



**HAL**  
open science

## Une seule fonction de demande ?

Julien Boelaert

► **To cite this version:**

Julien Boelaert. Une seule fonction de demande ? : Une enquête sur la stabilité des préférences par mélanges discrets de réseaux de neurones. *Revue Economique*, 2014, Vol. 65 (4), pp.515-535. 10.3917/reco.654.0515 . hal-03921556

**HAL Id: hal-03921556**

**<https://cnrs.hal.science/hal-03921556>**

Submitted on 3 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Une seule fonction de demande? Une enquête sur la stabilité des préférences par mélanges discrets de réseaux de neurones.

Julien Boelaert \*

Résumé : Nous présentons un nouveau test de stabilité des préférences, construit à partir d'un modèle de mélange discret de réseaux de neurones artificiels, conçu pour détecter des sous-populations caractérisées par des fonctions de consommation différentes. La procédure est testée sur données simulées et appliquée à deux jeux de microdonnées de consommation des ménages canadiens (1969-1986 et 2004-2008). Nous trouvons que chacun de ces jeux de données rejette la stabilité des préférences. Nous trouvons ensuite, à l'aide de données simulées, que les modèles de demande simples suffisent à la plupart des applications pratiques.

Abstract : We introduce a new test of preference stability across consumers, based on a discrete mixture model of feedforward neural networks, designed to detect clusters of consumers following distinct consumption functions. The procedure is tested on simulated data and applied to two sets of Canadian household consumption microdata (1969-1986 and 2004-2008). On both these datasets we find that two clusters best explain the data, thus rejecting preference stability. Finally we address elasticity estimation, and find on simulated data that for most practical uses simple demand systems suffice.

JEL : C45, D12

---

\*Université Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne

Maison des Sciences Economiques, 106-112 boulevard de l'Hôpital, 75013 Paris

julien.boelaert@gmail.com, 0144078235

Ce travail a bénéficié de l'ANR MALDI 06-BLAN-0140. Nous remercions Pr. Simon Langlois du département de sociologie de l'Université Laval (Québec) pour l'accès aux données de consommation canadiennes, ainsi que Joseph Rynkiewicz, François Gardes, Philip Merrigan, Nicolas Pistoletti, les participants aux XXXèmes Journées de Microéconomie Appliquée et deux rapporteurs anonymes pour leurs remarques.

## Introduction

La stabilité des préférences est une des hypothèses fondamentales qui sous-tendent la grande majorité des travaux sur la demande des consommateurs à ce jour. Énoncée simplement, elle signifie que tous les consommateurs partagent les mêmes préférences, et suivent donc une unique fonction de consommation. Les tests utilisés à ce jour pour évaluer cette hypothèse forment deux familles : d'un côté les tests paramétriques, qui consistent à tester les conditions d'intégrabilité dans des systèmes complets de demande (voir Deaton et Muellbauer (1980a), Selvanathan (1993), Banks et collab. (1997)), de l'autre les tests non paramétriques fondés sur le principe des préférences révélées (voir Varian (2006) pour une revue de littérature). Chacune présente des faiblesses spécifiques : les tests paramétriques reposent sur des hypothèses fortes, comme en particulier la forme spécifique des fonctions d'utilité ou de coût, et ne testent l'intégrabilité que conjointement à ces autres hypothèses<sup>1</sup> ; les tests non paramétriques ne sont capables de falsifier l'hypothèse de stabilité qu'en présence d'intersections des contraintes budgétaires (Varian (2006)) c'est-à-dire de fortes variations de prix et de revenu, phénomène rare lorsque les données couvrent un faible intervalle de temps.

Le présent article présente un nouveau type de test d'universalité des préférences, pensé pour éviter les écueils des deux procédures existantes : il ne pose aucune hypothèse sur la forme des fonctions de demande, et peut rejeter l'universalité même en l'absence d'intersection des contraintes budgétaires. Le point de départ est un système de demande extrêmement flexible, permettant à un nombre prédéterminé de sous-populations de suivre des fonctions de demande très différentes, sans hypothèses préalables sur la forme de ces fonctions. Le principe du test est simple : si des modèles à deux sous-populations ou plus expliquent les données significativement mieux que des modèles à fonction de demande unique, la stabilité des préférences est rejetée. Si l'articulation de réseaux de neurones artificiels à une structure de mélange discret donne une grande flexibilité au modèle de demande, elle présente cependant des inconvénients : l'estimation est difficile, deux exécutions successives de l'algorithme donnent des fonctions de demande (souvent légèrement) différentes - un inconvénient habituel des méthodes d'apprentissage statistique - et les tests fondés sur les critères d'information ne sont pas très précis. L'application de la méthode à deux jeux de microdonnées canadiennes tend à favoriser des modèles à au moins deux sous-populations, et donc à rejeter l'hypothèse de stabilité des préférences.

S'il existe en effet différentes fonctions de consommation qui cohabitent au sein d'une population, un modèle de mélange rend compte bien plus précisément des comportements de consommation que ne le fait un modèle à fonction unique, puisqu'il permet des variations de comportement le long de variables inobservées. Or, le principal but pratique de la modélisation de la consommation étant l'estimation d'élasticités de la demande, il semble important de se demander si des mélanges de modèles flexibles peuvent amélio-

---

1. Cette critique rejoint celle faite par Blackorby et collab. (1978) à l'égard des tests paramétriques de séparabilité de la fonction d'utilité. Reprise par Diewert et Parkan (1985) (traduction de l'auteur) : "la littérature économétrique des tests de l'hypothèse de séparabilité est sérieusement déficiente, en ce que les tests économétriques tendent à imposer des restrictions indésirables sur la forme fonctionnelle paramétrique, en plus de l'hypothèse de séparabilité à tester."

rer cette estimation. Nos résultats sur données simulées montrent que si l'on cherche à estimer des élasticités moyennes, les modèles de mélanges flexibles n'améliorent pas les estimations de modèles plus simples à fonction unique (telles les spécifications log-log ou QUAIDS).

L'article est organisé comme suit. Nous exposons en première section la question de la stabilité des préférences. La deuxième section explique brièvement le fonctionnement des modèles de réseaux de neurones artificiels et des mélanges discrets, leur articulation, et la manière dont ils peuvent servir à tester l'universalité des préférences. Dans la troisième section nous évaluons d'abord la pertinence du modèle sur données simulées, et l'appliquons ensuite aux données canadiennes. La quatrième section est dédiée à la comparaison, sur données simulées, des élasticités estimées par modèles de mélange flexibles à celles estimées à l'aide de spécifications standards. La cinquième section offre quelques commentaires de conclusion.

## Une hypothèse centrale de l'analyse de la demande.

Enoncée simplement, l'hypothèse de préférences universelles signifie que les consommateurs ont les mêmes goûts, à travers le monde et à travers les âges, quels que soient leur culture, leur statut social, leur sexe ou leur âge. Ou, comme l'écrivent Stigler et Becker (Stigler et Becker (1977), traduction de l'auteur) :

“les goûts ne se discutent pas pour la même raison que les montagnes Rocheuses ne se discutent pas - tous deux sont là, le seront également l'année prochaine, et sont les mêmes pour tous les hommes.”

L'hypothèse paraîtrait certainement farfelue à tout autre qu'un économiste, mais elle est un présupposé standard de l'analyse de la demande, et a été justifiée par quelques-uns de ses plus grands théoriciens. Il existe une raison pratique à l'importance de cette hypothèse, puisqu'elle simplifie grandement les travaux théoriques et empiriques. Les préférences déterminent de façon unique les fonctions d'utilité ordinales ; si les préférences changent au fil du temps ou entre les agents, alors différents agents suivent des fonctions de demande différentes, ce qui complique l'agrégation des comportements dans les travaux théoriques, et l'estimation de systèmes de demande empiriques. Si au contraire les préférences sont stables, alors tous les agents partagent une même fonction d'utilité, et donc une même fonction de demande ; dans ce cas les différences de comportements viennent uniquement des contraintes, supposées observables en des variables comme les prix et revenus, et le comportement d'une population entière peut être expliqué par celui d'un agent représentatif (voir Deaton et Muellbauer (1980b)).

Stigler et Becker (1977) justifient la stabilité des préférences en la plaçant dans le contexte de la “nouvelle théorie du consommateur” amorcée par Becker (1965), Michael et Becker (1973) et Muth (1966). Dans ce contexte le consommateur ne dérive pas son utilité directement des biens ou services marchands, mais plutôt de *biens composites* résultant d'un effort de production de la part de l'agent. Cette production domestique combine d'une part les biens marchands et d'autre part le temps, le travail et le savoir qui

interviennent dans la consommation des biens. On remplace alors dans la théorie classique du consommateur les prix et revenus monétaires par des prix et revenus “complets”, qui les généralisent en incluant les coût du temps, du savoir et d’autres facteurs. Stigler et Becker avancent que cette approche permet d’attribuer des différences apparentes de goûts (que l’on retrouve dans des situations comme l’addiction, les habitudes, la publicité et les modes) à des simples différences dans les intrants de la fonction de production domestique, et de les concilier avec la stabilité des préférences si ces préférences portent sur les biens composites et non les biens marchands.

L’approche de Becker-Muth est une approche théorique très puissante, capable de réconcilier de nombreux aspects du comportement humain à la théorie du choix rationnel. Cependant sa portée empirique nous semble limitée, puisque d’une part les biens composites ne sont jamais mesurés (et leur définition précise est au mieux problématique) et d’autre part les intrants de la fonction de production domestique sont difficiles à mesurer (lorsqu’ils peuvent être clairement définis).

Theil et collab. (1981) donnent une autre justification, qui semble de plus grande utilité pratique. Ils avancent que, s’il est certain que les préférences varient lorsqu’elles concernent des catégories restreintes de biens (et que l’on peut s’attendre, par exemple, à ce que les Indiens dépensent peu pour l’achat de viande) elles sont cependant stables lorsqu’elles concernent de larges catégories agrégées de biens. La plupart de leurs travaux empiriques sur données de consommation internationales emploie une décomposition en huit postes : alimentation, habillement, loyer, mobilier, soins médicaux, transports, loisirs, et autres. Il semble raisonnable de supposer que ces grandes catégories représentent les désirs fondamentaux communs à l’humanité entière, et que les consommateurs de tous temps et tous lieux partagent à leur égard un même ordre de préférence<sup>2</sup>.

En outre, Theil apporte une réponse au problème de l’agrégation, par ce qu’on appelle “l’approche par convergence” (voir Clements et collab. (1992)). La question difficile de l’agrégation de la demande (*ie* comment exprimer la demande agrégée d’un groupe d’agents sous la forme d’une fonction conforme à la théorie microéconomique, chaque agent ayant potentiellement une fonction de demande différente) y est résolue en supposant que les consommateurs individuels suivent des fonctions de demande dont les paramètres sont identiquement et indépendamment distribués au sein de la population, et admettent une espérance et une variance finies. Ces conditions suffisent à assurer l’agrégabilité de systèmes complets de demande tels le système de Rotterdam (Barten (1964)), de sorte que la demande macro peut être exprimée par une unique fonction conforme à la théorie micro. Il est intéressant de noter que dans cette approche les préférences individuelles ne sont pas identiques pour tous les consommateurs, mais suivent plutôt une même loi de probabilité (unimodale).

---

2. On peut cependant remarquer que les biens et services qui composent dans les faits ces grandes catégories changent au cours du temps. Par exemple, les dépenses de communication ne servent pas les mêmes désirs à notre époque d’internet et de téléphones mobiles qu’au cours des années 1970. C’est là une des raisons pour lesquelles nous appliquons notre modèle séparément à deux sous-ensembles de données canadiennes, couvrant chacun une période relativement courte.

Il existe deux familles de tests de stabilité des préférences<sup>3</sup>, paramétriques et non paramétriques. Les tests non paramétriques sont fondés sur le principe samuelsonien des préférences révélées ; voir par exemple Dowrick et Quiggin (1994) qui ne rejettent pas la stabilité des préférences sur des données internationales de moyennes par pays, ou Diaye et collab. (2008) qui ne la rejettent pas sur un panel de ménages polonais (au moins pour les ménages dont les conditions de choix n’ont pas changé trop radicalement au cours de la période d’étude). Les tests paramétriques sont construits à partir de systèmes complets de demande, qui peuvent être contraints à respecter les conditions d’intégrabilité (comme dans Deaton et Muellbauer (1980b) ou Selvanathan (1993)).

Dans le présent article nous proposons une procédure alternative de test de la stabilité des préférences, fondée sur l’utilisation de mélanges discrets de réseaux de neurones artificiels. Le but de cette approche est de déterminer si le comportement de consommation d’une population donnée peut être expliqué correctement par une unique fonction, ou si au contraire plusieurs sous-populations suivent des fonctions de demande différentes et ont donc des préférences différentes (ou, en suivant l’approche de Theil, présentent des fonctions de demande dont les paramètres suivent des lois de probabilité différentes, ou encore une loi multimodale).

## Mélanges discrets et perceptrons multicouches

### Perceptrons multicouches

Les perceptrons multicouches (*multilayer perceptrons*, MLP) sont un cas particulier de réseaux de neurones artificiels<sup>4</sup> apparus dans les années 1980 et particulièrement adaptés aux tâches de régression et de classification. Leur succès dans de nombreux domaines est dû au fait qu’ils peuvent, en théorie, approximer n’importe quelle fonction à intrants et extrants multivariés discrets et/ou continus, tout en ayant une forme générale relativement simple. La spécification la plus simple et la plus commune (MLP à unique couche cachée, celle utilisée dans la suite de l’article) est donnée par l’équation 1 :

$$O_l = g \left( b_{0l} + \sum_{j=1}^{N_{hidden}} b_{jl} \cdot f \left( a_{0j} + \sum_{i=1}^{N_{input}} a_{ij} \cdot I_i \right) \right) \quad (1)$$

où  $O_l$  est la  $l$ ème variable de sortie,  $N_{hidden}$  est le nombre prédéterminé de neurones de la couche cachée,  $N_{input}$  est le nombre de variables d’entrée,  $I_i$  est la  $i$ ème variable d’entrée,  $f$  est une fonction non linéaire, habituellement une sigmoïde (pour toute la suite nous choisissons la tangente hyperbolique :  $f(x) = \frac{1-e^{-x}}{1+e^{-x}}$ ),  $g$  une fonction (linéaire ou non, dans toute la suite nous choisissons l’identité,  $g(x) = x$ ), et les  $a_{ij}$  et  $b_{jl}$  sont des paramètres à estimer<sup>5</sup>

3. Ou de façon équivalente les tests d’intégrabilité des fonctions de demande, puisque ces méthodes lient étroitement les deux concepts : en testant l’intégrabilité l’on suppose la stabilité des préférences, et inversement.

4. Pour un exposé détaillé des réseaux de neurones artificiels et leur histoire voir Rojas (1996).

5. Il est intéressant de noter que si l’on choisit pour  $f$  et  $g$  des fonctions linéaires, le MLP se réduit à un modèle linéaire général. En outre, si  $f$  et  $g$  sont respectivement une sigmoïde et l’identité, et que le réseau

L'estimation des paramètres est faite par minimisation d'une fonction de coût, habituellement la somme d'un terme d'erreur de prédiction quadratique moyenne et d'un terme de régularisation qui pénalise les grandes valeurs absolues des paramètres afin d'éviter le sur-apprentissage (comme c'est le cas dans les modèles *ridge regression*). Le vecteur de paramètres estimés  $\tilde{\alpha}$  (qui est simplement la concaténation des  $a_{ij}$  et  $b_{jl}$  de l'équation 1) est donné par :

$$\tilde{\alpha} = \arg \min_{\alpha} \left\{ \beta \times \|\alpha\|^2 + \frac{1}{N} \sum_{n=1}^N \sum_l (O_l^n(x^n, \alpha) - y_l^n)^2 \right\}$$

où  $\beta$  est un nombre réel positif déterminé à l'avance et appelé paramètre de régularisation,  $\|\alpha\|^2$  est la somme des carrés des éléments du vecteur  $\alpha$ ,  $N$  est le nombre total d'observations, l'exposant  $n$  dénote les observations,  $y_l^n$  est la lème variable de sortie observée pour l'observation  $n$ ,  $x^n$  est le vecteur de variables d'entrée de l'observation  $n$ , et  $O_l^n(x^n, \alpha)$  est la lème variable de sortie prédite par le réseau de neurones pour l'observation  $n$ , une fonction d' $\alpha$  et du vecteur d'entrée  $x^n$ . Ce programme de minimisation n'ayant pas, dans le cas général, de solution exacte, les paramètres sont habituellement estimés par un algorithme de descente de gradient itérative<sup>6</sup>, appelée dans ce contexte *rétropropagation* ; dans toute la suite nous utilisons l'algorithme Rprop Riedmiller et Braun (1993).

Des travaux récents (McAlee et collab. (2008), Boelaert (2012)) ont montré que les réseaux de neurones artificiels peuvent être un puissant outil d'analyse de la demande. Ils sont plus flexibles que les systèmes de demande traditionnels, s'ajustent mieux aux données (même avec moins de paramètres), ne sont pas nécessairement (mais peuvent être) contraints par des hypothèses sur le comportement des consommateurs, et donnent de bonnes estimations de courbes d'Engel et d'élasticités de la demande<sup>7</sup>. Dans la suite de cet article nous utilisons le réseau simple étudié dans Boelaert (2012) plutôt que l'architecture hybride (combinant modèle économétrique standard et réseau de neurones) de McAlee et collab. (2008), puisque notre but premier est de libérer les tests de stabilité des préférences de restrictions indésirables sur la forme des fonctions de demande.

---

ne comporte qu'un seul neurone caché, le MLP se réduit à une régression logistique. Le modèle peut donc être vu comme une généralisation de ces deux techniques communes. Cependant, contrairement aux modèles paramétriques habituels, le but d'une estimation par réseau de neurones est de maximiser le pouvoir explicatif du modèle, et non d'estimer correctement ses paramètres pour interprétation. En effet la valeur des paramètres et des neurones cachés n'a généralement pas de signification interprétable, ce qui vaut parfois à ces modèles la qualification de "boîte noire".

6. En partant d'une initialisation aléatoire des éléments d' $\alpha$  (tirée d'une distribution normale d'espérance nulle et de variance faible) on calcule à chaque itération le gradient de la fonction de coût par rapport à  $\alpha$ , et  $\alpha$  est déplacé dans la direction opposée à ce gradient, dans l'espoir d'atteindre un point plus bas de la surface de coût. Cette procédure est normalement répétée jusqu'à convergence, c'est-à-dire lorsqu'un minimum local de la fonction de coût est atteint.

7. Les élasticités sont facilement calculées à partir du modèle 1 si  $f$  et  $g$  sont dérivables.

## Mélange discret de perceptrons multicouches

Les modèles de mélange discrets (Everitt et Hand (1981)) sont des modèles probabilistes dans lesquels on suppose qu'il existe, au sein de la population étudiée, des sous-populations (ou composantes) qui suivent des lois de probabilité différentes. La densité de probabilité de la population entière est supposée être une combinaison convexe des densités des sous-populations. Formellement, supposons un nombre  $K$  de composantes  $(S_1, S_2, \dots, S_K)$ , chaque composante  $S_k$  étant caractérisée par une fonction de densité  $f_k$  (sujette à un vecteur de paramètres  $\theta_k$ ) et un paramètre  $\pi_k$  égal à la part de la population totale occupée par la composante  $S_k$ . La densité  $f$  de la population totale, en notant  $\Theta$  la concaténation des  $\theta_k$ , est alors donnée par :

$$f(y|\Theta) = \sum_{k=1}^K \pi_k f_k(y|\theta_k)$$

La probabilité a posteriori qu'une observation  $n$  appartienne au groupe  $S_k$  est donnée par :

$$P(n \in S_k) = \frac{\pi_k f_k(y_n)}{\sum_{j=1}^K \pi_j f_j(y_n)} \quad (2)$$

Cette architecture générale est facilement adaptée au contexte de la régression en conditionnant  $f$  et les  $f_k$  aux régresseurs, ce qui permet au modèle d'identifier différents comportements pour les différentes composantes<sup>8</sup>. Contrairement aux techniques de *clustering* habituelles, un mélange de modèles de régression capte les différences non dans l'espace des variables mais dans celui des relations entre variables. Ceci constitue à notre avis une voie intéressante et largement inexplorée pour la modélisation de l'hétérogénéité non observée, l'un des principaux points focaux de l'économétrie contemporaine (Heckman (2004)).

Les modèles de mélange discrets sont presque toujours estimés à l'aide de l'algorithme EM (*expectation-maximization*) exposé par Dempster et collab. (1977) et qui permet d'estimer simultanément les  $\pi_k$  et les paramètres  $\theta_k$ . Il s'agit d'un algorithme itératif qui alterne entre deux étapes, la *E-step* au cours de laquelle on calcule la valeur de la log-vraisemblance du modèle à partir des estimations de  $\theta_k$ , et la *M-step* où l'on estime les  $\theta_k$  à partir des estimations de l'appartenance de groupe de l'équation 2. Lorsque l'estimation par maximum de vraisemblance est impossible dans l'étape *M*, comme c'est le cas dans le présent travail, l'algorithme est appelé EM Généralisé (GEM, voir Dempster et collab. (1977)). Son application aux mélanges de perceptrons multicouches demande, pour le calcul de la vraisemblance, une hypothèse concernant la loi suivie par les erreurs. Nous choisissons une distribution gaussienne<sup>9</sup>, sans contrainte sur la matrice de covariance de

---

8. Les modèles de régression utilisant cette architecture sont parfois appelés modèles à classe latente, voir Aitkin et Rubin (1985).

9. Le choix d'une loi normale est certes le plus courant, mais n'est pas forcément justifié. Nous avons trouvé lors d'expériences de simulation non reproduites ici que l'hypothèse de normalité des erreurs est souvent rejetée pour les perceptrons multicouche. La portée de ce problème est cependant réduite par le fait que l'algorithme GEM tient plus de la descente approximative de gradient que de la procédure statistique exacte.



façon à permettre à la loi des erreurs de varier d'une composante à l'autre (voir Celeux et Govaert (1995) pour un traitement détaillé).

Les mélanges de perceptrons multicouches constituent une technique de régression extrêmement flexible, capable de simultanément détecter de multiples sous-populations et d'estimer une fonction de régression différente pour chacune d'elles, sans en spécifier la forme a priori. Cependant leur estimation correcte est problématique, puisqu'ils présentent deux sources d'optima locaux (et donc d'ajustement sous-optimal). Pris séparément, ni les algorithmes d'optimisation de réseaux de neurones ni l'algorithme EM ne présentent de garantie quant à leur convergence vers un minimum global de leur fonction de coût respective; tous deux dépendent des choix d'initialisation, et nécessitent donc de nombreuses exécutions avec différentes initialisations pour l'obtention de bons résultats. La combinaison de ces deux algorithmes prompts à converger vers des solutions sous-optimales fait de l'estimation de mélanges de perceptrons multicouches une tâche gourmande en capacité de calcul, nécessitant de nombreuses exécutions de l'algorithme complet. En outre, la structure correcte du réseaux de neurones (*ie* le nombre de neurones cachés) est inconnue a priori, et une spécification correcte pourrait nécessiter une structure différente pour des réseaux de neurones associés à des sous-populations différentes.

## Déterminer le nombre optimal de composantes

Le nombre  $K$  de composantes est fixé à l'avance lors de l'estimation d'un mélange discret, mais le nombre véritable de sous-populations est évidemment inconnu a priori. L'estimation du  $K$  optimal présente donc un grand intérêt. La solution intuitive que serait un test de ratio de vraisemblance n'est pas appropriée, en raison du problème des bornes des paramètres<sup>10</sup> (voir Cameron et Trivedi (2005)).

Une possibilité alternative est l'utilisation de l'architecture bayésienne exposée dans Aitkin et Rubin (1985), mais la majorité des travaux appliqués de modèles de mélange repose sur l'utilisation des critères d'information. Le principe intuitif de ces critères est simple : à mesure que l'on augmente le nombre de modèles (ou la complexité des modèles) pour expliquer un jeu de données, l'ajustement devient mécaniquement meilleur, l'étape ultime étant l'utilisation d'un modèle distinct pour chaque observation - soit un ajustement parfait. Or cet extrême serait absolument contraire au but de l'analyse, qui est de donner une compréhension parsimonieuse du processus générateur de données. Les critères d'information cherchent à trouver un équilibre entre ces deux objectifs contradictoires, en ajoutant à une mesure de qualité d'ajustement (généralement calculée à partir de la vraisemblance) un terme pénalisant la complexité du modèle (en général mesurée par le nombre de paramètres libres). Idéalement, dans le contexte des mélanges discrets, le critère détermine le nombre minimal de composantes permettant d'expliquer correctement les données.

En particulier, le critère d'information bayésien de Schwartz (BIC) est très largement utilisé pour déterminer le nombre de composantes d'un mélange discret (voir Cameron

---

10. Une intuition du problème est donnée par le fait que si un  $\pi_k$  est nul, les paramètres  $\theta_k$  de la composante correspondante n'existent pas.

et Trivedi (2005)). Il est calculé à partir de la vraisemblance  $\widehat{L}$  du modèle, du nombre  $k$  de paramètres du modèle, et du nombre  $n$  d'observations :

$$BIC = -2\ln(\widehat{L}) + k\ln(n).$$

Dans le contexte des mélange de perceptrons multicouches, Olteanu et Rynkiewicz (2008) prouvent que le BIC est convergent pour l'estimation de modèles de Markov cachés. Dans la section suivante nous évaluons à l'aide de données simulées la pertinence de ce critère pour déterminer le nombre de fonctions de consommation présentes dans une population, avant de l'appliquer à des microdonnées canadiennes.

## Résultats

### Evaluation du BIC sur données de consommation simulées

Comme expliqué dans la section précédente, il n'existe pas de méthode parfaite pour déterminer le bon nombre de composantes d'un mélange, et la plupart des applications reposent sur des critères d'information tels le BIC. Cependant la prudence s'impose, puisqu'il n'existe pas à notre connaissance de résultat théorique sur la convergence de ce critère dans le cadre d'estimation que nous utilisons. Nous nous tournons donc vers des simulations pour évaluer la pertinence du BIC dans un environnement contrôlé, conçu pour ressembler aux microdonnées canadiennes étudiées par la suite.

Les données sont simulées de la façon suivante : on choisit un nombre de sous-populations (de 1 à 3) et leur prévalence dans la population totale. On génère des dépenses totales et des prix suivant des lois log-normales, en gardant la même loi de probabilité pour toutes les sous-populations. On génère alors aléatoirement pour chaque sous-population un ensemble de paramètres de fonction de demande, qui donnent donc une fonction de demande distincte par sous-population, et qui permettent de calculer des coefficients budgétaires simulés. Pour toutes les simulations nous utilisons des fonctions de demande QUAIDS<sup>11</sup> (Banks et collab. (1997)) à cinq catégories de biens. Les paramètres des modèles, générés aléatoirement, sont de grandeurs comparables à celles trouvées dans l'article original, et satisfont aux contraintes d'additivité, d'homogénéité et de symétrie. On ajoute à chaque coefficient budgétaire un bruit blanc, suivant une loi normale d'espérance nulle et d'écart-type égal à un dixième de l'écart-type des simulations du coefficient budgétaire en question. On génère, suivant ce protocole, un grand nombre d'échantillons simulés de 5000 observations, en faisant varier les méta-paramètres (nombre de sous-populations, prévalence, distance entre les fonctions de consommation<sup>12</sup>). On estime sur

---

11. QUAIDS, ou Quadratic Almost Ideal Demand System, est un modèle de demande fréquemment utilisé, prenant comme point de départ une fonction de coût de type *piglog*, et qui étend le modèle original Almost Ideal Demand System de Deaton et Muellbauer (1980a) pour donner aux courbes d'Engel une forme quadratique.

12. Nous définissons la distance entre deux fonctions de consommation comme la distance euclidienne moyenne entre les structures budgétaires obtenues, pour un ensemble donné de vecteurs d'intrants (dépense totale et prix), par prédiction selon les deux fonctions de demande (une structure de consommation étant le vecteur des coefficients budgétaires associés à une observation).

chaque échantillon un ensemble de mélanges discrets de réseaux de neurones, en faisant varier de 1 à 6 le nombre de composantes (20 essais pour chaque nombre de composantes), et on calcule le BIC de chaque modèle estimé.

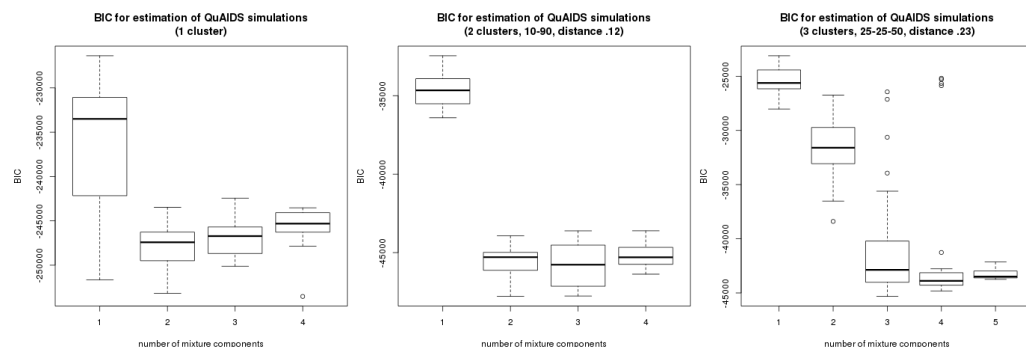


FIGURE 1: Distribution du BIC pour des mélanges discrets de réseaux de neurones, avec différents choix de nombre de composantes, estimés sur données simulées dont différentes sous-population suivent des fonctions de consommation QUAIDS distinctes. Gauche : une seule sous-population, l'ensemble de l'échantillon simulé suit la même fonction de consommation. Centre : deux sous-populations, représentant 10% et 90% de la population totale, avec une distance moyenne de 0,12. Droite : trois sous-populations, représentant 25%, 25% et 50% de la population totale, avec une distance moyenne de 0,23 entre sous-populations. Source : simulations

Les résultats sont qualitativement très similaires pour les différents types de simulations. Le détail exhaustif des résultats est trop volumineux pour être reporté ici, mais est librement disponible sur demande à l'auteur. Les mélanges détectent presque toujours le bon découpage des données : l'indice de Rand ajusté<sup>13</sup> (Hubert et Arabie (1985)) est approximativement égal à 1 lorsque l'on suppose le bon nombre de composantes, et reste proche de l'unité pour un nombre de composantes erroné. La figure 1 montre des résultats typiques de l'expérience en termes de BIC : l'application stricte d'un critère consistant à choisir le BIC le plus faible conduit à favoriser des modèles comportant une composante de trop (par exemple trois composantes alors que les données simulées ne comprennent que deux sous-populations véritablement distinctes). Pour les données à deux sous-populations ou plus, on note une différence claire entre les BIC de modèles à unique composante et ceux de modèles à deux composantes ou plus. Pour les données composées d'une seule fonction de demande il y a une différence statistiquement significative<sup>14</sup> entre les modèles à une et deux composantes, mais cette différence est relativement faible, et les deux distributions se recoupent.

Ces résultats de simulation nous permettent de formuler une règle de décision (quelque

13. Une mesure habituelle de similarité entre deux découpages en sous-populations, qui donne une valeur absolue proche de 1 lorsque les découpages sont très similaires, et proche de 0 s'ils sont très dissemblables. Nous comparons ici le découpage donné par le mélange au vrai découpage, connu, des données simulées.

14. Le test statistique suppose la normalité de la distribution des BIC, hypothèse probablement incorrecte.

peu vague) : le nombre correct de composantes présentes dans les données est donné par le modèle de mélange donnant le dernier “saut” qualitatif de BIC par rapport aux modèles à plus petit nombre de composantes. Les résultats tendent en revanche à indiquer que le BIC est loin d’être un outil parfait dans la détermination du nombre de sous-populations, et qu’il serait souhaitable de pénaliser plus sévèrement la complexité du modèle. Le BIC est pourtant, de ce point de vue, plus sévère que les autres critères d’information habituels (critères d’Akaike et d’Hanna-Quinn), et semble donc le plus approprié.

## Résultats sur microdonnées canadiennes

Nous utilisons les microdonnées de neuf enquêtes de Statistique Canada : l’Enquête sur les dépenses des familles pour les années 1969, 1978, 1982 et 1986, et l’Enquête sur les dépenses des ménages pour les années 2004 à 2008. Ces données de consommation ont été compilées par le professeur Simon Langlois de l’Université Laval de Québec. Les indices de prix sont tirés de l’indice des prix à la consommation de Statistique Canada, combiné aux indices comparatifs des prix des biens et services à la consommation de 2008 afin d’harmoniser les niveaux de prix entre provinces. Nous agrégeons les postes de dépenses suivant la nomenclature COICOP (classification fonctionnelle des dépenses) du Système de comptes nationaux des Nations Unies. Cette nomenclature comporte douze catégories : Alimentation à domicile et boissons non alcoolisées ; Boissons alcoolisées, tabac et narcotiques ; Habillement et chaussures ; Loyer, eau, électricité gaz et autres combustibles ; Mobilier, équipement ménager et maintenance routinière du logement ; Santé ; Transports ; Communication ; Loisirs et culture ; Education ; Restaurants et hôtels ; autres biens et services. Nous désagrégeons les dépenses de Transports entre transports privés et services de transports. Le modèle de mélange tend, sur cette nomenclature, à diviser les ménages selon qu’ils aient ou non une dépense en éducation nulle, ce qui semble une division trop simpliste ; nous excluons donc les dépenses en éducation de notre étude<sup>15</sup>, en recalculant les coefficients budgétaires pour qu’ils se somment à 1. Nous aboutissons donc à un total de douze catégories. Afin d’éliminer les valeurs aberrantes, nous excluons des données les ménages à dépense totale négative, ceux présentant pour au moins un poste une dépense négative, et ceux dont les quantiles de dépense totale et de revenu différent de plus de 30 points. Ces opérations de filtrage éliminent moins de 10% des observations de chaque enquête, et donnent des tailles d’échantillon variant de 8000 à 14000 pour chaque année.

Nous estimons des mélanges discrets de perceptrons multicouches séparément sur deux jeux de données, l’un construit en regroupant les données des années 1969, 1978, 1982 et 1986 (42529 observations), l’autre regroupant les données des années 2004 à 2008 (61448 observations). Notre but est de déterminer si les données sont adéquatement expliquées par une seule fonction de demande (soit un modèle de mélange à une seule composante), afin de tester la stabilité des préférences, entre larges catégories de dépenses agrégées, sur chacun des jeux de données. Le choix de regrouper les données d’années différentes est

---

15. L’exclusion des dépenses d’éducation n’apporte pas de changement qualitatif des résultats en termes de nombre de sous-populations, ce qui indique une robustesse inattendue de la procédure d’estimation.

guidé par la nécessité de prendre en compte les effets de prix, qui seraient insuffisamment estimés année par année (puisque pour chaque année les prix ne varient qu'entre les cinq groupes de provinces). Le choix de la séparation des données en deux sous-périodes relativement courtes est guidé par la nécessité d'étudier des catégories de dépense stables et bien définies ; en effet le contenu de ces catégories a subi des changements profonds entre les années 1970 et les années 2000, l'exemple évident étant le poste Communications. Cette séparation nous permet en outre de déterminer si la structure des préférences a évolué au cours du temps.

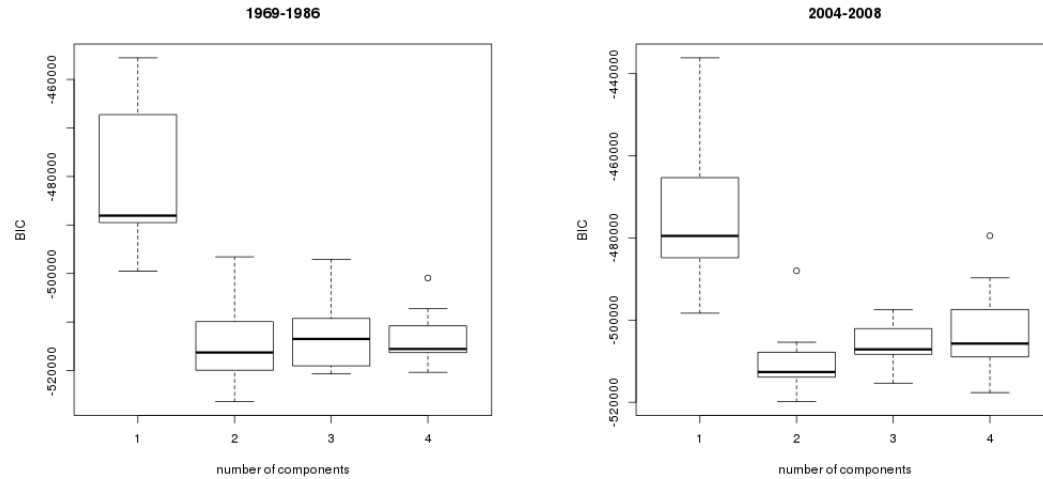


FIGURE 2: Distribution du BIC pour des mélanges discrets de réseaux de neurones appliqués aux données de consommation canadiennes, avec différents nombres de composantes, années 1969, 1978, 1982, 1986 (gauche) et 2004 à 2008 (droite). Source : microdonnées Statistique Canada

Nous estimons les mélanges discrets décrits plus haut en supposant différents nombres de composantes. Pour chaque modèle de mélange, on estime des réseaux de neurones (à cinq neurones cachés) par 100 itérations de l'algorithme Rprop, en appliquant un paramètre de régularisation de 0,2 à la moyenne quadratique des paramètres (hors paramètres de constantes notés  $a_{0j}$  et  $b_{0l}$  dans l'équation 1). L'algorithme GEM est exécuté avec un nombre maximal de 50 itérations. Les variables d'entrée des réseaux de neurones sont la dépense totale logarithmique, les prix logarithmiques, l'âge de la personne de référence du ménage, et des variables indicatrices de l'année d'enquête, le niveau d'éducation de la personne de référence, le type de ménage, la propriété du logement, et la résidence urbaine ; les variables de sortie sont les coefficients budgétaires des douze postes retenus. Pour chaque nombre de composantes choisi, la procédure d'estimation complète est exécutée 20 fois, en tirant chaque fois aléatoirement un échantillon (de 10000 observations, afin de réduire le temps de calcul) et des paramètres initiaux différents. Nous reportons la distribution du BIC pour chaque configuration. Le BIC est calculé en comptant au nombre des paramètres les coefficients estimés des matrices de covariance des erreurs,

mais sans prendre en compte le fait que la régularisation des paramètres des réseaux de neurones diminue le nombre effectif de degrés de liberté du modèle.

La figure 2 montre la distribution du BIC pour des modèles de mélange comportant d'une à quatre composantes, pour chacun des jeux de données. Pour chaque période le BIC favorise les modèles à deux composantes, qui donnent les plus faibles moyennes et médianes, et présentent le seul "saut" qualitatif. Les modèles à deux composantes donnent un BIC moyen significativement différent de celui des modèles à une seule composante (le test T de Welch donnent des valeurs p de 0 pour les deux jeux de données), et le recouvrement des deux distributions est négligeable. Ce résultat montre que les données sont significativement mieux expliquées par deux fonctions de consommation plutôt qu'une seule, et mène donc à rejeter pour chaque période l'hypothèse de stabilité des préférences. Il est intéressant de noter que ce résultat rejoint celui récemment obtenu par Crawford et Pendakur (2013) avec une méthode très différente. A partir d'une approche de préférences révélées sur données en coupe, Crawford et Pendakur trouvent que si une unique fonction de demande ne suffit pas à expliquer les données, "quatre ou cinq types suffisent, [...] deux tiers des données sont compatibles avec un type unique, et deux types suffisent à modéliser 85% des observations" (traduction de l'auteur).

Une interprétation alternative de ce résultat, plutôt que de rejeter l'hypothèse de stabilité des préférences en tant que telle, serait de considérer qu'il existe des variables inobservées qui exercent une grande influence sur les choix de consommation. Cette seconde interprétation signifie cependant elle aussi que l'hypothèse de stabilité est abusive dans un contexte d'économétrie appliquée, puisque le modèle présenté dans le présent travail est construit sur un choix de variables explicatives largement partagé dans la littérature, et que la stabilité est habituellement supposée conditionnellement à ces variables uniquement.

On pourrait objecter aux résultats présentés ci-dessus que le fait de regrouper les données relatives à des années d'enquête différentes conduit à confondre les dimensions temporelle et individuelle, et que la méthode proposée est susceptible de rejeter l'hypothèse de stabilité même si l'instabilité ne concerne en réalité qu'une seule de ces dimensions. En particulier, si les préférences sont universelles en coupe instantanée mais qu'elles varient dans le temps, le modèle proposé rejettera la stabilité. Afin de contrôler cette possibilité nous appliquons le protocole d'estimation décrit dans la section précédente à chaque année d'enquête prise séparément, en modifiant uniquement la taille d'échantillon (5000 observations par exécution de l'algorithme GEM). La table 1 reporte les résultats de BIC de ces estimations, qui indiquent clairement que le rejet de la stabilité des préférences tient encore pour chaque enquête : les modèles à deux composantes ou plus donnent toujours un BIC significativement plus faible que leurs équivalents à une seule composante. Cette table fournit en outre un résultat étonnant (et inexpliqué) : l'estimation des modèles à plusieurs composantes est beaucoup moins robuste pour les années récentes, dont le BIC présente un écart-type généralement plus élevé que pour les enquêtes plus anciennes.

En retenant le modèles à deux composantes donnant le meilleur BIC, nous pouvons déterminer les caractéristiques des sous-populations détectées. Pour la période 1969-1986 le meilleur modèle donne des parts de 40% et 60% aux deux sous-populations, et pour

Année	1 composante	2 composantes	3 composantes	4 composantes
1969	-266884.0 <i>8031.4</i>	-284119.1 <i>4909.7</i>	-287579.6 <i>4240.1</i>	-293198.5 <i>11588.6</i>
1978	-269153.5 <i>6163.2</i>	-288619.6 <i>3255.1</i>	-293392.3 <i>2862.4</i>	-296254.4 <i>13303.7</i>
1982	-271198.7 <i>8434.9</i>	-289176.0 <i>4102.0</i>	-292878.8 <i>2771.6</i>	-294464.1 <i>11754.4</i>
1986	-267084.7 <i>7756.8</i>	-287736.1 <i>3661.2</i>	-289446.3 <i>3114.1</i>	-290497.8 <i>14185.1</i>
2004	-259167.4 <i>5463.9</i>	-283196.7 <i>3345.7</i>	-301885.6 <i>28235.6</i>	-305088.1 <i>31127.2</i>
2005	-257473.0 <i>9325.6</i>	-282929.8 <i>12735.1</i>	-312463.2 <i>29216.2</i>	-315594.9 <i>30038.6</i>
2006	-257689.5 <i>9019.9</i>	-282914.3 <i>6319.7</i>	-301063.3 <i>27671.1</i>	-322860.6 <i>26915.0</i>
2007	-256002.7 <i>10849.6</i>	-315978.8 <i>32409.4</i>	-333004.1 <i>30003.2</i>	-347643.9 <i>4398.0</i>
2008	-261034.2 <i>9128.0</i>	-296984.3 <i>24823.9</i>	-331693.5 <i>28869.3</i>	-347093.7 <i>10393.4</i>

TABLE 1: BIC et son écart-type pour des mélanges discrets de perceptrons multicouches à différents nombres de composantes, estimés séparément pour chaque année d'enquête (écarts-types en italiques). Source : microdonnées Statistique Canada

2004-2008 des parts de 49% et 51%. Il est intéressant de noter que pour chaque période les variables d'entrée sont distribuées de façon très similaire d'une composante à l'autre : comme indiqué plus haut, un mélange discret de régressions détecte des différences dans les relations entre variables, contrairement aux techniques de clustering habituelles qui détectent des différences de position dans l'espace des variables. Comme attendu, les variables de sortie montrent, elles, une certaine disparité entre sous-populations. La figure 3 montre, à titre d'illustration, les courbes d'Engel estimées de deux postes de dépenses pour lesquels elles diffèrent beaucoup entre les deux composantes.

Cependant une analyse des coefficients budgétaires moyens ou des élasticités moyennes par sous-population détectée n'a pas beaucoup de sens, parce que la méthode ne permet d'identifier les sous-populations de façon robuste (raison pour laquelle nous ne reportons pas ces moyennes dans ce texte). En effet, la prévalence des sous-populations varie grandement d'une estimation à l'autre : pour chaque période, au cours des 20 exécutions de l'algorithme, la prévalence de la plus petite composante varie de 5% à 50%. En exécutant l'algorithme de nombreuses fois sur un même échantillon on obtient un indice de Rand (ajusté) moyen de 0,30 ce qui confirme le manque de robustesse dans la détection des sous-populations. Ce manque de robustesse a deux sources : la première est le double problème de minima locaux décrit plus haut, qui rend difficile l'estimation correcte du

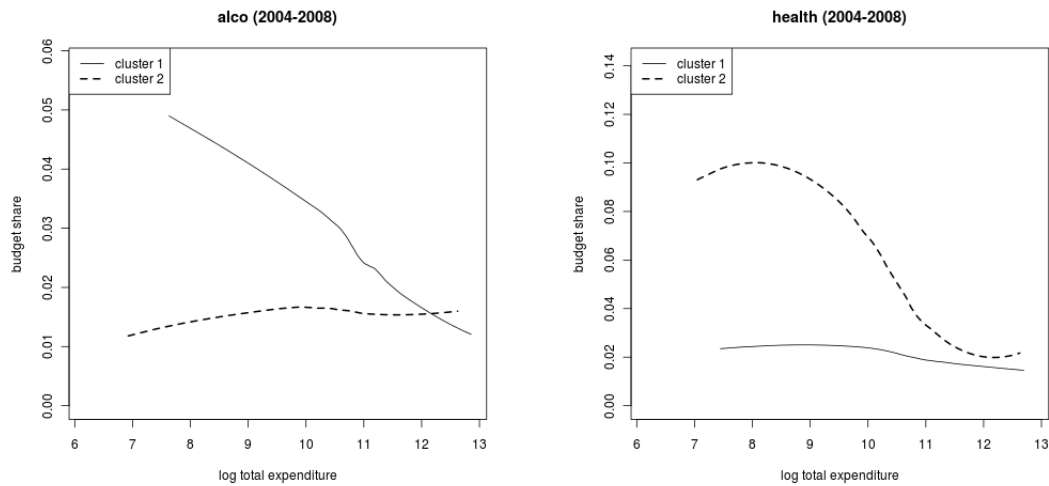


FIGURE 3: Courbes d'Engel pour les deux composantes détectées sur le jeu de données 2004-2008, postes Alcool et Tabac (gauche) et Santé (droite). Lissage par lowess (paramètre 0,4). Source : microdonnées Statistique Canada

modèle. La seconde est un problème d'identification inhérent à l'application de modèles mélange discrets à des tâches de régression, illustré par la figure 4 : le modèle étant libre de s'adapter aux données avec un minimum de restrictions, il est incapable de choisir entre deux modélisations différentes si elles donnent une qualité d'ajustement équivalente. Ainsi des statistiques calculées séparément pour chaque sous-population n'auront-elles que peu de valeur ; nous nous tournons dans la section suivante vers des statistiques calculées comme moyennes entre sous-populations, qui tendent à être plus robustes.

## Estimation d'élasticités de la demande

Le principal objectif pratique de l'estimation de systèmes de demande est le calcul d'élasticités-prix et d'élasticités-revenu. Le modèle flexible à plusieurs fonctions exposé plus haut est potentiellement un outil puissant d'estimation des élasticités, en particulier parce qu'il traite naturellement l'hétérogénéité non observée. En outre, s'il existe réellement plus d'une fonction de consommation au sein de la population observée, comme tendent à l'indiquer nos résultats sur données canadiennes, il semble intuitivement juste que des modèles à fonction unique soient moins bons que des modèles à multiples fonctions. Dans la présente section nous étudions cette question à l'aide de données simulées, qui nous offrent le confort incomparable de la connaissance des vraies élasticités.

En utilisant le même protocole de simulation que précédemment, on génère des jeux de 5000 observations divisées en sous-populations caractérisées par des fonctions de demande différentes (spécifiées comme des fonctions QUAIDS à cinq postes de dépenses, aux paramètres aléatoires), les variables d'entrée étant distribuées identiquement sur la population entière. On ajuste alors différents modèles à ces données :



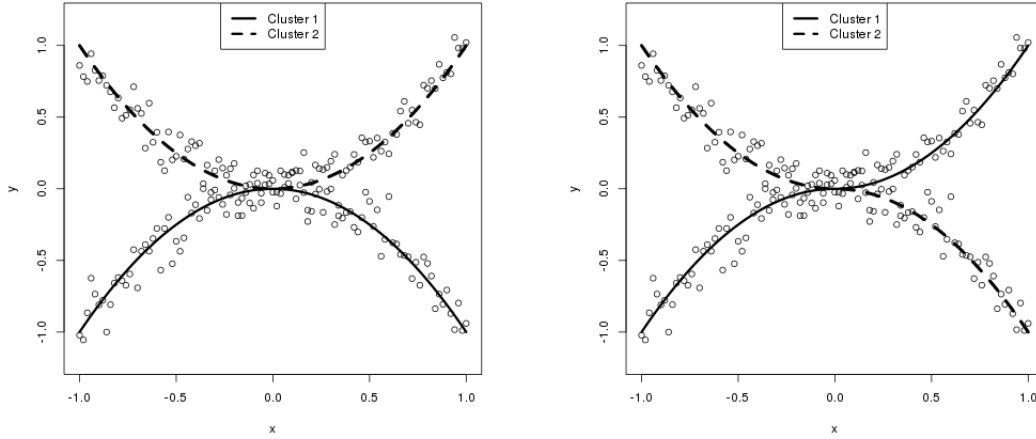


FIGURE 4: Exemple du problème d'identification inhérent aux mélanges mixtes de régressions flexibles : l'algorithme est incapable de décider si les deux fonctions détectées se croisent (droite) ou non (gauche). Des exécutions successives alterneront entre ces deux résultats. Source : données simulées

- une spécification log-log (où  $q_i$  est la dépense pour le poste  $i$ ,  $y$  représente la dépense totale,  $p_j$  le prix du bien  $j$ , et  $\varepsilon_i$  le terme d'erreur)

$$\ln(q_i) = \alpha_i + \beta_i \ln(y) + \sum_j \gamma_{ij} \ln(p_j) + \varepsilon_i$$

- une spécification QUAIDS (Banks et collab. (1997)) (où  $w_i$  est le coefficient budgétaire du poste  $i$ ,  $y$  la dépense totale,  $p_j$  le prix du bien  $j$ , et  $\varepsilon_i$  le terme d'erreur)

$$w_i = \alpha_i + \sum_{j=1}^n \gamma_{ij} \ln(p_j) + \beta_i \ln\left(\frac{y}{a(\mathbf{p})}\right) + \frac{\lambda_i}{b(\mathbf{p})} \left(\ln\left(\frac{y}{a(\mathbf{p})}\right)\right)^2 + \varepsilon_i$$

$$\ln a(\mathbf{p}) = \alpha_0 + \sum_{i=1}^n \alpha_i \ln(p_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} \ln(p_i) \ln(p_j)$$

$$b(\mathbf{p}) = \prod_{i=1}^n p_i^{\beta_i}$$

- un réseau de neurones artificiels, de la même forme que ceux utilisés dans les composantes du modèle de mélange
- un mélange discret de perceptrons multicouches, en supposant le bon nombre de sous-populations

	poste 1	poste 2	poste 3	poste 4	poste 5
vraie (sous-pop. 1)	0.92	0.94	1.34	0.96	1.00
vraie (sous-pop. 2)	1.11	1.53	0.86	0.85	0.91
vraie (moyenne)	1.05	1.36	1.00	0.88	0.93
log-log	1.04	1.33	1.04	0.88	0.93
QUAIDS	1.04	1.26	0.95	0.87	0.94
MLP	1.05	1.27	0.94	0.87	0.94
mélange MLP	1.05	1.35	1.00	0.88	0.94

TABLE 2: Élasticité-revenu (ou dépense totale) de données QUAIDS simulées (deux sous-populations couvrant 30% et 70% de la population). Vraie élasticité par sous-population, élasticité moyenne sur la population entière, et élasticités moyennes estimées par spécifications log-log, QUAIDS, perceptron multicouche (MLP), et mélange discret de réseaux de neurones. Source : données simulées

	poste 1	poste 2	poste 3	poste 4	poste 5
vraie (sous-pop 1)	1.32	0.93	0.84	1.27	1.12
vraie (sous-pop 2)	0.74	1.32	0.98	1.13	1.01
vraie (sous-pop 3)	1.00	1.10	1.22	0.99	0.92
vraie (moyenne)	0.93	1.17	1.09	1.07	0.97
log-log	0.91	1.19	1.12	1.07	0.93
QUAIDS	0.92	1.18	1.07	1.04	0.93
MLP	0.92	1.16	1.10	1.04	0.93
mélange MLP	0.92	1.16	1.09	1.07	0.98

TABLE 3: Élasticité-revenu (ou dépense totale) de données QUAIDS simulées (trois sous-populations couvrant 10%, 40% et 50% de la population). Vraie élasticité par sous-population, élasticité moyenne sur la population entière, et élasticités moyennes estimées par spécifications log-log, QUAIDS, perceptron multicouche (MLP), et mélange discret de réseaux de neurones. Source : données simulées

Les tables 2 et 3 montrent les élasticités moyennes estimées sur des simulations à respectivement deux et trois sous-populations. Les résultats sont très stables entre différentes exécutions de la simulation. Tous les modèles donnent de bons résultats, le mélange de MLP étant souvent légèrement plus proche des vraies élasticités. Les spécifications QUAIDS et perceptron multicouche simple donnent des résultats très proches. La spécification log-log, de très loin la plus simple puisqu'elle impose une même élasticité à toutes les observations, donne remarquablement des résultats légèrement meilleurs que ces deux spécifications sophistiquées, probablement parce que l'estimation d'une élasticité moyenne est précisément le but de ce modèle.

L'un des avantages théoriques des modèles semi-paramétriques, et a fortiori des mélanges de spécifications semi-paramétriques, vis-à-vis de modèles de demande traditionnels est leur capacité à estimer plus précisément des élasticités locales : puisqu'ils im-

Quintile 1	poste 1	poste 2	poste 3	poste 4	poste 5
true (mean)	0.91	1.39	0.92	1.01	0.96
log-log	0.91	1.19	1.12	1.07	0.93
QUAIDS	0.85	1.44	1.22	0.95	0.90
MLP	0.92	1.17	1.11	1.04	0.94
mélange MLP	0.95	1.17	0.90	1.07	0.96
Quintile 3	poste 1	poste 2	poste 3	poste 4	poste 5
true (mean)	0.92	1.15	1.10	1.08	0.98
log-log	0.91	1.19	1.12	1.07	0.93
QUAIDS	0.92	1.16	1.06	1.05	0.93
MLP	0.92	1.16	1.10	1.04	0.93
mélange MLP	0.91	1.16	1.12	1.08	0.98
Quintile 5	poste 1	poste 2	poste 3	poste 4	poste 5
true (mean)	0.95	1.01	1.24	1.12	0.98
log-log	0.91	1.19	1.12	1.07	0.93
quaid	0.99	0.96	0.93	1.13	0.96
MLP	0.93	1.15	1.09	1.04	0.93
mélange MLP	0.91	1.15	1.22	1.07	0.99

TABLE 4: Élasticité-revenu (ou dépense totale) de données QUAIDS simulées (trois sous-populations couvrant 10%, 40% et 50% de la population), moyennes par quintile de dépense totale. Vraie élasticité par sous-population, élasticité moyenne sur la population entière, et élasticités moyennes estimées par spécifications log-log, QUAIDS, perceptron multicouche (MLP), et mélange discret de réseaux de neurones. Source : données simulées

posent moins de contraintes à la fonction de consommation, ils devraient permettre aux élasticités de varier plus librement entre les observations. Subséquemment, ils devraient donner de meilleures estimations d'élasticités pour des sous-groupes de la population tels que des quintiles de dépense totale. La table 4 présente les élasticités moyennes pour trois quintiles, à partir du même protocole de simulation et d'estimation que pour la table 3. Étonnamment, les résultats n'indiquent pas clairement de vainqueur : pour chaque spécification la qualité des élasticités estimées varie d'un poste à l'autre. Le modèle de mélange donne des résultats presque parfaits pour les postes 3 et 5 uniquement, le QUAIDS tend à exagérer les variations d'élasticités, le réseau de neurones simple tend au contraire à les aplanir, et la simple spécification log-log reste compétitive malgré la rigidité de ses élasticités constantes.

Ces simulations montrent que bien que les mélanges de réseaux de neurones soient faits pour pouvoir traiter de multiples sous-populations, ils ne font pas mieux que les spécifications traditionnelles à fonction unique quand il s'agit d'estimer des élasticités moyennes. En outre leurs estimations seraient probablement moins robustes sur des données réelles, dont le découpage en sous-populations serait moins clair que celui de nos données simulées. Les modèles à unique fonction de demande se montrent performants

même en présence de données divisées en sous-populations distinctes. Il pourrait être intéressant d'apporter une variation au modèle présenté ici, en articulant à l'architecture de mélange discret une spécification moins flexible que le réseau de neurones artificiels : la simplification de la procédure d'estimation et le supplément de structure pourraient grandement améliorer la précision et la robustesse de l'estimation des élasticités. Un tel modèle dépasse néanmoins le cadre du présent article, dont le but était de dépouiller la modélisation de la demande d'un maximum d'hypothèses afin de tester l'hypothèse de stabilité des préférences.

## Conclusion

La stabilité des préférences est une hypothèse centrale de la théorie du consommateur, communément supposée dans les travaux appliqués mais difficile à tester. La présente contribution tente de dépasser les défauts des procédures de test existantes, à l'aide de puissants outils numériques originaires du domaine de l'apprentissage statistique. Les modèles de mélange discrets de perceptrons multicouches permettent une modélisation extrêmement flexible, mais sont en retour difficiles à estimer correctement. En outre la procédure de test par BIC proposée est moins stable que les procédures existantes.

Nous rejetons l'hypothèse de stabilité des préférences sur deux jeux de microdonnées canadiennes : tant en 1969-1986 qu'en 2004-2008 il semble exister plus d'une fonction de consommation au sein de la population canadienne.

Les modèles traditionnels à unique fonction de consommation ne sont certes pas capables de rendre compte de cette diversité de comportements, mais nos résultats sur données simulées semblent indiquer qu'ils sont des outils suffisamment performants d'estimation d'élasticités moyennes. Les modèles de demande flexibles à plusieurs fonctions semblent donc d'une utilité pratique limitée, même s'ils permettent une meilleure compréhension de la structure de la demande. Des modèles légèrement plus contraints pourraient en revanche dépasser les performances des spécifications traditionnelles en termes d'estimation d'élasticités, et permettre en retour des applications plus fines de micro-simulation.

## Références

- Aitkin, M. et D. B. Rubin. 1985, «Estimation and hypothesis testing in finite mixture models», *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 47, n° 1, p. 67–75.
- Banks, J., R. Blundell et A. Lewbel. 1997, «Quadratic engel curves and consumer demand», *Review of Economics and Statistics*, vol. 79, n° 4, p. 527–539.
- Barten, A. P. 1964, «Consumer demand functions under conditions of almost additive preferences», *Econometrica*, vol. 32, p. 1–38.

- Becker, G. 1965, «A theory of the allocation of time», *The Economic Journal*, vol. 75, n° 299, p. 493–517.
- Blackorby, C., D. Primont et R. R. Russell. 1978, *Duality, separability, and functional structure : theory and economic applications*, North-Holland, New York.
- Boelaert, J. 2012, *La convergence internationale et intranationale des structures de consommation des ménages*, thèse de doctorat, Université Paris 1 Panthéon-Sorbonne, Centre d'économie de la Sorbonne.
- Cameron, A. C. et P. K. Trivedi. 2005, *Microeconometrics, methods and applications*, Cambridge University Press.
- Celeux, G. et G. Govaert. 1995, «Gaussian parsimonious clustering models», *Pattern Recognition*, vol. 28, n° 5, p. 781–793.
- Clements, K. W., E. A. Selvanathan et S. Selvanathan. 1992, «Henri Theil's contributions to demand analysis», dans *Contributions to consumer demand and econometrics, essays in honour of Henri Theil*, édité par R. Bewley et T. Van Hoa, Macmillan, p. 74–104.
- Crawford, I. et K. Pendakur. 2013, «How many types are there?», *The Economic Journal*, vol. 123, n° 567, p. 77–95.
- Deaton, A. et J. Muellbauer. 1980a, «An almost ideal demand system», *The American Economic Review*, vol. 70, n° 3, p. 312–326.
- Deaton, A. et J. Muellbauer. 1980b, *Economics and Consumer Behavior*, Cambridge University Press.
- Dempster, A. P., N. M. Laird et D. B. Rubin. 1977, «Maximum likelihood from incomplete data via the EM algorithm», *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, n° 1, p. 1–38.
- Diaye, M. A., F. Gardes et C. Starzec. 2008, «GARP violation, economic environment distortions and shadow prices : evidence from household expenditure panel data», *Annales d'économie et de statistique*, vol. 90, p. 3–33.
- Diewert, W. E. et C. Parkan. 1985, «Tests for the consistency of consumer data», *Journal of Econometrics*, vol. 30, p. 127–147.
- Dowrick, S. et J. Quiggin. 1994, «International comparisons of living standards and tastes : a revealed-preferences analysis», *The American Economic Review*, vol. 84, n° 1, p. 332–341.
- Everitt, B. et D. Hand. 1981, *Finite mixture distributions*, Chapman & Hall.
- Heckman, J. J. 2004, «Micro data, heterogeneity and the evaluation of public policy : Part 1», *The American Economist*, vol. 48, n° 2, p. 3–25.

- Hubert, L. et P. Arabie. 1985, «Comparing partitions», *Journal of Classification*, p. 193–218.
- McAleer, M., M. C. Medeiros et D. Slottje. 2008, «A neural network demand system with heteroskedastic errors», *Journal of Econometrics*, vol. 147, p. 359–371.
- Michael, R. et G. S. Becker. 1973, «On the new theory of consumer behavior», *Swedish Journal of Economics*, vol. 75, n° 4, p. 378–396.
- Muth, R. F. 1966, «Household production and consumer demand functions», *Econometrica*, vol. 34, n° 3, p. 699–708.
- Olteanu, M. et J. Rynkiewicz. 2008, «Estimating the number of components in a mixture of multilayer perceptrons», *Neurocomputing / EEG Neurocomputing*, vol. 71, n° 7-9, p. 1321–1329.
- Riedmiller, M. et H. Braun. 1993, «A direct adaptive method for faster backpropagation learning : The rprop algorithm», dans *Proceedings of the IEEE International Conference on Neural Networks (ICNN) 1993*, édité par H. Ruspini, p. 586–591.
- Rojas, R. 1996, *Neural networks*, Springer-Verlag.
- Selvanathan, S. 1993, *A System-Wide Analysis of International Consumption Patterns*, Kluwer Academic Publishers.
- Stigler, G. J. et G. S. Becker. 1977, «De gustibus non est disputandum», *The American Economic Review*, vol. 67, n° 2, p. 76–90.
- Theil, H., F. E. Suhm et J. F. Meisner. 1981, *International Consumption Comparisons : a System-Wide Approach*, Elsevier.
- Varian, H. 2006, «Revealed preference», dans *Samuelsonian economics and the 21st century*, édité par M. Szenberg, L. Ramrattan et A. A. Gottesman, Oxford university press.