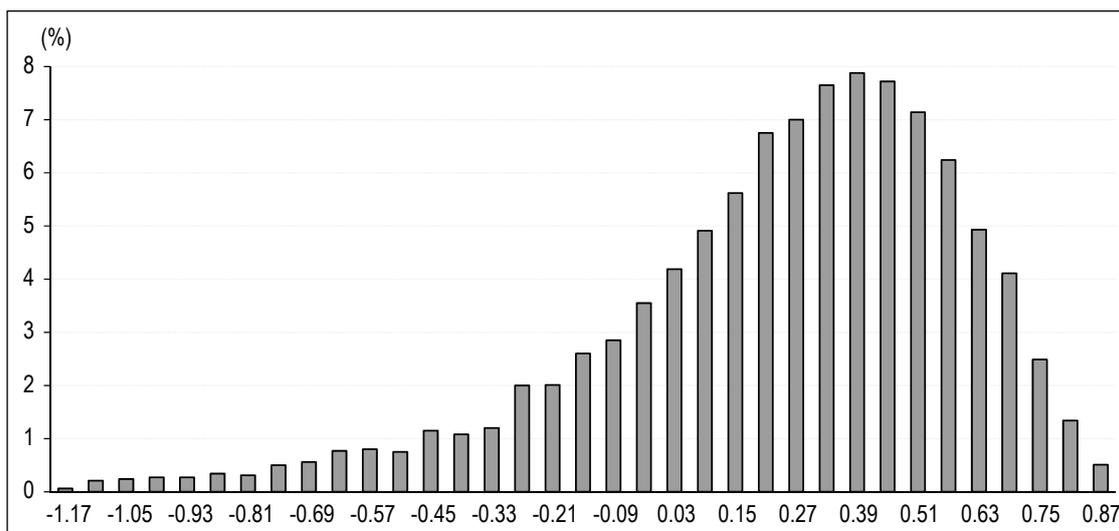
Tableau A1-2 – Taux d'épargne, biens durables inclus, selon l'aisance financière déclarée du ménage

(En %)

Situation du ménage face à son budget	Taux d'épargne
1 - Vous êtes à l'aise	32.3
2 - Ça va	27.4
3 - C'est juste mais il faut faire attention	21.8
4 - Vous y arrivez difficilement	16.7
5 - Vous ne pouvez pas y arriver sans faire de dettes	13.6

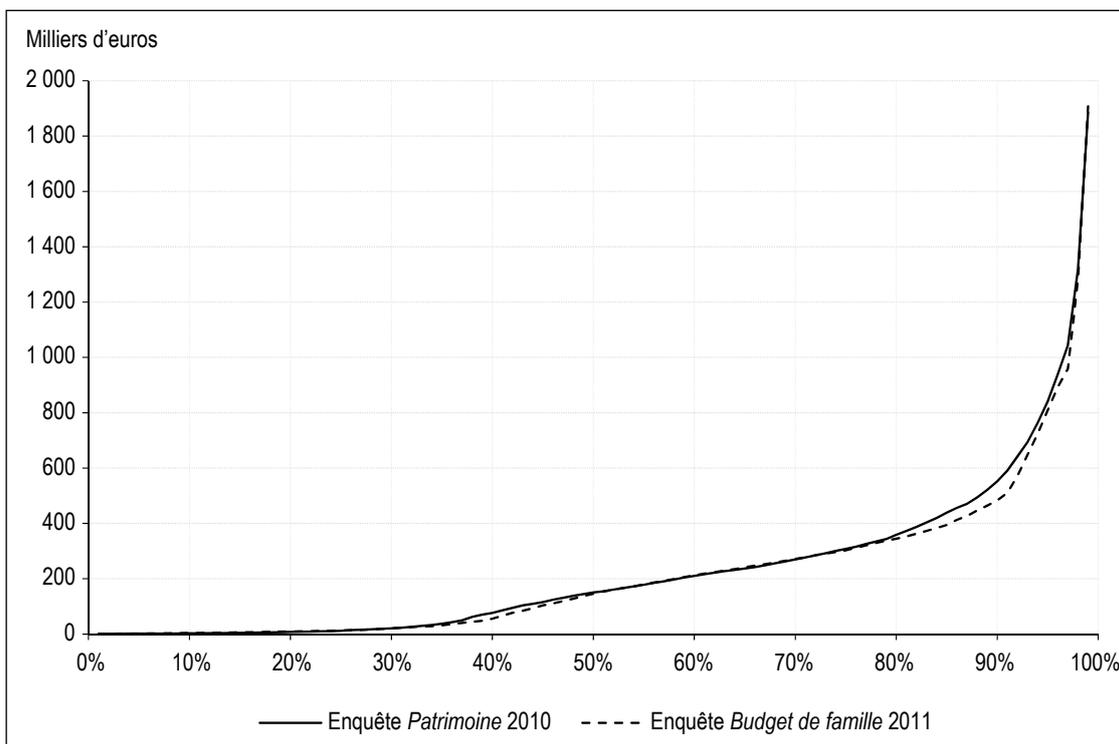
Source : Insee, enquête *Budget de famille* 2011.

Figure A1-I – Distribution pondérée des taux d'épargne sur l'échantillon tronqué



Source et champ : Insee, enquête *Budget de famille* 2011. Ensemble des ménages, à l'exclusion des 1 % de taux d'épargne extrêmes et des ménages dont la personne de référence est membre du clergé, étudiant, chômeur n'ayant jamais travaillé ou inactif autre que retraité.

Figure A1-II – Distribution de la variable « patrimoine brut » dans les enquêtes *Budget de famille* 2011 et *Patrimoine* 2010



Note : la distribution est représentée jusqu'au 99^e centile.
Sources : Insee, enquête *Budget de famille* 2011 et enquête *Patrimoine* 2010.

ANNEXE 2

STATISTIQUES DESCRIPTIVES ET RÉGRESSIONS

Tableau A2-1 – Revenu disponible, consommation et taux d'épargne

(En euros courants)

Quintiles de revenu disponible	Q1	Q2	Q3	Q4	Q5	Ensemble
Revenu disponible médian	12 280	18 748	26 218	36 113	54 555	26 218
Revenu disponible moyen	11 777	18 852	26 276	36 355	63 776	31 405
Consommation médiane	9 827	14 901	18 647	24 041	32 423	18 570
Consommation moyenne	10 710	15 851	19 850	25 414	35 717	21 507
Taux d'épargne médian (%)	20.0	20.5	28.9	33.4	40.6	29.2
Taux d'épargne moyen (%)	9.1	15.9	24.5	30.1	44.0	31.5

Source : Insee, enquête *Budget de famille* 2011.

Tableau A2-2 – Régression directe du taux d'épargne sur le diplôme

	Revenu courant	Écart-type
Constante	35.4***	(1.2)
<i>Diplôme de la personne de référence</i>		
2 ^e ou 3 ^e cycle universitaire	3.6***	(1.1)
BTS, DUT et 1 ^{er} cycle universitaire	-1.3	(1.2)
Bac, CAP, BEP, Brevet professionnel	-1.0	(0.9)
BEPC, CEP	-2.3	(1.4)
Aucun diplôme	réf.	réf.
<i>Diplôme du conjoint</i>		
2 ^e ou 3 ^e cycle universitaire	2.2*	(1.3)
BTS, DUT et 1 ^{er} cycle universitaire	-0.4	(1.3)
Bac, CAP, BEP, Brevet professionnel	-2.3**	(1.0)
BEPC, CEP	-4.9***	(1.5)
Aucun diplôme	réf.	réf.
Nombre d'observations	11 780	

Note : ***, **, * : significatif aux seuils respectifs de 1 %, 5 % et 10 %. Les variables de contrôle suivantes ont été introduites dans les régressions : âge moyen, type de ménage, sexe de la personne de référence, urbain/rural, maladie, héritage. Ménage de référence : couple de deux actifs avec enfant(s), d'âge moyen compris entre 40 et 49 ans, urbain, n'ayant pas connu de période de maladie et n'ayant pas perçu d'héritage.

Source : Insee, enquête *Budget de famille* 2011.

Tableau A2-3 – Régression du revenu courant sur le diplôme (1^{ère} étape)

	Revenu courant	Écart-type
Constante	31.1***	(1.9)
<i>Âge moyen du ménage</i>		
Moins de 30 ans	-11.8***	(0.7)
30 à 39 ans	-5.3***	(0.6)
40 à 49 ans	réf.	réf.
50 à 59 ans	1.5**	(0.6)
60 à 69 ans	-1.4**	(0.7)
Plus de 70 ans	-2.8***	(0.7)
<i>Nationalité de la personne de référence</i>		
Française	réf.	réf.
Autre	-3.6***	(0.8)
<i>Sexe de la personne de référence</i>		
Homme	réf.	réf.
Femme	-1.5***	(0.4)
<i>Diplôme de la personne de référence</i>		
2 ^e ou 3 ^e cycle universitaire	20.0***	(0.7)
BTS, DUT, Bac+2 pro	10.7***	(0.7)
1 ^{er} cycle universitaire	9.5***	(1.6)
Bac général, brevet supérieur	9.2***	(0.8)
Bac pro ou techno	7.6***	(0.8)
CAP, BEP, Brevet professionnel	3.4***	(0.6)
BEPC	4.3***	(0.8)
Certificat d'études	1.0	(0.7)
Aucun diplôme	réf.	réf.
<i>Diplôme du conjoint</i>		
2 ^e ou 3 ^e cycle universitaire	23.4***	(1.9)
BTS, DUT, Bac+2 pro	13.7***	(1.9)
1 ^{er} cycle universitaire	11.0***	(2.7)
Bac général, brevet supérieur	9.1***	(2.0)
Bac pro ou techno	8.3***	(2.0)
CAP, BEP, Brevet professionnel	6.0***	(1.9)
BEPC	5.9***	(2.0)
Certificat d'études	3.1	(2.0)
Aucun diplôme	0.9	(1.9)
Pas de conjoint	réf.	réf.
<i>Type de ménage</i>		
Personne seule	-14.1***	(1.9)
Famille monoparentale	-8.3***	(1.9)
Couple sans enfant	-7.9***	(0.5)
Couple avec enfant(s)	réf.	réf.
Autre type de ménage	-0.8	(1.6)
Nombre d'observations	11 780	
R ²	0.37	

Note : ***, **, * : significatif aux seuils respectifs de 1 %, 5 % et 10 %.
Source : Insee, enquête Budget de famille 2011.

Tableau A2-4 – **Quantifications alternatives de la probabilité de chômage**

Probabilité de chômage	Estimation de base	Variante 1	Variante 2	Variante 3	Variante 4
1 - Non, il n'y a aucun risque	0	0	0	0	0
2 - C'est possible, mais le risque est faible	0.3	0.5	0.1	0.2	0.2
3 - C'est possible, et le risque est moyen	0.5	0.5	0.5	0.4	0.5
4 - C'est possible, et le risque est élevé	0.7	0.5	0.8	0.6	0.8
5 - Oui, c'est quasiment inévitable	0.9	0.9	0.9	0.8	1
Coefficient de la variance du revenu	3.2***	2.7***	3.8***	3.9***	3.9***
Surplus d'épargne (en %)	6.4	6.3	6.5	6.6	6.3

Note : ***, **, * : significatif aux seuils respectifs de 1 %, 5 % et 10 %.
 Source : Insee, enquête *Budget de famille* 2011.

Quatre décennies d'achats alimentaires : évolutions des inégalités de qualité nutritionnelle en France, 1971-2010

Four Decades of Household Food Purchases: Changes in Inequalities of Nutritional Quality in France, 1971-2010

France Caillavet*, Nicole Darmon**, Flavie Létaille*
et Véronique Nichèle*

Résumé – Les inégalités socioéconomiques affectent tous les domaines de la consommation. Les disparités de consommation alimentaire ont des conséquences en termes nutritionnels, pouvant contribuer aux inégalités sociales de santé. En nous appuyant sur 40 années de données représentatives au niveau des ménages (1971-2010), nous étudions l'évolution des grands groupes d'aliments et de la qualité nutritionnelle de la consommation au domicile selon le revenu et le niveau d'éducation. Dans une tendance globale d'amélioration de la qualité nutritionnelle sur la période, nous constatons une évolution favorable pour tous les quartiles de revenu et pour quatre niveaux d'éducation. Les inégalités, importantes en début de période, s'amenuisent dans les années 2000 : très marquées selon le niveau d'éducation dans les années 1970, elles semblent près de disparaître en 2010 ; selon le niveau de revenu, elles sont modérées et décroissent jusqu'en 2010 mais subsistent néanmoins.

Abstract – Socioeconomic inequalities affect all areas of consumption. Disparities in food consumption have nutritional consequences that may contribute to social inequalities in health. Drawing on 40 years of representative data at the household level (1971-2010), this paper examines changes in the major food groups and nutritional quality in at-home consumption by income and education. In a global trend of improving nutritional quality over the period, the study provides evidence of a positive trend for all income quartiles and for four levels of education. Inequalities were significant at the beginning of the period but on the decline in the 2000s: they were very pronounced between education levels in the 1970s but appear to be on the verge of disappearing by 2010; according to income level, they were limited and on the decline until 2010, but still persist.

Codes JEL / JEL Classification : D12, I24

Mots-clefs : inégalités socioéconomiques, achats alimentaires, qualité nutritionnelle, revenu, éducation

Keywords: socioeconomic inequalities, food purchases, nutritional quality, income, education

* Université Paris-Saclay, INRAE, UR ALISS (france.caillavet@inrae.fr ; flavie.letaille@inrae.fr ; veronique.nichele@inrae.fr) ; ** Université de Montpellier, UMR MOISA - INRAE, CIHEAM-IAMM, CIRAD, Montpellier SupAgro (nicole.darmon@inrae.fr)

Nous tenons à remercier Adélaïde Fadhuile (Université Grenoble Alpes), Christine Boizot-Szantai et Nicolas Guinet (INRAE, UR ALISS) pour leur aide sur les données. Merci également à Anne Lhuissier (Centre Maurice Halbwachs) et aux deux rapporteurs anonymes pour leurs commentaires et suggestions.

Reçu le 11 mai 2019, accepté après révisions le 24 juin 2019.

Citation : Caillavet, F., Darmon, N., Létaille, F. & Nichèle, V. (2019). Four Decades of Household Food Purchases: Changes in Inequalities of Nutritional Quality in France, 1971-2010. *Economie et Statistique / Economics and Statistics*, 513, 69–89. <https://doi.org/10.24187/ecostat.2019.513.2003>

L'aggravation et/ou la persistance des inégalités est une question d'actualité, pour laquelle l'analyse de long terme peut fournir des éléments de contexte. Plus fréquente dans les domaines du pouvoir d'achat et de l'emploi, la mise en évidence des inégalités socioéconomiques doit être interrogée dans tous les secteurs, y compris alimentaire : les disparités de consommation alimentaire sont en effet socialement emblématiques et peuvent contribuer aux inégalités sociales de santé.

Plusieurs décennies de données sur la consommation alimentaire permettent de mettre en évidence des transformations profondes, qui ont accompagné celles des modes de vie. Des enquêtes de consommation alimentaire existent ainsi en France depuis la fin du 18^e siècle et ont comporté dès le départ des préoccupations d'ordre social (Desrosières, 2003 ; Lhuissier, 2007). Le regard sur cette consommation a également beaucoup évolué, à mesure que les connaissances scientifiques permettaient de comprendre son rôle majeur dans l'incidence de certaines pathologies de santé (GBD, 2019).

Le rôle des facteurs socioéconomiques dans les inégalités nutritionnelles et de santé a été mis en évidence dans plusieurs travaux (Darmon & Drewnowski, 2008 ; Mackenbach *et al.*, 2008), et les disparités de comportements alimentaires selon le revenu, l'éducation ou la catégorie socioprofessionnelle (PCS) sont étudiées depuis longtemps. Elles concernent les dépenses, avec la décroissance de la part budgétaire de l'alimentation à mesure de l'élévation du niveau de vie (la classique loi d'Engel), vérifiée aussi bien en coupe instantanée que dans le temps. En France, la part budgétaire de l'alimentation a ainsi régulièrement baissé, passant d'environ 40 % en 1950 (Sauvy, 1959) à 34 % en 1960, 19.7 % en 1979 (Larochette & Sanchez-Gonzales, 2015) et 15.6 % en 2017 (Ferret & Demoly, 2019). Les ménages les moins aisés y consacrent une part plus élevée que les plus aisés : en 2011, 19 % pour ceux du premier quintile de revenu (celui du bas de la distribution) contre 13.8 % pour le dernier quintile (Accardo *et al.*, 2013). Elle varie aussi selon le niveau d'éducation et la PCS (professions et catégories socioprofessionnelles), la composition des ménages et d'autres caractéristiques sociodémographiques (Caillavet *et al.*, 2009 ; Buron *et al.*, 2014 ; Ferret & Demoly, 2019).

Dans le domaine de la statistique économique, les catégories de recueil de la consommation alimentaire sont structurées selon le lieu de

consommation (au domicile ou en-dehors) et la provenance des aliments (achats, autoconsommation ou autofourniture). Au sein de l'alimentation dans son ensemble, les différences socioéconomiques concernent assez peu la répartition du budget entre grands groupes alimentaires. Elles sont plus sensibles pour la part budgétaire consacrée à l'alimentation hors-domicile ou les quantités consommées au domicile pour certaines catégories d'aliments, comme les produits de la mer ou les fruits et légumes, en particulier frais, qui ont pu apparaître comme marqueurs des inégalités sociales (Caillavet *et al.*, 2009 ; Castetbon, 2014 ; Plessz & Gojard, 2015 ; Bocquier *et al.*, 2015). Ces disparités peuvent conduire à des inégalités sociales nutritionnelles et de santé (Inserm, 2014). Mais pour les analyser, il faut procéder à une analyse précise et détaillée du contenu de l'alimentation. En effet, pour saisir les inégalités découlant des disparités de consommation, il est nécessaire de passer à la dimension nutritionnelle, ce qui oblige à travailler au niveau fin des quantités, même si le coût de l'alimentation est évidemment un facteur déterminant des choix alimentaires (Darmon & Drewnowski, 2015).

Les travaux couvrant une période longue en France sont très rares, notamment pour des questions de disponibilité de données. En effet, le recueil des données de consommation alimentaire procède de deux champs disciplinaires différents. D'un côté, les enquêtes économiques se sont attachées à relever les dépenses alimentaires au sein de l'ensemble des dépenses, dans un souci budgétaire et de connaissance du coût de la vie. C'est la logique des enquêtes *Budget de famille* de l'Insee par exemple. De l'autre, les enquêtes à visée épidémiologique s'intéressent spécifiquement à l'alimentation dans une perspective sanitaire et de santé publique, qui implique de mesurer les quantités consommées par les individus et la qualité nutritionnelle, en mobilisant des tables de composition des produits consommés.

C'est sous cet angle de la qualité nutritionnelle que nous examinons ici l'évolution sur quatre décennies (1971-2010) des disparités de consommation alimentaire en France. Pour cela, il faut disposer de séries de données permettant de calculer des indicateurs de qualité nutritionnelle et d'appréhender des caractéristiques socioéconomiques des ménages. Or la réalisation d'enquêtes représentatives dans ce domaine n'a pas été continue. En particulier, l'enquête *Consommation alimentaire* de l'Insee, qui permet de mesurer les quantités, a été arrêtée

en 1991 ; elle n'a pas été compensée par les enquêtes *Budget de famille*, qui ne relèvent pas les quantités consommées. Un processus d'enquêtes nutritionnelles représentatives au niveau national a été initié à la fin des années 1990, l'Étude individuelle nationale des consommations alimentaires (INCA), réalisée jusqu'ici que tous les sept ans. En outre, des changements de méthodologie empêchent la comparabilité des deux dernières éditions (INCA2 et INCA3). Aussi, pour les années postérieures à 1991, nous mobilisons des données de panel du secteur privé (Kantar Worldpanel, voir annexe 1), afin de construire une continuité – dans la mesure du possible – avec les données de l'enquête de l'Insee sur la consommation alimentaire. Après un important travail de raccordement, nous pouvons calculer des indicateurs de qualité nutritionnelle pour toute la période.

La première section rappelle les grandes évolutions de la consommation alimentaire au niveau international. On examine en section 2, pour la France, le profil de l'évolution générale de la consommation alimentaire, en termes d'apports caloriques et de qualité nutritionnelle, puis, en section 3, l'évolution des disparités selon le revenu et le niveau d'éducation.

1. Le contexte international : transformations de la consommation alimentaire et qualité nutritionnelle

L'évolution de la consommation alimentaire sur le long terme a fait l'objet de nombreuses études dans plusieurs pays, en se fondant principalement sur les données de bilans alimentaires de la FAO. Celles-ci fournissent une estimation assez imprécise de la consommation alimentaire à partir des données de production corrigées par celles du commerce extérieur. Il s'agit en fait d'une mesure des disponibilités alimentaires. En revanche, l'apport des données de source microéconomique, relevées directement auprès des individus ou des ménages, est moins fréquent mais beaucoup plus précis. Nous nous concentrons ici sur ces travaux. On dégage tout d'abord les tendances globales afin d'y situer l'analyse des inégalités relatives.

On relève deux principales caractéristiques de l'évolution de la consommation d'aliments. Tout d'abord, sur le plan des grands groupes alimentaires, dans un contexte global de transition nutritionnelle, les aliments d'origine animale ont progressivement remplacé les aliments d'origine végétale (Drewnowski & Popkin,

2014). Cependant, des schémas plus complexes peuvent être observés à un niveau plus fin. Par exemple, Marques-Vidal *et al.* (2006) constatent au Portugal une diminution de la consommation de poisson et de fruits et un accroissement de la viande, du lait et des féculents entre 1987 et 1999. Aux États-Unis, on relève des quantités consommées décroissantes de lait, porc et boeuf et croissantes de snacks salés, pizzas et sodas entre 1977 et 1996 (Nielsen *et al.*, 2002).

Ensuite, au niveau du degré de transformation des aliments, on assiste à la forte croissance des aliments transformés et à la diminution des aliments bruts. Cette tendance est décrite pour les États-Unis en 1977-1996 (Nielsen *et al.*, 2002), pour le Canada de 1938 à 2011 (Moubarac *et al.*, 2014), le Brésil de 1987 à 2003 (Monteiro *et al.*, 2013), et la Suède entre 1960 et 2010 (Juul & Hemmingson, 2015). Elle est observée aussi en France entre 1969 et 2001 (Nichèle *et al.*, 2008).

Cependant, ces fortes transformations de la structure de consommation ne peuvent être jugées favorables ou défavorables qu'à l'aune d'une analyse plus poussée. En prolongement de ces analyses au niveau des groupes d'aliments, se pose la question de l'évolution en termes nutritionnels de la consommation. Dans un contexte de développement de maladies chroniques liées à la nutrition, telles que les maladies cardio-vasculaires, l'obésité, certaines formes de cancers, il est important d'évaluer l'impact des changements de structure de consommation sur le contenu nutritionnel de l'alimentation, en termes de quantités (apports caloriques) et aussi de qualité. Ainsi des travaux relient la progression du poids corporel à l'augmentation de la disponibilité calorique et aux apports caloriques (Dave *et al.*, 2016). Aux États-Unis, les apports caloriques ont augmenté entre 1977 et 1996 (Nielsen *et al.*, 2002), alors qu'ils sont restés relativement stables en France de 1999 à 2007 (Lioret *et al.*, 2010), en Espagne entre 2000 et 2005 (Valdés *et al.*, 2009) et en Suisse entre 1993 et 2006 (Marques-Vidal *et al.*, 2015). Notre étude antérieure sur la France (Nichèle *et al.*, 2008) constatait des apports caloriques en diminution entre 1969 et 2001 à partir des achats pour la consommation au domicile.

En termes d'évolution de la qualité nutritionnelle de la consommation des ménages au cours des décennies passées, les conclusions varient dans la littérature. Il convient tout d'abord de souligner les difficultés de comparaison de ces études dont la méthodologie est très variable, aussi bien

au niveau de l'utilisation de différents scores de qualité nutritionnelle que de la population de référence, ou du champ de la consommation. Des études constatent une amélioration globale de 1990 à 2010 pour les pays à niveau de vie élevé (Imamura *et al.*, 2015), pour les États-Unis entre 1965 et 1996 (Popkin *et al.*, 2003), entre 1989 et 2008 (Beatty *et al.*, 2014) et entre 1999 et 2010 (Wang *et al.*, 2014), pour l'Australie entre 1992 et 2007 (Arabshahi *et al.*, 2011). En revanche, des travaux sur des pays méditerranéens décrivent une baisse de la qualité nutritionnelle : au Portugal entre 1990 et 2000 (Rodrigues *et al.*, 2008) et en Espagne entre 1987 et 2005 (Bach-Faig *et al.*, 2010) et entre 2000 et 2005 (Valdés *et al.*, 2009).

En ce qui concerne les inégalités socioéconomiques, peu d'études permettent d'en connaître l'évolution. En effet les séries longitudinales prenant en compte le statut socioéconomique sont rares, ou portent sur des décennies récentes. La plupart des travaux disponibles ont été réalisés sur des données des États-Unis. Parmi eux, une seule étude fait état d'une amélioration de la qualité nutritionnelle selon le niveau d'éducation entre 1965 et 1996, mais sans différences significatives selon le revenu (Popkin *et al.*, 2003). En revanche, plusieurs des travaux nord-américains constatent des inégalités persistantes concernant la densité énergétique de l'alimentation (c'est-à-dire la teneur en calories dans 100 g consommés) et le risque d'obésité selon l'éducation et le revenu entre 1997 et 2002 (Kant & Graubard, 2007). Ces travaux mettent également en avant une aggravation des inégalités sociales de qualité nutritionnelle entre 1999 et 2010 (Sugiyama *et al.*, 2014 ; Wang *et al.*, 2014) et des disparités dans les apports nutritionnels selon l'ethnicité, l'éducation, et le revenu entre 1999 et 2012 (Rehm *et al.*, 2016). Ici aussi, la variété des indicateurs nutritionnels et des périodes considérées peut sans doute expliquer en partie la diversité des résultats pour un même pays. Une étude australienne constate une meilleure qualité nutritionnelle associée à un meilleur statut socioprofessionnel chez les hommes entre 1992 et 2007, mais pas d'association avec le niveau d'éducation, ni chez les femmes (Arabshahi *et al.*, 2011). Les travaux dans un contexte européen sont plus rares. Au Danemark, Groth *et al.* (2014) constatent quelques améliorations nutritionnelles entre 1995 et 2008 pour tous les niveaux d'éducation. En Finlande, Prättälä *et al.* (1992) observent une diminution des écarts dans l'observation des recommandations nutritionnelles entre 1979 et 1990 selon le niveau d'éducation.

Pour la France, on relève de très rares travaux, portant sur une période réduite. Ainsi, une analyse des apports nutritionnels des enfants n'observe pas d'évolution, à huit ans d'intervalle, des disparités selon le niveau d'éducation du parent (Lioret *et al.*, 2010). L'association est à chaque fois positive entre le niveau d'éducation du parent enquêté et la consommation d'aliments reconnus favorables à la santé tels que les fruits et légumes, et négative entre éducation du parent et apports de féculents, snacks, sucre et confiserie.

2. En France, une profonde transformation de la structure de la consommation entre 1969 et 2010 et l'amélioration de la qualité nutritionnelle des achats

Pour disposer sur le long terme des données permettant le calcul et le suivi d'indicateurs de qualité nutritionnelle depuis les années 1970, un travail important de constitution de séries a été nécessaire. Nous nous appuyons sur les enquêtes des quatre dernières décennies (1969-2010) qui à la fois fournissent l'information permettant de mesurer la qualité nutritionnelle des achats alimentaires des ménages, enregistrent des caractéristiques socioéconomiques, et dont le champ et la méthodologie pouvaient être rapprochés : les enquêtes *Consommation alimentaire* de l'Insee depuis 1969 (ensuite « série Insee ») malheureusement abandonnées en 1991, et les données de Kantar Worldpanel (ensuite « séries Kantar »). Ces deux sources, leur méthodologie, et le travail de raccordement, sont détaillés dans l'annexe 1. Notons que les données ainsi constituées, si elles fournissent des données permettant de mesurer les inégalités, ne permettent pas d'analyse en termes de causalité¹.

Le champ commun de ces deux sources couvre les achats alimentaires pour le domicile. Il exclut donc non seulement la restauration hors-domicile, mais aussi l'autoconsommation. Dans ces deux domaines, on manque de données fiables de long terme sur lesquelles baser des indicateurs de qualité nutritionnelle.

Pour l'alimentation hors-domicile, les données sont éparpillées, mais (bien que non comparables

1. Les données Kantar Worldpanel ont été partiellement utilisées (1998-2010) dans une modélisation économétrique des inégalités liées à la mise en place de mesures fiscales sur l'alimentation (Caillavet *et al.*, 2016, 2019).

entre les sources) indiquent toutes une forte augmentation : en termes monétaires, sur la base des comptes nationaux, elle représentait 14 % des dépenses alimentaires des ménages en 1960 et 26 % en 2014 (Larochette & Sanchez-Gonzales, 2015). À partir de l'enquête *Budget de famille*, elle est évaluée à 21.9 % du budget alimentaire des ménages en 2001, 23 % en 2006 (Caillavet *et al.*, 2009) et 25 % en 2011 (Buron *et al.*, 2014). Sur la base des enquêtes *Consommation alimentaire*, on constate également une hausse notable, la part budgétaire consacrée à ce poste passant de 13.7 % en 1980 à 17.9 % en 1991 (Manon, 1993). Mais l'enregistrement continu du nombre de repas pris hors-domicile (tout comme celui des quantités consommées) n'existe plus depuis l'arrêt des enquêtes *Consommation alimentaire*. Concernant l'auto-consommation, le manque de données au niveau national est également critique depuis 1991. Si, jusqu'aux années 1990, on voyait baisser cette source d'approvisionnement au profit des produits du commerce, les études sociologiques suggèrent un renouveau depuis plusieurs années. De même que l'alimentation hors-domicile, ce phénomène n'est pas neutre sous l'angle des disparités sociales : le hors-domicile représente une part croissante du budget à mesure que le revenu s'élève et l'auto-consommation est davantage associée à de faibles niveaux d'éducation et certaines PCS (Caillavet *et al.*, 2009). Malheureusement, les données dont nous disposons sont trop parcellaires pour faire des hypothèses raisonnables qui nous permettraient de compléter les données d'achats pour l'alimentation à domicile sur longue période.

2.1. Évolution du contenu calorique des achats pour la consommation à domicile

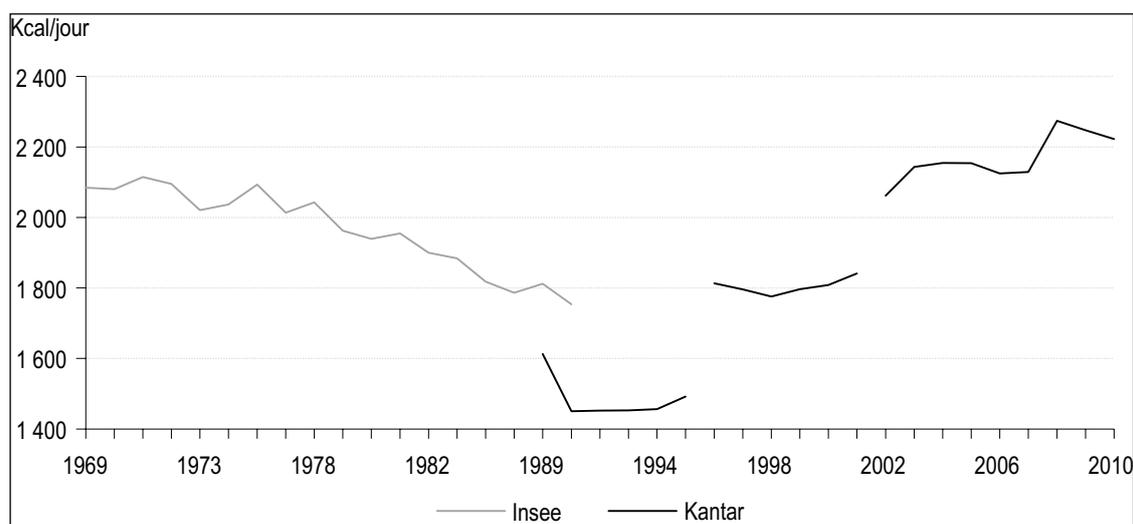
Nous commençons par examiner l'évolution du contenu calorique des achats. Pour cela, les achats sont convertis en calories en utilisant une table de composition (voir encadré). Ici et dans toute la suite, on ne considère que les apports sans alcool.

Sur l'ensemble de la période, nous constatons une augmentation du contenu calorique des achats, avec des moyennes par personne par jour de 2 084 kcalories en 1969 et de 2 222 kcalories en 2010 (figure I).

Cette augmentation générale recouvre en fait une première phase de décroissance importante entre 1969 et 1991 (série Insee), des résultats « décrochés » jusqu'à 1995 (première série Kantar) puis stationnaires de 1996 à 2001 (deuxième série Kantar), suivis d'une phase de croissance modérée entre 2002 et 2009 (troisième série Kantar). Ainsi nos deux sources (Insee et Kantar) et les différents segments au sein de Kantar montrent une différence de niveau calorique et de tendance, que l'on peut relier aux importantes disparités méthodologiques d'échantillonnage et de recueil des données. Celles-ci sont détaillées dans l'annexe 1.

Soulignons en particulier que les changements de méthodologie au sein des séries Kantar (mode de collecte, composition de la population couverte, taille de l'échantillon, nomenclature des produits

Figure I – Contenu calorique des achats pour l'alimentation à domicile en moyenne par personne par jour



Note : consommation à domicile, hors alcool.

Source : Insee, enquêtes *Consommation alimentaire* 1969-1991 ; panels de consommateurs Kantar WorldPanel 1989-2010.

alimentaires – voir annexe 1) ont été multiples au cours de la période et rendent délicate une analyse de long terme. On note un écart important entre la source Insee et la première série Kantar (1989-1995), celle-ci étant de niveau inférieur. La deuxième série Kantar se situe au même niveau que les chiffres Insee et la troisième série marque une augmentation. En ce qui concerne la première série Kantar, elle ne couvre pas encore les achats des ménages constitués d'un homme seul, et certains produits ne sont pas relevés. Ces deux facteurs minorent sans doute les apports caloriques. Dans la série suivante, le passage à un mode de recueil des données à l'aide d'une « scannette » en 1996 a favorisé des oublis d'enregistrement affectant les produits sans code-barres (notamment des produits frais tels que viande, poisson, fruits et légumes). Ces changements de méthodologie affectent différemment les divers groupes alimentaires, donc leur équilibre respectif dans les achats et l'évaluation de la qualité nutritionnelle. Pour cette raison, les évolutions sont présentées sans relier entre elles les différentes séries de données, y compris au sein d'une même source (en l'occurrence les trois séries Kantar, voir annexe 1). La diminution des apports caloriques observée au cours de la première phase (série Insee) pourrait refléter un déplacement vers le hors-domicile : le nombre de repas pris hors-domicile est en effet passé de 1.9 par personne par semaine en 1967 (Villeneuve & Bigata, 1975) à 2.8 en 1991 (Manon, 1993). On ne dispose pas de données similaires pour la deuxième phase correspondant à la période des données Kantar. Sur la base des deux enquêtes nationales représentatives INCA enregistrant l'ensemble des prises alimentaires au niveau individuel, les apports caloriques hors-domicile restent stables : 19 % de l'apport calorique total en 1999 vs 20 % en 2006-2007 (AFSSA, 2009).

2.2. Évolution de la structure des apports caloriques par grands groupes d'aliments

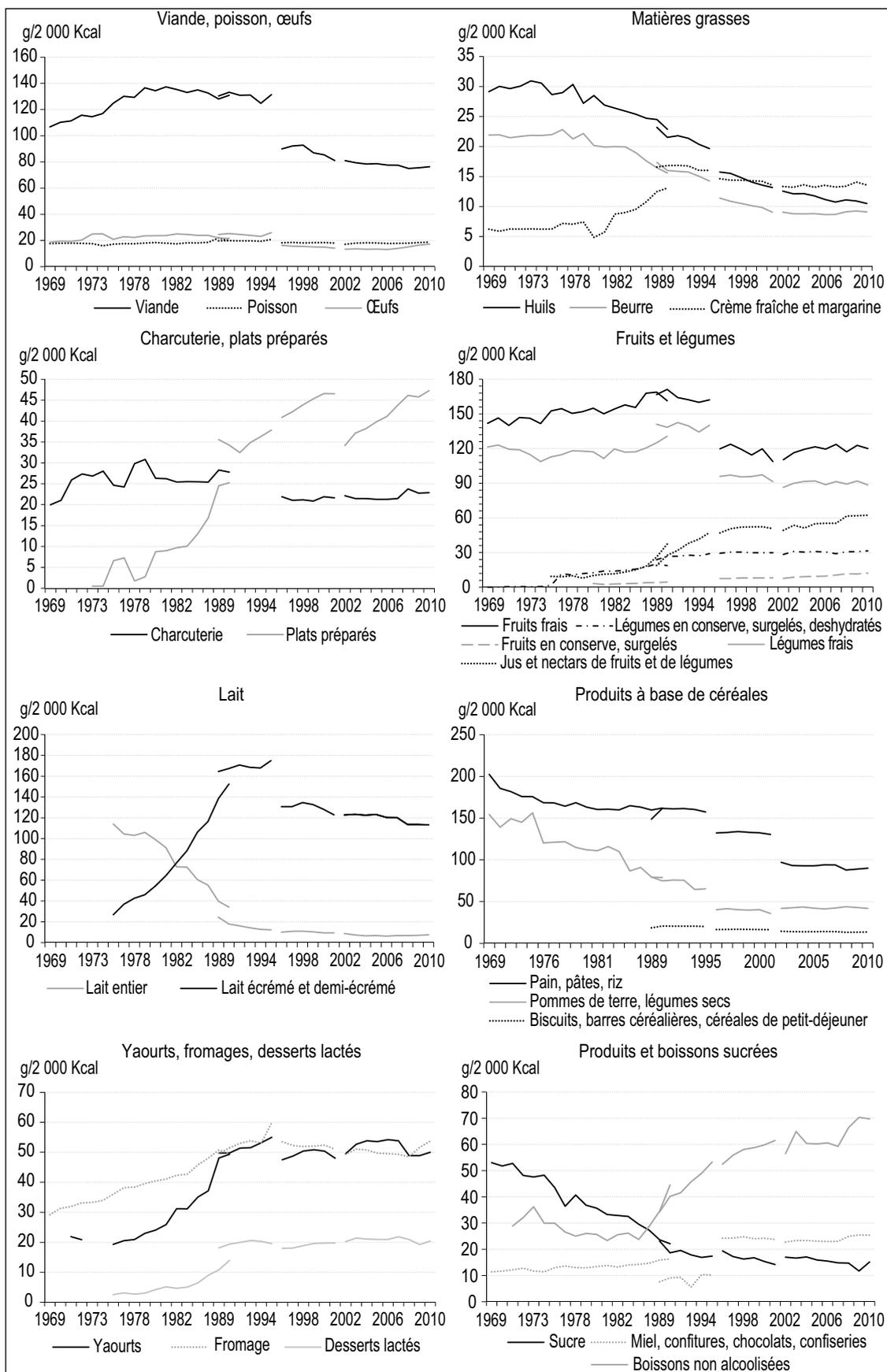
Les denrées fraîches ont toujours dominé les achats dans les groupes/catégories viande, volaille, fruits, légumes (Nichèle *et al.*, 2008). Cependant les transformations de la structure des achats alimentaires s'avèrent importantes pour tous les grands groupes d'aliments, dans une logique globale de redistribution entre produits bruts et produits transformés. La part des catégories d'aliments traditionnels (pain/pâtes/riz, viandes fraîches, pommes de terre/légumineuses, sucre, beurre, huiles, lait entier) a fortement diminué, et celle des catégories

transformées (produits laitiers, plats préparés, sodas, confiserie) a augmenté. Les substitutions s'opèrent également au sein des catégories : ainsi les légumes transformés augmentent aux dépens des légumes frais, les jus aux dépens des fruits frais, les plats préparés aux dépens des viandes fraîches. Et ce au-delà de la sous-déclaration des achats de produits frais (sans code-barres) qu'a pu provoquer pendant quelques années le passage à la « scannette ». À un niveau plus fin, la redistribution entre les formes de lait selon leur teneur en matières grasses est frappante : le lait entier, qui constituait le principal achat de lait jusqu'en 1981 (58.6 %), ne représente plus que 6.2 % des achats en 2010, remplacé par des laits écrémé et demi-écrémé. Globalement, dans la mesure où le contenu calorique des achats est supérieur en 2010 par rapport à 1969, la croissance des achats de produits transformés a plus que compensé la diminution des produits bruts, du moins à partir de 2002, lorsque le contenu calorique des achats pour le domicile augmente clairement (cf. figure I).

L'évolution de la part des différents groupes d'aliments dans les achats ne rend toutefois pas directement compte de l'évolution de la structure des apports caloriques de l'alimentation, dans la mesure où les apports caloriques de l'ensemble des achats varient au cours du temps. Pour nous affranchir des variations du contenu calorique global (de l'ensemble des achats) et analyser l'évolution de la structure des apports caloriques et de la qualité nutritionnelle, les achats des différents groupes d'aliments sont, dans la suite de l'article, exprimés pour 2 000 kcal, valeur de référence correspondant au besoin énergétique journalier d'un individu moyen. Rappelons que l'on ne peut pas tenir compte ici d'éventuels changements au cours de la période du contenu des aliments en calories ou en nutriments.

La figure II présente l'évolution du contenu calorique, exprimé en grammes/2 000 kcal, par catégorie de produits. On observe sur la période la décroissance des viandes fraîches (de 106.8 à 76.4 g/2 000 kcal), alors que la charcuterie progresse (de 20.0 à 22.9 g/2 000 kcal) et les plats préparés sont en forte augmentation entre 1973 et 2010 (de 0.5 à 47.3 g/2 000 kcal). Une tendance semblable s'observe aussi pour les fruits et légumes, avec la diminution des achats de frais (respectivement de 141.9 à 120.0 g/2 000 kcal pour les fruits et de 121.5 à 88.6 g/2 000 kcal pour les légumes), alors que celle des jus d'une part (de 9.5 en 1976 à 62.3 g/2 000 kcal) et des conserves et surgelés d'autre part progressent (respectivement de

Figure II – Achats de divers groupes et sous-groupes de produits, en grammes par 2 000 kcal



Source : Insee, enquêtes *Consommation alimentaire* 1969-1991 ; panels de consommateurs Kantar WorldPanel 1989-2010.

ENCADRÉ – Des achats alimentaires aux mesures nutritionnelles

Pour mesurer le contenu calorique et nutritionnel des achats alimentaires, nous avons utilisé une matrice construite à partir de la table de composition nutritionnelle REGAL, issue du Répertoire Général des ALiments du Ciqual (Centre d'information sur la qualité des aliments) de l'année 1992. Celle-ci permet de convertir les 314 produits alimentaires définis dans la nomenclature analytique de l'enquête Insee en leur contenu calorique et nutritionnel. On peut ainsi mesurer l'évolution des substitutions entre catégories de notre nomenclature alimentaire (mais pas tenir compte de la possible évolution du contenu nutritionnel de ces catégories sur la période).

La valeur énergétique et les teneurs en nutriments (macronutriments, micronutriments, vitamines et minéraux, fibres et cholestérol) se rapportent à 100 g de la partie comestible de l'aliment en question. Un facteur de conversion propre à chaque aliment, appelé fraction comestible, est utilisé pour passer du poids de l'aliment tel qu'acheté en poids d'aliment comestible.

Contenu calorique et nutritionnel

Les contenus caloriques et nutritionnels des achats alimentaires sont exprimés en moyenne par personne et par jour. L'alcool, l'eau, les boissons diététiques, le thé et le café sont exclus. En pratique, on calcule :

$$\begin{aligned} \text{Energie}_{jh} &= q_{jh} (nrj_i/100) \\ \text{Nutriment}_{jh} &= q_{jh} (\text{nutri}_j/100) \end{aligned}$$

où q_{jh} est la quantité achetée et transformée en quantité consommable, du produit alimentaire i par le ménage h pendant une période d'observation (semaine d'enquête pour la série Insee, ensemble des semaines d'activité pour les séries Kantar) et nrj_i et nutri_j sont respectivement la teneur en énergie et en un nutriment j pour 100 g du produit alimentaire i consommable.

Qualité nutritionnelle

La qualité nutritionnelle des achats de produits alimentaires est estimée par le *Mean Adequacy Ratio* (MAR). Il s'agit d'un indicateur composite calculé comme le

pourcentage moyen d'apports nutritionnels conseillés pour quinze nutriments calculé selon :

$$\begin{aligned} \text{MAR} &= \frac{\sum_{i=1}^{15} \text{ratio}_i}{15} * 100 \\ \text{ratio}_i &= \begin{cases} 1 & \text{si } N_i \geq \text{ANC}_i \\ \frac{N_i}{\text{ANC}_i} & \text{sinon} \end{cases} \end{aligned}$$

où N_i est l'apport journalier observé pour le nutriment i et ANC_i est l'apport nutritionnel conseillé pour le nutriment i . Les 15 nutriments pris en compte sont les protéines, les fibres, le rétinol (vit A), la thiamine (vit B1), la riboflavine (vit B2), la niacine (vit B3), la vitamine B6, les folates (vit B9), les vitamines C, E et D, le calcium, le fer, le magnésium et le potassium.

Les scores sont ramenés à un standard calorique, afin que la disparité du contenu calorique des achats n'affecte pas les différences de qualité nutritionnelle.

Calcul des déviations à la moyenne

Pour chaque produit alimentaire i et chaque quartile de revenu (respectivement du niveau d'éducation) Niv , nous avons calculé le pourcentage de déviation D_{iNiv} par rapport à la moyenne nationale des achats par personne et par an^(a) comme suit :

$$D_{iNiv} = \left(\frac{q_{iNiv}}{q_i} - 1 \right) * 100$$

où q_i est la quantité moyenne de produit alimentaire i achetée par personne et par an^(a) par l'ensemble des ménages et q_{iNiv} est la quantité moyenne de produit alimentaire i achetée par personne et par an^(a) par les ménages du quartile de revenu (respectivement du niveau d'éducation) Niv .

(a) Pendant la semaine de l'enquête pour la série Insee, pendant l'ensemble des semaines d'activité pour les séries Kantar, voir annexe 1.

3.2 à 12.3 g/2 000 kcal pour les fruits et de 0.1 à 31.5 g/2 000 kcal pour les légumes). L'écart observé entre produits bruts et transformés est sans doute accru ponctuellement par le passage à la « scannette » en 1996 qui favorise les oublis de déclaration de produits frais.

2.3. L'évolution de la qualité nutritionnelle

On enregistre une diminution importante depuis 1969 des achats de produits à base de

céréales, pain/pâtes/riz, et de pommes de terre et légumineuses. Les achats de pain frais ne sont pas enregistrés dans Kantar, à la différence du pain pré-emballé, aussi nous avons imputé la quantité achetée dans la dernière année de la série Insee à toutes les années Kantar (cf. annexe 1), ce qui entraîne une diminution *a minima*. Enfin, les achats de sucre en tant qu'aliment ont été divisés par presque 3.5 entre 1969 et 2010 alors que ceux de produits sucrés tels que confitures, chocolats, confiserie ont plus que doublé. Parallèlement, la croissance

des boissons sucrées sans alcool est forte depuis 1987. Pour les produits laitiers, on assiste à une augmentation globale sur la période, plus prononcée entre 1970 et 1990. Fromages et notamment yaourts et desserts lactés sont en progression (respectivement de 29.2 à 53.7 g/2 000 kcal et de 19.3 à 50.0 g/2 000 kcal). Présentes dans les statistiques à partir de 1976, les formes allégées de lait (écrémé et demi-écrémé) se développent fortement (de 26.8 à 113.2 g/2 000 kcal) aux dépens du lait entier (de 114.0 à 7.5 g/2 000 kcal). Enfin, les achats de matières grasses, notamment d'huiles et de beurre, ont fortement diminué depuis 1969. Dans cette catégorie, seules la margarine et la crème fraîche ont augmenté².

Ces variations importantes au niveau des groupes et sous-groupes de produits et en leur sein s'accompagnent d'une évolution globale de la qualité nutritionnelle. Celle-ci est mesurée par un score d'adéquation aux recommandations nutritionnelles, appelé MAR (*Mean Adequacy Ratio*, cf. encadré). Ce score augmente globalement sur la période (figure III). Cela traduit une amélioration de la qualité nutritionnelle, qui peut s'expliquer notamment par la diminution de produits denses en calories et pauvres en nutriments, comme le sucre ou les matières grasses. Cependant on

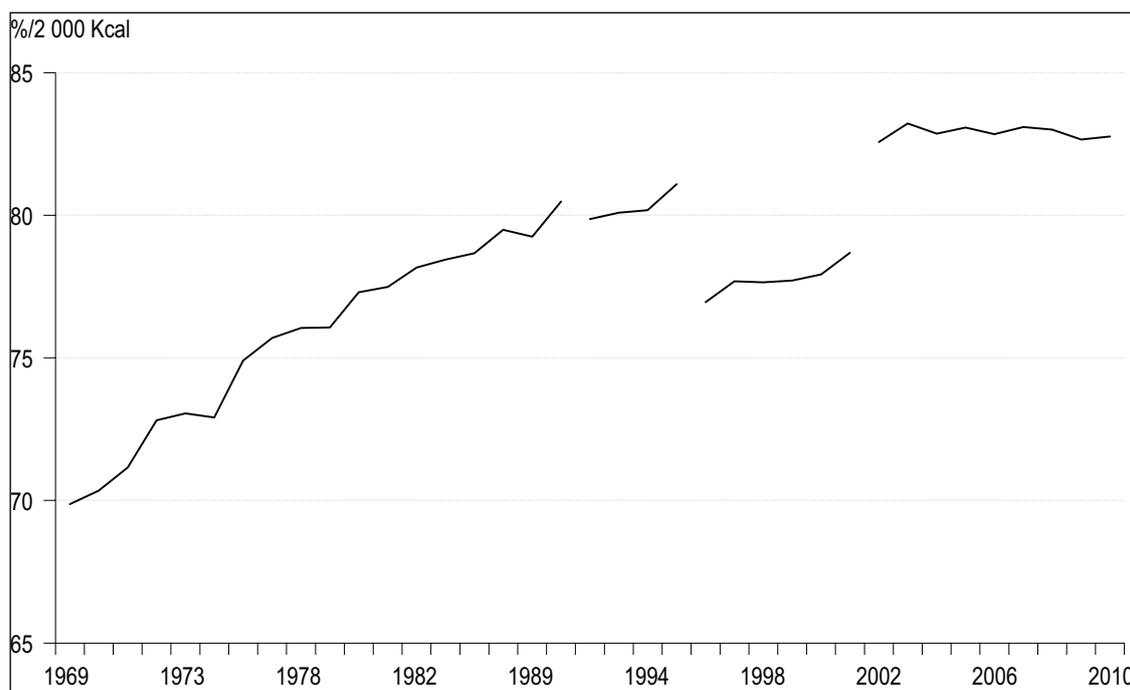
observe, dans l'évolution générale du MAR, une première phase de croissance soutenue jusqu'en 2003³ (observée dans la série Insee jusqu'à 1991, puis la première série Kantar jusqu'en 1995, puis au début de la troisième série Kantar en 2002 et 2003), suivie par une deuxième phase de stagnation depuis cette date.

Le décrochage entre 1996 et 2002 correspond à une modification de méthodologie de la série Kantar, qui minore notamment les enregistrements de fruits et légumes, viande et poisson sans code-barres (produits frais). Or ces aliments sont sources de nutriments entrant de façon positive dans la construction du score nutritionnel (cf. encadré et annexe 1).

Notons qu'en dehors de cette période où le changement de méthodologie est majeur, les différences de niveau constatées au niveau calorique entre les séries Insee et Kantar ne semblent pas affecter la cohérence et la continuité du score nutritionnel.

2. Le commentaire détaillé pour toutes les catégories de produits des figures I et II peut être trouvé dans Caillavet et al. (2018).
3. Le décrochage entre 1996 et 2002 correspond à une modification de méthodologie de la série Kantar Worldpanel (voir annexe 1).

Figure III – Évolution du score nutritionnel MAR des achats pour l'alimentation à domicile



Source : Insee, enquêtes *Consommation alimentaire* 1969-1991 ; panels de consommateurs Kantar WorldPanel 1989-2010.

3. Les inégalités socioéconomiques de qualité nutritionnelle de l'alimentation s'amenuisent mais restent notables

Les inégalités socioéconomiques sont appréhendées à partir de deux variables construites au niveau ménage : le quartile de revenu par équivalent-adulte et le niveau d'éducation. Le revenu du ménage par équivalent-adulte est calculé avec l'échelle d'équivalence actuellement employée par l'Insee⁴. Notons qu'il ne s'agit pas ici d'une mesure du niveau de vie, car le calcul porte sur l'ensemble du revenu et non le revenu disponible. En outre, l'information dont on dispose est d'origine déclarative, or son caractère approximatif et le biais de sous-déclaration sont bien connus ; nous espérons en minimiser les conséquences en considérant les quartiles de revenus.

Le niveau d'éducation est mesuré par le niveau d'études le plus élevé atteint par la personne de référence du ménage ; quatre niveaux sont distingués : études primaires, études secondaires, baccalauréat et études supérieures. On ne dispose d'information sur l'éducation qu'à partir de 1978. À la différence des quartiles de revenu, critère relatif, le niveau d'éducation est un critère absolu. Ainsi, du fait de l'augmentation du niveau général d'éducation au cours de la période, le groupe le plus faible s'amenuise au cours du temps et le niveau le plus élevé s'accroît. La complémentarité de mesures du statut socio-économique à partir d'indicateurs monétaires et non monétaires a été soulignée dans plusieurs études (Galobardes *et al.*, 2007 ; Lalluka, 2007). On sait que le revenu et l'éducation sont liés et interagissent (ainsi qu'avec des caractéristiques démographiques, notamment l'âge, la composition du ménage, la profession, la localisation géographique, etc.). Nos données ne permettent toutefois pas de prendre en compte ces interactions, et nous nous en tenons à une présentation séparée des évolutions par niveau de revenu et par niveau d'éducation.

Pour étudier la variation des inégalités indépendamment de la variation globale de la qualité nutritionnelle, nous calculons, pour chaque quartile de revenu et chaque niveau d'éducation, le pourcentage de déviation par rapport à la moyenne nationale du contenu énergétique puis du MAR, en prenant en compte les achats de chaque produit alimentaire (voir encadré). Les résultats sont présentés pour cinq points de la période : 1971, 1981, 1991, 2002, 2010, soit environ tous les dix ans. Le choix de ces dates

permet de minimiser les disparités de méthodologie puisque nous mobilisons ainsi seulement deux des quatre séries. Les trois premières dates de la période sont issues de la série Insee (1971, 1981, 1991) et les deux dernières (2002 et 2010) de la troisième série Kantar.

Les inégalités de qualité nutritionnelle sont apparentes aussi bien en début qu'en fin de période, et pour les deux variables socioéconomiques considérées : le MAR est toujours le plus élevé pour les ménages dont le niveau de revenu ou dont le niveau d'éducation est le plus élevé (figure IV).

Très marquées en début de période, les différences des caractéristiques énergétiques et nutritionnelles des achats entre niveaux de revenu et entre niveaux d'éducation ont tendance à s'amenuiser mais certaines inégalités persistent. Nous les examinons maintenant plus en détail.

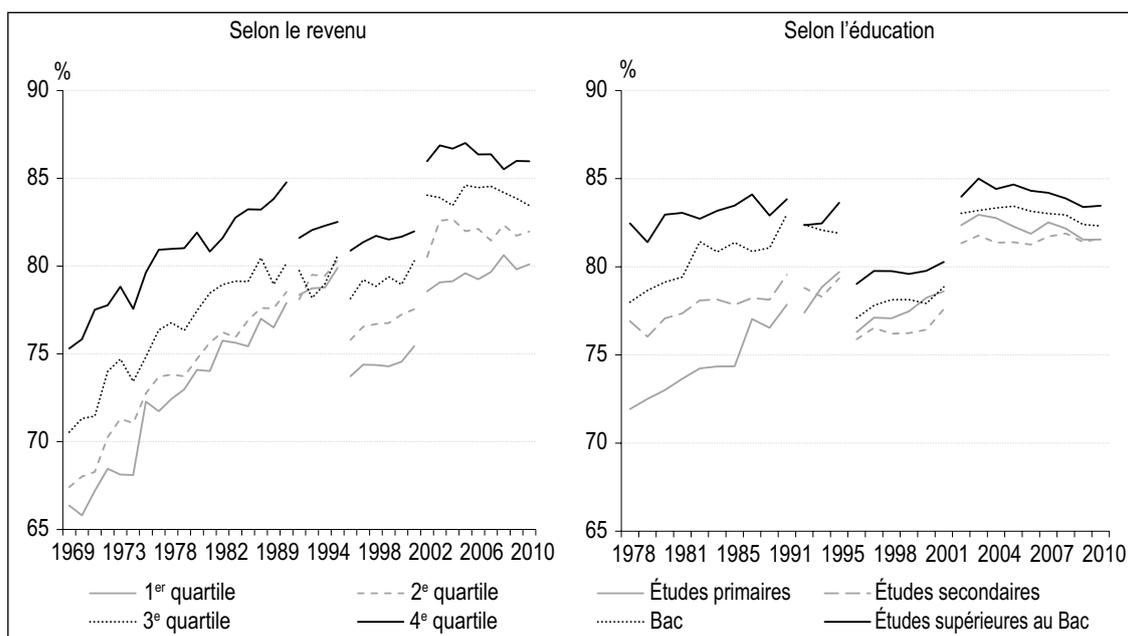
3.1. Selon le revenu : inégalité de la qualité nutritionnelle plus que du contenu calorique

Sur le plan nutritionnel, il existe une hiérarchie entre les aliments en fonction de leur rapport coût/qualité nutritionnelle : les aliments denses en énergie, généralement riches en sucres et en graisses, sont des sources de calories bon marché mais contribuent peu aux apports en micronutriments protecteurs et sont donc moins favorables à la santé. Cette hiérarchie tend à orienter les ménages contraints financièrement vers des aliments moins sains, favorisant ainsi la plus forte prévalence d'obésité et de pathologies associées dans les populations les moins favorisées (Darmon & Drewnowski, 2015). D'où l'importance du contenu calorique de l'alimentation dans la question des inégalités. Mais il ne peut être dissocié d'une analyse qualitative, étant donné qu'on ne connaît pas le nombre de repas auquel correspond ce niveau de calories dans les achats.

Concernant le contenu calorique des achats, on distingue deux phases : une première phase (années 1971, 1981 et 1991 - données Insee), où il se révèle supérieur à la moyenne pour les quartiles de revenu les plus faibles (Q1, Q2 et, dans une moindre mesure, Q3), et inférieur à la moyenne pour le quartile de revenu le plus élevé Q4 (figure V) ; une deuxième phase (2002, 2010 - données Kantar) où la tendance est inversée :

4. Échelle dite OCDE-modifiée, également appliquée pour les statistiques européennes. Pour les données des séries Kantar, dans lesquelles on ne dispose que de revenus par tranche, l'échelle d'équivalence a été appliquée au centre de tranche.

Figure IV – Évolution du score nutritionnel MAR des achats pour l'alimentation à domicile, selon le revenu et selon l'éducation



Source : Insee, enquêtes *Consommation alimentaire* 1969-1991 ; panels de consommateurs Kantar WorldPanel 1989-2010.

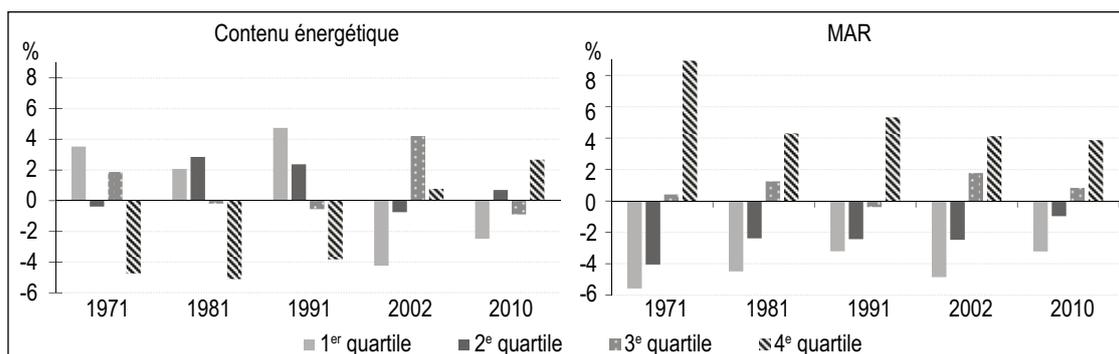
Q1 se trouve au-dessous de la moyenne, alors que Q4 est désormais au-dessus en 2002 comme en 2010. Ces deux phases différentes s'insèrent au sein de la première phase de décroissance du contenu calorique, et dans un deuxième temps au sein d'un contenu calorique supérieur et relativement croissant dans les années 2000. En même temps, comme on l'a vu plus haut, elles correspondent à deux sources de données différentes, dont l'homogénéité n'est garantie qu'au sein de chaque sous-période. Ainsi il s'avère délicat d'interpréter cette inversion de tendance. Dans tous les cas, l'amplitude des déviations par rapport au contenu calorique moyen est relativement faible, et s'amenuise ; ainsi, au niveau de chaque quartile, ces déviations par rapport à la moyenne dépassent à peine 5 % en valeur absolue sur tous les points de la période et restent inférieures à 2.5 % en 2010. Cependant, les quartiles extrêmes révélant des déviations de signe opposé par rapport à la moyenne, on observe un écart entre Q1 et Q4 de 8.3 points en 1971, qui demeure de 5.4 points en 2010.

Si l'on s'attache aux résultats de la première phase (1971-1981-1991, données Insee), le contenu calorique des achats plus élevé parmi les ménages les plus modestes que parmi les plus aisés peut recouvrir plusieurs phénomènes. D'une part, il peut correspondre à une consommation moins fréquente hors-domicile. On a constaté par exemple que les ménages des déciles de revenu

les plus bas consacraient une part inférieure de leur budget alimentaire à la consommation hors-domicile (14 % pour le premier décile de revenu vs 30 % pour le dernier décile, cf. Caillavet *et al.*, 2009). Ce niveau calorique plus élevé des achats au domicile des ménages modestes fait aussi écho au gradient social de surpoids et d'obésité (i.e. la prévalence d'autant plus forte que le statut social est faible). Mais cela peut également refléter des pratiques de sociabilité différente (invitations à domicile vs au restaurant), des contraintes de temps de la journée active (pause méridienne au domicile vs au restaurant ou sur le lieu de travail, cf. Lhuissier *et al.*, 2020), que sous-tendent évidemment les logiques économiques de coût du repas.

En ce qui concerne la légère inversion de tendance apparente à partir de 2002 (troisième série Kantar), elle coïncide avec la stabilisation des apports caloriques de la consommation au domicile aux alentours de 80 % selon les estimations des enquêtes alimentaires individuelles INCA1 (1998-1999) et INCA2 (2006-2007) (cf. AFSSA, 2009), mais on ne dispose pas d'éléments sur la différenciation sociale de ce chiffre. Le contexte général de l'évolution des inégalités de niveau de vie en France semble cependant conforter l'hypothèse de deux phases distinctes : ces inégalités ont baissé continûment dans les années 1970 et ont augmenté dans les années 2000 (Boiron, 2016).

Figure V – Déviation par rapport à la moyenne (%) du contenu énergétique et du MAR des achats pour l'alimentation à domicile, selon le revenu



Note : consommation à domicile hors alcool.

Source : Insee, enquêtes *Consommation alimentaire* 1971-1981-1991 ; panels de consommateurs Kantar WorldPanel 2002-2010.

En termes de qualité nutritionnelle (figure V), l'évolution selon le revenu est sans équivoque : pour tous les points de la période, quelle que soit la source de données, les déviations du MAR sont inférieures à la moyenne pour les quartiles de revenu les plus faibles (Q1 et, dans une moindre mesure, Q2). Parallèlement, le quartile de revenu le plus élevé (Q4) jouit toujours d'une relative meilleure qualité nutritionnelle. Comme pour le niveau de calories, l'écart entre Q1 et Q4 s'amenuise avec le temps : il atteignait 9 points en 1971, il demeure à 6 points en 2010.

3.2. Selon le niveau d'éducation, des inégalités qui tendent à disparaître

Les déviations par rapport à la moyenne des teneurs en calories des achats selon le niveau d'éducation (figure VI) sont plus prononcées qu'elles ne le sont en fonction du revenu. En effet, le niveau d'éducation le plus faible (niveau d'études primaires) montre sur toute la période des teneurs en calories supérieures à la moyenne, les trois autres niveaux d'éducation restant en-dessous. Les déviations par rapport à la moyenne sont fortes puisque l'écart en valeur absolue entre le niveau d'éducation le plus faible et le plus élevé atteint 13 points en 1978 (première année pour laquelle on dispose du niveau d'éducation). En revanche, elles sont près de disparaître en 2010 (troisième série Kantar) avec moins d'1 point d'écart.

Comme pour les niveaux de revenus les plus faibles, le fait que l'apport calorique des achats soit supérieur à la moyenne pour des ménages dont la personne de référence est moins éduquée peut refléter une consommation alimentaire plus fréquente au domicile. S'il n'y a pas de

travaux sur les différences de fréquence de la consommation à domicile selon le niveau d'éducation, des études montrent que les catégories socio-professionnelles élevées telles que les cadres supérieurs mangent plus fréquemment à l'extérieur du foyer (4.6 repas par semaine vs 2.8 en moyenne en 1991, cf. Manon, 1993).

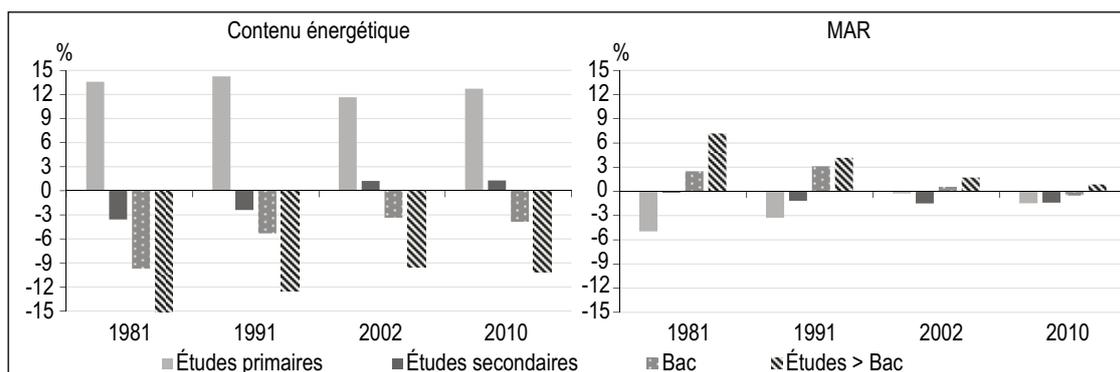
Concernant la qualité nutritionnelle globale des achats (figure VI), on note en début de période des inégalités franches reflétées par d'importantes différences de déviations du MAR par rapport à la moyenne selon l'éducation, à l'instar du revenu. Mais ces inégalités persistantes selon le revenu ne le sont pas selon l'éducation, où elles diminuent tout au long de la période étudiée pour disparaître pratiquement en 2010 (troisième série Kantar).

3.3. Malgré la convergence, des inégalités subsistent

Entre les différents niveaux socioéconomiques, qu'ils soient appréhendés par le revenu ou par le niveau d'éducation, on observe sur les quatre décennies étudiées une convergence de la qualité nutritionnelle globale des achats. Cependant, des disparités persistent pour certaines caractéristiques nutritionnelles.

Tout d'abord le contenu calorique de ces achats reste supérieur pour les ménages les moins éduqués. Le contenu en macronutriments (lipides, glucides, protéines) suit la tendance générale à la convergence avec des déviations par rapport à la moyenne selon le revenu et l'éducation ne dépassant pas 5 % à la fin de la période en 2010 (voir annexe 2). En ce qui concerne les lipides et les glucides, les déviations de leur contribution au contenu calorique des achats restent faibles sur

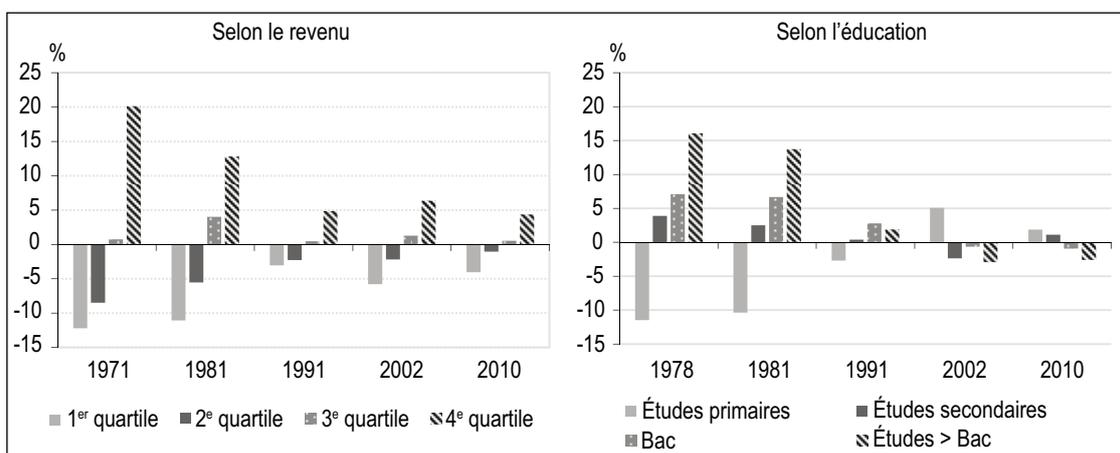
Figure VI – Déviation par rapport à la moyenne (%) du contenu énergétique et du MAR des achats pour l'alimentation à domicile, selon l'éducation



Note : consommation à domicile hors alcool.

Source : Insee, enquêtes *Consommation alimentaire* 1971-1981-1991 ; panels de consommateurs Kantar WorldPanel 2002-2010.

Figure VII – Déviation par rapport à la moyenne (%) du contenu en protéines animales des achats pour l'alimentation à domicile, selon le revenu et selon l'éducation



Source : Insee, enquêtes *Consommation alimentaire* 1971-1981-1991 ; panels de consommateurs Kantar WorldPanel 2002-2010.

toute la période, mais ont tendance à s'inverser dans les années les plus récentes, qui correspondent à la troisième série Kantar. Le contenu en lipides, avec un écart entre les déviations par rapport à la moyenne marqué, demeure un facteur d'inégalité entre niveaux d'éducation (voir annexe 2).

La déviation à la moyenne de la contribution des protéines s'amenuise au cours du temps. En revanche, pour les protéines animales, la déviation par rapport à la moyenne met en évidence de fortes disparités selon le revenu (figure VII) : en début de période, les ménages aux revenus les plus élevés montrent des apports de plus de 20 % supérieurs à la moyenne en 1971, puis cet écart se réduit à 4.4 % en 2010. On note une évolution différente selon le niveau d'éducation, avec en début de période un contenu des achats

en protéines animales beaucoup plus élevé parmi les ménages les plus éduqués que parmi ceux ayant un faible niveau d'éducation (déviations de +16.1 % et -11.5 % respectivement en 1978), soit un écart entre déviations de 27.6 points. Cette tendance semble s'inverser dans la troisième série Kantar à partir de 2002, avec des apports en protéines animales supérieurs pour le plus bas niveau d'éducation (déviation de +5.1 % en 2002) ; mais la déviation devient très faible (+1.9 %) en 2010. Si cette inversion de tendance correspond au changement de série (Kantar), on peut noter cependant que ce résultat n'est pas observé selon le niveau de revenu, pour lequel on observe une tendance continue.

Ainsi, plus que les protéines totales, les sources de protéines apparaissent comme un élément clivant en début de période, ce qui peut avoir

une incidence sur le plan nutritionnel et sur la durabilité de l'alimentation. Le rôle de marqueur social traditionnellement dévolu à la consommation de protéines animales (Grignon, 1996), s'estompe complètement dans la troisième série Kantar. On décèle même une légère inversion de tendance pour les ménages les plus éduqués en 2002⁵. Cette évolution peut refléter l'impact de l'information nutritionnelle sur les effets néfastes d'une surconsommation de certains produits animaux. Des travaux soulignent en effet l'influence différenciée de l'information selon le statut social et, en particulier, la plus grande sensibilité aux recommandations nutritionnelles des individus ayant un statut socioéconomique supérieur (Régnier & Masullo, 2009). Cela contribuerait à expliquer, pour la consommation à domicile, le rapprochement observé entre les ménages ayant les niveaux d'éducation les plus et les moins élevés.

* *
*

Cet article s'appuie sur la constitution d'une base de données originale sur les achats alimentaires pour la consommation au domicile sur quatre décennies, couvrant la période de 1969 à 2010, à partir d'enquêtes directes auprès des ménages. Il n'existe pas de séries chronologiques longues en France sur la consommation alimentaire, et les données mobilisées dans cette étude sont les seules à pouvoir renseigner sur la période au niveau des ménages. La constitution de cette base de données a nécessité un long et minutieux travail de mise en cohérence de deux sources différentes, les enquêtes *Consommation alimentaire* de l'Insee, qui couvrent les années 1969 à 1991, puis les séries Kantar-Worldpanel jusqu'en 2010. Nous parlons à dessein de « mise en cohérence » et non d'harmonisation, car le champ et la méthodologie des sources comportent des différences importantes, y compris au sein même des séries Kantar. C'est pourquoi nous avons choisi de présenter les évolutions sans relier les séries. Les contraintes des données conduisent par ailleurs à adopter un nombre limité d'indicateurs pour analyser les inégalités entre ménages et leur évolution ; nous avons retenu le quartile de revenu par équivalent-adulte, et le niveau d'éducation, approché par celui de la personne de référence du ménage.

Sur la base de ces données, l'article propose une étude descriptive de l'évolution de la structure

de la consommation sous l'angle des apports énergétiques et de la qualité nutritionnelle. L'analyse repose sur le calcul d'un score de qualité nutritionnelle (le MAR) ; ce score est ramené à un standard calorique, afin que la disparité du contenu calorique des achats n'affecte pas l'évaluation de la qualité nutritionnelle. Nos résultats indiquent d'abord une tendance générale à l'amélioration de la qualité nutritionnelle (hausse du MAR). On examine ensuite comment les inégalités de consommation alimentaire, sous cet angle nutritionnel, ont évolué entre les ménages selon leur niveau de revenu et leur niveau d'éducation. Au long des quatre décennies étudiées (1971-2010), les ménages dont le revenu ou le niveau d'éducation est le plus élevé ont constamment bénéficié d'une alimentation au domicile de meilleure qualité nutritionnelle, contribuant ainsi aux inégalités sociales de santé. Cependant, le constat global sur longue période est celui d'une réduction des inégalités observée à la fois dans la série Insee et la troisième série Kantar : les écarts de MAR entre quartiles de revenu extrêmes passent de 9 points en 1971 à 6 points en 2010, alors que les écarts de MAR entre niveaux d'éducation extrêmes passent de 13 points en 1978 à moins d'1 point en 2010. Rappelons ici que l'évolution des inégalités entre niveaux de revenu et niveaux d'éducation ne peut pas s'interpréter de façon identique : en effet, les groupes de revenus (les quartiles) ont une importance relative constante au cours du temps. Il n'en va pas de même des niveaux d'éducation, mesurés en termes « absolus » : du fait de l'élévation du niveau d'éducation dans la population dans la période étudiée, leur poids relatif dans la population a changé de façon très importante – en particulier, la proportion de ménages au niveau le plus faible a été divisée par dix, tandis qu'à l'autre extrême, celle de ménages au niveau le plus élevé a plus que doublé, modifiant ainsi la signification sociale de ces inégalités. De fait, la période étudiée et le recours à des sources de méthodologies différentes répondent à la vacance statistique de l'appareil national, alors qu'une analyse en termes de cohortes combinant âge et éducation permettraient de prendre les effets générationnels en compte.

Rappelons pour finir que les évolutions observées ne concernent que les achats alimentaires pour le domicile connus à travers des enquêtes auprès des ménages. Ne sont donc pas pris en

5. Cela n'est pas forcément le cas dans tous les pays, par exemple dans un pays comme le Danemark où la consommation de viande reste plus importante pour les ménages plus favorisés à la même période (Smed & Jensen, 2007).

compte d'une part l'autoconsommation et le hors-domicile, d'autre part les populations non prises en compte dans les enquêtes auprès des ménages – celles des résidences collectives et celles sans logement. D'autres spécificités des achats alimentaires comportent des enjeux de santé publique et d'inégalités sociales : accès aux lieux d'achat dans la mesure où l'offre alimentaire peut déterminer les univers de choix du consommateur (Chaix *et al.*, 2012 ; Drewnowski *et al.*, 2014 ; Caillavet *et al.*, 2015), ou aux caractéristiques de qualité des produits, les prix supérieurs rendant les produits avec labels de qualité ou biologiques moins accessibles pour les ménages aux revenus plus faibles (Marette *et al.*, 2012 ; UFC, 2017). Mais on dispose ici encore moins de données de long terme. Sur la base de la structure des achats alimentaires pour la consommation à domicile nos résultats montrent qu'en 2010, les inégalités de qualité

nutritionnelle selon le niveau d'éducation semblent près de disparaître, mais les inégalités selon le revenu subsistent. On pourrait espérer que la stagnation de la qualité nutritionnelle des achats constatée au début du 21^e siècle, la dernière période de notre analyse, laisse la place à une amélioration. En effet des politiques d'information officielles ont été mises en place (Plan national nutrition santé, PNNS) avec à la fois des campagnes générales et des signaux au niveau des produits (le « NutriScore »). Les consommateurs semblent sensibilisés à la qualité nutritionnelle de leur alimentation, comme le montre le succès croissant des outils digitaux (Ifop, 2019). Des travaux récents suggèrent cependant que leur mode d'utilisation reproduit les inégalités sociales d'accès à l'information (Régnier & Chauvel, 2018). Politiques alimentaires et de santé restent donc d'actualité pour enrayer les inégalités sociales. □

BIBLIOGRAPHIE

Accardo, J., Kranklader, E. & Place, D. (2013). Les comportements de consommation en 2011. *Insee Première* N° 1458. <https://www.insee.fr/fr/statistiques/1281036>

AFSSA (2009). *Étude Individuelle Nationale des Consommations Alimentaires 2 (INCA2) (2006-2007)*. <https://www.anses.fr/fr/system/files/PASER-Ra-INCA2.pdf>

Arabshahi, S., Lahmann, P. H., Williams, G. M., Marks, G. C. & van der Pols, J. C. (2011). Longitudinal Change in Diet Quality in Australian Adults Varies by Demographic, Socio-Economic and Lifestyle Characteristics. *The Journal of Nutrition*, 141(10), 1871–1879. <https://doi.org/10.3945/jn.111.140822>

Bach-Faig, A., Fuentes-Bol, C., Ramos, D., Carrasco, J. L., Roman, B., Bertomeu I. F., Cristià, E., Geleva, D. & Serra-Majem, L. (2010). The Mediterranean diet in Spain: adherence trends during the past two decades using the Mediterranean Adequacy Index. *Public Health Nutrition*, 14(4), 622–628. <https://doi.org/10.1017/S1368980010002752>

Beatty, T. K. M., Lin, B.-H. & Smith T. A. (2014). Is Diet Quality Improving? Distributional Changes in the United States, 1989-2008. *American Journal of Agricultural Economics*, 96(3), 769–789. <https://doi.org/10.1093/ajae/aat104>

Bocquier, A., Vieux, F., Lioret, S., Dubuisson, C., Caillavet, F., & Darmon, N. (2014). Socio-economic characteristics, living conditions and diet quality associated with food insecurity in France. *Public Health Nutrition*. 18(16), 2952–2961. <http://doi.org/10.1017/S1368980014002912>

Boiron, A. (2016). Évolution des inégalités de niveau de vie entre 1970 et 2013. *Insee Références – Les revenus et le patrimoine des ménages*, pp. 55–70. <https://www.insee.fr/fr/statistiques/2017609?sommaire=2017614>

Buron, M. L., Kranklader, E. & Ribera, J. (2014). Les dépenses des ménages en 2011. *Insee Résultats* N° 158. <https://www.insee.fr/fr/statistiques/2015691>

Caillavet, F., Kyureghian, G., Nayga Jr., R. M., Ferrant, C. & Chauvin, P. (2015). Does Healthy Food Access Matter in a French Urban Setting? The Role of Food Retail Structure. *American Journal of Agricultural Economics*, 97(5), 1400–1416. <https://doi.org/10.1093/ajae/aav031>

- Caillavet, F., Darmon, N., L  toile, F. & Nich  le, V. (2018).** Is nutritional quality of food-at-home purchases improving? 1969-2010: 40 years of households consumption surveys in France. *European Journal of Clinical Nutrition*, 72, 220–227. <https://doi.org/10.1038/s41430-017-0041-6>
- Caillavet, F., Fadhuile, A. & Nich  le, V. (2016).** Taxing animal-based foods for sustainability: environmental, nutritional and social perspectives in France. *European Journal of Agricultural Economics*, 43(4), 537–560. <https://doi.org/10.1093/erae/jbv041>
- Caillavet, F., Fadhuile, A. & Nich  le, V. (2019).** Assessing the distributional effects of carbon taxes on food: Inequalities and nutritional insights in France. *Ecological Economics*, 163(C), 20–31. <http://doi.org/10.1016/j.ecolecon.2019.04.020>
- Caillavet, F., Lecogne, C. & Nich  le, V. (2009).** La fracture alimentaire : des in  galit  s persistantes mais qui se r  duisent. *Insee R  f  rences – Cinquante ans de consommation en France*, pp. 49–62. <https://www.insee.fr/fr/statistiques/1372378?sommaire=1372388>
- Castetbon, K. (2014).** Position socio  conomique et alimentation. In: *In  galit  s sociales de sant   en lien avec l'alimentation et l'activit   physique. Rapport*. Paris : Les   ditions Inserm. <http://hdl.handle.net/10608/6522>
- Chaix, B., Bean, K., Daniel, M., Zek, S. N., Kestens, Y., Charreire, H. et al. (2012).** Associations of Supermarket Characteristics with Weight Status and Body Fat: A Multilevel Analysis of Individuals within Supermarkets (RECORD Study). *Plos One*, 7(4). <https://doi.org/10.1371/journal.pone.0032908>
- Darmon, N. & Drewnowski, A. (2008).** Does social class predict diet quality? *American Journal of Clinical Nutrition*, 87(5), 1107–1117. <https://doi.org/10.1093/ajcn/87.5.1107>
- Darmon, N. & Drewnowski, A. (2015).** Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. *Nutrition Reviews*, 73(10), 643–660. <https://doi.org/10.1093/nutrit/nuv027>
- Desrosi  res, A. (2003).** Du travail    la consommation : l'  volution des usages des enqu  tes sur le budget des familles. *Journal de la Soci  t   Statistique de Paris*, 144(1-2), 75–111. http://www.numdam.org/item/JSFS_2003__144_1-2_75_0/
- Drewnowski, A. & Popkin, B. M. (1997).** The Nutrition Transition: New Trends in the Global Diet. *Nutrition Reviews*, 55(2), 31–43. <https://doi.org/10.1111/j.1753-4887.1997.tb01593.x>
- Favier, J. C., Ireland-Ripert, J., Toque, C. & Feinberg, M. (1995).** *R  pertoire g  n  ral des aliments : table de composition*. Paris : Technique & Documentation / INRA.
- Ferry, A. & Demoly, E. (2019).** Les comportements de consommation en 2017. *Insee Premi  re* N   1749. <https://www.insee.fr/fr/statistiques/4127596>
- Galobardes, B., Lynch, J., & Smith, G. D. (2007).** Measuring socioposition in health research. *British Medical Bulletin*, 81-82(1), 21–37. <https://doi.org/10.1093/bmb/ldm001>
- GBD (2019).** Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017, *The Lancet*, online 3 avril 2019. [http://dx.doi.org/10.1016/S0140-6736\(19\)30041-8](http://dx.doi.org/10.1016/S0140-6736(19)30041-8)
- Grignon, C. (1996).**   volution de la consommation de viande en France depuis 30 ans. In : *Devenons-nous v  g  tariens ? Opinions, croyances et consommation. Cahiers de l'Ocha* N   7, pp. 40–43. <http://www.lemangeur-ocha.com/chapitre/devenons-nous-vegetariens-opinions-croyances-et-consommation/>
- Groth, M. V., Sorensen, M. R., Matthiessen, J., Fagt, S., Landvad, N. & Knudsen, V. K. (2014).** Disparities in dietary habits and physical activity in Denmark and trends from 1995 to 2008. *Scandinavian Journal of Public Health*, 42(7), 611–620. <https://doi.org/10.1177/1403494814547668>
- Ifop (2019).** Usage et impact des applications alimentaires sur l'alimentation des Fran  ais -   tude Ifop pour Charal. <https://www.charal.fr/wp-content/uploads/2019/10/CHARAL-Etude-IFOP-Usage-et-impact-des-applications-alimentaires.pdf>
- Imamura, F., Micha, R., Khatibzadeh, D. et al. (2015).** Dietary quality among men and women in 187 countries in 1990 and 2010: a systematic assessment. *Lancet Global Health*, 3(3), 132–142. [https://doi.org/10.1016/S2214-109X\(14\)70381-X](https://doi.org/10.1016/S2214-109X(14)70381-X)

Inserm (2014). *Inégalités sociales de santé en lien avec l'alimentation et l'activité physique*. Paris : Les éditions Inserm. <http://hdl.handle.net/10608/6522>

Juul, F. & Hemmingsson, E. (2015). Trends in consumption of ultra-processed foods and obesity in Sweden between 1960 and 2010. *Public Health Nutrition*, 18(17), 3096–3107. <http://dx.doi.org/10.1017/S1368980015000506>

Kant, A. K. & Graubard B. (2007). Secular trends in the association of socio-economic position with self-reported diary attributes and biomarkers in the US population: National Health and Nutrition Examination Survey (NHANES) 1971–1975 to NHANES 1999–2002. *Public Health Nutrition* 10(2), 158–167. <https://doi.org/10.1017/S1368980007246749>

Lalluka, T., Laaksonen, M., Rakkonen, O., Roos, E., & Lahti, E. (2007). Multiple socio-economic circumstances and healthy food habits. *European Journal of Clinical Nutrition*, 61, 701–710. <https://doi.org/10.1038/sj.ejcn.1602583>

Larochette, B. & Sanchez-Gonzalez, J. (2015). Cinquante ans de consommation alimentaire : une croissance modérée, mais de profonds changements. *Insee Première* N° 1568. <https://www.insee.fr/fr/statistiques/1379769>

Lhuissier, A. (2007). *Alimentation populaire et réforme sociale. Les consommations ouvrières dans le second XIX^e siècle*. Versailles : Maison des Sciences de l'Homme/ éditions Quae.

Lhuissier, A., Caillavet, F. & Chen, S. Y. (2020). La pause méridienne des actifs: modes et lieux de restauration en temps contraint. In : *Les cantines se mettent à table : commensalité et identité sociale*, pp. 23–44. Versailles : Éditions Quae.

Lin, B. H. & Guthrie, J. (2012). Nutritional quality of food prepared at home and away from home, 1977-2008. USDA ERS *Economic Information Bulletin* N° 105. <https://www.ers.usda.gov/publications/pub-details/?pubid=43699>

Lioret, S., Dubuisson, C., Dufour, A., Touvier, M., Calamassi-Tran, G., Maire, B. et al. (2010). Trends in food intake in French children from 1999 to 2007: results from the INCA (étude Individuelle Nationale des Consommations Alimentaires) dietary surveys. *British Journal of Nutrition*, 103(4), 585–601. <http://dx.doi.org/10.1017/s0007114509992078>

Mackenbach, J. P., Stirbu, I., Roskam, A. J. R., Schaap, M. M., Menvielle, G., Leinsalu, M. & Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine* 2008, 358, 2468–2481. <http://doi.org/10.1056/NEJMsa0707519>

Manon, N. (1993). Repas à l'extérieur du domicile en 1991. *Insee Résultats – Consommation - Modes de vie* N° 56. <http://www.epsilon.insee.fr/jspui/handle/1/65499>

Marette, S., Messéan, A. & Millet, G. (2012). Consumers' willingness to pay for eco-friendly apples under different labels: evidences from a lab experiment. *Food Policy*, 37, 151–161. <https://doi.org/10.1016/j.foodpol.2011.12.001>

Marques-Vidal, P., Ravasco, P., Dias, C. M. & Camilo, M. (2006). Trends of food intake in Portugal, 1987-1999: results from the National Health Surveys. *European Journal of Clinical Nutrition*, 60, 1414–1422. <https://doi.org/10.1038/sj.ejcn.1602472>

Martin, A. (2006). *Apports nutritionnels conseillés pour la population française*. Paris: Tec & Doc Lavoisier.

Moubarac, J. C., Batal, M., Martins, A. P., Claro, R., Levy, R. B., Cannon, G. B. et al. (2014). Processed and Ultra-Processed Food Products: Consumption Trends in Canada from 1938 to 2011. *Canadian Journal of Diet Practice and Research*, 75(1), 15–21. <https://doi.org/10.3148/75.1.2014.15>

Nichèle, V., Andrieu, E., Boizot-Szantai, C., Caillavet, F. & Darmon, N. (2008). L'évolution des achats alimentaires: 30 ans d'enquêtes auprès des ménages en France. *Cahiers de Nutrition et de Diététique*, 43(3), 123–130. <https://www.em-consulte.com/en/article/174350>

Plessz, M. & Gojard, S. (2015). Fresh is Best? Social Position, Cooking and Vegetable Consumption in France. *Sociology*, 49(1), 172–190. <https://doi.org/10.1177%2F0038038514521715>

Popkin, B. M., Zizza, C. & Siega-Riz, A. M. (2003). Who is leading the change? US dietary quality comparison between 1965 and 1996. *American Journal of Preventive Medicine*, 25(1), 1–8. [https://doi.org/10.1016/S0749-3797\(03\)00099-0](https://doi.org/10.1016/S0749-3797(03)00099-0)

- Prättälä, R., Berg, M. A. & Puska, P. (1992).** Diminishing or increasing contrasts? Social class variation in Finnish food consumption patterns, 1978-1990. *European Journal of Clinical Nutrition*, 46(4), 279–287. <http://europepmc.org/abstract/med/1600925>
- Régnier, F. & Masullo, A. (2009).** Obésité, goûts et consommation. Intégration des normes d'alimentation et appartenance sociale. *Revue française de sociologie*, 50(4), 747–773. <https://doi.org/10.3917/rfs.504.0747>
- Régnier, F. & Chauvel, L. (2018).** Digital inequalities in the use of self-tracking diet and fitness apps : interview study on the influence of social, economic and cultural factors, *JMIR Mhealth Uhealth*, 6(4). <https://doi.org/10.2196/mhealth.9189>
- Rehm, C. D., Penalvo, J. L., Afshin, A. & Mozaffarian, D. (2016).** Dietary Intake Among US Adults, 1999-2012. *JAMA*, 315(23), 2542–2553. <https://dx.doi.org/10.1001%2Fjama.2016.7491>
- Rodrigues, S. S. P., Caraher, M., Trichopoulou, A. & Almeida, M. D. V. (2008).** Portuguese households' diet quality (adherence to Mediterranean food pattern and compliance with WHO population dietary goals): trends, regional disparities and socioeconomic determinants. *European Journal of Clinical Nutrition*, 62, 1263–1272. <https://doi.org/10.1038/sj.ejcn.1602852>
- Sauvy, A. (1959).** Changements de consommation des ménages de 1950 à 1958. *Population*, 14(4), 740–742. https://www.persee.fr/doc/pop_0032-4663_1959_num_14_4_6397
- Smed, S., Jensen, J. D. & Denver, S. (2007).** Socio-economic characteristics and the effect of taxation as a health policy instrument. *Food Policy*, 32, 624–639. <https://doi.org/10.1016/j.foodpol.2007.03.002>
- Sugiyama, T. & Shapiro, M. F. (2014).** The Growing Socioeconomic Disparity in Dietary Quality. *JAMA Intern Med*, 174(10), 1595–1596. <https://doi.org/10.1001/jamainternmed.2014.3048>
- UFC (2017).** Fruits et légumes bio : les sur-marges de la grande distribution. Que Choisir 29 août 2017. <https://www.quechoisir.org/action-ufc-que-choisir-fruits-et-legumes-bio-les-sur-marges-de-la-grande-distribution-n45900/>
- Valdés, J., Grau, M., Subirana, I., Marrugat, J., Covas, M.-I. & Schröder, H. (2009).** Secular Trends in Energy Intake and Diet Quality in a Mediterranean Population. *Annals of Nutrition and Metabolism*, 54, 177–183. <https://doi.org/10.1159/000217814>
- Villeneuve, A. & Bigata, G. (1975).** *Repas à l'extérieur et repas au domicile en 1971*. Paris: Insee.
- Wang, D. D., Leung, C. W., Li Y. et al. (2014).** Trends in Dietary Quality Among Adults in the United States, 1999 Through 2010. *JAMA Intern Med*, 174(10), 1587–1595. <https://doi.org/10.1001/jamainternmed.2014.3422>
- World Cancer Research Fund International (2007).** *Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective*. Washington, DC: World Cancer Research Fund/American Institute for Cancer Research.
-

ANNEXE 1
CONSTRUCTION DES DONNÉES DE CONSOMMATION ALIMENTAIRE

Pour étudier l'évolution des consommations alimentaires sur quatre décennies, nous avons construit des données micro-économiques sur les approvisionnements alimentaires des ménages à partir de deux sources : l'enquête *Consommation alimentaire* de l'Insee et le panel de ménages Kantar WorldPanel, anciennement Secodip.

Chacune de ces enquêtes est réalisée sur des échantillons représentatifs de ménages résidant en France métropolitaine dans des logements « ordinaires » (hors habitations mobiles, foyers de travailleurs, maisons de retraite, résidences universitaires, maisons de détention, etc.). Des pondérations sont fournies dans chacune des sources.

L'enquête *Consommation alimentaire*

Elle a été réalisée annuellement de 1969 à 1983 (sauf en 1975 du fait du recensement démographique), puis tous les deux ans jusqu'en 1991, date de l'arrêt définitif de cette enquête. Nous disposons donc de 18 enquêtes portant sur 6 000 à 8 000 ménages selon les années. Le mode de collecte, en deux visites, est resté constant.

Nombre d'observations des échantillons harmonisés

Année	Nombre de ménages	Nombre d'individus
1969	8 178	25 488
1970	8 149	25 150
1971	8 047	24 820
1972	7 980	24 534
1973	6 551	19 658
1974	7 524	22 677
1976	8 715	25 869
1977	7 660	22 776
1978	7 934	23 257
1979	7 644	22 166
1980	7 872	22 805
1981	8 406	23 895
1982	8 841	24 952
1983	8 877	24 851
1985	7 288	20 443
1987	6 938	19 103
1989	3 202	8 781
1991	6 353	16 906
Total	136 159	398 131

Enquêtes Insee, *Consommation alimentaire*

Chaque ménage est enquêté une seule fois pendant sept jours consécutifs. Pour tenir compte des aspects saisonniers de la consommation alimentaire, l'enquête se déroule tout au long de l'année. Plus précisément, l'année est divisée en huit périodes de six semaines chacune, auxquelles s'ajoutent deux quinzaines sans enquête : la première quinzaine d'août et la deuxième quinzaine de décembre. Lors d'une première visite, l'enquêteur recueille les caractéristiques sociodémographiques du ménage interrogé, en même temps que des renseignements sur la disposition d'un jardin ou d'un élevage familial, l'équipement du ménage en appareils de cuisine et de réfrigération et ses habitudes d'approvisionnement. À l'issue de cette interview, l'enquêteur remet au membre du ménage s'occupant des approvisionnements alimentaires un carnet dans lequel devront être consignés au jour le jour pendant une semaine l'ensemble des

approvisionnements en produits alimentaires destinés à la consommation à domicile, qu'ils proviennent d'un achat, d'un cadeau reçu, d'un prélèvement sur la production du ménage (autoconsommation) ou sur le stock du magasin (autofourniture), ainsi que les dépenses alimentaires hors du domicile (repas pris dans un restaurant, une cantine, consommations dans un bar, etc.)

Parallèlement, le contenu des enquêtes a évolué. Deux changements importants pour notre étude sont à souligner. L'un en 1978 concerne l'introduction d'une nouvelle variable, le niveau d'études de chaque individu du ménage. L'autre concerne la variable de nomenclature analytique permettant d'identifier les produits alimentaires, qui s'est enrichie : elle décrivait 170 produits différents en 1969, et 320 en 1991. Un travail d'harmonisation des variables au sein de la série Insee a en conséquence été réalisé sur les données des niveaux ménages, individus et produits.

Les séries Kantar Worldpanel (anciennement Secodip).

Kantar (ex-SECODIP) est une société privée qui produit des données sur les achats des ménages afin de répondre aux intérêts principaux commerciaux des entreprises agroalimentaires. Ces données, payantes, sont à usage confidentiel et strictement contrôlé. Le protocole de collecte répond aux besoins des clients de Kantar et peut évoluer de manière assez importante d'une année sur l'autre. Afin de limiter la charge pour les ménages interrogés, les ménages ne consignent pas les achats de tous les produits alimentaires.

Les séries Kantar sont des données de panel, c'est-à-dire constituées d'enregistrements répétés auprès des mêmes ménages. Les ménages panélistes enregistrent leurs achats pendant quatre années en moyenne. Les données dont nous disposons couvrent l'année 1989 puis toutes les années de 1991 à 2010, soit 20 enquêtes. Kantar suit un très grand nombre de ménages toute l'année, et sélectionne parmi ces ménages ceux qui répondent de façon satisfaisante tout au long de l'année, tout en gérant les phénomènes d'attrition liés aux panels. Les ménages retenus sont appelés « ménages actifs ». Ils représentent environ 3 000 ménages entre 1989 et 1995, atteignent 6 000 entre 1996 à 2001, et 10 000 en fin de période.

Les enregistrements portent sur les achats de produits alimentaires pour la consommation à domicile. L'autofourniture, l'autoconsommation et la restauration hors foyer ne sont pas collectées. Les informations sociodémographiques et quelques autres caractéristiques des ménages (disposition d'un jardin, d'un verger, équipement électro-ménager, etc.) sont recueillies une fois par an à l'aide d'un questionnaire.

La méthodologie de collecte des données au cours de la période étudiée ici a subi plusieurs changements. On peut distinguer trois grandes périodes, liées à des changements dans le protocole de collecte, des changements dans la nomenclature et des changements dans la taille de l'échantillon.

De 1989 à 1995 : collecte par carnets de bord

Deux panels indépendants P1 et P2 coexistent, chacun composé d'environ 3 000 ménages. Afin de mieux répartir la charge de la collecte entre les ménages, le panel P2 est divisé en deux sous-panels P3 et P4. Chaque panel consigne les achats des produits suivants :

- P1 : épicerie, yaourts, desserts, boissons non alcoolisées et matières grasses ;
- P2 : viande, charcuterie, poisson, produits congelés, lait, fromage, boissons alcoolisées, légumes en conserve, fruits et légumes frais ;

- P3 : fruits frais ;
- P4 : légumes frais.

Certains produits alimentaires ne sont pas enregistrés dans les achats (par exemple les liqueurs), et les ménages composés d'hommes seuls ne sont pas enquêtés.

1996-2008 : collecte par « scannette »

Trois panels coexistent : un panel d'environ 5 000 ménages actifs (GC), et deux sous-panels (VP et FL) d'environ 3 000 ménages actifs. Chaque panel consigne les achats des produits suivants :

- GC : produits congelés, produits laitiers, lait, épicerie, eau, alcools autres que le vin, œufs, fromage, sucre et pâtisseries ;
- VP : viande, charcuterie, poisson frais et vin ;
- FL : fruits et légumes frais.

La couverture des produits consignés est plus grande et tous les types de ménages sont enquêtés, quelle que soit leur composition, y compris ceux composés d'hommes seuls.

2009 : fusion des deux sous-panels VP et FL en PF, GC restant le même

Les sous-panels VP et FL fusionnés comptent respectivement environ 10 000 et 6 000 ménages actifs par an.

La mise en commun des deux sources de données

Constituer des séries longues de consommation moyenne de produits alimentaires nécessite l'adoption de définitions communes. Celles-ci concernent, d'une part la liste des produits alimentaires étudiés et, d'autre part, la mesure de quantité moyenne adoptée. Les champs différents couverts par les deux sources de données nous imposent le choix d'une définition commune de la consommation reposant sur les seuls achats de produits alimentaires pour la consommation à domicile. L'autoconsommation, l'autofourniture et l'alimentation hors-domicile ne sont donc pas prises en compte.

Le passage à la « scannette » a pu amener des ménages à « oublier » d'enregistrer certains achats, la procédure pour les produits sans code-barres (principalement les achats de produits frais) étant plus lourde que pour les autres produits. Entre 1995 (dernière année du carnet de bord) et 1996 (première année d'utilisation de la scannette), on constate effectivement une baisse significative des achats de plusieurs produits, tels que les pommes de terre, les légumes frais, les fruits frais, la volaille et le poisson frais. Nous avons testé une procédure de correction en appliquant, aux produits frais concernés par le passage à la scannette, un coefficient de correction de la moyenne des achats par personne et par jour. Pour les années 1996 à 2001, ces moyennes ont été corrigées par un coefficient représentant le rapport entre les achats moyens de 1995 et ceux de 1996 (les coefficients sont de 1.33 pour les pommes de terre, de 1.21 pour les légumes frais, de 1.14 pour les fruits frais, de 1.19 pour la volaille et de 1.23 pour le poisson frais). Toutefois, le changement de nomenclature de 2002 a rendu difficile la poursuite de cette correction. En outre, le volume des achats sans code-barres est appelé à diminuer au fil du temps. Ces incertitudes sur la qualité des coefficients calculés sur la base de données datant du milieu des années 1990 nous ont amenées à renoncer à cette correction.

Les données d'achat des panels Kantar ont été recodées selon la nomenclature analytique des produits de l'enquête *Consommation alimentaire* 1991 de l'Insee. Nous nous sommes basées sur cette définition commune des produits alimentaires pour constituer une nouvelle nomenclature. Celle-ci est le résultat d'un compromis entre

les conventions de classification utilisées dans les publications économiques de l'Insee sur l'alimentation, les limites fixées par les sources de données que nous exploitons et notre intérêt pour l'évaluation des conséquences nutritionnelles de la consommation alimentaire.

Il en découle un regroupement des produits à deux niveaux d'agrégation. Au niveau le plus agrégé, 18 grandes catégories de produits alimentaires sont considérées : 1) produits à base de céréales ; 2) pommes de terre ; 3) légumes ; 4) fruits ; 5) viandes et produits à base de viande ; 6) volailles, lapin, gibier ; 7) œufs ; 8) poissons ; 9) plats préparés ; 10) lait ; 11) yaourts et desserts lactés ; 12) fromages ; 13) matières grasses ajoutées ; 14) produits sucrés ; 15) eaux en bouteille ; 16) boissons non alcoolisées ; 17) boissons alcoolisées ; 18) café, thé, plantes d'infusion. Au niveau le plus fin, 78 groupes de produits ont été constitués. Néanmoins, le suivi de l'évolution des achats de chacun de ces groupes n'est pas toujours exhaustif. En effet, le champ couvert par les deux sources de données peut différer du fait de la modification de leur nomenclature : par exemple, les achats de pain ne sont pas présents dans les données Kantar. Aussi, les résultats de calculs nutritionnels issus des données Kantar sont obtenus en ajoutant 100 g de pain aux achats alimentaires quotidiens (selon les données d'achats de pain issues de l'enquête Insee 1991). À l'inverse, les quantités des conserves de poissons ne sont pas mesurées dans les enquêtes Insee, les desserts lactés ne sont identifiables en tant que tels qu'à partir de 1976 et l'huile de tournesol à partir de 1979. Nous n'avons pas fait d'hypothèses à cet égard.

Dans les deux sources de données, les méthodes de collecte des quantités achetées permettent d'évaluer les différents produits alimentaires au niveau de chaque ménage, mais sans possibilité d'aller au niveau de l'individu. Pour éliminer les effets d'échelle liés à la taille des ménages, les estimations sont établies par personne. Cela revient à faire l'hypothèse d'une répartition identique pour tous les membres du ménage. Nous retenons donc la mesure de la consommation moyenne établie par l'Insee dans le cadre de l'enquête *Consommation alimentaire*, à savoir la consommation annuelle par personne au domicile (hors autoconsommation et autofourniture) définie comme le rapport des achats totaux annuels des ménages enquêtés et du nombre total d'individus dont ils se composent. En pratique, cela revient à calculer :

- pour les enquêtes Insee :

$$q_i = \frac{\sum_h q_{ih} * poids_h * 52}{\sum_h npers_h * poids_h}$$

- pour les panels Kantar :

$$q_i = \frac{\sum_h q_{ih} * poids_h * (52/nbsem_h)}{\sum_h npers_h * poids_h}$$

où q_{ih} est la quantité totale de bien i achetée par le ménage h pendant la semaine d'enquête pour l'Insee ou pendant l'ensemble des semaines d'activité pour Kantar, $poids_h$ est le coefficient de redressement du ménage h , $npers_h$ est le nombre de personnes composant le ménage h et $nbsem_h$ est le nombre de semaines d'activité du ménage h dans le panel Kantar. On dispose donc d'un point par année.

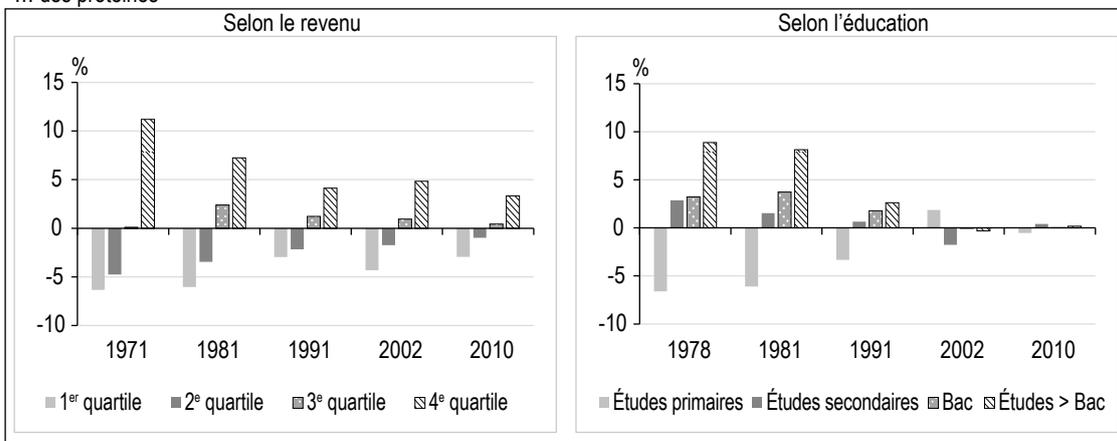
Dans l'article, nous avons choisi de ne pas relier entre eux les résultats issus des deux sources de données ni les séries de méthodologie de collecte différente au sein de Kantar Worldpanel, pour garder en évidence les décrochages dus aux changements de méthodologie.

ANNEXE 2

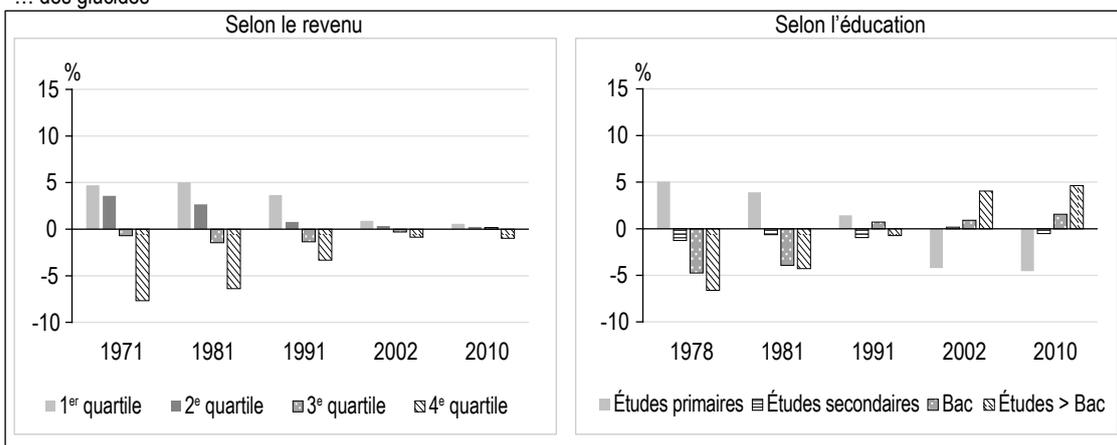
INÉGALITÉ DES CONTRIBUTIONS DES MACRONUTRIMENTS AU CONTENU CALORIQUE
SELON LE REVENU ET L'ÉDUCATION

Déviations par rapport à la moyenne, en (%) de la contribution au contenu calorique...

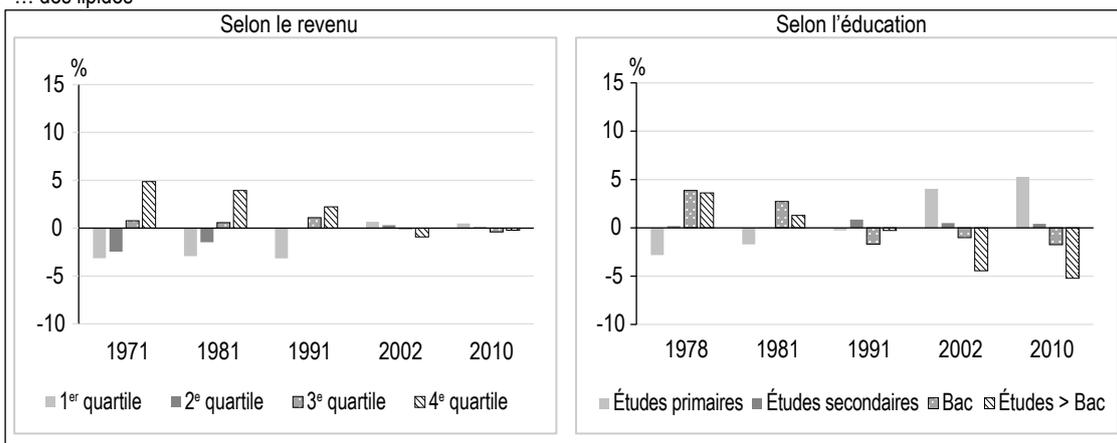
... des protéines



... des glucides



... des lipides



Source : Insee, enquêtes *Consommation alimentaire* 1971-1981-1991; panels de consommateurs Kantar WorldPanel 2002-2010.

L'élasticité-prix de la demande d'électricité en France

Price Elasticity of Electricity Demand in France

Stéphane Auray*, Vincenzo Caponi** et Benoît Ravel***

Résumé – La concurrence et le changement climatique étant des facteurs de plus en plus importants pour la production et la consommation d'électricité, les opérateurs du marché et les pouvoirs publics ont un besoin accru de prévisions fiables de prix et de consommation, pour planifier les investissements et définir les politiques de réglementation. Pour établir des prévisions pertinentes, il est essentiel de comprendre la réaction des consommateurs face aux variations de prix, synthétisée par le concept d'élasticité. Grâce à un jeu de données unique rassemblant des millions de relevés semestriels de compteurs d'électricité en France entre 2007 et 2015, nous estimons l'élasticité-prix des dépenses en électricité des ménages. Nous proposons trois spécifications : une spécification canonique régressant la consommation d'électricité à hauteur d'un prix par kilowatt/heure, avec laquelle nous obtenons une élasticité de -0.8 (remarquablement similaire à celle de la littérature existante et la corroborant), une estimation suivant le modèle de système de demande quasi idéal (AIDS) de Filippini (1995), qui donne des résultats très proches des siens, et une extension de celle-ci intégrant des élasticités saisonnières, qui montre que la demande en électricité est plus élastique pendant la période estivale.

Abstract – Competition and climate becoming more important for electricity production and consumption, market operators are increasingly interested in reliable forecasts of electricity prices and consumption for planning their investments and regulating policies. Key for good forecasts is understanding the consumers' reaction to price changes, synthesized by the concept of elasticity. Using a unique dataset of millions of bi-annual meter readings of electricity consumption in France from 2007 to 2015, we estimate the price elasticity of electricity expenditure of private households. We propose three specifications: a canonical one that regresses electricity consumption on a price per kilowatt/hour, where we find an elasticity equal to -0.8, a result remarkably in line and corroborating previous literature; a specification that follows Filippini's (1995) model of an Almost Ideal Demand System (AIDS), in which we substantially replicate his results; and finally, an extension of the latter that allows elasticities to be season-dependent that shows the demand of electricity being more elastic in summer.

Codes JEL / JEL Classification : Q4, Q41, C5, D12

Mots-clés : demande en électricité, élasticité-prix

Keywords: electricity demand, price elasticity

* CREST-Ensaï, UCLO (stephane.auray@ensai.fr) ; ** CREST-Ensaï, IZA (vincenzo.caponi@ensai.fr) ; *** DataStorm (benoit.ravel@datastorm.fr)

Nous remercions Électricité Réseau de France (aujourd'hui Enedis) pour nous avoir communiqué leurs données et nous avoir permis de les utiliser, ainsi que Mathieu Bordigoni, Jean-Baptiste Galland et Laurent Gilotte pour leur aide.

Reçu le 19 juillet 2017, accepté après révisions le 20 janvier 2019.

Traduit de la version originale en anglais.

Citation : Auray, S., Caponi, V. & Ravel, B. (2019). Price Elasticity of Electricity Demand in France. *Economie et Statistique / Economics and Statistics*, 513, 91–103. <https://doi.org/10.24187/ecostat.2019.513.2002>

L'électricité est considérée comme un « service au public » et pour des raisons principalement historiques son prix est fixé, la plupart du temps, en fonction du coût de production. Aujourd'hui, la concurrence et le changement climatique étant des facteurs de plus en plus importants, il est plus utile que jamais pour les régulateurs et les opérateurs du marché de l'électricité d'analyser la réaction des consommateurs face aux variations de prix. Les opérateurs de réseau (gestionnaires de réseaux de transport et distributeurs), notamment, doivent planifier leurs investissements en tenant compte de leurs prévisions de variations de prix, ainsi que de la façon dont la demande évoluera en conséquence. Nous utilisons un large jeu de données tirées du marché français pour estimer l'élasticité de la consommation d'électricité. Ces données uniques, de grande envergure, nous permettent de reproduire une partie des résultats déjà acquis dans la littérature, qui font état d'une élasticité proche de -1 en France, et de les corroborer. Nous reproduisons également, sur des données d'un autre pays, les résultats de Filippini (1995) pour un modèle à deux tarifs et nous complétons cette approche, avec la même stratégie de modélisation, en introduisant un modèle saisonnier. Notre contribution est double : d'une part, nous corroborons les résultats de la littérature existante grâce à un jeu de données largement représentatif et, d'autre part, la richesse de nos données permet de diviser l'échantillon afin d'identifier les différences entre la consommation constatée à différentes saisons de l'année.

Nos données ont deux avantages principaux : elles couvrent plus de 95 % de la consommation privée d'électricité en France métropolitaine et, grâce aux relevés de compteurs, permettent d'observer les prix réels par kWh sans qu'il soit nécessaire de calculer un prix moyen en divisant les dépenses totales par la consommation totale (sachant que les dépenses totales incluent les coûts de livraison fixes, etc.). Notre analyse est menée en deux étapes. Premièrement, nous utilisons toutes les informations tirées des relevés de compteurs pour créer un nouveau jeu de données fusionnant, à des niveaux géographiques plus fins, ces informations avec d'autres données comportant des variables économiques et géographiques, provenant de l'Insee pour la plupart, mais incluant également des variables météorologiques. En même temps, nous établissons des données mensuelles à partir des observations semestrielles, en répartissant la consommation d'électricité individuelle tout au long du semestre selon les coefficients tirés

du système de profilage officiel d'ERDF, afin d'estimer la courbe de charge de chaque acheteur. Deuxièmement, pour des échantillons du jeu principal de données, nous fusionnons avec des données fournissant d'autres variables et des données mensuelles pour mener l'analyse économétrique.

Nous proposons trois spécifications différentes pour étudier les élasticités-prix. La première spécification, la plus canonique, régresse la consommation d'électricité sur un prix par kilowatt/heure calculé à partir des prix réels pour les clients ne payant qu'un seul tarif, ou à partir de la moyenne pondérée de différents prix pour les clients payant des tarifs différents à différents moments de la journée. Dans notre deuxième spécification, nous suivons Filippini (1995) et présentons un modèle de système de demande quasi idéal (*Almost Ideal Demand System*, ou AIDS). Notre troisième et dernière spécification est une extension de la deuxième intégrant des élasticités saisonnières qui sont différentes en été et en hiver. Dans tous les modèles, nous contrôlons les effets fixes d'année et de mois, ainsi que des effets des variables météorologiques et d'un autre ensemble de variables économiques au niveau du département. Avec la première estimation, nous obtenons une élasticité de la consommation d'électricité par rapport au prix égale à -0.8, soit un résultat remarquablement similaire à ceux de la littérature existante. Avec les modèles AIDS, nous obtenons également des résultats très proches de ceux obtenus par Filippini, notamment des élasticités-prix de -1.46 et -1.86 pour les prix en heures pleines et en heures creuses respectivement (Filippini fait état de -1.41 et -2.57). Avec notre modèle saisonnier, nous obtenons des élasticités de -1.45 et -1.85 en hiver et des valeurs absolues un peu plus fortes en été à -1.61 et -2.08.

Notre article s'organise comme suit : dans la première section, nous présentons brièvement la littérature existante ; la section 2 est consacrée au traitement préliminaire de notre principal jeu de données ; la section 3 présente de façon détaillée notre stratégie d'estimation et la section 4 les résultats.

1. Revue de littérature

La littérature sur l'estimation de l'élasticité-prix de la demande d'électricité est abondante. Elle se répartit en trois grands courants selon les données utilisées pour cette estimation : données agrégées en séries chronologiques (les recherches

les plus nombreuses dans ce domaine), données transversales, et données de panel. Les données transversales et les données de panel peuvent être plus ou moins agrégées, selon que les observations portent sur des ménages (niveau le plus désagrégé) ou sont agrégées au niveau de régions (par exemple, Nakajima, 2010, tire ses estimations de données de panel comprenant des préfectures japonaises) ou au niveau de pays (voir par exemple Bernstein & Madlener, 2011, pour un panel de pays de l'OCDE).

1.1. Résultats empiriques à partir de séries chronologiques et de données de panel longue période

La plupart des études relatives à l'élasticité-prix de la demande d'électricité reposent sur la variation de la consommation d'électricité et de son prix dans le temps. Ces études sont effectuées à partir soit de séries chronologiques soit de données de panel de longue période. Les panels longs contiennent habituellement des données agrégées à un niveau large, comme des pays et des régions, et dont les observations couvrent de nombreuses années. En termes méthodologiques, ces études utilisent habituellement des méthodes d'estimation basées sur la cointégration avec des retards distribués auto-régressifs (ARDL), dans la mesure où les séries chronologiques des prix et les niveaux de consommation sont des séries intégrées. Cette méthode permet de calculer des élasticités à court et à long terme, c'est-à-dire la réaction aux variations de prix dans les années suivant immédiatement la variation concernée, et la réaction qui se manifesterait pendant une période plus longue si le prix reste relativement stable après la variation concernée. Dans le contexte de la demande d'électricité, cette information est extrêmement importante car non seulement les ménages mais aussi les entreprises et les sites industriels peuvent procéder à des ajustements plus importants dans la durée que lors des années qui suivent immédiatement la variation de prix. De fait, l'élasticité-prix à long terme de la demande d'électricité est généralement estimée comme étant supérieure à l'élasticité à court terme. Okajima & Okajima (2013) offrent un examen approfondi des études utilisant des séries chronologiques ou des données de panel long, où ils présentent les estimations obtenues pour plusieurs pays, à savoir l'Australie, la Turquie, l'Afrique du Sud, les États-Unis (six études) et le Japon (deux études). En règle générale, l'élasticité à court terme est relativement faible tandis que celle à long terme est beaucoup plus importante. Narayan & Smyth (2005) font

état d'une élasticité de 0.26 à court terme et de 0.54 à long terme en Australie. L'échantillon s'étale de 1959 à 1972. Pour la Turquie, Halicioglu (2007) utilise des données allant de 1968 à 2005 et estime à 0.33 l'élasticité à court terme et à 0.52 à long terme. Ros (2017), avec des données provenant de compagnies d'électricité américaines dans un panel long de 1972 à 2009, obtient des élasticités de même ordre, comprises entre 0.48 et 0.61 selon le modèle utilisé (statique ou dynamique). Dans le même article, Ros estime des équations de prix pour différents types de clients ; il est intéressant – mais pas surprenant – de constater que les prix de l'électricité sont généralement inférieurs dans les États où la concurrence est plus importante, et que les clients industriels en profitent beaucoup plus que les particuliers. En outre, il constate également que la productivité totale des facteurs est associée à des prix moins élevés.

Dergiades & Tsoulfidis (2008), avec des séries chronologiques allant de 1965 à 2006 pour les États-Unis, estiment une élasticité de 1.07 à long terme. Pour l'Afrique du Sud entre 1978 et 2005, Ziramba (2008) obtient une demande d'électricité complètement inélastique, avec des élasticités estimées à 0.02 et à 0.04 sur le court et le long terme respectivement. Nakajima & Hamori (2010b) constatent également une demande relativement inélastique aux États-Unis, estimant l'élasticité à long terme à 0.33 à partir de données de panel long agrégées au niveau régional sur une période de 1993 à 2008. En revanche, au Japon, Nakajima (2010) sur une période de 1975 à 2005, en utilisant des séries chronologiques, obtient une élasticité à long terme de 1.13. D'autres études basées sur des séries chronologiques ou des données de panel long utilisent un modèle d'ajustement partiel. Notamment, pour les États-Unis, Kamerschen & Porter (2004) obtiennent 0.13 et 1.89 entre 1973 et 1998, Paul *et al.* (2009) obtiennent environ 0.17 et 0.35 entre 1990 et 2006 et Alberini & Filippini (2011) obtiennent 0.12 et 0.2 entre 1995 et 2007. Pour finir, pour le Japon, Okajima & Okajima (2013) obtiennent des estimations de 0.4 et de 0.49 pour le court et le long terme, à partir d'un échantillon de données d'un panel de grande envergure rassemblant des préfectures japonaises et s'étalant entre 1990 et 2007.

1.2. Résultats empiriques à partir de données transversales et de données de panel de grande envergure

Dans la littérature, les études basées sur des données transversales ou des données de panel de

grande envergure sont plus rares. Deux raisons à cela : premièrement, il est plus difficile de trouver des données désagrégées et, deuxièmement, le prix marginal de l'électricité est souvent le même pour une grande partie de tout échantillon dont il est possible de disposer. Même si la consommation d'électricité varie d'un ménage à l'autre selon diverses caractéristiques pour une section transversale de ménages, dans la plupart des cas, tous les ménages payent exactement le même prix pour leur électricité, de sorte qu'il est difficile d'estimer l'élasticité-prix. En outre, si le prix marginal ne change pas d'un ménage à l'autre, il ne fait habituellement pas partie des données. Pour cette raison, la plupart des études se basent sur les prix moyens, c'est-à-dire qu'elles reposent sur des données relatives aux dépenses d'électricité et sur le prix moyen calculé en fonction de la consommation réelle. Bien que l'utilisation de prix moyens soit justifiée principalement par les données disponibles, le consensus s'accorde sur le fait que c'est le prix marginal qui compte lorsque les ménages font leur choix en matière de consommation d'électricité – voir Ito (2014) et Alberini *et al.* (2011b), entre autres. Parmi ces quelques études, Krishnamurthy & Kriström (2015) estiment l'élasticité-prix et l'élasticité-revenu de la demande d'électricité des ménages à partir d'un panel de onze pays de l'OCDE, et constatent une sensibilité importante de la consommation aux variations du prix moyen et une moindre sensibilité aux variations des revenus. L'élasticité-prix va de -0.27 en Corée du Sud à -1.4 en Australie, en passant par -0.96 en France. Alberini & Filippini (2011) se concentrent sur la demande en électricité dans les différents États américains et présentent un modèle économétrique dynamique indiquant les élasticités à long et à court terme. Ces estimations s'établissent à environ -0.15 sur le court terme et vont de -0.44 à -0.73 sur le long terme, selon la méthodologie utilisée. Alberini & Filippini (2011) accordent une attention particulière à deux problèmes critiques dans ces types d'estimations : d'une part, le fait que, dans les modèles basés sur des panels, la variable dépendante retardée est endogène, d'autre part le fait que les prix de l'électricité, communiqués par l'État sous forme de moyenne, sont mal mesurés. Le premier problème est résolu par l'estimateur des moindres carrés à variables muettes (LSDV) de Kiviet et les procédures de Blundell-Bond, tandis que le deuxième est résolu à l'aide de variables instrumentales. Filippini (2011) mène une analyse semblable à celle d'Alberini & Filippini (2011), mais avec des données sur la Suisse, et identifie les élasticités heures creuses

et heures pleines. Il constate également que les consommateurs passent des heures creuses aux heures pleines en fonction des grilles tarifaires. Toutes les études susmentionnées, ainsi que celles qu'elles citent, supposent que les ménages sont des « preneurs de prix » car ils peuvent ajuster leur consommation en fonction d'un prix donné. Reiss & White (2005) ont mis au point un modèle incluant un classement endogène associé à une grille tarifaire non linéaire afin de tenir compte de la possibilité pour différents ménages de choisir parmi les différentes grilles tarifaires proposées par les fournisseurs de services aux collectivités locales. Cette étude estime un modèle de demande en électricité des ménages pouvant être utilisé pour évaluer une conception tarifaire alternative. Ce modèle se concentre sur l'hétérogénéité des élasticités de la demande des ménages et sur leur relation avec les appareils détenus par les ménages et d'autres caractéristiques de ces derniers, ainsi que sur la façon dont elles influencent la consommation des ménages face à des variations complexes (non linéaires) des grilles tarifaires. Reiss & White (2005) constatent que les élasticités moyennes estimées sont légèrement supérieures à celles qui auraient été obtenues à l'aide de méthodes d'estimation plus traditionnelles.

2. Données et traitement préliminaire

Compte tenu de la nature des données dont nous disposons, nous menons notre analyse en deux étapes. Premièrement, nous travaillons avec le jeu de données initial fourni par ERDF afin de générer des observations mensuelles et de le rendre cohérent en vue de fusionner les données avec d'autres variables tirées de données de l'Insee. Deuxièmement, nous constituons un sous-échantillon à partir du jeu initial de données, et fusionnons d'autres variables à un niveau géographique plus fin. Nos données proviennent de relevés de compteurs représentant plus de 95 % de la consommation privée d'électricité en France métropolitaine. Les compteurs étant relevés environ tous les six mois, la consommation d'électricité est enregistrée entre deux dates de relevé. Notre point de départ est un volume d'électricité réellement consommé durant une période donnée, selon les chiffres d'un compteur correspondant habituellement à un ménage. Les consommateurs d'électricité font partie de trois groupes différents selon leur contrat. Les ménages payant un prix par kWh unique sont les clients de « BASE ». Les clients payant deux prix différents pour les heures de pointe (journée) et hors pointe (nuit) sont dénommés « HP/HC ».

La troisième catégorie est celle des « TEMPO », avec six prix différents pour le kWh, combinant l'option HP/HC avec une série de trois types de journées codées ROUGE, BLANC et BLEU de la plus chère à la moins chère. Les clients diffèrent également en termes d'abonnement, c'est-à-dire le volume de kW pouvant être consommé à tout moment de la journée. Plus le volume de l'abonnement est élevé et plus le coût fixe contractuel correspondant est élevé. Les options BASE et HP/HC n'imposent pas d'abonnement minimal (sachant que le minimum de tout contrat est de 3 kW), tandis que l'option TEMPO l'exige. Pour cette raison, les clients TEMPO sont censés avoir une consommation d'électricité plus importante mais ne représentent qu'un petit échantillon du marché de l'électricité dans son ensemble. Pour chaque compteur, notre jeu de données enregistre un numéro d'identification indiquant le site (ou le compteur), la date à laquelle le relevé débute et celle à laquelle il prend fin. En conséquence, les relevés sont enregistrés pour chaque segment de consommation (heures pleines, heures creuses et chaque type de journée pour les clients TEMPO) et la consommation en kWh par type. Le jeu de données contient 36 390 648 compteurs relevés durant une période de huit ans à compter de 2007, pour plus de 800 millions d'observations. Un autre jeu d'observations par compteur permet d'identifier le contrat, y compris l'abonnement d'électricité, ainsi que les prix par kWh pour chaque segment de consommation. Les segments de consommation étant différents selon la localisation en France, les données indiquent donc les heures exactes des segments pour chaque compteur.

L'une des principales difficultés rencontrées avec ces données vient du fait que les dates de relevé des compteurs varient. Pour cette raison, bien que tous les compteurs soient relevés tous les six mois et que la consommation d'électricité soit enregistrée durant ces mois, ces relevés ne peuvent pas être comparés immédiatement entre différents compteurs. Avant de procéder à l'analyse économétrique, nous devons donc faire en sorte que les observations puissent être comparées d'un compteur à l'autre. La sous-section suivante décrit la méthodologie que nous avons utilisée pour obtenir des observations comparables.

2.1. Harmonisation des observations sur la charge d'électricité

Les observations sur la charge sont harmonisées à l'aide de coefficients fournis directement par

RTE et Enedis, les opérateurs du réseau d'électricité en France. Ces coefficients sont calculés pour un panel représentatif de consommateurs pour lesquels l'électricité est mesurée toutes les dix minutes. Dans la pratique, ces coefficients servent à extrapoler l'utilisation d'électricité pour tous les compteurs observés à partir de l'utilisation observée pour le panel. Le panel est riche du point de vue de la fréquence des observations mais moins en termes de covariables telles que les variables géographiques. Les coefficients sont alors calculés par profil, c'est-à-dire selon que le compteur correspond à un contrat BASE, HP/HC ou TEMPO. Le coefficient de chaque profil est ensuite enrichi grâce à des variables météorologiques, afin de tenir compte de la variation potentielle de la consommation due aux variations de la température durant la journée.

Nous définissons alors les coefficients en tenant compte du climat et des profils $C(j, w, d, h, t)$, où j représente le profil, w, d, h et t représentent la semaine, le jour, l'heure (par demi-heure) et une classification temporelle. Nous pouvons, compte tenu de la consommation annuelle moyenne d'un profil, établir la consommation par demi-heure en multipliant la moyenne annuelle par le coefficient. Appelons la consommation par demi-heure ainsi obtenue $PM(j, w, d, h)$:

$$PM(j, w, d, h) = PM_Y(j) \cdot C(j, w, d, h, t)$$

où $PM_Y(j)$ est la consommation moyenne durant une année donnée, que nous ne connaissons pas, et où les conditions météorologiques dépendent du jour de l'année et de l'heure. La consommation d'électricité effectivement enregistrée en kWh pour toute période P peut s'écrire comme suit :

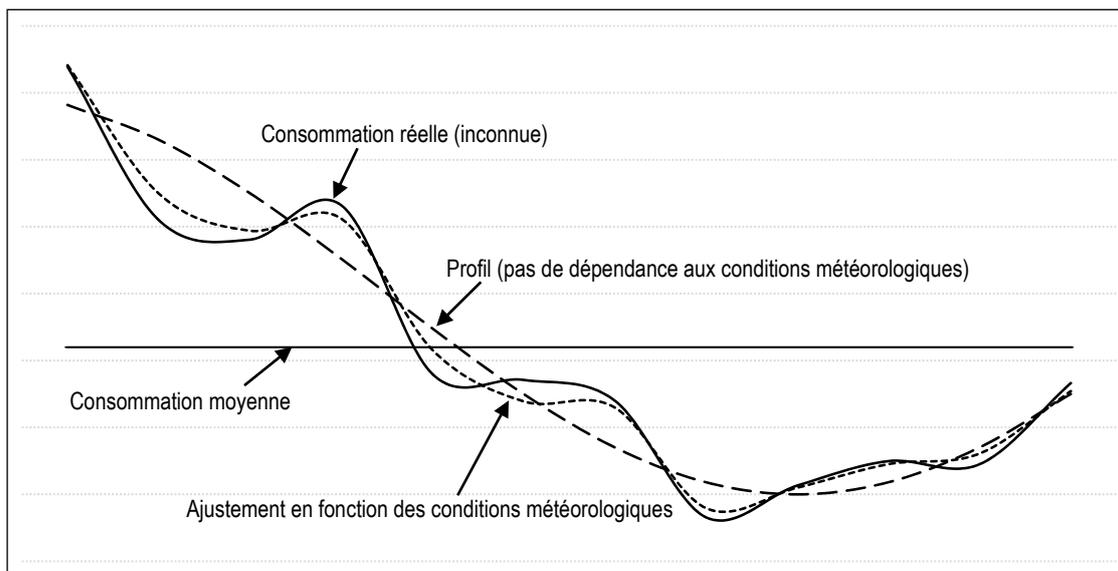
$$Q(j, P) = \frac{1}{2} \sum_{i \in P} PM(j, i) = PM_Y(j) \cdot C(j, i)$$

où l'indice $i = (w, d, h)$ contient toutes les informations relatives à l'heure et aux conditions météorologiques à une fréquence de 30 minutes (raison pour laquelle la somme est divisée par deux afin d'identifier la consommation horaire en kWh). Nous pouvons alors obtenir la consommation annuelle moyenne représentée par :

$$PM_Y(j) = \frac{2Q(j, P)}{\sum_{i \in P} C(j, i)}$$

La figure I montre une illustration de cette procédure : on y voit la consommation moyenne durant la période d'observation, à savoir six mois, la consommation réelle inconnue, la consommation

Figure I – Illustration d'un profil imputé



Source : illustration des auteurs, à partir de RTE (2015), Annexe F du dispositif de reconstitution des flux.

imputée tirée de l'application des coefficients associés au profil, et tenant compte des conditions météorologiques. Cette dernière est censée être le meilleur prédicteur de la consommation réelle à tout moment dans le temps.

Une fois que nous connaissons la consommation moyenne par an et les coefficients $C(j, i)$, nous pouvons calculer la consommation par demi-heure pour chaque compteur de notre jeu de données et agréger de la façon requise pour obtenir la consommation quotidienne, hebdomadaire, mensuelle ou semestrielle. Nous obtenons ainsi un jeu de données dans lequel nous avons enregistré l'identifiant du compteur, une variable indiquant si la consommation a eu lieu pendant les heures pleines ou creuses, le mois et la consommation totale durant le mois. Pour la période couverte, nous disposons d'environ deux milliards d'observations.

2.2. Constitution d'un sous-échantillon pour l'analyse

Après avoir harmonisé les observations de façon à ce qu'une période d'observation corresponde à la même période pour tous les compteurs, compte tenu du grand volume d'observations dont nous disposons, nous extrayons du jeu de données un échantillon aléatoire de 1 % de toutes les observations. Sachant que les indicateurs géographiques des compteurs sont à plus petite échelle, nous fusionnons pour cet échantillon des données fournissant plusieurs autres

variables économiques telles que l'indice des prix à la consommation, ainsi que des indicateurs de l'activité économique des zones géographiques (dont la part de la population active, le niveau d'éducation moyen, etc.).

La première constatation est une grande différence entre les contrats TEMPO et les autres. Si, dans le cadre des contrats à prix unique et à deux prix, les prix varient de façon déterministe au fil du temps et uniquement en l'espace d'une journée, dans le cadre des contrats TEMPO les prix peuvent au contraire varier d'une journée à l'autre et, fait plus important encore, le prix appliqué à chaque journée est choisi par le fournisseur d'électricité avec un préavis de quelques heures. De fait, les fournisseurs d'électricité fixent stratégiquement des prix plus élevés durant les journées où ils s'attendent à une demande en électricité plus importante (par exemple en hiver). Cela engendre une forte endogénéité du prix pour les clients TEMPO à la différence des autres clients. Pour cette raison, et sachant qu'ils ne représentent qu'une petite partie du marché global, nous excluons les clients TEMPO de la suite de l'analyse.

3. Analyse

Nous proposons trois spécifications différentes pour étudier les élasticités-prix. La première spécification, plus canonique, régresse la consommation d'électricité à hauteur d'un prix par

kilowatt/heure calculé à partir des prix réels pour les clients ne payant qu'un seul tarif, ou à partir de la moyenne pondérée de différents prix pour les clients payant des tarifs différents à différents moments de la journée. La deuxième spécification suit Filippini (1995) avec un modèle de système de demande quasi idéal (AIDS). La troisième et dernière spécification est une extension de la deuxième intégrant des élasticités saisonnières. Dans tous les modèles, nous contrôlons les effets fixes d'année et de mois, ainsi que les variables météorologiques et un autre ensemble de variables économiques au niveau du département, dont : le nombre de jours par mois durant lesquels la température est supérieure à 15 °C (seuil de « confort » en-dessous duquel il est probablement nécessaire de mettre le chauffage) ; le nombre de jours effectifs dans le mois ; la part des logements déclarés comme résidence principale ; la part des logements construits avant 1990 ; la part des maisons dans le nombre total de logements. Nous ajoutons toutes les variables qui contribuent à contrôler les facteurs pouvant influencer la consommation d'électricité et qui, surtout dans leur dimension temporelle, pourraient également être corrélées au prix de l'électricité. Nous ajoutons également des variables telles que l'âge moyen de la population, la part d'actifs et la part ayant fait des études supérieures.

3.1. Fixation des prix en France

Il est difficile d'estimer l'élasticité de la demande d'un bien ou d'un service, car le prix et la quantité sont habituellement des variables d'équilibre et sont déterminés simultanément. Pour cette raison, dans un modèle de régression simple comme celui que nous utilisons dans le présent article, nous sommes confrontés à un problème d'endogénéité susceptible de biaiser les estimations. C'est pour cela que, souvent, d'autres modèles sont utilisés pour corriger ce biais potentiel, notamment avec des variables instrumentales. En ce qui nous concerne, nous avons toutefois de bonnes raisons de penser que les prix de l'électricité en vigueur sur le marché français présentent une forte exogénéité en raison des règles imposées par l'État au principal fournisseur d'électricité en termes de fixation des prix.

En France, l'électricité est produite principalement par EDF, une société à participation publique chargée par l'État, depuis 1946, de produire et de distribuer de l'électricité dans un régime de quasi-monopole (c'est-à-dire excluant certaines très grandes entreprises), en tant que service public. Ce régime a été légèrement

modifié en 2007 avec l'introduction d'un marché concurrentiel pour la fourniture d'électricité, ainsi que d'une distinction entre fourniture et distribution d'électricité. La société ERDF, devenue Enedis par la suite, a été créée sous un régime de monopole total, sous la houlette de l'État, pour la distribution d'électricité. Parallèlement, aux côtés d'EDF où l'État conservait une participation et un contrôle majoritaires, d'autres entreprises ont été autorisées à fournir de l'énergie au client final en passant par Enedis pour la distribution. Toutefois, la concurrence s'est avérée asymétrique car le pouvoir de fixation des prix d'EDF est toujours entièrement déterminé par l'État tandis que les autres entreprises peuvent offrir des grilles tarifaires différentes. Ces entreprises payent néanmoins les mêmes prix qu'EDF à la source, de sorte qu'elles ne peuvent exercer leur concurrence qu'en offrant des grilles tarifaires différentes pour les prix fixes, ainsi que des tarifs heures pleines ou heures creuses. Chez EDF, la fixation des prix est transparente : la partie variable reflète le coût marginal de la production d'électricité, tandis que le coût fixe est calculé afin de couvrir les investissements nécessaires pour maintenir les capacités de production et de livraison d'électricité. Ainsi, nous sommes convaincus que les tarifs d'EDF peuvent être considérés comme exogènes dans notre analyse, mais nous ne sommes pas certains qu'il en soit de même pour les clients qui s'en remettent aux prix « du marché » faisant concurrence à ceux d'EDF. Heureusement, bien que nos données couvrent une période allant de 2007 à 2015 (c'est-à-dire après l'ouverture du marché), seule une petite partie des clients français avait choisi la concurrence jusqu'en 2015. En 2014, la part de ceux ayant choisi les prix du marché n'était que de 6.7 % (alors qu'elle passera à 13 % en 2017). Ainsi, la plupart de nos observations reposent sur des prix régulés fixés par convention entre EDF et l'État¹.

3.2. Modèle à un prix

Pour estimer l'élasticité-prix, notre spécification préférée est un modèle de régression à effets fixes dans lequel nous contrôlons les variables temporelles, c'est-à-dire les années et les mois (pour les effets saisonniers autant que pour les effets par année). Les prix et la consommation sont mesurés au niveau des compteurs. Nous incluons également des variables économiques et démographiques, par localisation, qui, selon

1. Pour une description complète, voir <https://www.cre.fr/Electricite/marche-de-detail-de-l-electricite>.

nous, peuvent influencer la relation entre la consommation et le prix de l'électricité. Ces variables sont collectées au niveau du département et associées aux compteurs en fonction du lieu où ils sont situés. Le prix moyen dont le client de base bénéficie est calculé en fonction de la composante variable du prix réellement payé. Pour les clients payant deux prix variables pour la consommation heures pleines et heures creuses, la moyenne est calculée en pondérant la part de la consommation totale à ce prix. Nous définissons la consommation au prix P_i comme étant C_i , et la consommation totale comme étant C , de sorte que :

$$C = \sum_i^n C_i$$

avec $n = 2$, puis nous définissons le prix variable moyen comme :

$$P = \sum_i^n W_i P_i$$

Avec :

$$W_i = \frac{C_i}{C}$$

Tous les prix sont exprimés en euros 2005 constants déflatés avec l'indice des prix à la consommation.

3.3. Modèle à deux prix

Afin de conserver des informations intéressantes sur le comportement des ménages, à savoir leur réaction aux écarts de prix entre différents segments temporels de la journée, nous avons également estimé des modèles de système de demande quasi idéal (AIDS). Nous suivons Filippini (1995) et reproduisons leur étude faite pour les clients suisses à l'aide de notre jeu de données, qui est beaucoup plus important². Pour que nos estimations puissent être comparées à celles de Filippini, nous construisons notre variable dépendante de façon à représenter la part des dépenses en électricité durant les heures pleines et creuses. Au lieu de calculer la consommation brute d'électricité, nous calculons ainsi le total des dépenses variables en électricité puis la part correspondante durant les deux segments temporels de la journée, comme suit :

$$m = \sum_i^2 C_i P_i$$

$$w_i = \frac{C_i P_i}{m}$$

où m représente le total des dépenses variables en électricité.

Pour nos variables indépendantes, nous utilisons le log des prix des deux segments temporels et le log des dépenses totales en électricité en termes réels. Nous répétons l'estimation pour la totalité de l'échantillon, faisant également la différence entre l'hiver et l'été. Ce modèle estime les élasticités partielles de la demande en électricité dans les deux segments temporels, à condition que la consommation totale d'électricité soit maintenue à un niveau constant. Ainsi, il nous donne des informations supplémentaires sur la façon dont les clients payant deux prix répartissent leur consommation dans un segment ou dans l'autre lorsque le prix relatif varie. Ces modèles, contrairement au modèle à un prix, n'indiquent pas la variation globale de la consommation d'électricité au regard de son prix.

Les équations que nous avons estimées prennent la forme suivante :

$$w_i = \mu_i + \sum_j \gamma_{ij} \log(P_{ij}) + \beta_p \log\left(\frac{m}{P}\right) + X' \theta$$

où $i = hp, hc$ et $j = hp, hc$ pour les heures pleines et creuses et P est l'indice de référence du prix de l'électricité :

$$P = \sum_j w_j \log(P_j)$$

et où pour finir $X' \theta$ est un ensemble de facteurs de variation de la demande pouvant influencer la demande en électricité. En outre, l'homogénéité et la symétrie sont imposées à l'estimation en limitant les paramètres, de sorte que :

$$\sum_i \gamma_{ij} = 0 \text{ et } \gamma_{ij} = \gamma_{ji}$$

L'élasticité-prix propre et croisée peut être calculée comme suit :

$$\hat{\epsilon}_{ij} = -1 + \frac{\hat{\gamma}_{ij}}{w_i} - \hat{\beta}_m$$

$$\hat{\epsilon}_{ij} = \frac{\hat{\gamma}_{ij}}{w_i} - \hat{\beta}_m \frac{\hat{w}_j}{w_i}$$

où la part des dépenses en électricité peut être estimée en calculant la moyenne de l'échantillon.

² Naturellement, nous limitons notre échantillon aux clients qui payent deux prix et excluons ceux qui payent un prix unique, ainsi que les clients TEMPO.

Pour finir, l'élasticité de substitution est obtenue de la façon suivante :

$$\hat{\sigma}_{ij} = 1 + \frac{\hat{\gamma}_{ij}}{\hat{w}_i \hat{w}_j}$$

4. Résultats

Le tableau 1 présente les résultats du modèle à un prix. L'élasticité-prix de la demande d'électricité s'établit à environ -0.8. Notre résultat semble correspondre aux estimations obtenues dans d'autres études, surtout dans les pays européens. Par exemple, Krishnamurthy & Kriström (2015) obtiennent une élasticité de -0.96 pour la France, c'est-à-dire un résultat très proche du nôtre, en utilisant des données très différentes. Notons également que la corrélation entre la consommation d'électricité et son prix fixe est positive. Ce résultat découle de la structure des contrats, faisant payer un prix supérieur aux ménages qui ont besoin d'une plus grande puissance absorbée

et qui vont donc inévitablement consommer plus. Pour cette raison, et sachant qu'il est impossible de séparer cet effet de l'effet du prix sur la demande, nous incluons le prix fixe afin de contrôler l'effet de l'abonnement d'électricité, mais nous n'interprétons pas ce coefficient comme un effet du prix sur la demande. Cela suggère également que, si l'on utilise le prix moyen pour estimer l'élasticité-prix de la demande d'électricité, cela implique un biais baissier car la composante fixe du prix moyen tend à compenser la relation négative entre le prix par kW et la consommation d'électricité.

Dans le tableau 2, nous reproduisons le modèle précédent mais pour la consommation saisonnière. Dans cette optique, nous séparons les mêmes données entre la consommation d'hiver et celle d'été et examinons l'élasticité durant ces deux saisons. Comme nous pouvons le voir dans ce tableau, l'élasticité est plus forte en hiver qu'en été. Dans une certaine mesure, cela peut sembler illogique car, durant les mois d'hiver,

Tableau 1 – Consommation d'électricité (modèle à un prix)

Variable	Coefficient	Écart-type
Constante	0.7769	0.0117
Log (népérien) du prix variable moyen	0.7997	0.0031
Log (népérien) du prix fixe	1.1044	0.0006
Nombre de jours durant lesquels la température est inférieure à 15 °C	0.0002	0.0000
Nombre de jours enregistrés durant le mois	-0.0035	0.0001
Indicatrices temporelles	Oui	
R2	0.2989	

Note : la variable dépendante est le logarithme népérien de la consommation.
Source : données Enedis ; calculs des auteurs.

Tableau 2 – Consommation d'électricité (modèle saisonnier à un prix)

Variable	Hiver		Été	
	Coefficient	Écart-type	Coefficient	Écart-type
Constante	-0.7053	0.0225	0.9075	0.0150
Log (népérien) du prix variable moyen	-1.1611	0.0050	-0.6358	0.0039
Log (népérien) du prix fixe	1.2279	0.0009	1.0089	0.0007
Nombre de jours durant lesquels la température est inférieure à 15 °C	0.0002	0.0000	0.0003	0.0000
Indicatrices temporelles	Oui		Oui	
R2	0.3054		0.2630	

Note : la variable dépendante est le logarithme népérien de la consommation.
Source : données Enedis ; calculs des auteurs.

les clients consomment plus d'électricité pour se chauffer. Toutefois, l'énergie utilisée pour le chauffage peut venir de sources différentes, telles que le fioul ou le gaz, et d'ailleurs le marché offre plus d'options de chauffage que d'autres types de consommation d'énergie. Cela explique probablement pourquoi, durant l'hiver, les clients sont plus sensibles au prix de l'électricité. L'été, la demande en énergie est habituellement moins importante mais il est plus difficile d'y répondre avec des sources alternatives.

4.1. Système de demande quasi idéal

Dans la mesure où nos données enregistrent la consommation réelle d'électricité et les prix réels, variables directement liées à la consommation en heures pleines et en heures creuses, nous pouvons reproduire, grâce à notre jeu de données représentatif de grande envergure, le modèle AIDS utilisé dans Filippini (1995) et l'appliquer à un modèle saisonnier. Le modèle AIDS nous fournit des informations

supplémentaires sur la façon dont les clients changent leur consommation d'un segment temporel à un autre lorsque le prix relatif de la consommation change dans ces segments. Ainsi, nous obtenons de précieuses informations sur le comportement des consommateurs.

Le tableau 3 présente les résultats du modèle de régression générale et le tableau 4 les élasticités implicites. Nos résultats sont immédiatement comparables avec les estimations de Filippini car, à l'exception des variables de contrôle, la méthodologie est exactement la même. Bien que nos données portent sur un pays et une période différente, nos estimations sont remarquablement semblables à celles de Filippini (tableau 4) : notamment, l'élasticité-prix des heures pleines est de -1.47 dans notre étude et de -1.41 dans celle de Filippini. Nos estimations sont inférieures pour l'élasticité heures creuses, mais celle-ci reste supérieure à l'élasticité heures pleines. Ce résultat n'est pas surprenant car les heures creuses correspondent à une demande moins importante et les clients décident de

Tableau 3 – Part de la consommation d'électricité pendant les heures de pointe (modèle AIDS à deux prix)

Variable	Coefficient	Écart-type
Constante	0.1443	0.0009
logP {pointe}	0.3025	0.0002
logP {hors pointe}	0.3025	0.0002
Log(m/P)	0.0087	0.0001
Log (népérien) du prix fixe	0.0328	0.0001
Nombre de jours durant lesquels la température est inférieure à 15 °C	0.0001	0.0000
Nombre de jours enregistrés durant le mois	0.0031	0.0000
Indicatrices temporelles	Oui	
R2	0.2974	
Nombre d'observations	16 133 468	

Note : procédure SYSLIN, méthode itérative SURE (*Seemingly Unrelated Regression Estimation*).
Source : données Enedis ; calculs des auteurs.

Tableau 4 – Élasticité-prix de la demande d'électricité (modèle à deux prix)

	Cette étude	Filippini (1995a)
Élasticité-prix, pointe	-1.47	-1.41
Élasticité-prix, hors pointe	-1.87	-2.57
Élasticité-prix croisée, pointe / hors pointe	0.46	0.41
Élasticité-prix croisée, hors pointe / pointe	0.85	1.57
Élasticité de substitution	2.32	2.98

Source : données Enedis ; calculs des auteurs.

Tableau 5 – Élasticité-prix de la demande d'électricité (modèle saisonnier à deux prix)

	Cette étude		Filippini (1995a)
	Hiver	Été	
Élasticité-prix, pointe	-1.42	-1.63	-1.41
Élasticité-prix, hors pointe	-1.80	-2.11	-2.57
Élasticité-prix croisée, pointe / hors pointe	0.41	0.61	0.41
Élasticité-prix croisée, hors pointe / pointe	0.78	1.08	1.57
Élasticité de substitution	2.20	2.72	2.98

Source : données Enedis et Filippini (1995a) ; calculs des auteurs.

passer des heures pleines aux heures creuses afin de profiter de prix moins élevés. Globalement, l'élasticité de substitution suggère que, dans le cadre de nos estimations, les deux segments sont légèrement moins substituables que dans l'étude de Filippini, mais la substitution reste importante³. Le tableau 5 montre les résultats du modèle saisonnier, c'est-à-dire que les estimations ne concernent que les mois d'hiver ou d'été. Dans ce cas, nous constatons que les estimations ne sont pas très différentes durant les deux saisons mais observons néanmoins des élasticités légèrement plus fortes en été qu'en hiver. Le modèle à un prix indique que l'élasticité-prix globale de la demande en électricité, s'agissant de l'unique prix variable moyen, est plus importante en hiver qu'en été. En revanche, le modèle à deux prix indique que, dans un contexte de réaction plus prononcée au prix moyen en hiver, l'allocation entre les heures pleines et creuses à ce moment est plus rigide.

* *
*

La capacité à prévoir les prix de façon plus précise, pour la consommation d'énergie et plus particulièrement pour la consommation d'électricité, suscite de plus en plus d'intérêt. Pour le marché de l'électricité, cette question a une importance primordiale car l'électricité produite ne peut pas être stockée et il est donc important d'éviter toute surproduction, tout en garantissant un flux suffisant pour tous les consommateurs. Pour établir des prévisions pertinentes, il est essentiel de comprendre la réaction des consommateurs face aux variations de prix, résumée par le concept d'élasticité de la consommation par rapport au prix. Dans le contexte de l'abondante littérature qui existe à ce sujet, notre principale contribution est double : d'une part, nous corroborons les résultats trouvés

dans cette littérature grâce à un jeu de données largement représentatif et, d'autre part, compte tenu de la richesse de nos données, nous divisons l'échantillon afin de tenir compte des différences entre la consommation constatée à différentes saisons de l'année.

Dans le présent article, nous utilisons des données relatives à la consommation d'électricité en France entre 2007 et 2015 et estimons l'élasticité-prix des dépenses en électricité des ménages privés. Nous proposons trois spécifications différentes pour étudier l'élasticité-prix. Premièrement, nous régressons la consommation d'électricité, sur la base du prix par kilowatt/heure, et obtenons une élasticité par rapport au prix égale à -0.8, soit un résultat remarquablement similaire à ceux de la littérature existante. Dans notre deuxième spécification, nous suivons Filippini (1995) et estimons un modèle de système de demande quasi idéal (AIDS), obtenant des résultats très semblables malgré nos données différentes, notamment des élasticités-prix de -1.46 et -1.86 pour les prix en heures pleines et en heures creuses respectivement (Filippini fait état de -1.41 et -2.57). Notre troisième et dernière spécification est une extension du modèle AIDS intégrant des élasticités saisonnières qui sont différentes en été et en hiver. Avec ce modèle saisonnier, nous obtenons des élasticités de -1.45 et -1.85 en hiver et des valeurs absolues un peu plus élevées en été à -1.61 et -2.08. Dans tous les modèles, nous contrôlons les effets fixes par année et par mois, ainsi que les variables météorologiques et un autre ensemble de variables économiques au niveau du département.

3. La différence peut s'expliquer par le fait que la part du chauffage électrique dans la consommation d'électricité totale est moins importante en Suisse qu'en France (en pourcentage), tandis que la composante « consommation » de l'électricité due au chauffage semble être la moins élastique parmi les ménages.

Notre article soulève également des questions sur la façon dont il serait possible d'améliorer notre compréhension du sujet. Les différences constatées entre les élasticités saisonnières suggèrent qu'il pourrait y avoir des écarts considérables entre les différentes régions de

France (nord/sud, par exemple). Des recherches dans cette direction mettraient probablement en évidence un modèle plus différencié dans le temps et l'espace, qui devrait permettre de mieux estimer et prévoir la consommation d'électricité. □

BIBLIOGRAPHIE

- Alberini, A. & Filippini, M. (2011).** Response of residential electricity demand to price: The effect of measurement error. *Energy Economics*, 33(5), 889–895. <https://doi.org/10.1016/j.eneco.2011.03.009>
- Alberini, A., Gans, W. & Velez-Lopez, D. (2011).** Residential consumption of gas and electricity in the U.S.: The role of prices and income. *Energy Economics*, 33(5), 870–881. <https://doi.org/10.1016/j.eneco.2011.01.015>
- Bernstein, R. & Madlener, R. (2011).** Responsiveness of Residential Electricity Demand in OECD Countries: A Panel Cointegration and Causality Analysis. FCN *Working Paper* N° 8/2011. <http://dx.doi.org/10.2139/ssrn.1887044>
- Dergiades, T. & Tsoulfidis, L. (2008).** Estimating residential demand for electricity in the United States, 1965–2006. *Energy Economics*, 30(5), 2722–2730. <https://doi.org/10.1016/j.eneco.2008.05.005>
- Filippini, M. (1995).** Electricity demand by time of use: An application of the household AIDS model. *Energy Economics*, 17(3), 197–204. [https://doi.org/10.1016/0140-9883\(95\)00017-0](https://doi.org/10.1016/0140-9883(95)00017-0)
- Filippini, M. (2011).** Short- and long-run time-of-use price elasticities in Swiss residential electricity demand. *Energy Policy*, 39(10), 5811–5817. <https://doi.org/10.1016/j.enpol.2011.06.002>
- Halicioğlu, F. (2007).** Residential electricity demand dynamics in Turkey. *Energy Economics*, 29(2), 199–210. <https://doi.org/10.1016/j.eneco.2006.11.007>
- Ito, K. (2014).** Do Consumers Respond to Marginal or Average Price? Evidence from Non-linear Electricity Pricing. *American Economic Review*, 104(2), 537–63. <https://doi.org/10.1257/aer.104.2.537>
- Kamerschen, D. R. & Porter, D. V. (2004).** The demand for residential, industrial and total electricity, 1973–1998. *Energy Economics*, 26(1), 87–100. [https://doi.org/10.1016/S0140-9883\(03\)00033-1](https://doi.org/10.1016/S0140-9883(03)00033-1)
- Krishnamurthy, C. K. B. & Kriström, B. (2015).** A cross-country analysis of residential electricity demand in 11 OECD-countries. *Resource and Energy Economics*, 39, 68–88. <https://doi.org/10.1016/j.reseneeco.2014.12.002>
- Nakajima, T. (2010).** The residential demand for electricity in Japan: An examination using empirical panel analysis techniques. *Journal of Asian Economics*, 21(4), 412–420. <https://doi.org/10.1016/j.asieco.2010.03.005>
- Nakajima, T. & Hamori, S. (2010).** Change in consumer sensitivity to electricity prices in response to retail deregulation: A panel empirical analysis of the residential demand for electricity in the United States. *Energy Policy*, 38(5), 2470–2476. <https://doi.org/10.1016/j.enpol.2009.12.041>
- Narayan, P. K. & Smyth, R. (2005).** Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests. *Energy Policy*, 33(9), 1109–1116. <https://doi.org/10.1016/j.enpol.2003.11.010>
- Okajima, S. & Okajima, H. (2013).** Estimation of Japanese price elasticities of residential electricity demand, 1990–2007. *Energy Economics*, 40, 433–440. <https://doi.org/10.1016/j.eneco.2013.07.026>
- Paul, A., Myers, E. & Palmer, K. (2009).** A Partial Adjustment Model of US Electricity Demand by Region, Season, and Sector. Resources for the Future, *Technical Report*. <http://dx.doi.org/10.2139/ssrn.1372228>

Reiss, P. C. & White, M. W. (2005). Household Electricity Demand, Revisited. *The Review of Economic Studies*, 72(3), 853–883. <https://doi.org/10.1111/0034-6527.00354>

Ros, A. J. (2017). An Econometric Assessment of Electricity Demand in the United States Using Utility-specific Panel Data and the Impact of Retail Competition on Prices. *The Energy Journal*, 38(4). <https://doi.org/10.5547/01956574.38.4.aros>

Réseau de Transport d'électricité (RTE) (2015). Règles relatives à la programmation, au mécanisme d'ajustement et au dispositif de responsable d'équilibre. <https://www.cre.fr/content/download/12998/158245>

Ziramba, E. (2008). The demand for residential electricity in South Africa. *Energy Policy*, 36(9), 3460–3466. <https://doi.org/10.1016/j.enpol.2008.05.026>

RÉSULTATS DÉTAILLÉS DES RÉGRESSIONS

Tableau A-1 – Régression complète pour le tableau 1

Variable	Coefficient	Écart-type
Constante	0.7681	0.0118
Log (népérien) du prix variable moyen	-0.7992	0.0031
Log (népérien) du prix fixe	1.1044	0.0006
Nombre de jours durant lesquels la température est inférieure à 15 °C	0.0002	0.0001
Nombre de jours enregistrés durant le mois	-0.0035	0.0001
Part d'actifs	0.4117	0.0066
Âge moyen de la population	-0.0083	0.0001
Part des logements utilisés comme résidence principale	1.2524	0.0033
Part des maisons dans le nombre total de logements	0.4005	0.0013
Part ayant fait des études supérieures	0.0968	0.0032
Part des logements construits avant 1990	-0.5468	0.0106
Prix pétrole	0.0002	0.0000
Éffets fixes temporels	Oui	
Nombre d'observations	19 768 361	
R2	0.2989	

Note : la variable dépendante est le logarithme népérien de la consommation.
Source : données Enedis ; calculs des auteurs.

Tableau A-2 – Régression complète pour le tableau 2

Variable	Hiver		Été	
	Coefficient	Écart-type	Coefficient	Écart-type
Constante	0.7054	0.0225	0.9075	0.0150
Log (népérien) du prix variable moyen	-1.1611	0.0050	0.6358	0.0040
Log (népérien) du prix fixe	1.2279	0.0009	1.0089	0.0007
Nombre de jours durant lesquels la température est inférieure à 15 °C	0.0003	0.0000	0.0003	0.0000
Nombre de jours enregistrés durant le mois	0.0000	0.0005	0.0021	0.0002
Part d'actifs	0.7482	0.0105	0.1825	0.0085
Âge moyen de la population	0.0086	0.0001	0.0081	0.0001
Part des logements utilisés comme résidence principale	1.3117	0.0053	1.2192	0.0043
Part des maisons dans le nombre total de logements	0.4476	0.0021	0.3686	0.0017
Part ayant fait des études supérieures	0.0299	0.0051	0.1396	0.0042
Part des logements construits avant 1990	0.6773	0.0168	0.4890	0.0136
Prix pétrole	0.0002	0.0000	0.0002	0.0001
Éffets fixes temporels	Oui		Oui	
Nombre d'observations	8 455 612		11 312 749	
R2	0.3054		0.2630	

Note : la variable dépendante est le logarithme népérien de la consommation.
Source : données Enedis ; calculs des auteurs.

**N° 510-511-512 (2019) – NUMÉRO SPÉCIAL 50^{ème} ANNIVERSAIRE / 50th ANNIVERSARY
SPECIAL ISSUE**

- Éditorial / *Editorial* – Jean-Luc Tavernier
- Cinquante ans de résumés d'Economie et Statistique / *Fifty Years of Abstracts in the Journal Economie et Statistique* – Julie Djiriguan & François Sémécurbe
- Préface – Les temps ont changé / *Preface – Times Have Changed* – Daniel Cohen

ÉVOLUTIONS DE LONGUE PÉRIODE / CHANGES OVER THE LONG TERM

- Évolutions de la part du travail dans les pays de l'OCDE au cours des deux dernières décennies / *Labour Share Developments in OECD Countries Over the Past Two Decades* – Mathilde Pak, Pierre-Alain Pionnier & Cyrille Schweltnus
- La part du travail sur le long terme : un déclin ? / *The Labor Share in the Long Term: A Decline?* – Gilbert Cette, Lorraine Koehl & Thomas Philippon
- Croissance économique et pouvoir d'achat des ménages en France : les principales évolutions depuis 1960 / *Economic Growth and Household Purchasing Power in France: Key Changes Since 1960* – Didier Blanchet & Fabrice Lenseigne
- Inégalités de revenus et de richesse en France : évolutions et liens sur longue période / *Income and Wealth Inequality in France: Developments and Links over the Long Term* – Bertrand Garbinti & Jonathan Goupille-Lebret
- Les grandes transformations du marché du travail en France depuis le début des années 1960 / *The Major Transformations of the French Labour Market Since the Early 1960s* – Olivier Marchand & Claude Minni
- Égalité professionnelle entre les femmes et les hommes en France : une lente convergence freinée par les maternités / *Gender Equality on the Labour Market in France: A Slow Convergence Hampered by Motherhood* – Dominique Meurs & Pierre Pora
- Quarante ans d'évolution de l'offre et de la demande de travail par qualification – Progrès technique, coût du travail et transformation sociale / *Forty Years of Change in Labour Supply and Demand by Skill Level – Technical Progress, Labour Costs and Social Change* – Dominique Goux & Éric Maurin

NOUVEAUX ENJEUX / NEW CHALLENGES

- Intelligence artificielle, croissance et emploi : le rôle des politiques / *Artificial Intelligence, Growth and Employment: The Role of Policy* – Philippe Aghion, Céline Antonin & Simon Bunel
- Quelle valeur donner à l'action pour le climat ? / *What Value Do We Attach to Climate Action?* – Alain Quinet

MÉTHODES : ÉVALUATION ET MICROSIMULATION / METHODS: EVALUATION AND MICROSIMULATION

- Inférence causale et évaluation d'impact / *Causal Inference and Impact Evaluation* – Denis Fougère & Nicolas Jacquemet
- L'émergence et la consolidation des modèles de microsimulation en France / *The Emergence and Consolidation of Microsimulation Methods in France* – François Legendre

N° 509 (2019) – BIG DATA ET STATISTIQUES 2^{ème} PARTIE / BIG DATA AND STATISTIQUES PART 2

- Introduction – La chaîne de valeur des données de caisse et des données moissonnées sur le Web / *Introduction – The Value Chain of Scanner and Web Scraped Data* – Jens Mehrhoff
- Les données de caisse : avancées méthodologiques et nouveaux enjeux pour le calcul d'un indice des prix à la consommation / *Scanner Data: Advances in Methodology and New Challenges for Computing Consumer Price Indices* – Marie Leclair, Isabelle Léonard, Guillaume Rateau, Patrick Sillard, Gaëtan Varlet & Pierre Vernédal
- Mesure de l'inflation avec des données de caisse et un panier fixe évolutif / *Inflation Measurement with Scanner Data and an Ever Changing Fixed Basket* – Can Tongur
- Comparaison des indices de prix des vêtements et des chaussures à partir de données de caisse et de données moissonnées sur le Web / *Comparing Price Indices of Clothing and Footwear for Scanner Data and Web Scraped Data* – Antonio G. Chessa & Robert Griffioen
- Écart spatial de niveaux de prix entre régions et villes françaises avec des données de caisse / *Spatial Differences in Price Levels between French Regions* – Isabelle Léonard, Patrick Sillard, Gaëtan Varlet & Jean-Paul-Zoyem

Economie et Statistique / Economics and Statistics

Objectifs généraux de la revue

Economie et Statistique / Economics and Statistics publie des articles traitant de tous les phénomènes économiques et sociaux, au niveau micro ou macro, s'appuyant sur les données de la statistique publique ou d'autres sources. Une attention particulière est portée à la qualité de la démarche statistique et à la rigueur des concepts mobilisés dans l'analyse. Pour répondre aux objectifs de la revue, les principaux messages des articles et leurs limites éventuelles doivent être formulés dans des termes accessibles à un public qui n'est pas nécessairement spécialiste du sujet de l'article.

Soumissions

Les propositions d'articles, en français ou en anglais, doivent être adressées à la rédaction de la revue (redaction-ecostat@insee.fr), en format MS-Word. Il doit s'agir de travaux originaux, qui ne sont pas soumis en parallèle à une autre revue. Un article standard fait environ 11 000 mots en français, y compris encadrés, tableaux, figures, annexes et bibliographie, non compris éventuels compléments en ligne. Aucune proposition initiale de plus de 12 500 mots (11 000 pour les soumissions en anglais) ne sera examinée.

La soumission doit comporter deux fichiers distincts :

- Un fichier d'une page indiquant : le titre de l'article ; le prénom et nom, les affiliations (maximum deux), l'adresse e-mail et postale de chaque auteur ; un résumé de 160 mots maximum (soit environ 1 050 signes espaces compris) qui doit présenter très brièvement la problématique, indiquer la source et donner les principaux axes et conclusions de la recherche ; les codes JEL et quelques mots-clés ; d'éventuels remerciements.
- Un fichier anonymisé du manuscrit complet (texte, illustrations, bibliographie, éventuelles annexes) indiquant en première page uniquement le titre, le résumé, les codes JEL et les mots-clés.

Les propositions retenues sont évaluées par deux à trois rapporteurs (procédure en « double-aveugle »). Les articles acceptés pour publication devront être mis en forme suivant les consignes aux auteurs (accessibles sur <https://www.insee.fr/fr/information/2410168>). Ils pourront faire l'objet d'un travail éditorial visant à améliorer leur lisibilité et leur présentation formelle.

Publication

Les articles sont publiés en français dans l'édition papier et simultanément en français et en anglais dans l'édition électronique. Celle-ci est disponible, en accès libre, sur le site de l'Insee, le jour même de la publication ; cette mise en ligne immédiate et gratuite donne aux articles une grande visibilité. La revue est par ailleurs accessible sur le portail francophone Persée, et référencée sur le site international Repec et dans la base EconLit.

Main objectives of the journal

Economie et Statistique / Economics and Statistics publishes articles covering any micro- or macro- economic or sociological topic, either using data from public statistics or other sources. Particular attention is paid to rigor in the statistical approach and clarity in the concepts and analyses. In order to meet the journal aims, the main conclusions of the articles, as well as possible limitations, should be written to be accessible to an audience not necessarily specialist of the topic.

Submissions

Manuscripts can be submitted either in French or in English; they should be sent to the editorial team (redaction-ecostat@insee.fr), in MS-Word format. The manuscript must be original work and not submitted at the same time to any other journal. The standard length of an article is of about 10,000 words (including boxes if any, tables and figures, appendices, list of references, but not counting online complements if any). Manuscripts of more than 11,000 words will not be considered.

Submissions must include two separate files:

- A one-page file providing: the title of the article; the first name, name, affiliation-s (at most two), e-mail et postal addresses of each author; an abstract of maximum 160 words (about 1050 characters including spaces), briefly presenting the question(s), data and methodology, and the main conclusions; JEL codes and a few keywords; acknowledgements.
- An anonymised manuscript (including the main text, illustrations, bibliography and appendices if any), mentioning only the title, abstract, JEL codes and keywords on the front page.

Proposals that meet the journal objectives are reviewed by two to three referees ("double-blind" review). The articles accepted for publication will have to be presented according to the guidelines for authors (available at <https://www.insee.fr/en/information/2591257>). They may be subject to editorial work aimed at improving their readability and formal presentation.

Publication

The articles are published in French in the printed edition, and simultaneously in French and in English in the online edition. The online issue is available, in open access, on the Insee website the day of its publication; this immediate and free online availability gives the articles a high visibility. The journal is also available online on the French portal Persée, and indexed in Repec and EconLit.

Economie Statistique **ET**

Economics **AND** Statistics

Prochain numéro :
Jeunes et
transitions vers l'âge adulte

Next Issue:
Youth and
Transitions to Adulthood

