

1 Hétérogénéité des préférences, perception de la qualité et choix
2 d'alimentation en eau potable en France
3 Un logit multinomial à facteurs d'échelle

4 Olivier Beaumais et Paul Veyronnet

5 April 28, 2016

6 **Abstract**

7 Nous explorons les déterminants des choix individuels d'alimentation en eau potable en France. Parmi
8 ces déterminants, nous nous concentrons plus particulièrement sur la qualité perçue de l'eau du robinet
9 mesurée par le niveau de satisfaction déclaré par les individus.

10 Cette variable étant susceptible d'être endogène, nous proposons une méthodologie novatrice, com-
11 binant l'estimation d'un modèle logit multinomial à facteurs d'échelle (*scaled multinomial logit*) et une
12 correction de l'endogénéité par la méthode d'inclusion des résidus en deux étapes (*two-stage residual*
13 *inclusion*).

14 En utilisant une coupe instantanée de 4003 individus, nous montrons que le niveau de satisfaction
15 est bien un déterminant des choix individuels d'alimentation en eau potable. Nous montrons également
16 que la correction du problème d'endogénéité change sensiblement les résultats relatifs aux sources de
17 l'hétérogénéité des comportements individuels d'alimentation en eau potable.

18 **1 Introduction**

19 Quels sont donc les déterminants des choix individuels d'alimentation en eau potable en France ? L'eau de
20 boisson¹, bien de première nécessité, peut en effet apparaître pour l'économiste comme entourée d'un halo de
21 mystère. Depuis trente ans, en France, la consommation d'eau en bouteille a été multipliée par trois. Or, le
22 prix de l'eau robinet est environ cent fois moins élevé que le prix de son proche substitut, l'eau en bouteille.

¹En France, l'eau de boisson représente environ 1% de la consommation domestique d'eau.

23 Pour explorer cette question, nous disposons d'une base de données originale résultant d'une enquête
24 menée en 2013 par le Commissariat général au développement durable auprès d'un échantillon de 4003
25 individus représentatif de la population française, sur la perception des ressources en eau des français et sur
26 leurs habitudes consommation d'eau potable (CGDD, 2014)².

27 Cette enquête montre que 60% de la population ne connaît pas le prix de l'eau du robinet ; seulement
28 20% de la population semble capable d'indiquer un prix compatible avec le prix réellement pratiqué (environ
29 3,4 € par m³). La différence de prix entre l'eau du robinet et l'eau en bouteille est sous-estimée par environ
30 la moitié de la population, 27% étant incapables de se prononcer sur la différence en question.

31 Pour le régulateur public, cela pose la question de l'instrument pertinent pour agir sur les comporte-
32 ments individuels de consommation d'eau potable. Outre l'impact sur les ressources en eau, l'achat d'eau
33 en bouteille s'accompagne d'une pollution importante, tant en amont qu'en aval de leur production ; le
34 polyéthylène téréphtalate (PET), matière plastique utilisée très majoritairement pour le conditionnement
35 de l'eau en bouteille, est issu de la transformation du pétrole brut, avec un taux de recyclage faible³, et
36 des impacts en termes de pollution visuelle, de pollution des cours d'eau et des océans très significatifs.
37 Compte tenu de l'ignorance répandue du prix de l'eau au sein de la population française, le recours à la
38 tarification pour infléchir les choix les individuels d'alimentation en eau potable paraît peu opérant. Plus
39 que son prix, les caractéristiques perçues de l'eau du robinet, comme bien d'expérience, jouent certainement
40 un rôle central dans les choix d'alimentation en eau potable, ouvrant des possibilités de régulation publique
41 par l'information du consommateur. Ainsi, la qualité perçue de l'eau du robinet, comme déterminant du
42 choix d'alimentation en eau potable est au centre de notre analyse. Ce choix étant de nature discrète, ici
43 selon trois modalités (eau du robinet, eau en bouteille, eau du robinet filtrée), cela nous conduit logique-
44 ment à mobiliser les techniques d'économétrie des variables qualitatives pour explorer les données dont nous
45 disposons.

46 Aussi, afin de modéliser finement l'hétérogénéité des comportements individuels, nous recourons à un
47 modèle de portée très générale : le modèle logit multinomial à facteurs d'échelle (*scaled multinomial logit*,
48 Fiebig *et al.*, 2010 ; Greene et Hensher, 2010). Ainsi que nous le soulignerons, l'intérêt de ce modèle réside
49 dans son aptitude à capturer l'hétérogénéité observée et non-observée des individus. Comme mesure de la
50 qualité perçue de l'eau du robinet, nous disposons du niveau de satisfaction vis-à-vis de l'eau du robinet
51 déclaré par les répondants de l'enquête menée par le Commissariat général au développement durable.
52 Cette mesure est une variable attitudinale susceptible, comme déterminant du choix d'alimentation en eau

²Nous remercions vivement le CGDD pour nous avoir donné accès gracieusement à ces données de grande qualité.

³En France, le taux de recyclage des matières plastiques s'établit à seulement 25% (Ademe, 2015).

53 potable, de poser un problème d'endogénéité. Nous traitons ce problème d'endogénéité en appliquant une
54 méthode novatrice, notamment présentée par Terza et *al.* (2008) : la méthode d'inclusion des résidus en
55 deux étapes (*two-stage residual inclusion*, 2SRI). Cette méthode, adaptée aux modèles non-linéaires tel
56 que le modèle logit multinomial à facteurs d'échelle, nous permet notamment d'établir rigoureusement que
57 l'essentiel de l'hétérogénéité des comportements de choix d'alimentation en eau potable peut être reliée à des
58 caractéristiques observées des individus.

59 En somme, cet article contribue tant du point de vue empirique que méthodologique à la littérature
60 existante. D'un point de vue empirique, il s'agit, à notre connaissance, de la contribution la plus aboutie
61 sur les choix d'alimentation en eau potable. D'un point de vue méthodologique, nous présentons un modèle
62 novateur, le modèle logit multinomial à facteurs d'échelle, ainsi qu'une méthode appropriée pour corriger
63 l'endogénéité dans ce type de modèle, d'application potentielle à bien d'autres questions de recherche que
64 celle qui nous occupe ici.

65 Le reste du papier est organisé comme suit : la section 2 présente une rapide revue de la littérature, suivie
66 dans la section 3 par une présentation des données dont nous disposons. La section 4 décrit le modèle logit
67 multinomial à facteurs d'échelle, ainsi que la méthode de correction de l'endogénéité. La section 5 présente
68 les résultats, tandis que la section 6 apporte des éléments de conclusion.

69 2 Revue de la littérature

70 En France, l'unique contribution sur le choix d'alimentation en eau potable dont nous ayons connaissance
71 est celle de Nauges et Bontemps (2009). Les auteurs y étudient l'impact des facteurs environnementaux
72 et socioéconomiques sur la décision de boire ou non de l'eau du robinet. Ils mettent en évidence que la
73 qualité de l'environnement, mesurée à partir d'un indice dit de « mauvaise qualité des eaux brutes », est un
74 déterminant important de cette décision. L'effet marginal de l'indice de mauvaise qualité des eaux brutes
75 (-0,63 sur la probabilité de consommer de l'eau du robinet) est modéré par un effet de revenu ou encore
76 par le statut du chef de famille : les ménages aisés et les ménages ayant comme chef de famille un retraité
77 préfèrent, relativement, consommer de l'eau en bouteille. La limite principale de cette contribution tient au
78 fait que les auteurs n'ayant pas d'information directe sur le choix du mode de consommation de l'eau des
79 individus de leur panel, cette information a été construite *ex post* à partir du volume d'achat de boissons
80 non-alcoolisées, volume dont ils disposaient. Les individus achetant moins de 0,5 litre par jour de boisson
81 non-alcoolisée ont ainsi été considérés comme buveurs d'eau du robinet. Un test de sensibilité sur ce seuil,
82 ramené à 0,2 litre pour être en cohérence avec la proportion nationale de buveurs d'eau du robinet, conduit

83 à une remise en cause du résultat principal, l'indice de mauvaise qualité des eaux brutes n'apparaissant plus
84 comme un déterminant significatif du choix d'alimentation en eau potable.

85 Pour les États-Unis, Abrahams *et al.* (2000) s'intéressent aux choix d'alimentation en eau potable dans
86 l'État de Géorgie. Ils mettent en évidence le rôle des notifications données par les autorités locales lorsqu'il y a
87 un incident concernant la distribution ou la qualité de l'eau du robinet, des facteurs organoleptiques, du risque
88 perçu de la consommation d'eau du robinet de mauvaise qualité, ainsi que des facteurs socioéconomiques
89 sur le choix d'alimentation en eau potable (eau du robinet, eau du robinet filtrée ou eau en bouteille).
90 Concernant l'eau en bouteille leurs conclusions sont les suivantes :

91 - L'effet marginal des facteurs organoleptiques sur la probabilité de boire de l'eau en bouteille s'élève à
92 environ 0,23 point de probabilité, ce qui indique que les ménages préfèrent consommer de l'eau en bouteille
93 lorsque l'eau du robinet présente un goût, une couleur ou une odeur trop prononcés ;

94 - L'effet marginal du risque perçu s'élève environ 0,18 point de probabilité. À mesure que la satisfaction
95 des individus vis-à-vis de l'eau du robinet se dégrade, ces derniers préfèrent consommer de l'eau en bouteille
96 ou de l'eau du robinet filtrée pour se prémunir d'un risque sanitaire ;

97 - Enfin, seuls l'âge et l'appartenance ethnique sont des facteurs socioéconomiques significatifs. Plus l'âge
98 de l'individu est avancé, plus la probabilité de boire de l'eau du robinet est importante ; lorsque l'individu
99 appartient à une minorité ethnique cela augmente la probabilité qu'il consomme de l'eau en bouteille.

100 Concernant la décision d'utiliser un système de filtration domestique, les variables significatives sont
101 l'information, avec un effet marginal sur la probabilité de boire de l'eau filtrée d'environ 0,13 point de
102 probabilité, la perception du risque (0,07 point de probabilité) et le revenu (0,002 point de probabilité). La
103 non-significativité de la variable relative aux facteurs organoleptiques pourrait s'expliquer par le fait que
104 l'achat et l'utilisation d'un système de filtration domestique est une forme de comportement défensif ou de
105 prévention (*averting behaviour*).

106 Um *et al.*, (2002) s'intéressent quant-à-eux aux déterminants du choix d'alimentation en eau potable en
107 Corée du Sud et utilisent un modèle de type probit. Plus les individus jugent l'eau du robinet polluée, plus
108 la probabilité de choisir une alternative à l'eau du robinet augmente (la variable qui code la perception de
109 la pollution est significative pour chaque alternative). Comme dans l'étude d'Abrahams *et al.* (2000), le
110 revenu agit positivement sur la probabilité d'acheter / d'utiliser un système de filtration domestique puisque
111 cela correspond à un investissement coûteux. La consommation d'eaux directement captées dans le milieu
112 est influencée positivement lorsque le logement est une maison individuelle. À mesure que l'âge augmente, la
113 probabilité de choisir une alternative à l'eau du robinet augmente elle aussi. Un des intérêts de cet article est
114 la prise en compte du risque objectif, mesuré à partir de données observées par les autorités sanitaires, dans

115 le calcul du consentement à payer pour une amélioration de la qualité de l'eau du robinet. Le consentement
116 à payer pour une réduction de la quantité de particules en suspension dans l'eau du robinet est estimé à
117 partir d'un modèle Tobit. Les valeurs trouvées par les auteurs sont sensiblement plus importantes lorsqu'ils
118 utilisent la spécification intégrant le risque perçu et non le risque objectif.

119 Enfin, Lloyd-Smith *et al.* (2014) mettent en évidence sur données canadiennes que le risque perçu agit
120 négativement sur la probabilité de boire de l'eau du robinet. Leur propos, cependant, se concentre sur la
121 question des dépenses de protections pour se protéger des risques sanitaires liés à la consommation d'eau du
122 robinet.

123 En pratique, l'analyse des choix individuels d'alimentation en eau potable s'est longtemps heurtée, en
124 France et ailleurs, à l'absence de données pertinentes. Pour la France, celles-ci ont été rendues disponibles
125 par l'enquête menée par le Commissariat général au développement durable en 2013, dont nous présentons
126 les caractéristiques principales dans la section suivante.

127 **3 Présentation des données**

128 **3.1 L'échantillon de données**

129 Les données dont nous disposons sont issues d'une enquête réalisée en 2013 par l'Institut français d'opinion
130 publique (IFOP) pour le compte du Service de l'Economie, de l'Evaluation de l'Intégration du Développement
131 Durable (SEEIDD) du Commissariat Général au Développement Durable (CGDD). L'échantillon se compose
132 de 4003 individus, âgés de 18 à 90 ans, répartis sur l'ensemble du territoire métropolitain. La représentativité
133 de l'échantillon est assurée par la méthode des quotas au regard des critères de sexe, d'âge, de profession,
134 de catégories d'agglomération, de région et de statut d'occupation du logement, le tout basé sur les données
135 du recensement INSEE de 2010. La base de données comprend un ensemble de variables relatives à la
136 perception des individus vis-à-vis des ressources en eau, aux habitudes de consommation d'eau potable
137 et aux caractéristiques socioéconomiques des individus. Nous l'avons complétée par des variables exogènes
138 relatives à la qualité physico-chimique et microbiologique de l'eau du robinet, au niveau de pollution (nitrates
139 et pesticides) des eaux brutes ainsi qu'à la pluviométrie moyenne régionale et départementale. Ces variables
140 sont issues de la base de données Eider du Ministère de l'Ecologie, du Développement Durable et de l'Energie,
141 ainsi que de l'Observatoire national des services d'eau et d'assainissement (ONSEA).

Tableau 1: Statistiques descriptives de l'échantillon

Variables	Moyenne
Consommation d'eau du robinet	40%
Âge	47,5 ans
Sexe (Femme)	52,37%
Diplôme	
Sans diplôme	3%
Brevet, Certificat d'étude, CAP, BEP	26%
Baccalauréat	19%
Bac +1, +2, +3	34%
Bac +4 et plus	19%
Habitat en zone rurale (moins de 2 000 habitants)	23%
Revenu moyen	1 900€
Statut d'occupation du logement	
Locataire	63%
Propriétaire	37%

3.2 Statistiques descriptives

Nous présentons les statistiques descriptives principales de l'échantillon dans le tableau 1. Dans la mesure où la représentativité de l'échantillon est bien établie, ces statistiques descriptives ne nous paraissent pas devoir faire l'objet d'un commentaire spécifique.

Mode de consommation d'eau potable

Le principal mode d'alimentation en eau potable choisi par les Français (tableau 2) est l'eau du robinet (40%). L'eau en bouteille est le deuxième mode de consommation préféré par les Français (39%). Enfin l'eau filtrée arrive en troisième place avec 21% de buveurs.

Tableau 2: Mode d'alimentation en eau potable

Mode de consommation	Part de buveurs
Eau en bouteille	39%
Eau filtrée	21%
Eau du robinet	40%

Mode de consommation et satisfaction vis-à-vis de l'eau du robinet

Le choix du mode de consommation dépend-t-il de la qualité perçue de l'eau du robinet ? Dans une première approche, nous présentons le choix d'alimentation en eau potable en fonction de la satisfaction des individus vis-à-vis de l'eau du robinet (tableau 3). Ainsi, la part de buveurs d'eau du robinet passe de 11% lorsqu'ils sont non satisfaits à 40% lorsqu'ils sont très satisfaits, soit une augmentation de 29 points de pourcent. La

155 dépendance entre les deux variables (l'une étant nominale, l'autre étant ordonnée) est confirmé par un test
 156 de χ^2 ($\chi^2(4) = 703,4986$, $p - value = 0.000$).

Tableau 3: Mode d'alimentation en eau potable en fonction de la satisfaction

	Eau en bouteille	Eau filtrée	Eau du robinet
Non satisfait	67%	22%	11%
Satisfait	50%	23%	27%
Très satisfait	39%	21%	40%

157 **La localisation géographique : un déterminant du mode de consommation d'eau potable et de**
 158 **la satisfaction**

159 L'échantillon dont nous disposons regroupe les individus dans neuf grandes régions françaises :

- 160 - Le Nord de la France,
- 161 - la région parisienne,
- 162 - le Bassin parisien Est (BPE),
- 163 - le Bassin parisien Ouest (BPO),
- 164 - l'Est,
- 165 - l'Ouest,
- 166 - le Sud-Est,
- 167 - le Sud-Ouest,
- 168 - la Méditerranée.

Tableau 4: Mode d'alimentation en eau potable et localisation régionale des individus

Grandes régions françaises	Mode de consommation (%)			Total
	Eau en bouteille	Eau filtrée	Eau du robinet	
<i>Région parisienne</i>	43.72	18.89	37.38	100
<i>Nord</i>	67.06	16.27	16.67	100
<i>Est</i>	39.71	20.29	40.00	100
<i>Bassin parisien est</i>	46.79	24.68	28.53	100
<i>Bassin parisien Ouest</i>	40.27	29.33	30.40	100
<i>Ouest</i>	41.07	25.05	33.89	100
<i>Sud-Ouest</i>	29.98	17.45	52.57	100
<i>Sud-Est</i>	24.27	19.46	46.28	100
<i>Méditerranée</i>	37.76	19.22	49.02	100
Total	39.02	21.06	39.92	100

169 Les plus fortes proportions de consommateurs d'eau du robinet se trouvent dans les régions du Sud de la
 170 France (Sud-Ouest : 52,57% ; Méditerranée : 49,02% ; Sud-Est : 46,28%). *A contrario*, le Nord de la France

171 (16,67%) et les bassin parisien Est et Ouest (28,53% et 30,40%) sont les régions les moins consommatrices
 172 d'eau du robinet. L'étude de Janvier et Roy (2001) trouve des tendances régionales similaires.

173 4 Stratégie économétrique

174 La question sur le choix d'alimentation en eau potable était libellée comme suit :

175 "Chez vous, lorsque vous buvez de l'eau, le plus souvent vous buvez ... ?

176 (Une seule réponse)

- 177 1. principalement de l'eau en bouteille
- 178 2. principalement de l'eau du robinet filtrée
- 179 3. principalement de l'eau du robinet non filtrée"

180 Le choix de l'une des trois alternatives de consommation d'eau peut donc se modéliser dans le cadre
 181 flexible du modèle d'utilité aléatoire, tel que proposé initialement par McFadden (1974).

182 4.1 Hétérogénéité des préférences et modèle logit multinomial à facteurs d'échelle

183 En reprenant la présentation de Train (2009), dans le modèle de comportement de base de choix discret,
 184 l'utilité qu'un individu i tire du choix de l'alternative j dans un ensemble de J alternatives est :

185 $U_{ij} = \beta x_{ij} + \epsilon_{ij}$ ou x_{ij} est un vecteur de dimension k de caractéristiques observées de l'alternative j , β
 186 est un vecteur de coefficients, qui s'interprètent comme des poids dans la fonction d'utilité, supposés, dans
 187 cette version simple, homogènes d'un individu à l'autre. Le terme d'erreur ϵ_{ij} capture l'effet des facteurs qui
 188 influencent U_{ij} mais ne sont pas inclus dans βx_{ij} . L'individu i choisit alors l'alternative k , si et seulement si
 189 $U_{ik} > U_{ij}, \forall j \neq k$.

190 En conséquence, la probabilité que l'individu i choisisse l'alternative k est :

$$\begin{aligned}
 191 \quad P_{ik} &= \Pr(U_{ik} > U_{ij}, \forall j \neq k) \\
 192 \quad &= \Pr(\beta x_{ik} + \epsilon_{ik} > \beta x_{ij} + \epsilon_{ij}, \forall j \neq k) \\
 193 \quad &= \Pr(\epsilon_{ij} - \epsilon_{ik} < \beta x_{ik} - \beta x_{ij}, \forall j \neq k)
 \end{aligned}$$

194 En posant $\epsilon'_i = (\epsilon_{i1}, \dots, \epsilon_{iJ})$ et $f(\epsilon_i)$ la densité multivariée associée à ce vecteur de termes d'erreur, cette
 195 probabilité cumulée s'écrit :

$$196 \quad P_{ik} = \int_{\epsilon} I(\epsilon_{ij} - \epsilon_{ik} < \beta x_{ik} - \beta x_{ij}, \forall j \neq k) f(\epsilon_i) d_{\epsilon_i}$$

197 où la fonction indicatrice $I(\cdot)$ prend la valeur 1 lorsque l'expression entre parenthèses est vraie, et 0 sinon.
 198 Le modèle logit multinomial de base repose sur l'hypothèse que le terme d'erreur ϵ_{ij} suit une loi de Gumbel
 199 (ou loi extremum type I). Dans ce cas, en effet, l'intégrale multiple précédente a une solution analytique, ce
 200 qui explique, ainsi que le rappelle D. McFadden dans sa leçon de réception du prix Nobel (McFadden, 2000),
 201 le succès du modèle logit multinomial en économétrie du choix discret.

202 L'hypothèse d'un vecteur β homogène d'un individu à l'autre est cependant très restrictive. L'une
 203 des façons d'introduire de l'hétérogénéité observée, déterministe, dans la fonction d'utilité U_{ij} consiste à
 204 introduire des variables d'interaction entre les caractéristiques socio-économiques observées des individus,
 205 et les caractéristiques observées des alternatives. Ainsi, si nous disposons d'une variable *gender* codant le
 206 genre (1 pour les hommes, 0 pour les femmes), celle-ci peut être croisée avec les caractéristiques observées
 207 des alternatives pour capter l'effet du genre sur les choix, soit :

$$208 \quad U_{ij} = \beta x_{ij} + \beta_g (\text{gender}_i \times x_{ij}) + \epsilon_{ij}$$

209 En supposant que les variables d'interaction $(\text{gender}_i \times x_{ij})$ sont significatives, le vecteur de poids des
 210 caractéristiques x_{ij} des alternatives serait alors $(\beta + \beta_g)$ pour un homme et β pour une femme.

211 Le modèle logit multinomial hétéroscédastique suppose plus en avant que le vecteur β peut se déformer
 212 vers le haut ou vers le bas, par rapport à une valeur moyenne, suivant un facteur d'échelle lié aux caractéris-
 213 tiques des individus, et inversement proportionnel à la variance du terme d'erreur :

$$214 \quad U_{ij} = (\beta \mu_i) x_{ij} + \epsilon_{ij}$$

215 où μ_i est le facteur d'échelle, qui s'écrit à son tour comme suit :

216 $\mu_i = \exp(\theta' y_i)$ où y_i est un vecteur de variables spécifiques aux individus et θ est un vecteur de
 217 paramètres qui reflètent l'influence de ces variables sur la variance du terme d'erreur. Selon cette spéci-
 218 fication, l'hétérogénéité non observée dépend donc de caractéristiques individuelles.

219 Il est également possible d'introduire une dimension aléatoire, non reliée aux caractéristiques individuelles
 220 observées (Train, 2009), dans la spécification du facteur d'échelle :

$$221 \quad \mu_i = \exp(-\tau^2/2 + \tau \zeta_i + \theta' y_i)$$

222 où ζ_i est une variation individuelle aléatoire, $\zeta_i \sim N(0, 1)$, et la constante $-\tau^2/2$ est choisie telle que
 223 $E[\mu_i] = 1$.

224 Le modèle qui en découle est connu sous l'appellation de *scaled multinomial logit model* (S-MNL, voir
 225 Fiebig *et al.*, 2010 ; Greene and Hensher, 2010), que nous traduisons ici par modèle logit multinomial à
 226 facteurs d'échelle⁴. Dans ce cas, l'intégrale multiple précédente n'a plus de solution analytique, et le modèle

⁴Le pluriel de facteur vise ici à souligner qu'il y a bien plusieurs facteurs d'échelle, dont l'une des composantes est reliée aux caractéristiques des individus, l'autre étant aléatoire.

227 doit être estimé par une technique de simulation, par exemple le maximum de vraisemblance simulé (Train,
228 2009). Soulignons qu'il est bien sûr possible d'introduire une hétérogénéité déterministe des préférences (des
229 goûts) dans le S-MNL via le recours à des variables d'interaction, comme nous l'avons indiqué précédemment
230 pour le logit multinomial de base, et comme nous le ferons dans la suite de ce papier.

231 Au final, nous disposons ainsi d'un outil très généraliste capable de mesurer l'influence de sources
232 d'hétérogénéité observées et non-observées sur les choix arrêtés par les individus.

233 Parmi les variables spécifiques aux individus qui peuvent influencer les choix, figurent des variables
234 catégorielles attitudinales, telle que la perception des enjeux environnementaux, ou encore, dans le cas qui
235 nous intéresse ici, la perception de la qualité de l'eau du robinet. Ces variables catégorielles attitudinales,
236 introduites dans le modèle comme variables indépendantes, posent cependant un problème d'endogénéité,
237 rarement traité de façon satisfaisante dans la littérature, particulièrement dans la littérature empirique sur
238 les choix d'alimentation en eau potable.

239 **4.2 Le traitement de l'endogénéité des variables attitudinales**

240 En effet, le traitement de l'endogénéité de variables discrètes, dans les modèles non-linéaires tels que le
241 S-MNL, reste complexe, et fait l'objet de développements en cours. Nous nous proposons d'appliquer ici
242 l'une des méthodes proposées par Terza *et al.* (2008) et reprise par Wooldridge (2014). Le papier de
243 Terza *et al.* (2008) compare en effet deux méthodes de traitement de l'endogénéité dans les modèles non-
244 linéaires : la méthode dite 2SPS (*two-stage predictor substitution*) et la méthode 2SRI (*two-stage residual*
245 *inclusion*). Comme leur nom l'indique, il s'agit de méthodes en deux étapes, dont la contrepartie, pour le
246 modèle linéaire, est la méthode 2SLS (*two-stage least squares ; doubles moindres carrés*). Les deux méthodes
247 sont des méthodes à variables instrumentales qui consistent à définir, au-delà de l'équation principale que
248 le chercheur cherche à estimer, des équations auxiliaires (non linéaires, dans le cas général) pour chacune
249 des variables soupçonnées d'endogénéité. Ces équations auxiliaires permettent le calcul de valeurs prédites
250 pour les variables potentiellement endogènes, ainsi que le calcul de résidus obtenus comme différence entre
251 les valeurs observées et prédites des variables potentiellement endogènes. La méthode 2SPS consiste alors
252 à remplacer les variables soupçonnées d'endogénéité dans l'équation principale par leurs valeurs prédites
253 par les équations auxiliaires. La méthode 2SRI consiste à conserver les variables soupçonnées d'endogénéité
254 dans l'équation principale, et à ajouter les résidus calculés après estimation des équations auxiliaires, comme
255 régresseurs dans l'équation principale. Dans le cas linéaire, les deux méthodes sont équivalentes aux doubles
256 moindres carrés (Terza *et al.* (2008)). Comme le rappelle Wooldridge (2014), il s'agit d'une approche par

fonction de contrôle (*control function approach*), qui présente l'avantage d'être d'application aisée. Terza *et al.* (2008) établissent en outre que la méthode 2SRI fournit des estimateurs convergents pour les paramètres d'intérêt, ce qui n'est pas le cas de la méthode 2SPS. Enfin, l'endogénéité des variables peut être testée simplement, à la seconde étape de l'estimation, par un test de Wald sur la nullité conjointe des paramètres associés aux résidus inclus à l'occasion de celle-ci (Wooldridge (2014)).

Ainsi que nous l'avons écrit plus haut, les données dont nous disposons décrivent le choix de chaque répondant i selon trois alternatives : eau en bouteille, eau du robinet filtrée, et eau du robinet non-filtrée. La variable correspondante, $choice_i$, est donc une variable catégorielle non-ordonnée, codée simplement de 1 (eau en bouteille), à 3 (eau du robinet non-filtrée). Parmi les prédicteurs que nous souhaitons intégrer dans le modèle de choix discret correspondant, figure le degré de satisfaction déclaré de chaque répondant. La variable correspondante, sat_i , est une variable catégorielle ordonnée, qui prend la valeur 1, 2, 3 pour les individus qui se déclarent respectivement non-satisfaits, satisfaits, et très satisfaits de l'eau du robinet. Cette variable déclarative attitudinale, en tant que reflet de l'expérience personnelle du choix discret au centre de notre analyse, est susceptible d'être endogène : d'une part les deux variables sont susceptibles d'être dépendantes, d'autre part elles sont probablement influencées par des variables inobservables (*confounders*) communes, causes couramment mises en avant comme sources d'endogénéité (Terza *et al.*, 2008). Pour saisir finement l'influence de la variable sat_i sur le choix d'alimentation en eau potable, celle-ci est discrétisée (pour $k = 1..3$, $d_sat_{ik} = 1$ si $sat_i = k$, 0 sinon).

Nous pouvons ainsi écrire, dans la logique de notre présentation du modèle d'utilité aléatoire, que l'utilité pour un individu i de choisir l'une des trois alternatives d'alimentation en eau potable de boisson s'écrit (pour $j = 1..3$) :

$$U_{ij} = \sum_{j=1}^3 (\beta_{ascj} \mu_i) asc_j + \sum_{j=1}^3 (\beta_{intj} \mu_i) asc_j \times xint_i + \sum_{j=1}^3 \sum_{k=1}^3 (\beta_{satjk} \mu_i) asc_j \times d_sat_{ik} + \epsilon_{ij}$$

où la variable asc_j est codée 1 si l'alternative j est choisie, 0 sinon, $xint_i$ est un vecteur de variables socio-démographiques mises en interaction avec la variable asc_j . L'estimation du modèle logit multinomial à facteurs d'échelle nécessite, comme dans le modèle logit multinomial de base, que l'on choisisse l'une des alternatives comme référence, les coefficients associés à cette alternative de référence étant alors nuls. De même l'une des catégories de la variable sat_i doit être choisie comme référence, ce qui implique que les coefficients de la binaire associée (par exemple d_sat_{i1} si la catégorie "non-satisfait" est prise comme référence) sont nuls.

La variable sat_i , susceptible d'être endogène, est une variable catégorielle ordonnée. La mise en œuvre de la méthode 2SRI suppose donc, lors de la première étape, d'estimer un modèle de type logit ou probit

288 ordonné. La probabilité qu'un individu se déclare non-satisfait, satisfait ou très satisfait de l'eau du robinet
 289 dépend tant des variables socio-démographiques qui interviennent comme variables indépendantes lors de la
 290 seconde étape, lors de l'estimation du modèle logit multinomial à facteurs d'échelle (le vecteur $xint_i$), que de
 291 variables instrumentales qui doivent être en nombre au moins équivalent au nombre de variables susceptibles
 292 d'être endogènes (Terza *et al.* (2008)). Notre choix, pour cette première étape, s'est porté sur un modèle
 293 logit ordonné ; la variable instrumentale choisie est une mesure de conformité physico-chimique de l'eau du
 294 robinet (voir *infra*).

295 Comme nous l'avons exposé plus haut, cette première étape permet de calculer des résidus qui seront
 296 introduits comme variables explicatives lors de la seconde étape (l'estimation logit multinomial à facteurs
 297 d'échelle). Suivant Geraci *et al.* (2014), nous retenons une version réduite des résidus (*standardized residuals*)
 298 : pour chaque individu i , le modèle logit ordonné estimé permet de calculer les probabilités prédites que
 299 l'individu se déclare non-satisfait, satisfait ou très satisfait de l'eau du robinet, que nous notons \hat{P}_{ij} . Les
 300 résidus réduits ($\hat{r}_{s_{ik}}$) sont alors calculés comme suit :

$$301 \quad \hat{r}_{s_{ik}} = \hat{P}_{ik}^{-1/2}(1 - \hat{P}_{ik})^{-1/2}(dk_sat_i - \hat{P}_{ik})$$

302 Geraci *et al.* (2014) montrent en effet, à l'aide de simulations de type Monte-Carlo, que la puissance du
 303 test d'endogénéité de la seconde étape de la méthode 2SRI est plus élevée en utilisant des résidus réduits
 304 plutôt que des résidus simples (*raw residuals*) obtenus comme $\hat{r}_{ik} = (dk_sat_i - \hat{P}_{ik})$.

305 5 Résultats

306 5.1 Première étape de la méthode 2SRI : le logit ordonné

307 Dans cette première étape, sat_i est la variable dépendante ; nous souhaitons modéliser la probabilité qu'un
 308 individu se déclare non-satisfait, satisfait, très satisfait de l'eau du robinet. Une variable instrumentale
 309 intervient ici, au côté de variables socio-démographiques telles que l'âge (*age*), le genre (*genre*, codé 1 pour
 310 les femmes, 0 pour les hommes), le revenu (*revenu*), le niveau de diplôme (*diplome*), le statut d'occupation
 311 du logement (*logement*, codé 1 pour les locataires, 0 pour les propriétaires) ou encore la localisation régionale
 312 : il s'agit de la variable *physico* qui mesure le taux de conformité physico-chimique, en pourcentage, de l'eau
 313 du robinet au niveau de chaque département français. Ce taux de conformité physico-chimique est compris,
 314 dans notre échantillon, entre 95% et 100%, pour une moyenne de 98,80% et un écart-type de 1,43 points de
 315 pourcentage.

316 Au-delà des coefficients estimés, le tableau 5 présente les effets marginaux de chacune des variables sur,

Tableau 5: estimation du logit ordonné

	Coefficients	Effets marginaux		
		Non satisfait	Satisfait	Très satisfait
age	0.010*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	0.002*** (0.000)
genre	-0.198** (0.002)	0.015*** (0.001)	0.032** (0.004)	-0.047** (0.002)
revenu	0.031* (0.020)	-0.002* (0.025)	-0.005* (0.018)	0.007* (0.019)
diplome	0.140*** (0.000)	-0.011*** (0.000)	-0.022*** (0.000)	0.033*** (0.000)
logement	-0.219** (0.002)	0.017** (0.001)	0.035** (0.006)	-0.052** (0.003)
physico	0.069** (0.002)	-0.005** (0.003)	-0.011** (0.003)	0.017** (0.002)
Nord	-0.587*** (0.000)	0.046*** (0.000)	0.094*** (0.000)	-0.140*** (0.000)
Est	0.226 (0.079)	-0.018 (0.101)	-0.036 (0.065)	0.054 (0.075)
BPE	-0.067 (0.619)	0.005 (0.613)	0.011 (0.624)	-0.016 (0.620)
BPO	-0.202 (0.106)	0.016 (0.088)	0.032 (0.124)	-0.048 (0.110)
Ouest	0.075 (0.502)	-0.006 (0.514)	-0.012 (0.492)	0.018 (0.500)
Sud-Ouest	0.380** (0.001)	-0.030** (0.005)	-0.061*** (0.000)	0.090*** (0.001)
Sud-Est	0.534*** (0.000)	-0.042*** (0.000)	-0.086*** (0.000)	0.127*** (0.000)
Méditerranée	0.239* (0.034)	-0.019 (0.052)	-0.038* (0.024)	0.057* (0.031)
cut1	5.645* (0.013)			
cut2	8.404*** (0.000)			
<i>N</i>	4003	4003	4003	4003

p-values in parentheses

p* < 0.05, *p* < 0.01, ****p* < 0.001

317 respectivement, la probabilité de se déclarer non-satisfait, satisfait, très satisfait de l'eau du robinet. Ces
318 effets marginaux sont calculé en référence à un profil précis : la région de référence (région parisienne), pour
319 une femme dont l'âge est fixé à la moyenne d'échantillon, propriétaire de son logement.

320 Soulignons tout d'abord que la variable instrumentale, est fortement significative, avec l'effet attendu
321 : la probabilité de se déclarer très satisfait de l'eau du robinet augmente avec la qualité physico-chimique
322 objective de l'eau (+0,017 point de probabilité). Rappelons que dans le système français, la qualité physico-
323 chimique de l'eau fait l'objet d'un rapport spécifique, joint à la facture d'eau, diffusé par affichage dans les
324 logements collectifs. Les résultats des contrôles sanitaires sont affichés en mairie, et nombre de sites web
325 recensent des informations relatives à la qualité de l'eau du robinet (www.eaufrance.fr, par exemple).

326 L'âge a un effet significatif et positif, mais faible sur la probabilité de se déclarer très satisfait de l'eau du
327 robinet. En revanche l'effet de genre a une certaine ampleur : le fait d'être une femme diminue significative-
328 ment la probabilité de se déclarer très satisfait de l'eau du robinet de 0.047 point. Cet effet est sans doute
329 à mettre en lien avec les recommandations pédiatriques usuelles, reçues notamment dans les maternités, sur
330 la préparation de l'alimentation des enfants en bas-âge. L'effet du revenu est positif, tout comme celui du
331 diplôme : lorsque le niveau de diplôme augmente, la probabilité de se déclarer très satisfait de l'eau du
332 robinet s'accroît (+0,033 point de probabilité). Le niveau de diplôme capte ici en partie l'aptitude à intégrer

333 l'information objective relative à la qualité de l'eau.

334 Par ailleurs, le fait d'être locataire agit négativement sur la probabilité de se déclarer très satisfait (-0,052
335 point de probabilité).

336 Plusieurs variables indicatrices régionales sont également significatives. Certaines régions ne se dis-
337 tinguent pas de la région de référence (région parisienne), ce qui est le cas des régions Est, bassin parisien
338 Est, bassin parisien Ouest, Ouest, tandis qu'une localisation dans Sud-Ouest, Sud-Est, Méditerranée agit
339 significativement et positivement sur la probabilité de se déclarer très satisfait de l'eau du robinet (+0,127
340 point de probabilité pour le Sud-Est), l'inverse étant obtenu pour une localisation au Nord (-0,140 point
341 de probabilité). Sans doute capte-t-on ici l'effet de la distribution régionale de la pollution de l'eau par
342 certaines substances, ainsi que le souligne l'analyse descriptive des données par le Commissariat général au
343 développement durable (CGDD, 2014).

344 Enfin, les variables de seuil (*cut1*, *cut2*) sont également significatives ce qui indique que les répondants
345 distinguent correctement les trois niveaux de satisfaction.

346 Au total, comme nous l'avons précisé dans la section précédente, l'estimation de ce modèle logit ordonné
347 nous permet de calculer les résidus prédits qui seront intégrés au modèle estimé lors de la seconde étape de
348 la méthode 2SRI pour tester et corriger le biais d'endogénéité lié à la variable de satisfaction.

349 **5.2 Seconde étape de la méthode 2SRI : le S-MNL et la correction de l'endogénéité**

350 Le tableau 6 présente les résultats de l'estimation du logit multinomial à facteurs d'échelle, avec et sans
351 correction de l'endogénéité. Notons tout d'abord que la présence d'endogénéité est confirmé par un test de
352 Wald (Wooldridge (2014), Geraci *et al.* (2014)) sur la nullité conjointe des coefficients associés aux résidus,
353 dans le modèle avec correction de l'endogénéité ($\chi^2(4) = 11,03$, $p - value = 0.026$). La comparaison des
354 résultats des deux modèles (avec et sans correction de l'endogénéité), montre une différence principale : une
355 fois l'endogénéité corrigée, le paramètre τ qui capture la dimension aléatoire de l'hétérogénéité non observé
356 des préférences est non significatif ($p - value = 0.263$). Aussi, l'absence de correction de l'endogénéité aurait
357 pu conduire à conclure faussement sur les sources d'hétérogénéité non-observée, qui se limitent ici à des effets
358 régionaux. Ainsi, une location dans le Sud-Est agit significativement et positivement sur le facteur d'échelle
359 (coefficient=0,184, $p - value = 0,001$), ce qui signifie que pour cette région, relativement à la région de
360 référence, le poids de la variable de satisfaction dans l'utilité est plus élevé.

361 Rappelons que l'alternative de référence, pour le choix d'alimentation en eau potable, est ici l'eau du
362 robinet. L'alternative 1 est l'eau en bouteille, l'alternative 2 l'eau du robinet filtrée. La significativité des

Tableau 6: estimation du S-MNL

	Sans correction de l'endogénéité	Avec correction de l'endogénéité
asc_1	1.653*** (0.000)	4.445*** (0.000)
asc_2	0.571*** (0.001)	2.469** (0.007)
$asc_1 X_{age}$	0.012*** (0.000)	0.015*** (0.000)
$asc_2 X_{age}$	0.009*** (0.000)	0.007* (0.033)
$asc_1 Xd_sat_2$	-1.537*** (0.000)	-5.342*** (0.000)
$asc_1 Xd_sat_3$	-3.556*** (0.000)	-5.866*** (0.000)
$asc_2 Xd_sat_2$	-1.021*** (0.000)	-3.706** (0.005)
$asc_2 Xd_sat_3$	-2.108*** (0.000)	-3.241*** (0.000)
$asc_1 X\hat{r}s_2$		1.996** (0.005)
$asc_1 X\hat{r}s_3$		1.294** (0.004)
$asc_2 X\hat{r}s_2$		1.386* (0.031)
$asc_2 X\hat{r}s_3$		0.582 (0.086)
Facteurs d'échelle		
Nord	-0.932*** (0.000)	-0.279* (0.011)
Est	0.115 (0.270)	0.024 (0.536)
BPE	-0.221 (0.089)	-0.079 (0.100)
BPO	-0.066 (0.573)	-0.011 (0.798)
Ouest	-0.040 (0.675)	-0.031 (0.378)
Sud-Ouest	0.432*** (0.001)	0.156** (0.003)
Sud-Est	0.439*** (0.001)	0.184** (0.001)
Méditerranée	0.347** (0.003)	0.132** (0.004)
τ	-0.545* (0.021)	-0.131 (0.263)
N	12009	12009

p-values in parentheses

p* < 0.05, *p* < 0.01, ****p* < 0.001

363 constantes (variables asc_1 et asc_2) implique qu'une partie de la population a une préférence pour l'un ou
 364 l'autre de ces deux choix d'alimentation en eau potable, ce qui est bien sûr en cohérence avec les statistiques
 365 descriptives de la section 3. Peu de variables socio-démographiques ont été trouvées significatives à cette
 366 étape. Ici l'âge influence positivement tant le choix de l'eau en bouteille que de l'eau du robinet filtré, ce qui
 367 est en cohérence, notamment, avec les résultats de Nauges et Bontemps (2009). On notera que le poids de
 368 l'âge est plus important pour la consommation d'eau en bouteille que pour l'eau du robinet (test d'égalité
 369 des coefficients associés à l'âge, $\chi^2(1) = 6,07$, *p* - value = 0.013).

370 Plus important, la variable de satisfaction agit dans le sens attendu, y compris lorsque l'endogénéité est
 371 corrigée. On notera que le poids accordé à cette variable est plus important pour l'eau en bouteille que pour
 372 l'eau du robinet filtrée. Que ce soit pour les répondants qui se déclarent satisfaits, ou très satisfaits, un test
 373 d'égalité des coefficients associés conduit à rejeter cette égalité ($asc_1 Xd_sat_2 = asc_2 Xd_sat_2$, $\chi^2(1) = 3,24$,
 374 *p* - value = 0.071 ; $asc_1 Xd_sat_3 = asc_2 Xd_sat_3$, $\chi^2(1) = 25,80$, *p* - value = 0.000 respectivement).

375 Aussi la satisfaction par rapport à l'eau du robinet apparaît être un déterminant plus important du choix
376 d'alimentation potable pour l'eau en bouteille que pour l'eau du robinet filtrée, relativement à l'eau du
377 robinet. Dans les deux cas, cependant, lorsque le niveau de satisfaction augmente, les consommations d'eau
378 en bouteille et d'eau du robinet filtrée diminuent. Ici, un test d'égalité des coefficients associés montre que le
379 niveau de satisfaction (satisfait, très satisfait, en référence à non-satisfait), n'importe pas ($asc_1Xd_sat_2 =$
380 $asc_1Xd_sat_3$, $\chi^2(1) = 0,29$, $p - value = 0.593$ pour l'eau en bouteille, $asc_2Xd_sat_2 = asc_2Xd_sat_3$,
381 $\chi^2(1) = 0,20$, $p - value = 0.652$ pour l'eau du robinet filtrée). Là encore, en l'absence de correction de
382 l'endogénéité, c'est exactement la conclusion inverse qui aurait été tirée de l'estimation du modèle ($\chi^2(1) =$
383 $27,08$, $p - value = 0.000$ pour l'eau en bouteille, $\chi^2(1) = 48,08$, $p - value = 0.000$ pour l'eau du robinet
384 filtrée).

385 6 Conclusion

386 Les deux étapes de notre stratégie économétrique apportent chacune des enseignements utiles à la com-
387 préhension fine des choix d'alimentation en eau potable. La première étape de la méthode 2SRI, l'estimation
388 d'un logit ordonné dont la variable dépendante est le niveau déclaré de satisfaction vis-à-vis de l'eau du
389 robinet, montre que les individus intègrent correctement l'information relative à la qualité physico-chimique
390 de l'eau du robinet. La seconde étape, et la comparaison des résultats de l'estimation d'un modèle logit à
391 facteurs d'échelle sans et avec correction de l'endogénéité, apporte trois résultats principaux. Tout d'abord,
392 l'endogénéité du niveau déclaré de satisfaction vis-à-vis de l'eau du robinet, lorsqu'il intervient comme pré-
393 dicteur du choix d'alimentation en eau potable, est établie. Ensuite, lorsque cette endogénéité est corrigée,
394 la qualité perçue de l'eau apparaît toujours comme un déterminant fort du choix d'alimentation en eau
395 potable. Pour le régulateur public, cela implique que l'infléchissement des comportements de consommation
396 d'eau potable passe sans doute par une meilleure information sur la qualité de l'eau. Enfin, en l'absence
397 de correction de l'endogénéité, l'estimation d'un modèle logit multinomial aurait pu conduire à conclure
398 faussement à la présence d'hétérogénéité non-observée de nature aléatoire, ne pouvant être reliée aux car-
399 actéristiques observées des individus. Nos résultats suggèrent que ce n'est pas le cas, offrant là encore de
400 véritable perspectives d'action, selon des leviers identifiés, au régulateur public. L'absence de correction de
401 l'endogénéité aurait pu nous conduire également à conclure faussement que le niveau de satisfaction agit de
402 façon différenciée sur le choix de l'eau en bouteille ou de l'eau du robinet filtrée. Là encore, il semble que ce
403 ne soit pas le cas ; les individus ont ici un comportement plus binaire, guidé par le fait d'être satisfaits ou
404 pas de l'eau du robinet.

405 La méthodologie que nous proposons, outre qu'elle permet d'analyser l'effet d'autres variables attitu-
406 dinales sur les choix d'alimentation en eau potable, par exemple l'effet de la perception de la qualité des
407 ressources en eau, est de portée générale. Elle forme un outil pertinent pour l'analyse de nombreux choix
408 discrets impliquant des variables attitudinales déclaratives, telles qu'on peut les trouver en économie de la
409 santé, en économie du travail ou encore en marketing quantitatif.

410 **Références bibliographiques**

411 Abrahams N., Hubell B. et Jordan J. (2000), "Joint production and averting expenditure measures of will-
412 ingness to pay: Do water expenditures really measure avoidance costs?", *American Journal of Agricultural*
413 *Economics*, 82(2), p. 427 - 437.

414 Ademe (2015), Chiffres-Clés Déchets, édition 2015, Agence de l'environnement et de la maîtrise de
415 l'énergie.

416 CGDD (2014), « Ressources en eau : perception et consommation des français. Résultats d'enquête »,
417 *Études et Documents*, n°106, Commissariat général au développement durable.

418 Geraci A., Fabbri D., Monfardini C. (2014), "Testing exogeneity of multinomial regressors in count data
419 models: does two stage residual inclusion work?", *Quaderni - Working Paper DSE n°921*.

420 Greene W.H., Hensher D.A., (2010), "Does scale heterogeneity across individuals matter? An empirical
421 assessment of alternative logit models", *Transportation*, 37: 413-428.

422 Fiebig D.G., Keane M.P., Louviere J.J., Wasi N., (2010), "The generalized multinomial logit model:
423 Accounting for scale and coefficient heterogeneity", *Marketing Science*, 29: 393 – 421.

424 Janvier C., et Roy A., (2001) « Les Français jugent leur environnement de proximité », *Les données de*
425 *l'environnement*, Ifen, 70, p. 1 - 4.

426 Lloyd-Smith P., Schram C., Adamowicz W., Dupont D. (2014), « Endogeneity of risk perceptions in
427 water expenditure models », *Working Paper*, University of Alberta.

428 McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior", in P. Zarembka, ed.,
429 *Frontiers in Econometrics*, Academic Press, New York, p. 105 – 142.

430 McFadden, D. (2000), *Economic Choices*, Prize Lecture, December 8.

431 Nauges C., et Bontemps C., (2009), « Carafe ou bouteille ? Le rôle de la qualité de l'environnement dans
432 la décision du consommateur », *Économie et Prévisions*, 188, p. 61 - 79.

433 Terza J., Basu A., et Rathouz P., (2008), "Two-stage residual inclusion estimation : Addressing endo-
434 geneity in health econometric modelling", *Journal of Health Economics*, 27(3), p. 531 - 543.

- 435 Train K., (2009), *Discrete Choice Methods with Simulation*, Second edition, Cambridge University Press.
- 436 Um M-J., Kwak S-J., et Kim T-Y., (2002), "Estimating the willingness to pay for improved drinking water
437 quality using averting behavior method with perception measure", *Environmental and Resource Economics*,
438 21, p. 287 – 302.
- 439 Wooldridge J.M., (2014), "Quasi-maximum likelihood estimation and testing for nonlinear models with
440 endogenous explanatory variables", *Journal of Econometrics*, 182, p. 226 - 234.