

Université Paris X - Nanterre
UFR SEGMI

Etude des déterminants de la consommation d'eau à Paris

Sous la direction de M. Bernard Barraqué

Encadrant du stage : Julien Souriau

Présenté et soutenu par Jeanne Dequesne

Master 2 Recherche Economie du Développement Durable, de l'Environnement et de l'Energie

Parcours Environnement

Année universitaire 2011-2012

L'université de Paris X Nanterre n'entend donner aucune approbation, ni désapprobation aux opinions émises dans ce mémoire ; elles doivent être considérées comme propres à leur auteur.

Table des matières

1	Introduction	5
2	Revue de la littérature	9
2.1	Etude de la baisse de la consommation d'eau à Paris	9
2.2	Modélisation de la demande en eau et élasticité-prix	13
3	Présentation des données	17
3.1	Présentation des données pour l'étude de la consommation journalière d'eau à Paris	17
3.2	Présentation des données pour l'étude de la consommation d'eau potable des ménages Parisiens	18
3.2.1	Présentation et construction de la base de données	18
3.2.2	Etude des Iris supprimés : statistiques descriptives et comparaisons	25
4	Etude économétrique de la consommation journalière d'eau à Paris	31
4.1	Objectifs	31
4.2	Méthodologie empirique	32
4.3	Résultats économétriques	34
4.4	Autres variables testées	38
5	Etude économétrique de la consommation d'eau potable des ménages parisiens	40
5.1	Objectifs	40
5.2	Méthodologie empirique	41
5.3	Résultats économétriques	43
5.3.1	Sélection des variables explicatives pertinentes du modèle	44
5.3.2	Résultats économétriques	48

5.3.3	Etude de différents sous-échantillons	51
6	Conclusion	60
7	Bibliographie	62
8	Annexes	64
8.1	Présentation des différentes catégories disponibles dans la base de consommation d'Eau de Paris utilisées pour distinguer les différents usagers des services d'eau	64
8.2	Revenu net déclaré par foyer fiscal par arrondissement : données utilisées pour le découpage des groupes d'arrondissements par niveaux de revenus	65
8.3	Comparaison des estimations sur la base domestique et sur la base d'origine	67

1 Introduction

Les problématiques concernant les ressources en eau sont multiples. Dans les pays en développement, les principales difficultés rencontrées concernent la rareté des ressources (dans certains pays, notamment en Afrique Noire) et leur mauvaise qualité. Les ressources en eau sont plus abondantes et de meilleure qualité dans les pays développés. Cependant, ces pays riches sont confrontés à de nouveaux problèmes concernant principalement la gestion des services d'eau (distribution, assainissement, etc.). Depuis une trentaine d'années, une littérature assez conséquente s'est développée à ce sujet en France et aux Etats-Unis. De nombreux articles se sont également intéressés aux comportements de consommation des populations des pays développés (estimation d'une fonction de demande d'eau des ménages, calcul d'une élasticité prix et/ou revenu, etc.). Ce mémoire se place dans cette littérature puisqu'il s'agira pour nous d'étudier les variables qui influencent la consommation d'eau à Paris.

Alors que la consommation d'eau à Paris a été en croissance quasi continue jusque la fin des années 1980, elle a connu une forte baisse du début des années 1990 à aujourd'hui (décroissance qui semble ne pas s'atténuer ces dernières années).¹ Notons que cette évolution, caractéristique des grandes villes françaises, n'a pas connu la même tendance que les consommations d'autres ressources naturelles telles que le pétrole ou le gaz.

Notre travail se propose d'étudier et de modéliser la consommation d'eau. En effet, nous chercherons à déterminer les variables qui influencent la consommation d'eau de la ville de Paris. Ce sujet est assez novateur car très peu d'études ont été menées en France pour tenter de modéliser et d'expliquer la demande (et donc la consommation) en eau. Les études existantes sont souvent relativement anciennes et peu d'entre elles

¹Cette évolution continue à la baisse cause des difficultés en termes de financement des services d'eau et donc de renouvellement et entretien des infrastructures. De plus, si les niveaux consommés continuent de diminuer, les infrastructures actuelles risquent de s'avérer inadaptées car elles laisseraient l'eau « stagner » trop longtemps dans les tuyaux et réservoirs ce qui serait préjudiciable à la qualité de l'eau potable et dangereux à long terme.

traitent le cas de Paris. Bien que de nombreuses études portent sur les Etats-Unis, celles-ci ne peuvent se généraliser au cas parisien. En effet, la tarification et les usages de l'eau² étant très différents, ces études seraient inappropriées à expliquer la consommation d'eau en France, et précisément à Paris (où la consommation se fait quasiment exclusivement pour des usages intérieurs). De plus, aux Etats-Unis, la majorité des habitations sont équipées de compteurs individuels ce qui rend les études sur la consommation des ménages beaucoup plus précises que celles qui peuvent être menées dans les grandes villes urbaines françaises (et notamment Paris).

Comprendre et expliquer les déterminants de la consommation d'eau devrait aussi permettre de mieux appréhender l'avenir et ainsi pouvoir adapter le fonctionnement et les infrastructures des services d'eau à Paris afin de garantir une bonne qualité de l'eau à long terme et un système pérenne. Ceci est d'autant plus important que la consommation d'eau relève en partie de la satisfaction des besoins vitaux.

Dans le constat général de baisse de la consommation totale d'eau à Paris, il est important de savoir quels sont les principaux usagers à l'origine de cette évolution. Or, sur ce point précis il semble y avoir un désaccord puisque certaines études se contredisent. Selon Guy Poquet, la consommation domestique joue un rôle très faible puisque les consommations des habitations collectives et celles des maisons individuelles sont relativement stables sur la période 1983-2002 pour la zone Ile de France, d'après un rapport du CREDOC de Novembre 2003 (Consommation et Modes de Vie N°170, Guy Poquet).

Contrairement à cette étude, Bernard Barraqué et Alexis Nercessian expliquent dans un travail de 2008 que 56% de la baisse de la consommation d'eau à Paris observée ces dernières années s'explique par une légère baisse au sein de la très nombreuse catégorie "habitations". Il est alors difficile d'avoir un avis tranché sur la question et nous ne partirons pas dans cette voie. Ainsi, lorsque nous étudierons la consommation domes-

²Aux Etats-Unis les consommations d'eau pour des usages extérieurs sont importantes dans les banlieues des grandes villes.

tique d'eau potable, nous chercherons à comprendre ce qui explique les différences de niveaux de consommation d'eau entre différentes zones géographiques parisiennes sans chercher à expliquer une évolution dans le temps (les données dont nous disposons sont en coupe transversale, et non des séries temporelles ce qui rend impossible l'explication d'une évolution dans le temps).

Nous utiliserons différentes données afin de prendre compte le maximum de variables pertinentes pour expliquer la consommation d'eau. Ces différentes variables pourront aussi bien être temporelles qu'individuelles.

Nous commencerons par l'étude d'une base de données concernant les consommations journalières d'eau à Paris sur la période 1980-2010. Ces données seront couplées avec des variables climatiques et des variables concernant le type de jour (jour ouvré, jour de week-end, jour férié, jour de vacances scolaires, etc.) afin de comprendre les cycles de consommation d'eau infra-annuels (c'est à dire les variations à l'intérieur d'une année et qui se reproduisent d'année en année).

Ensuite nous étudierons un jeu de données individuelles, disponibles seulement sur une année (2011 pour les relevés de consommation d'eau, et 2007 pour les données sociodémographiques INSEE), contenant les relevés des compteurs d'eau pour chaque compteur parisien (environ 90 000 compteurs). Les données concernant les consommations d'eau par compteur seront regroupées par Iris, ainsi elles pourront être associées à de nombreuses variables sociodémographiques disponibles elles aussi à l'échelle de l'Iris.

Nous commencerons ce mémoire par une revue de la littérature concernant la consommation et/ou la demande en eau potable (ou non potable) principalement en France afin de cibler notre travail. Nous présenterons en détail les données utilisées, leurs sources, les modifications apportées à ces données, les regroupements effectués etc. Nous analyserons enfin les deux études économétriques réalisées, les difficultés ren-

contrées, les solutions trouvées et discuterons des résultats obtenus ainsi que de leur robustesse ; avant de conclure.

2 Revue de la littérature

Comme énoncé dans l'introduction, peu de travaux traitent de la consommation d'eau en France, plus précisément de l'étude (économétrique/statistique) de ses déterminants. Certains travaux semblent cependant donner de bonnes pistes de réflexion pour comprendre les principaux groupes de variables susceptibles d'influencer les volumes d'eau consommés. Notamment, certains articles permettent de comprendre, au moins en partie, la baisse de la consommation d'eau observée ces 20 dernières années à Paris. Nous présenterons ici les principaux travaux et résultats concernant ce champ d'étude, organisés en 2 axes :

- les études portant sur la baisse de la consommation à Paris
- les tentatives de modélisation de la demande d'eau, notamment les études portant sur l'estimation d'une élasticité-prix (et/ou revenu)

2.1 Etude de la baisse de la consommation d'eau à Paris

Le travail le plus notable concernant l'étude de la consommation d'eau à Paris, et probablement l'un des plus récents, est le rapport final proposé par B. Barraqué et A. Nercessian (2008) « Mieux comprendre comment évolue la consommation d'eau à Paris ». Ce rapport analyse en détail les consommations d'eau par catégories, notamment en distinguant les petits, moyens et gros consommateurs, ainsi que différents « lieux » de consommation (habitations, enseignement, usagers municipaux, divers, commerces, hôtels/restaurants, bureaux, sports/loisirs, médecine).

Ce rapport permet de comprendre quels sont les usagers responsables de la plus grande part de la baisse de la consommation d'eau à Paris depuis 1990. De même, il

donne des informations sur les types de consommateurs qui connaissent les plus fortes variations infra-annuelles (cycles hebdomadaires notamment).

- Le résultat le plus important est que plus de la moitié (56%) de la baisse de la consommation d'eau à Paris est expliquée par une baisse « diffuse » à l'intérieur de la catégorie « habitations » ; en effet, une très grande part de ces compteurs enregistre une faible baisse, mais au vu du nombre de ces compteurs, la baisse globale en devient remarquable.
- Deuxièmement, l'étude de cas sur certains compteurs particuliers permet de confirmer la saisonnalité des consommations d'eau à Paris ; notamment, les écoles et certains services d'hôpitaux ferment en été, les bureaux et administration le week-end et en été également alors que les consommations dans les habitations sont plus stables.

« Vercingétorix : analyse des consommations d'eau dans un quartier de Paris » publié dans la revue Techniques-Sciences-Méthodes en 2000 par E. Jaskulke, J.P. Maugendre et S. Cambon-Grau est un bon exemple d'une étude de cas s'intéressant à la baisse de la consommation d'eau à Paris. Cet article étudie les comportements de consommation en eau des habitants d'un quartier précis (d'un îlot) situé dans le 14ème arrondissement de Paris entre la rue Vercingétorix, la rue Raymond Losserand et le Boulevard Brune. Les principales variables utilisées informent sur la nature du patrimoine bâti et le comportement des occupants. Les données utilisées concernent les années 1994, 1995 et 1996. Les résultats sont répartis en deux catégories :

- concernant le secteur de l'habitat, « le facteur le plus discriminant pour expliquer les différences de consommation d'eau par occupant est l'existence, ou non,

d'un contrat d'entretien de la robinetterie. » (les immeubles sans un tel contrat consomment en moyenne 20% de plus que ceux qui en possèdent un)

- concernant les autres secteurs, divers résultats sont énoncés, dont le plus important semble être l'impact notable des comportements d'économies d'eau (notamment les opérations « anti-fuites ») des gros consommateurs publics comme privés.

« L'influence des gros consommateurs sur l'évolution des consommations d'eau à Paris », publié dans la revue Techniques sciences méthodes, génie urbain génie rural (2000, n°2, pp. 19-22) étudie les évolutions des consommations d'eau d'un échantillon de gros consommateurs de la rive droite parisienne sur la période 1991-1998. Cet échantillon représente 1% des clients et 15% des consommations de cette zone géographique. Les principaux résultats qui ressortent de ce travail sont les suivants :

- la baisse observée sur l'échantillon est responsable d'environ 32% de la baisse de la rive droite sur la période étudiée
- les plus gros consommateurs sont ceux qui représentent les baisses de niveaux consommés les plus significatives
- les baisses les plus remarquables sont constatées dans les catégories « banques et assurances », « grands magasins » et « fontaines et piscines » alors que les catégories d'habitations (copropriétés et bailleurs sociaux) ont connu des baisses beaucoup plus relatives alors qu'elles représentent plus des 2/3 des consommations totales en eau de la ville de Paris

Le rapport du CREDOC datant de 2005 : « Etude de l'évolution des consommations d'eau potable à Paris » rédigé par Poquet Guy, H. Henriquez, L. Pouquet et M. Mazars

fait également office de référence dans le domaine des recherches sur les causes de la baisse de la consommation d'eau à Paris. Les conclusions majeures de ce rapport sont les suivantes :

- le niveau des consommations d'eau parisiennes de 2005 est comparable à celui de 1957, ceci signifie qu'après une période d'accroissement continu de la demande en eau, celle-ci a sensiblement diminué depuis le début des années 1990. Notons que cette évolution est comparable à celle de nombreuses autres grandes villes à l'échelle européenne
- la modification du « tissu économique » de l'agglomération parisienne est en partie responsable de cette évolution. En effet, les principaux facteurs qui sont présentés sont le vieillissement de la population et la « gentrification » (phénomène d'embourgeoisement) qui poussent à des comportements plus responsables et donc économes en eau
- la démographie (la population notamment) stable sur les 30 dernières années n'est pas responsable de la diminution de la consommation d'eau à Paris
- les prévisions de ce rapport tablent sur une baisse des consommations d'eau au moins jusqu'en 2008 (année où elles devraient n'atteindre que 198 millions de m^3); cette prévision est aujourd'hui avérée

Notons que le CREDOC a publié trois autres rapports antérieurs (1999, 2002 et 2003) dans lesquels, au niveau de la région Ile de France, sont révélés plusieurs résultats importants et notamment l'influence de la tertiarisation de l'économie et des variations de la population active sur la consommation d'eau dans cette région. De même le rôle du vieillissement de la population est remarqué cependant les comportements d'économies d'eau des ménages semblent peu significatifs au niveau global.

Les variables climatiques sont d'après ces rapports peu pertinentes pour expliquer les variations des niveaux consommés. L'effet du prix est quant à lui discuté bien qu'aucune conclusion ne soit réellement tirée. En effet, il y est expliqué que l'élasticité prix de la demande en eau est faible et que les niveaux de consommation et de prix sont liés, corrélés, mais le sens de la causalité n'est pas établi dans ce rapport.

2.2 Modélisation de la demande en eau et élasticité-prix

Les premiers travaux économétriques et statistiques visant à modéliser la consommation/demande en eau des ménages datent de la fin des années 1960. De nombreux travaux concernent les Etats-Unis et utilisent des méthodes économétriques assez développées et plutôt bien adaptées aux données utilisées dans ces années (données agrégées en coupe transversale, méthode des Moindres Carrés Ordinaires).

Il s'agit principalement des travaux de Howe et Linaweaver (1967), Gibbs (1978) et Danielson (1979). Ces travaux concluent à une inélasticité « imparfaite » aux prix de la demande en eau potable des ménages (ce qui signifie que les ménages ne réagissent que très peu aux variations de prix, mais ils ne sont pas tout à fait indifférents). Ceci conduit à développer aux Etats-Unis des systèmes de tarifications par blocs (ou tarification progressive) avec une hausse globale notable des factures des abonnés. Dès les années 1980, les données disponibles évoluent et permettent (aux Etats-Unis) de travailler sur des séries à deux dimensions (individuelle et temporelle) : des données de Panel.

Ces données permettent de contrôler et donc de capter de nombreuses informations et variables qui influencent la demande en eau afin d'en obtenir une modélisation assez fine. Ainsi, se développent des méthodes économétriques complexes mais indispensables au travail sur de telles données. Le premier point qui apparaît concerne l'inadéquation de la méthode des MCO dans le cas de régresseurs endogènes (notamment lorsqu'on introduit le prix dans la liste des variables explicatives). Ainsi, la méthode des variables

instrumentales semble être la plus fréquemment utilisée. Cette méthode d'estimation permet d'estimer des modèles avec un ou plusieurs régresseurs endogènes (le plus souvent les prix) tout en obtenant des estimations non biaisées et robustes. Ce type de méthode est cependant parfois difficile à mettre en oeuvre puisque de nombreux critères strictes doivent être vérifiés et validés (exogénéité de l'instrument principalement).

Les principaux travaux en France remontent aux années 1990. Notamment, l'article de Boistard (1993) semble le précurseur dans ce domaine. Il utilise un panel national de 500 services de distribution d'eau potable. La variable observée de ce panel correspond à la consommation moyenne par équivalent habitant permanent, à laquelle est associée la variable « prix moyen pour une consommation de $100 m^3$ pour 1975, 1980, 1985 et 1990. Boistard estime ainsi une élasticité-prix de la demande en eau (entre $-0,10^3$ et $-0,20$ à court terme soit 2-3ans ; et entre $-0,25$ et $-0,35$ à long terme soit 5-10ans).

Cependant, Boistard ne propose pas dans cet article un modèle explicatif de la consommation d'eau ; il se limite simplement à l'estimation de l'élasticité prix de la demande en eau. De plus, la non prise en compte d'autres variables potentiellement pertinentes telles que des variables sociodémographiques risque de biaiser ses résultats.

Une étude assez similaire est menée par Point (1993) sur le département de la Gironde. Il estime ainsi une élasticité prix de la consommation d'eau de l'ordre de $-0,17$ en utilisant également la méthode des MCO. Enfin, le rapport du CREDOC (1997) : « L'eau et les usages domestiques » réalise un gros travail économétrique sur diverses données pour comprendre les facteurs explicatifs de la consommation d'eau. Notamment, trois catégories de variables sont utilisées :

- des variables climatiques
- des variables démographiques

³Une élasticité prix de $-0,10$ signifie qu'une augmentation du prix de l'eau de 10% entraîne une diminution de la consommation d'eau en moyenne de 1%.

- des variables économiques

Ce travail est le premier à prendre en compte ces différents jeux de variables afin de proposer un modèle explicatif complet de la consommation d'eau. Les données utilisées sont les compteurs individuels. Elles permettent ainsi de comprendre les décisions de consommation à un niveau microéconomique, celui du ménage.

Les résultats majeurs qui ressortent de ce rapport sont les suivants :

- une élasticité prix de court terme de la demande égale à -0,32
- une sensibilité de la demande au prix plus élevée en 1995 qu'en 1989
- une élasticité revenu faible car l'eau est un bien de première nécessité
- l'amélioration des équipements (et des taux d'équipements des ménages) explique une augmentation de la consommation d'eau jusqu'au début des années 1990
- le vieillissement de la population entraîne une baisse de la consommation
- l'accroissement de population entraîne une hausse de la consommation

Les études les plus récentes concernant la modélisation de la demande en eau, et l'estimation de son élasticité-prix sont principalement les travaux de Nauges et Reynaud (2001) : « Estimation de la demande domestique d'eau potable en France » qui concluent un impact significatif de variables telles que le prix, le revenu moyen, les caractéristiques des logements, la présence de baignoires dans les logements ou les précipitations estivales pour expliquer les comportements de consommation en eau potable en Moselle et en Gironde.

Un travail très intéressant est l'article de Dalhuisen Jasper M. et al. (2003) : « Price and income elasticities of residential water demand : a meta-analysis ». Celui-ci reprend

de nombreux travaux existants qui ont estimé des élasticités prix et/ou revenu de la demande résidentielle en eau. Cette méta-analyse repose sur le calcul de 296 élasticités prix de la demande en eau des ménages (issues de différentes régions du monde entier) et conclut à une élasticité prix moyenne de la demande en eau de -0,41. Cette valeur est sensiblement plus élevée que celles des estimations réalisées pour la France.

3 Présentation des données

Dans cette partie nous présenterons de manière assez détaillée les bases de données que nous avons utilisées pour les deux études économétriques. Nous expliquerons les modifications que nous avons apportées en les justifiant. Puis nous présenterons des statistiques descriptives en rapport au traitement de la base de données utilisée pour la deuxième partie de l'étude économétrique (l'étude de la consommation moyenne d'eau des ménages parisiens).

3.1 Présentation des données pour l'étude de la consommation journalière d'eau à Paris

La base de données disponible et utilisée pour ce travail se compose de la variable concernant le volume d'eau journalier distribué à Paris entre Janvier 1986 et Mai 2011. Nous avons traité cette variable pour en obtenir une bonne approximation du volume total d'eau consommé quotidiennement à Paris. Pour cela, nous disposons des informations concernant le pourcentage de pertes (estimées) au niveau annuel, pour toutes les années de la base de données. Nous avons alors réparti ce pourcentage de pertes sur les 365 jours (en moyenne) des différentes années de la base, afin d'obtenir la variable dépendante que nous étudierons : le volume total d'eau consommé quotidiennement à Paris. Cette approximation est jugée comme très juste car les pertes et fuites sont relativement faibles, et considérées comme réparties de manière homogène sur l'année.

Les variables disponibles pour tenter d'expliquer la consommation d'eau sont premièrement des variables climatiques :

- température maximale en °C
- température minimale en °C
- température moyenne en °C
- température maximale arrondie en °C

- humidité max (en pourcentage)
- humidité min (en pourcentage)
- humidité moyenne (en pourcentage)
- volume des précipitations journalières (en mm)
- insolation : nombre de minutes d'ensoleillement par jour
- vitesse moyenne du vent en km/h
- indice humidex⁴ moyen

Ensuite, nous disposons d'informations concernant le type de jour, notamment les jours de vacances scolaires, les jours fériés, les jours de week-end (samedi et dimanche), et bien entendu un calendrier qui informe de la date précise de chaque jour de format jj/mm/aaaa.

A noter que la base de données n'est complète que pour la période Janvier 1990-Décembre 2010, nous retiendrons donc ces 20 années pour l'étude économétrique qui nécessite des données complètes pour être plus précise et pertinente.

3.2 Présentation des données pour l'étude de la consommation d'eau potable des ménages Parisiens

3.2.1 Présentation et construction de la base de données

Pour ce travail, les données utilisées proviennent de différentes sources.

La base de données concernant les consommations d'eau provient d'Eau de Paris, il s'agit d'une base comportant les relevés des consommations d'eau (journalières moyennes, consommations facturées) des compteurs parisiens (environ 91 000 compteurs) pour l'année 2011. Il s'agit d'une base recensant tous les compteurs parisiens qu'ils soient des compteurs domestiques (habitations collectives ou individuelles avec compteurs le plus souvent collectifs mais aussi parfois individualisés), des compteurs d'administrations (mairies, écoles, organismes publics tels que l'ANPE, des entreprises publiques comme

⁴ « L'indice humidex est une mesure utilisée par les météorologues canadiens pour intégrer les effets combinés de la chaleur et de l'humidité. Il diffère de l'indice de chaleur utilisé aux États-Unis : celui-ci utilise l'humidité relative plutôt que le point de rosée. » Wikipedia.

par exemple des ministères), des compteurs de professionnels (bureaux, commerces, restaurants, hôtels etc.)

A chaque compteur est associé un numéro de contrat, un numéro de point de desserte, le diamètre du compteur, la puissance du compteur, le système de récupération des niveaux de consommation (télé-relève ou non), le statut de l'abonnement (notamment en-cours ou résilié), l'adresse précise du compteur (numéro, rue, arrondissement), le type d'abonnement (eau potable ou non), le nom (ou raison sociale) du client titulaire du contrat et du client payeur, le libellé de la catégorie d'abonnement (eau potable ordinaire, eau potable incendie, Paris Plage, etc.). D'autres informations sont également présentes dans cette base mais ne sont pas détaillées ici car elles n'ont pas été utilisées dans le cadre de ce travail. Ces données concernent toutes l'année 2011.

Pour plus de cohérence seules les consommations d'eau potable ont été étudiées car elles représentent une très grande partie de la consommation totale d'eau à Paris et surtout une très grande majorité des compteurs (domestiques) parisiens.

La variable dépendante que nous avons choisi d'étudier est la consommation annuelle moyenne (par habitant et par Iris) d'eau potable domestique. Pour « créer » cette variable, un simple calcul a été effectué à partir de la variable « consommation journalière moyenne sur l'année 2011 » que nous avons multiplié par 365 (soit le nombre de jours dans l'année). Nous obtenons ainsi une consommation annuelle totale pour chaque compteur parisien présent dans la base « Eau de Paris ». Certaines corrections ont dû être apportées « à la main », notamment dans le cas de niveaux de consommation aberrants (consommations négatives, ou extrêmement élevées alors qu'il s'agit de bornes incendie, ou très basses alors qu'il s'agit d'administrations...). Ces modifications ont principalement été réalisées par Julien Souriau (doctorant à Eau de Paris et tuteur de mon stage).

Ces données ont été croisées avec des données sociodémographiques obtenues auprès

de l'INSEE , elles concernent l'année 2007 (recensement de 2005), mais nous pouvons les associer aux données de consommation de 2011 car les données socio-démographiques sont relativement stables dans le temps (dans le cas d'une période de quelques années).

Les données de l'INSEE utilisées correspondent à cinq bases distinctes concernant le logement, la famille, la population, l'activité de la population et le niveau de formation/scolarisation de la population. Ces données sont disponibles par Iris, zones géographiques définies par l'INSEE comme des ensembles d'îlots contigus (environ 50 800 Iris en France). Ces cinq bases de données portent sur toute la France, nous avons donc conservé uniquement les données concernant la ville de Paris.

Techniquement, pour que le croisement des données de consommation par compteurs et des données sociodémographiques par Iris soit possible, nous avons regroupé les données de consommation d'eau disponibles par compteurs en données par Iris.

Le premier travail a donc été d'associer à chaque compteur un numéro d'Iris (via l'utilisation d'une carte de correspondance), sachant que chaque Iris contient en général entre 20 et 100 compteurs (mais certains Iris contiennent un unique compteur et d'autres plus d'une centaine). Ce travail a été réalisé par Pierre Mahieu (Géomaticien et administrateur fonctionnel d'Eau de Paris).

Ensuite, il a fallu regrouper les compteurs par Iris. Notons ici, que certains compteurs n'avaient pas pu être associés à l'aide de la carte de correspondance à un Iris particulier (environ 500). Un travail de recherche d'association entre ces compteurs et les Iris correspondants a donc été réalisé « à la main » notamment en approximant le numéro d'Iris des compteurs par leur adresse, le plus souvent en encadrant l'adresse par deux autres adresses proches et qui se trouvaient souvent dans le même Iris (exemple : recherchant le numéro d'Iris du compteur numéro 1138930 situé au 14 Rue Eugène Million, le numéro d'Iris qui lui a été associé est le 751155712 puisque le 12 et le 16 de cette même rue font partie de cet Iris). Pour compléter et vérifier que les compteurs

avaient bien été regroupés dans les « bons » Iris nous avons aussi réussi à obtenir une carte plus précise (mieux découpée) des Iris parisiens ce qui nous a permis de corriger certaines associations qui été erronées (notamment pour les Iris frontaliers de la Seine).

A ce stade, la base de consommation d'eau contenait les 91421 compteurs parisiens (tous usages et abonnés confondus).

Cependant, puisque les données INSEE dont nous disposons concernent exclusivement la population des Iris (et non des informations sur l'activité économique des Iris) nous avons jugé pertinent de s'intéresser uniquement à la consommation domestique d'eau potable (consommations des ménages vivant dans ces différents Iris).

Ainsi, un long travail a été nécessaire pour retenir uniquement les compteurs qui correspondent à des usagers domestiques et non à des consommations d'administrations publiques, de commerces, de bureaux, ou encore des hôtels et restaurants.

Notons tout d'abord que la base d'Eau de Paris contenait différentes informations susceptibles de simplifier ce travail. Dans un premier temps la base disposait d'une colonne indiquant si chaque compteur correspondait à une consommation d'eau potable ou d'eau non potable, nous avons ainsi pu retirer de la base les compteurs correspondants à des consommations d'eau non potable.

Précisons que les données des consommations d'eau par compteur proviennent d'Eau de Paris (entreprise publique de gestion et de distribution de l'eau pour la Ville de Paris) et n'ont pas été créées pour les étudier, pour en tirer partie d'un point de vue scientifique, mais seulement dans un objectif de suivi des compteurs en fonction de leurs niveaux de consommation (notamment distinction entre petits, moyens et gros consommateurs). En effet, historiquement, les services publics de distribution d'eau pratiquaient des tarifications différentes en fonction des niveaux de consommation (notamment des rabais pouvaient être accordés pour les consommations supérieures à $6000 m^3$ par an). Dans cette optique, l'information sur le type d'abonné présent derrière chaque

compteur importait peu au moment de la construction de la base. Ainsi des catégories ont été renseignées mais d'une manière très imprécise et donc avec un grand nombre d'erreurs. Cependant, les colonnes « Libellé qualité du client » et « Libellé NAF activité » ont été renseignées pour une très grande partie des compteurs et elles seront utilisées pour un premier travail de sélection des compteurs domestiques. La colonne la plus intéressante ici est la première mentionnée ci-dessus, elle comporte 6 catégories :

- Administration
- Habitation
- Habitation collective
- Individualisé SRU
- Habitation Individuelle
- Professionnel

Il a donc été décidé d'utiliser, pour commencer, ces catégories afin d'effectuer une première distinction entre compteurs domestiques et non domestiques. Nous nous sommes ainsi fixés certains critères de tri pour enlever la plus grande partie des compteurs non domestiques.

Ont été conservés les compteurs de la catégorie Administration si le code NAF associé été soit vide soit « Activités Immobilières » car ces compteurs correspondaient en grande majorité à des logements sociaux de la ville de Paris (et donc des consommations d'eau domestiques).

Dans la catégorie Habitation nous avons choisi de conserver les compteurs pour lesquels le code NAF est « Activités Immobilières » car il s'agit dans la majorité des cas d'habitations (souvent collectives bien qu'elles soient dans la catégorie Habitation) et des syndicats de copropriétaires. Pour cette même catégorie, un important travail de tri par compteur a été nécessaire. En effet, environ 16 000 compteurs de cette catégorie n'avaient pas de code NAF renseigné ce qui pouvait laisser penser qu'il s'agit

réellement d'une habitation, cependant cette catégorie très « fouillis » contient aussi bien des habitations, des bureaux, des commerces, des hôtels/restaurants etc. . . Ainsi nous avons décidé d'étudier en détails ces compteurs, notamment par le nom du client titulaire et/ou payeur et de l'adresse afin de déterminer s'il s'agissait réellement d'une habitation ou non. Pour cette partie du travail, les sites internet « MeilleursAgents.com » et la partie « Street View » de Google Maps ont été utilisés pour observer les façades et les compositions des immeubles lorsque la décision (compteur domestique ou non) n'avait pas pu être prise simplement à partir des catégories ou du nom du client titulaire/payeur.

Les catégories “Individualisé SRU” et “Habitation Individuelle” ont été conservées dans leur totalité car elles correspondent à des compteurs domestiques individuels (chaque compteur correspondant à un logement). Enfin, concernant la catégorie “Professionnel”, seuls les compteurs pour lesquels le code NAF renseigné est « Activités Immobilières » sont conservés car il s'agit très souvent de syndicats de copropriétaires ou d'agences immobilières propriétaires d'immeubles résidentiels.

Après avoir regroupé les compteurs jugés « domestiques », nous avons calculé la consommation moyenne par habitant, par Iris (à l'aide de la variable « population totale » disponible dans la base Population de l'INSEE au niveau de l'Iris). Nous avons alors remarqué que certains Iris étaient alors caractérisés par une consommation d'eau moyenne par habitant très élevée (supérieure à $100 m^3$ par an par personne) ou très basse (inférieure à $35 m^3$ par an et par personne) sachant que la consommation annuelle moyenne par habitant à Paris se situe historiquement aux alentours de $50 m^3$. Ces seuils ont été choisis afin de limiter des erreurs trop importantes en termes de consommations moyennes qui auraient pu nuire à l'étude économétrique. De même les Iris concernés n'étant alors que peu nombreux, nous pouvions prendre le temps de les retraiter.

Il a alors été décidé de reprendre l'ensemble des compteurs de ces Iris marqués par des consommations d'eau par habitant aberrantes afin de les « vérifier » à nouveau. Certains compteurs ont été à nouveau retirés de la base car correspondant à des compteurs non domestiques et pour certains Iris (ceux pour lesquels les consommations moyennes étaient à ce moment jugées comme anormalement basses) nous avons cherché parmi les compteurs jugés « non domestiques » certains compteurs qui auraient été « sortis » de la base domestique par erreur.

Nous avons donc enfin obtenu une base des consommations d'eau domestiques assez cohérente qui semblait ne pas contenir trop d'erreurs.

A partir de cette base domestique nous avons créé plusieurs variables dépendantes que nous pourrions potentiellement utiliser, à savoir la consommation domestique totale par Iris et la consommation moyenne par habitant (ces deux variables s'entendent au niveau annuel, pour l'année 2011).

Nous disposons donc de données sur les consommations domestiques d'eau potable à Paris au niveau de l'Iris, données que nous pouvons dès à présent croiser avec les données sociodémographiques (INSEE) disponibles elles aussi au niveau de l'Iris. A partir de ces différentes données nous tenterons de construire un modèle explicatif des déterminants de la consommation d'eau des ménages Parisiens.

Précisons que dans la base domestique conservée pour la suite du travail, certains Iris ont disparu (involontairement en supprimant tous les compteurs de ces Iris car jugés non domestiques, ou Iris volontairement supprimés car comportant des consommations d'eau élevées sans aucune population recensée dans ces Iris). Afin de justifier les décisions prises, nous présenterons dans la section suivante des statistiques descriptives de comparaison entre la base complète et les Iris « supprimés ».

3.2.2 Etude des Iris supprimés : statistiques descriptives et comparaisons

Comme énoncé plus haut, le choix de travailler sur la consommation domestique d'eau potable a entraîné un tri des compteurs parisiens issus de la base de données originale.

Ce tri (suppression de compteurs correspondants à des usages non domestiques de l'eau) a débouché sur la disparition ou suppression volontaire de tous les compteurs de certains Iris. La ville de Paris est découpée en 987 Iris, parmi lesquels 897 restent présents dans la base domestique, le tri a donc entraîné la suppression de 90 Iris.

Afin de vérifier que ce tri ne soit pas préjudiciable à l'étude économétrique, nous nous devons de montrer que les données triées (même si elles contiennent des erreurs, notamment des compteurs de consommation domestique qui auraient été sortis de la base par erreur et surtout des compteurs non domestiques conservés dans la base dite « domestique ») n'ont pas entraîné l'apparition d'un biais de sélection.

En effet, cette notion de « biais de sélection », introduite par Heckman (1979) désigne une erreur systématique faite lors de la sélection des sujets à étudier. Ce terme regroupe tous les biais pouvant conduire à ce que les sujets effectivement observés lors d'une enquête ne constituent pas un groupe représentatif des populations censées être étudiées (Wikipedia). Par exemple, certains Iris ont été supprimés volontairement lorsque les consommations par habitant étaient considérées comme aberrantes (notamment au dessus de $100 m^3$ par habitant par an, ou en dessous de $35 m^3$) après vérification des compteurs de ces Iris. Ainsi, ces Iris peuvent avoir été supprimés par erreur alors que les niveaux de consommation étaient peut être extrêmes pour une bonne raison (notamment des Iris très riches ou très pauvres). Finalement, le risque est de « sortir » de la base des populations très riches ou très pauvres et donc de se concentrer (sans le vouloir et sans le savoir) uniquement sur les consommations d'eau des classes moyennes.

Deux études statistiques sont alors réalisées pour tenter de vérifier qu'un tel biais n'est pas apparu suite à ce tri. Premièrement, nous vérifierons que les Iris supprimés

sont différents les uns des autres en termes de niveau de vie, de population et de caractéristiques des habitations. Deuxièmement, nous comparerons les caractéristiques moyennes des Iris supprimés aux caractéristiques moyennes de tous les Iris parisiens pris dans leur ensemble. Les deux graphiques ci-dessous présentent des statistiques descriptives pour les Iris supprimés lors de la sélection des compteurs d'eau domestiques.

Figure 1: Indicateurs de Richesse des Iris supprimés

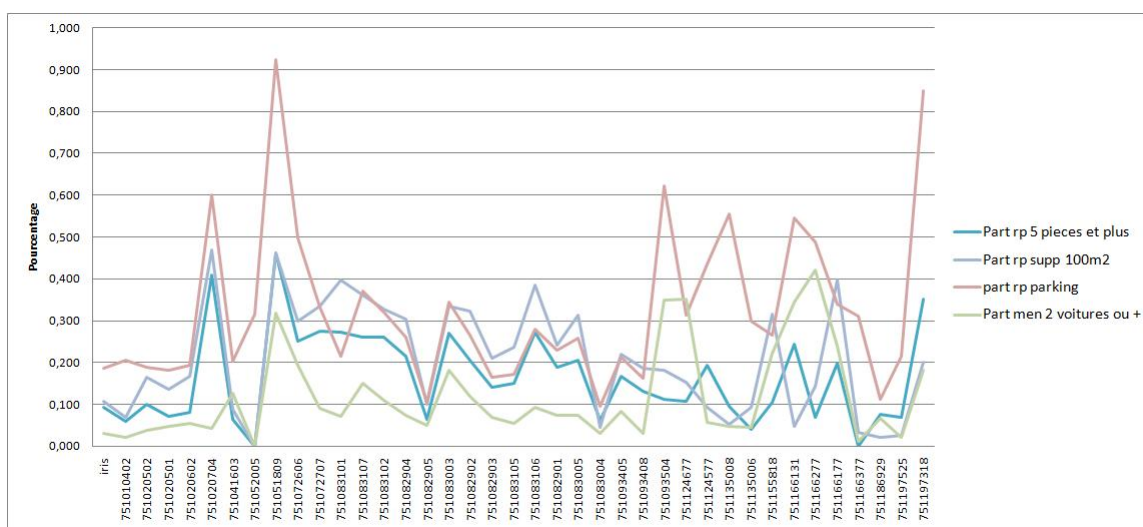
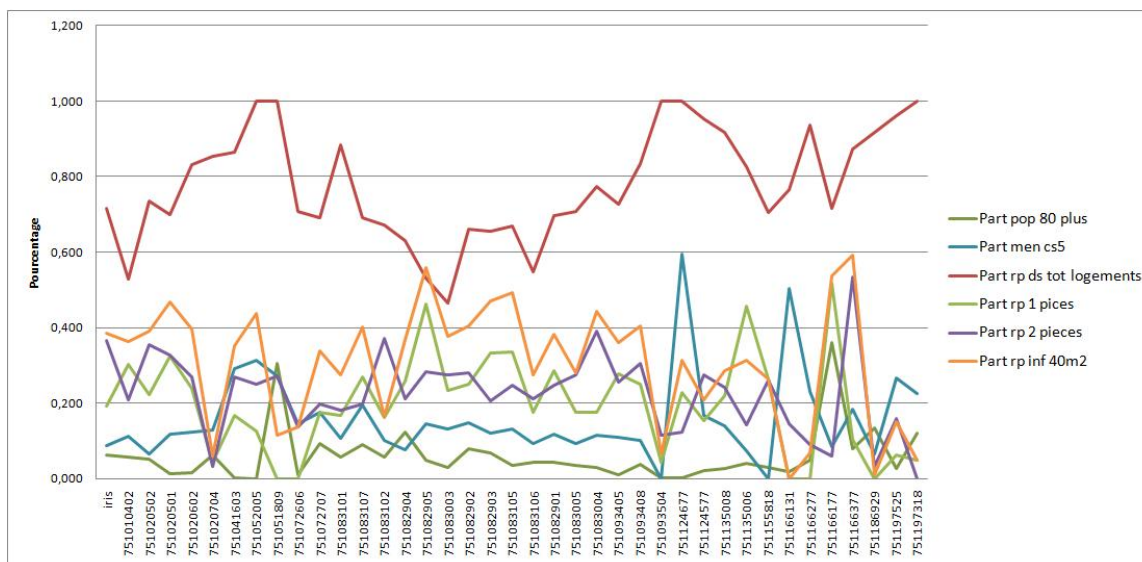


Figure 2: Statistiques Iris supprimés



Comme nous pouvons le voir dans les deux tableaux ci-dessus, les Iris supprimés sont assez différents en termes de caractéristiques des populations et des habitations. Les variables présentées permettent bien de vérifier que les Iris supprimés ne sont pas tous similaires en termes de taille des logements (indicateurs utilisés concernant la surface en mètres carrés avec différents seuils, le nombre de pièces notamment les résidences principales comportant 1 pièce, 2 pièces ou 5 pièces et plus, les résidences principales équipées d'au moins un parking), d'utilisation du logement (résidence principale ou non), ainsi que des informations sur la population (part des 80 ans ou plus) et sur le niveau de richesse de cette population (part des ménages en CS5, c'est à dire des actifs de 15 à 64 ans employés).

De plus, notons que les informations concernant les logements permettent d'approximer le niveau de richesse des populations notamment dans une ville comme Paris où le prix de l'immobilier est très élevé et donc très discriminant. Ainsi les deux tableaux présentés nous permettent de conclure que le tri n'a pas conduit à la suppression d'Iris similaires et donc le risque d'avoir écarté une certaine catégorie d'habitations et/ou de population

et créé un biais de sélection semble très faible.

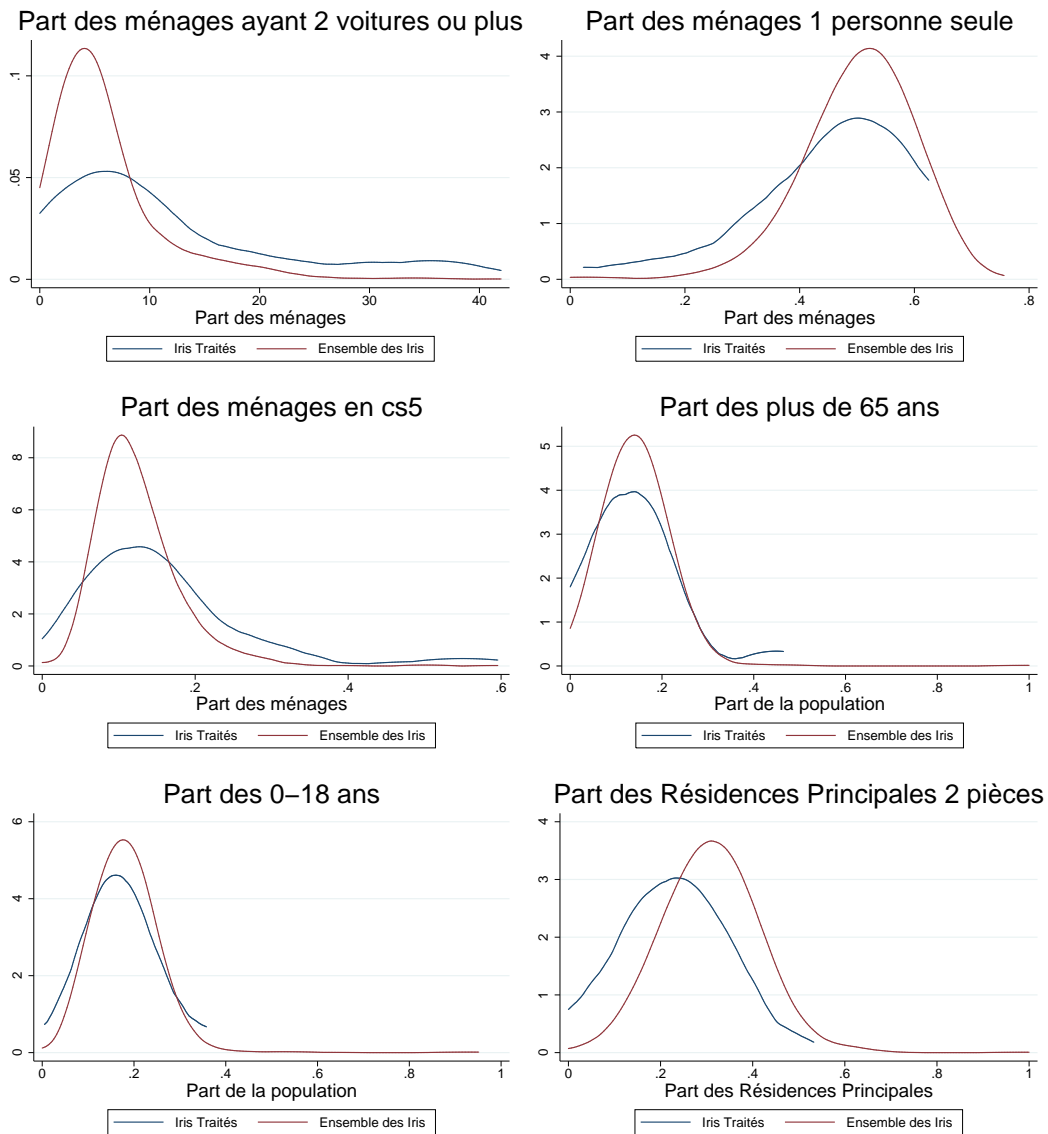
Cependant, pour corroborer cette conclusion et être certain de ne pas être confronté à un tel biais nous avons aussi réalisé certains tests statistiques de comparaison entre la base de données complète de la ville de Paris (tous les Iris parisiens) et celle comportant uniquement les Iris supprimés. Il s'agit ici de vérifier que les Iris supprimés sont en moyenne semblables aux Iris parisiens à nouveau en termes de richesse, de caractéristiques des populations et des habitations.

La figure 3 représente les graphes des densités de certaines variables explicatives pour l'échantillon des iris supprimés ainsi que pour l'échantillon original (complet). Nous pouvons ainsi comparer ces fonctions afin de vérifier que les deux échantillons sont relativement "proches" en termes de distribution. Cette analyse nous permet à nouveau de confirmer que le tri réalisé n'a pas été préjudiciable en terme de potentiel biais de sélection. En effet, comme nous le voyons dans la figure 3, les fonctions de densité sont assez semblables pour les deux échantillons concernés (pour toutes les variables présentées ici), bien que les iris supprimés représentent des distribution relativement plus plates (ce qui se justifie par le fait que l'échantillon des iris supprimés est beaucoup plus petit, il se comprend donc aisément que les valeurs prises par les différentes variables atteignent moins de valeurs extrêmes). Précisons que nous avons réalisé ces graphes avec les densités des deux échantillons pour un nombre important de variables (mais nous ne pouvons tous les représenter ici) et les deux fonctions de densité sont dans tous les cas relativement semblables. Comme nous pouvons malgré tout remarquer que certaines des variables présentées dans la figure 3 ont des fonctions de densité assez différentes entre les deux échantillons, nous présentons un autre test afin de vérifier que les échantillons sont bien proches statistiquement.

Ainsi, nous avons mené un test de comparaison de différents échantillons (échantillon complet de la ville de Paris puis échantillon « Iris supprimés ») : le test de Kolmogorov-

Smirnov. Il s'agit d'un test d'hypothèse qui permet de savoir si deux échantillons proviennent d'une même loi statistique (de même variance, même moyenne etc.).

Figure 3: Exemples de densités pour certaines variables



La table 1 représente les résultats du Test de Kolmogorov-Smirnov qui permet de vérifier que la distribution de chaque variable suit la même loi pour nos deux échantillons

Table 1: Résultats des Tests de Kolmogorov-Smirnov

Variable Considérée	P-value	Décision
Part des résidences principales	0,318	Les deux échantillons suivent la même loi
Part des résidences principales 1 pièce	0,731	Les deux échantillons suivent la même loi
Part des RP avec Parking	0,363	Les deux échantillons suivent la même loi
Part de la population de 80 ans ou plus	0,183	Les deux échantillons suivent la même loi
Part des ménages en CS5 (employés)	0,130	Les deux échantillons suivent la même loi
Part population 0-18 ans	0,612	Les deux échantillons suivent la même loi

Notes : Nous acceptons systématiquement H_0 quand la p-value est supérieure à 0,05. Plus la p-value est élevée moins il subsiste de doute quant à son acceptabilité.

(échantillon des Iris supprimés et échantillon complet) et donc que le tri effectué n'a pas créé de biais de sélection qui pourrait porter préjudice à notre étude économétrique. Nous présentons ce test pour quelques variables uniquement (6 variables).

L'hypothèse nulle (H_0) du Test de Kolmogorov-Smirnov correspond au cas où les deux échantillons suivent bien une même loi statistique.

Comme nous pouvons le constater dans la Table 1, nous acceptons pour toutes les variables l'hypothèse nulle du test, indiquant que notre tri des compteurs domestiques n'a pas induit de biais de sélection (de même que pour les fonctions de densité, nous avons répliqué ce test d'hypothèse sur d'autres nombreuses variables pour lesquelles nous pouvons toujours accepter l'hypothèse nulle). Nous savons alors que nous avons indubitablement commis des erreurs dans le tri des compteurs, mais ces erreurs ne seront pas préjudiciables pour notre étude économétrique.

4 Etude économétrique de la consommation journalière d'eau à Paris

4.1 Objectifs

Il s'agit pour nous de comprendre quelles sont les variables (parmi celles disponibles dans la base de données) qui influencent le volume d'eau consommé à Paris chaque jour (variable estimée à partir du volume d'eau mis en distribution quotidiennement et des pertes annuelles moyennes). Nous voulons aussi comprendre le sens des effets observés et l'importance qu'ils peuvent prendre, expliquer quelles sont les variables qui ont un effet majeur et quelles sont celles qui ont un impact plus faible sur la variation de la variable dépendante.

Cependant, cette étude ne prétend pas expliquer en totalité les variations des volumes d'eau consommés, elle tente simplement de trouver le modèle le plus précis possible en utilisant les méthodes adaptées aux différents problèmes rencontrés.

En effet, comme nous l'avons constaté, la consommation d'eau a connu une baisse continue d'année en année depuis le début des années 1990. Or, les variables climatiques étant relativement stables sur la période, de même le nombre de jours fériés, de week-ends, et de jours de vacances scolaires n'a pas varié non plus sur la période il serait incohérent d'expliquer la baisse pluriannuelle constatée.

Nous tenterons donc d'expliquer les comportements de consommation d'eau potable au niveau infra annuel puisque nous pouvons observer que la consommation d'eau à Paris reproduit chaque année deux types de cycles :

- un premier cycle annuel en forme de M qui est marquée par des fortes baisses de la consommation d'eau lors des vacances scolaires (notamment en Juillet et en Août et lors des vacances de Noël)
- à l'intérieur de ce cycle annuel, des cycles hebdomadaires marqués en général par

une baisse notable de la consommation en fin de semaine. Cette baisse est liée au ralentissement de l'activité économique en fin de semaine (fermetures des écoles et administrations ainsi que les activités de bureaux).

4.2 Méthodologie empirique

Nous avons choisi de réaliser une étude basée sur des régressions économétriques d'une variable dépendante (le volume d'eau distribué chaque jour à Paris) sur de nombreuses variables explicatives (climatiques et informations concernant le type de jour).

L'équation économétrique qui sera estimée dans cette section est la suivante :

$$Y_t = \alpha_0 + \alpha_1 C_t + \alpha_2 J_t + \delta_A + \varepsilon_t \quad (1)$$

- Avec Y_t la variable dépendante : le volume total d'eau consommé chaque jour à Paris
- Avec C_t le vecteur des variables climatiques au pas de temps journalier, jour t , comprenant la pluie, la température moyenne, l'humidité moyenne, la vitesse moyenne du vent, l'indice humidex moyen et le temps d'ensoleillement
- Avec J_t le vecteur des variables concernant des informations sur le type de jour (jour « normal », jour férié, jour de week-end, jour de vacances scolaires, jour de Noël)
- Avec δ_A les variables indicatrices années (variables qui captent des effets non expliqués par le modèle mais qui sont stables à l'intérieur d'une année et qui varient entre chaque année). Ces effets permettent en partie de capter la baisse de la consommation moyenne annuelle observée depuis le début des années 1990 mais qui ne peut être expliquée par nos variables indépendantes comme dans la partie "objectifs"
- ε_t est le résidu de la régression économétrique

- α_0 , α_1 et α_2 représentent respectivement la constante et les vecteurs des coefficients associés aux différentes variables explicatives du modèle

Afin de mener une étude économétrique la plus pertinente, nous avons mené différents types de tests. Ils permettent de vérifier que la méthode d'estimation choisie est la plus adaptée et qu'elle réponde aux problèmes économétriques qui se posent du fait de l'utilisation de cette base de données particulière.

Nous avons commencé par mener un test de White afin de vérifier l'hypothèse d'homoscédasticité bien que les modèles hétéroscédastiques soient plus fréquemment rencontrés avec des données en coupe transversale qu'avec des séries chronologiques. Le résultat de ce test confirme bien la validité de l'hypothèse d'homoscédasticité des perturbations.

Ensuite nous avons mené un test de Durbin et Watson car dans le cas de séries temporelles, le problème d'autocorrélation des perturbations se pose souvent. Le résultat du test nous indique que nous sommes bien dans un cas d'autocorrélation des perturbations (ce qui était prévisible). Des lors, nous optons pour la méthode d'estimation (de régression) de Prais-Winsten qui correspond à des Moindres Carrés Ordinaires (MCO) en plusieurs étapes avec itérations répétées afin de corriger les biais liés à l'autocorrélation des résidus.

Enfin, nous avons mené deux tests pour tenter de vérifier la stationnarité de nos variables. En effet, les variables qui sont marquées par une forte saisonnalité sont souvent non stationnaires ce qui peut contraindre à utiliser des méthodes d'estimations particulières. Les tests menés sont le test de Phillips-Perron et le test de Dickey-Fuller augmenté, tous deux adaptés au cas d'autocorrélation des perturbations. Les résultats de ces deux tests nous révèlent que nos principales variables (variable dépendante et variables d'intérêt retenues) sont stationnaires (p-values très faibles qui nous conduisent

à rejeter l'hypothèse nulle de non stationnarité).

Ceci se comprend notamment pour nos variables climatiques par le fait que la moyenne sur chaque année de ces variables est relativement stable dans le temps, bien qu'à l'intérieur de chaque année ces variables connaissent de fortes variations saisonnières. Cependant, nous tenons à préciser ici qu'un léger doute subsiste, car bien que les résultats de nos tests soient formels (et que les procédures soient respectées à la lettre) il nous semble un peu étonnant que les variables soient toutes stationnaires alors qu'elles pourraient au contraire être toutes non stationnaires d'un point de vue purement théorique.

4.3 Résultats économétriques

Dans cette dernière section nous présenterons les résultats obtenus pour les différentes spécifications retenues afin de bien comprendre les effets des différentes variables explicatives sur la consommation d'eau journalière. Nous présenterons les résultats obtenus avec différentes méthodes d'estimation afin de bien vérifier que nos choix en termes de méthode de régression sont judicieux et cohérents.

La table 2 représente les résultats des estimations réalisées avec 5 spécifications différentes. Les deux premières spécifications correspondent à des régressions selon la méthode des MCO alors que les 3 dernières utilisent la méthode de Prais-Winsten, méthode adaptée au problème d'autocorrélation que nous avons détecté.

Les spécifications (1) et (3) n'utilisent comme variables explicatives que les variables climatiques alors que les spécifications (2), (4) et (5) incluent les variables climatiques et les variables informant le type de jour. La spécification (5), la plus complète, correspond à une estimation par la méthode de Prais Winsten, qui inclut en plus la variable endogène retardée (le volume d'eau distribué le jour précédent) comme variable explicative. Ceci se justifie par la présence d'une certaine inertie dans les habitudes de consommation qui est plus qu'admise (cependant cette inertie peut être plus faible

Table 2: Résultats économétriques

Variable explicative	(1)	(2)	(3)	(4)	(5)
Volume eau dist (t-1)	-	-	-	-	0,16*** (0,01)
Pluie	-0,85** (0,26)	-0,12 (0,10)	-0,17* (0,09)	-0,07 (0,051)	-0,10** (0,05)
Vent Moyen	1,35*** (0,27)	-0,14 (0,11)	-0,04 (0,12)	-0,07 (0,07)	-0,09 (0,07)
Humidité Moyenne	2,31*** (0,14)	-0,01 (0,06)	-0,17* (0,073)	-0,09** (0,43)	-0,10** (0,04)
Température Moyenne	1,68* (0,87)	0,81** (0,34)	1,11*** (0,39)	1,18*** (0,23)	0,80*** (0,22)
Indice humidex Max	1,11** (0,50)	1,85*** (0,18)	1,01*** (0,18)	1,08*** (0,11)	1,21*** (0,11)
Insolation	0,09*** (0,01)	0,01*** (0,00)	0,01** (0,00)	0,01*** (0,00)	0,01*** (0,00)
Jour férié	-	-93,62*** (2,88)	-	-53,90*** (1,63)	-56,62*** (-1,62)
Week-end	-	-55,12*** (0,87)	-	-55,18*** (0,51)	-56,59*** (0,52)
Vacances	-	-37,98*** (1,13)	-	-7,11*** (1,31)	-7,64*** (1,27)
Vac Noel + Juil_Aout	-	-22,08*** (1,38)	-	-11,47*** (2,27)	-16,08*** (1,99)
Noël	-	-42,75*** (8,15)	-	-24,69*** (3,96)	-26,89*** (3,98)
1 er Mai	-	-7,08 (-7,76)	-	-1,53 (-3,77)	-2,44 (3,80)
Indicatrices Années	Non	Oui	Non	Oui	Oui
Estimation MCO	Oui	Oui	Non	Non	Non
Estimation Prais Winsten	Non	Non	Oui	Oui	Oui
R2 ajusté	0,0912	0,8628	0,0616	0,7039	0,7479
Observations	7814	7814	7814	7814	7814

Notes : ce tableau reporte les coefficients estimés pour cinq spécifications, (1) à (5), implémentées séparément en augmentant la précision à chaque étape. La variable dépendante est le volume d'eau distribué quotidiennement à Paris qui est assimilé au volume d'eau consommé chaque jour. Entre parenthèses sont reportés les écart-types estimés des coefficients.

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.

pour la consommation d'eau et surtout lorsque nous retenons un pas de temps journalier, mais comme nous pouvons le constater la variable dépendante retardée reste une variable pertinente et significative dans notre modèle).

Comme nous pouvons le constater dans le tableau ci-dessus, la spécification qui obtient le R2-ajusté le plus élevé (0,86) est celle utilisant une méthode de régression par MCO en incluant les variables climatiques ainsi que les variables de type de jour, et les indicatrices années. Cependant, cette régression ne peut être retenue pour nos conclusions car il ne s'agit pas de la méthode d'estimation la plus adaptée. En effet, en présence d'autocorrélation nous devons utiliser la méthode de Prais-Winsten. Cette spécification est implémentée et présentée à des fins de comparaison uniquement.

Notons que le R2 (ou R2-ajusté) peut s'interpréter comme la part de la variance de la variable dépendante expliquée par les variables explicatives retenues (variables climatiques et variables concernant le type de jour).

Nous avons implémenté différentes spécifications afin de pouvoir tirer des conclusions les plus robustes possible.

Ainsi, nous pouvons conclure que certaines variables sont significatives pour expliquer la variable endogène.

Ces variables sont :

- la température moyenne qui joue positivement sur la consommation d'eau
- l'insolation (temps d'ensoleillement) qui a également un impact positif mais faible
- l'humidité qui s'avère plus ou moins significative selon les spécifications et dont le coefficient peut être positif comme négatif ce qui nous laisse penser que nos conclusions quant à l'impact de cette variable doivent être précautionneuses car assez peu robustes. De plus, théoriquement l'humidité n'est pas une variable particulièrement pertinente pour expliquer la consommation d'eau

- l'indice humidex moyen qui exerce un effet à la hausse sur le volume d'eau distribué (et donc consommé)
- des variables indicatrices concernant le type de jour (jou férié, week-end, vacances scolaires, jour de Noël)

Les coefficients des variables indicatrices concernant le type de jour nous confirment bien qu'à Paris la consommation d'eau est fortement liée à l'activité économique et aux rythmes scolaires puisque la consommation d'eau diminue fortement et significativement les week-ends, les jours fériés (de manière très forte le jour de Noël), et les jours de vacances scolaires.

De plus, notons bien que la significativité de ces variables est très robuste et très élevée (à 1% pour chaque variable et pour toutes les spécifications), et que les coefficients sont élevés ce qui indique que ces variables ont beaucoup plus d'impact sur le volume total d'eau consommé que les variables climatiques.

Certaines variables apparaissent clairement non significatives :

- la vitesse du vent qui n'est significative que dans une spécification (qui n'est pas la plus adaptée ni la plus complète de nos estimations)
- la Fête du Travail (1er Mai) ne semble pas avoir d'incidence notable (plus qu'un jour férié quelconque du moins). Ceci peut se justifier par le fait que si ce jour n'est pas proche d'un week-end pour permettre de faire un pont, les parisiens ne partent pas pour une journée et donc la consommation d'eau baisse de la même manière que pour un jour férié non spécifié (du fait de la fermeture des écoles, entreprises et administrations notamment) mais pas dans une proportion significativement plus grande.

Cependant, concernant la variable Pluie, nous aurions du mal à conclure quant à son influence puisque sa significativité n'est pas toujours démontrée et pas toujours très élevée mais elle n'apparaît pas pour autant totalement non significative.

La pluviométrie peut avoir un impact théorique sur la consommation d'eau (impact négatif lié à des besoins plus faibles en arrosage des espaces verts extérieurs et jardins lors de périodes de fortes précipitations), cependant cet effet reste limité dans une ville comme Paris où l'habitat est principalement en immeuble (sans jardin ni piscine). Mais une certaine quantité d'eau est malgré tout consommée pour des usages externes, notamment les jardins publics.

4.4 Autres variables testées

Nous avons aussi pu disposer (plus tard dans notre étude) de données sociodémographiques telles que l'emploi parisien, la population, le nombre de ménages, le nombre d'établissements (données issues de l'INSEE) ainsi que de la variable prix de l'eau (variable disponible au pas de temps trimestriel) via des bases de données d'Eau de Paris. Ces variables auraient potentiellement pu nous permettre d'expliquer une partie de la baisse du niveau de la consommation d'eau constaté d'année en année depuis 1990. Cependant ces variables n'ont pas été retenues dans le modèle que nous avons présenté pour diverses raisons que nous allons expliciter ici. Nous ne présenterons pas de table de résultats dans cette partie, mais nous tenterons d'expliquer et détailler au mieux notre démarche et les raisons qui nous ont poussés à ne pas élargir le modèle en prenant en compte d'autres variables explicatives.

Concernant la variable prix de l'eau, un problème très fréquent d'endogénéité se pose. Ce problème est lié au fait que le prix de l'eau et la consommation sont déterminés de manière simultanée, pour être plus clair le prix de l'eau a une influence notable sur la consommation d'eau mais en même temps, le niveau de consommation d'eau peut influencer la fixation du prix (entreprise du service d'eau qui réagit à une baisse du

niveau de consommation par une augmentation du prix pour limiter les pertes par exemple).

Ce problème doit être impérativement traité sous peine d'obtenir des résultats biaisés et non robustes en utilisant une méthode d'estimation simple (MCO notamment). La méthode la plus adaptée est la méthode des variables instrumentales. Elle consiste à trouver une variable qui influence la variable explicative endogène (dans notre cas le prix de l'eau) sans pour autant être corrélée au résidu (c'est à dire avoir aucune influence sur la variable expliquée, à savoir la consommation d'eau). Ces variables sont souvent difficiles à trouver théoriquement, et surtout à se procurer en termes de données.

Nous disposons de certaines variables (notamment les charges d'exploitation et de personnels des services d'eau parisiens), qui théoriquement pourraient représenter de « bons » instruments. Cependant les résultats ne se sont pas avérés convaincants en raison principalement du nombre trop faible d'observations (en effet, nous ne disposons de ces variables « instruments » qu'au niveau annuel pour une quinzaine d'années). Ainsi, la corrélation observée entre la variable endogène (le prix de l'eau) et les instruments n'est pas apparue suffisamment robuste. Nous n'avons donc pas pu conclure quant à une élasticité prix de la demande en eau (les résultats étaient très peu robustes, trop variables en fonction des spécifications retenues et surtout avec un signe qui pouvait être aussi bien positif que négatif ce qui révèle la non robustesse des résultats).

Concernant les variables socio-éco-démographiques, nous avons choisi de ne pas les retenir car elles se sont révélées très peu significatives pour expliquer la consommation d'eau journalière, avec des coefficients très proches de 0 (dans les cas où ils étaient significatifs). De plus, ces variables n'améliorent pas la qualité du modèle, en effet son pouvoir explicatif reste le même avec ou sans l'introduction de ces variables (notamment R^2 et R^2 ajusté qui n'augmentent pas avec l'introduction de telles variables). Ainsi par principe de parcimonie nous avons préféré ne pas retenir ces variables pour ne pas compliquer le modèle alors qu'il n'en n'était pas pour autant amélioré.

5 Etude économétrique de la consommation d'eau potable des ménages parisiens

5.1 Objectifs

Dans cette partie, nous cherchons à expliquer la consommation d'eau potable domestique de la ville de Paris en fonction de variables sociodémographiques (variables de population, de richesse, de type de logements, etc.). Il s'agit donc d'une approche dite "clientèle" puisque nous nous intéressons à la consommation d'eau d'un certain groupe d'utilisateurs qui sont relativement homogènes (à savoir les ménages).

Notons cependant que ce travail est périlleux en raison des données dont nous disposons à l'origine (Eau de Paris) comme nous l'avons précisé dans la présentation des données. Ainsi il faut interpréter prudemment les résultats que nous pouvons obtenir et ne pas prendre pour résultat irréfutable toute conclusion que nous pourrions tirer suite à l'étude économétrique.

Nous chercherons plus précisément à expliquer une grande partie de la consommation d'eau domestique des parisiens avec peu de variables explicatives pertinentes. En effet, le principe de parcimonie nous incite à chercher un modèle simple mais ayant un bon pouvoir explicatif et relativement robuste.

Nous étudierons ces variables avec un découpage géographique relativement fin (l'Iris) afin de pouvoir utiliser les nombreuses variables INSEE disponibles à ce niveau, et surtout pour disposer de suffisamment « d'individus » (d'observations différentes) d'autant plus que nous ne pouvons pas utiliser de dimension temporelle. Ainsi nous aurons environ 900 observations (nombre d'Iris parisiens qui comportent au moins un compteur d'eau jugé domestique) ce qui est un nombre très acceptable.

5.2 Méthodologie empirique

En accord avec la littérature nous avons choisi comme variable dépendante la consommation moyenne (consommation moyenne par habitant par Iris). Il s'agit d'une variable calculée simplement à partir de la somme des consommations « domestiques » de chaque Iris, divisée par la population de ce même Iris (donnée disponible dans la base Population de l'INSEE).

Les variables explicatives utilisées sont alors des parts (lorsque cela est possible). Par exemple la part d'une fraction de la population (plus de 65 ans, moins de 19 ans, chômeurs, etc.) dans la population totale ou la part d'un certain type de logements dans le total des logements ou des résidences principales de l'Iris. Ceci a cependant pour conséquence de baisser mécaniquement les R2 d'où une importance limitée de ce coefficient qui sera malgré tout interprété (mais avec précaution). De même, l'utilisation de variables calculées en parts pour expliquer une variable telle que la consommation moyenne crée un possible problème d'endogénéité puisque la variable dépendante, ainsi que certaines variables explicatives sont divisées par le même dénominateur (la population de l'Iris).

Cependant, nos variables d'intérêt ne sont pas toutes concernées par cet éventuel biais et il semblerait que ce risque soit suffisamment limité et/ou ait un impact faible sur la robustesse des résultats pour qu'il ne soit pas traité spécifiquement dans la littérature actuelle. Notons également que nous spécifierons toujours nos variables en logarithme afin de pouvoir interpréter aisément les coefficients issus des différentes régressions. En effet, ces derniers se lisent ainsi directement comme des élasticités ce qui est bien plus simple pour la compréhension du modèle issu de ce travail.

Les données prennent donc la forme d'une coupe transversale puisque nous disposons de différents individus que nous étudions à un moment t précis. Les « individus » sont alors les différents Iris de l'agglomération parisienne (pour lesquels au moins un compteur d'eau potable est jugé « domestique »), qui sont étudiés sur l'année 2011 (niveau

annuel donc un seul instant). Avec ce type de données nous aurons plus de mal à obtenir un modèle très précis cependant nous avons alors l'avantage d'être confronté à beaucoup moins de problèmes économétriques qui imposeraient un traitement spécifique des données (notamment comme dans la première partie de ce mémoire avec les séries temporelles). Nous incluons dans nos régressions une option qui permet de corriger pour l'hétéroscédasticité (problème économétrique fréquent pour les données en coupe transversale, confirmé par le Test de White).

Le modèle étudié aura donc la forme fonctionnelle suivante :

$$C_i = \gamma_0 + \gamma_1 P_i + \gamma_2 L_i + \gamma_3 R_i + \mu_i \quad (2)$$

Avec :

- C_i la variable dépendante : consommation moyenne (par habitant) d'eau potable des ménages résidants dans l'Iris i en 2011
- γ_0 la constante
- γ_1, γ_2 et γ_3 les vecteurs des coefficients associés aux différents groupes de variables explicatives
- P_i le vecteur des variables explicatives concernant la population (part de la population de 65 ans ou plus, des moins de 20 ans, etc.) de l'Iris i
- L_i le vecteur des variables explicatives concernant les caractéristiques des logements de l'Iris i (superficie des logements, nombre de pièces, présence de parking ou non etc.)
- R_i le vecteur des variables explicatives permettant une approximation du niveau de richesse des habitants de l'Iris i (catégorie socioprofessionnelle, situation de locataire ou propriétaire etc.).
- μ_i le résidu de l'estimation économétrique

Le modèle économétrique estimé cherchera donc à expliquer la consommation moyenne d'eau potable en 2011 des ménages de chaque Iris parisien en utilisant différents types de variables. En effet, il semble important d'inclure des variables relatives au type de logement (nombre de logements, taille des logements, etc.), des variables de population (nombre d'habitants, âge de ces habitants, etc.) et des variables qui permettent d'approximer le niveau de richesse (et donc de vie) des habitants des différents Iris (notamment la catégorie socioprofessionnelle du chef de famille, la possession de voiture, ainsi que la taille du logement puisque dans une ville comme Paris où le prix de l'immobilier est très élevé nous pouvons déduire en partie de la taille du logement le niveau de richesse d'une famille, bien sur avec précaution). En outre, ces différentes variables permettent de capter différentes dimensions de la consommation d'eau.

5.3 Résultats économétriques

Dans cette dernière section nous présenterons les résultats obtenus pour différentes spécifications, avec différents groupes de variables et sur différents groupes d'arrondissements afin de bien comprendre les effets des différentes variables explicatives sur la consommation moyenne d'eau des ménages parisiens.

Nous présenterons notamment une première table (Table 3) dans laquelle beaucoup de variables sont introduites progressivement afin de sélectionner les plus pertinentes, expliquant au mieux la consommation moyenne d'eau à Paris.

La Table 4 présentera les résultats obtenus pour les variables d'intérêt retenues (à savoir 7 variables) avec différentes spécifications, notamment en incluant ou non des variables indicatrices arrondissements ou `iris_court`⁵ qui permettent de capter des habitudes de consommation propres à certains arrondissements ou zones géographiques plus fines de la Capitale qui ne peuvent pas être expliquées par les variables dont nous

⁵Nous avons appelé Iris-court une variable créée à partir du numéro complet d'Iris auquel nous avons enlevé les 2 derniers chiffres. Ainsi nous obtenons avec cette variable un découpage géographique de la ville de Paris plus fin que l'arrondissement mais plus large que l'Iris.

disposons (variables INSEE) mais qui doivent être prises en compte pour purger les coefficients de nos variables d'intérêt (et leur significativité).

La dernière sous section (en 3 sous tables) répliquera la Table 4 en décomposant l'échantillon en sous groupes d'arrondissements selon différents critères.

5.3.1 Sélection des variables explicatives pertinentes du modèle

Nous estimerons l'équation (2) en incluant progressivement différentes variables (un nombre assez important) afin de pouvoir sélectionner les variables les plus pertinentes pour expliquer notre variable dépendante. Notre modèle sera spécifié en log (variable dépendante et variables explicatives) en accord avec la littérature existante et surtout pour des raisons d'aisance dans l'interprétation des coefficients estimés. Nous choisirons les "variables pertinentes" du modèle sur la base de deux critères:

- le premier critère sera la significativité des coefficients estimés : nous conserverons les variables explicatives pour lesquelles la significativité est toujours au moins égale à 5%
- le second critère sera un critère d'adéquation entre le signe du coefficient et ce qui est admis théoriquement (en d'autres termes, un signe du coefficient qui peut s'expliquer simplement/logiquement)

Cette table représente donc l'estimation de l'équation (2) en incluant de nombreuses variables (de nos trois groupes, population, emploi, proxy de la richesse) afin de retenir celles qui semblent les plus pertinentes pour expliquer la consommation moyenne d'eau des ménages parisiens.

Nous ne pouvons pas inclure toutes les variables INSEE disponibles dans nos différentes bases mais nous avons essayé d'inclure des variables théoriquement pertinentes

Table 3: Sélection des variables explicatives “pertinentes”

Variable explicative	(1)	(2)	(3)
Part population 65 ans et plus	0,002 (0,045)	-0,106** (0,047)	-0,070** (0,032)
Part population 0-18 ans	0,217** (0,095)	0,265*** (0,090)	0,214*** (0,068)
Part familles avec enfant(s)	0,227*** (0,080)	0,149** (0,073)	0,208*** (0,071)
Nombre de personnes par ménage	-2,042*** (0,153)	-2,403*** (0,152)	-2,460*** (0,124)
Part de la population active	0,017 (0,157)	0,217 (0,189)	-
Part des chômeurs	-0,022 (0,040)	0,046 (0,031)	-
Part ménages 1 personne seule	-0,827*** (0,163)	-0,742*** (0,146)	-0,789*** (0,136)
Nombre logements	1,971*** (0,166)	1,835*** (0,155)	1,738*** (0,140)
Part RP construites après 2005	-0,575 (0,526)	-0,592 (0,576)	-0,675 (0,548)
Nombre de pièce(s) des RP	-	0,474*** (0,124)	0,629*** (0,112)
Part RP HLM	-0,009 (0,006)	-	-
Part population CS3 ou CS4	-	-0,139** (0,063)	-
Part population CS5 ou CS6	-	-0,056 (0,042)	-
Part ménages avec 2 voitures ou plus	-	0,0284 (0,0174)	-
Part ménages avec 1 voiture ou plus	-0,110** (0,054)	-	-0,164*** (0,043)
Part propriétaires	-0,037*** (0,013)	-0,044*** (0,014)	-0,050*** (0,013)
R2 ajusté	0,473	0,489	0,489
Observations	745	881	882

Notes : La variable dépendante pour cette table et les suivantes est la Consommation Moyenne d'eau potable (par habitant par Iris);

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.

et différentes les unes des autres en termes de contenu informationnel sur les habitations et populations des différents Iris de la ville de Paris.

Ainsi, cette table nous permet de sélectionner un groupe de variables “pertinentes” (terme que nous utiliserons par la suite pour désigner les variables d’intérêt retenues à cette étape).

Comme nous pouvons le voir dans la Table 3, les variables que nous retiendrons pour le reste de notre étude sont les suivantes (avec comme dimension individuelle l’Iris) :

- la part des 0-18 ans dans la population totale
- la part des familles avec au moins un enfant
- la part des ménages composés d’une seule personne
- le nombre de personnes par ménage
- le nombre de logements
- le nombre de pièces des résidences principales
- la part des résidences principales occupées par des propriétaires

Pour cette dernière variable notons que le signe du coefficient (négatif) peut étonner en premier abord. En effet, nous pourrions penser que les propriétaires parisiens sont plus aisés financièrement que les locataires et donc ils devraient consommer plus d’eau. Nous pouvons cependant expliquer ce résultat de deux manières :

- les propriétaires ont une probabilité plus élevée d’être équipés en compteurs individuels (d’autant plus que les maisons individuelles de la Capitale, certes très peu nombreuses, sont en général équipées de compteurs individuels et habitées par

des propriétaires) ce qui les incite à surveiller plus attentivement leurs consommations puisqu'ils paient exactement ce qu'ils consomment. Mais notons que cette justification est assez peu valable dans une ville comme Paris

- ils sont potentiellement plus sensibilisés aux problèmes environnementaux et consomment donc moins d'eau, toutes choses égales par ailleurs, que des locataires

Ainsi, nous ne conclurons pas nécessairement de manière très certaine quant à l'influence de cette variable sur la consommation d'eau puisque nous ne sommes pas certains que nos explications soient valables. De plus, cette variable ne semble pas être la plus importante pour expliquer la consommation d'eau des ménages parisiens.

Toutes les autres variables de la liste précédente (nos variables "pertinentes") ont un coefficient du signe théoriquement attendu :

- une part plus importante de jeunes (0-18 ans) tire la consommation moyenne d'eau potable vers le haut. Les plus jeunes consomment souvent plus d'eau que leurs aînés. De plus, la présence de jeunes dans les ménages élève probablement les volumes d'eau consommés en raison de l'entretien du linge (lessives) et de la maison.
- les familles avec enfant(s) contribuent à une augmentation de la consommation d'eau des ménages, ceci s'explique par les bains et soins donnés aux enfants qui sont en général très consommateurs d'eau. Cette variable est assez proche de la précédente donc les justifications présentées ci-dessus peuvent aussi s'appliquer ici.
- les ménages composés d'une seule personne consomment peu d'eau notamment en raison de la faible présence de ces habitants dans leurs logements surtout dans une ville animée comme Paris

- le nombre de personnes par ménage tire la consommation moyenne vers le bas ce qui s’explique par les “économies d’échelle” de la consommation d’eau
- le nombre de logements influence positivement la consommation
- le nombre de pièces des résidences principales contribue également à une forte consommation moyenne puisque ces plus grandes habitations nécessitent de fortes consommations en eau pour l’entretien (nettoyage). De plus, la taille des logements peut également être interprétée (au moins en partie) comme un indicateur de richesse des ménages parisiens. Or, les ménages aisés ont potentiellement des consommations d’eau plus élevées.

Nous n’avons pas retenu la variable “part de familles possédant une voiture ou plus” en raison de la non cohérence du signe du coefficient estimé. En effet, cette variable semblerait influencer négativement la consommation moyenne d’eau à Paris alors qu’il s’agit d’un indicateur de richesse ce qui signifierait que les plus riches consomment moins d’eau (élasticité revenu de la consommation d’eau négative alors qu’elle est admise dans la littérature positive ou non significative).

5.3.2 Résultats économétriques

La table 4, présentée dans cette section, correspond à l’estimation de l’équation (2) en utilisant uniquement les variables “pertinentes” sélectionnées dans la section précédente. La première spécification correspond à une estimation par la méthode des moindres carrés ordinaires (notons que nous “clusterisons” systématiquement les écarts-types afin de ne pas prendre le risque de surestimer la significativité de nos coefficients). La deuxième spécification inclut des variables indicatrices arrondissements puis la troisième des variables indicatrices iris-courts. Nous obtenons 20 variables indicatrices arrondissements (car Paris est composé de 20 arrondissements), ainsi que 78 variables indicatrices

iris-courts.

Le premier fait notable qui ressort de cette table de résultats est que toutes nos variables d'intérêt "pertinentes" (sauf une) restent significatives lorsque les estimations sont plus contraignantes (notamment lorsque nous incluons des variables indicatrices arrondissements ou iris-courts). La seule variable qui semble avoir une influence non significative sur la consommation moyenne d'eau est la part des familles avec enfant(s), ainsi nous ne concluons pas quant à un impact positif de cette variable puisque ce résultat semble peu robuste à ce stade de l'étude économétrique.

Notons également que la part des résidences principales occupées par des propriétaires semble avoir une influence significative mais très faible puisqu'une hausse de 10% de la part des propriétaires entraîne une très légère baisse de la consommation moyenne d'eau d'environ 0,4% ce qui est négligeable (d'autant plus que nous avons eu des difficultés pour interpréter ce coefficient, ce qui nous pousse à ne pas trop juger la significativité de cette variable). Quant aux autres variables "pertinentes", leur significativité semble relativement robuste ce qui est satisfaisant.

Il est important de remarquer également que les variables ayant un impact très important sur la consommation moyenne d'eau à Paris sont le nombre de logements (une hausse de 10% du nombre de logements entraînant une augmentation de plus de 10% de la consommation moyenne) ainsi que le nombre de personnes par ménage : plus les ménages sont composés de beaucoup de personnes plus la consommation d'eau moyenne est faible ce qui s'explique par les "économies d'échelle" de la consommation d'eau (une grande partie de la consommation d'eau est "incompressible" et n'augmente pas ou que très peu avec le nombre de personnes vivant dans un foyer).

Table 4: Estimations principales du modèle

Variable explicative	(1)	(2)	(3)
Part population 0-18 ans	0,259*** (0,067)	0,217*** (0,070)	0,193*** (0,064)
Part ménages 1 personne seule	-0,756*** (0,149)	-0,550*** (0,114)	-0,470*** (0,111)
Nombre de personnes par ménage	-2,183*** (0,115)	-1,735*** (0,160)	-1,335*** (0,152)
Part familles avec enfant(s)	0,181** (0,072)	0,044 (0,058)	-0,050 (0,055)
Nombre logements	1,895*** (0,150)	1,393*** (0,158)	1,102*** (0,162)
Nombre de pièces des résidences principales	0,179** (0,083)	0,288*** (0,094)	0,171* (0,091)
Part propriétaires	-0,049*** (0,013)	-0,044*** (0,012)	-0,024** (0,012)
Constante	5,927*** (0,295)	5,079*** (0,277)	5,165*** (0,213)
R2 ajusté	0,467	0,614	0,709
Observations	883	883	883
Variables Indicatrices Iris-Courts	Non	Non	Oui
Variables Indicatrices Arrondissements	Non	Oui	Non

Notes : Les résultats représentés dans cette table correspondent à l'estimation de l'équation (2) de notre modèle avec les variables jugées "pertinentes" grâce à la table précédente. Nous implémentons trois spécifications de la moins restrictive à la plus contraignante en ajoutant des variables indicatrices arrondissements puis iris-court.

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.

5.3.3 Etude de différents sous-échantillons

Etude économétrique sur trois groupes d'arrondissements en fonction du niveau de revenu

Nous présenterons ici les résultats de l'estimation de l'équation (2) avec les variables "pertinentes" du modèle, réalisée sur trois échantillons, à savoir trois groupes d'arrondissements en fonction du niveau moyen de richesse de la population de ces arrondissements. Les 3 groupes d'arrondissements sont les suivants:

- les arrondissements les plus riches : 6ème, 7ème, 8ème et 16ème avec un revenu net déclaré moyen par foyer fiscal supérieur à 66 000 euros en 2009⁶
- les arrondissements dont la population représente des niveaux de richesse moyens : 1er, 2ème, 3ème 4ème, 5ème, 9ème, 12ème, 14ème, 15ème et 17ème avec un revenu net déclaré moyen par foyer fiscal compris entre 30 000 euros et 48 000e uros en 2009.
- les arrondissements les plus pauvres : 10ème, 11ème, 13ème, 18ème, 19ème et 20ème caractérisés par un revenu net déclaré par foyer fiscal inférieur à 28 000 euros en 2009.

Cette distinction est certes arbitraire mais elle crée des groupes relativement proches en termes de niveaux de vie et dont les comportements des habitants peuvent se rapprocher dans une certaine mesure. Cependant comme nous pouvons des à présent le remarquer il s'agit de groupes de différentes tailles (seuls 4 arrondissements pour les plus riches contre 10 pour les "classes moyennes") ce qui se reproduira dans les données par Iris. En effet, la Table 5 révèle que le premier groupe d'arrondissements contient 138 observations (Iris), le second 368 et le dernier (les plus pauvres) 377. Il faudra donc

⁶Source des données sur les revenus fiscaux par arrondissement : INSEE

Table 5: Résultats pour trois groupes d'arrondissements (riches-moyens-pauvres)

Variable explicative	(1)	(2)	(3)	(4)	(5)	(6)
Part population 0-18 ans	0,073 (0,157)	0,203 (0,126)	0,219** (0,086)	0,078 (0,129)	0,213* (0,122)	0,217*** (0,082)
Part ménages 1 pers seule	-0,739*** (0,263)	-0,529** (0,212)	-0,281 (0,178)	-0,453 (0,275)	-0,615*** (0,213)	-0,278 (0,174)
Nb personnes par ménage	-2,829*** (0,286)	-1,832*** (0,329)	-1,085*** (0,229)	-1,891*** (0,298)	-1,477*** (0,327)	-1,071*** (0,221)
Part familles avec enfant(s)	0,089 (0,172)	0,028 (0,095)	-0,024 (0,071)	0,016 (0,175)	-0,088 (0,091)	-0,046 (0,068)
Nombre logements	1,874*** (0,230)	1,228*** (0,303)	0,947*** (0,271)	1,255*** (0,235)	1,206*** (0,317)	0,895*** (0,269)
Nombre pièces des RP	0,913*** (0,220)	0,575*** (0,179)	0,075 (0,137)	0,549** (0,224)	0,230 (0,177)	0,111 (0,138)
Part propriétaires	-0,159** (0,062)	-0,054** (0,024)	-0,007 (0,014)	-0,077 (0,058)	-0,040* (0,023)	-0,004 (0,014)
Constante	4,883*** (0,684)	4,581*** (0,437)	5,161*** (0,419)	5,403*** (0,698)	4,520*** (0,333)	4,894*** (0,391)
R2 ajusté	0,590	0,594	0,344	0,690	0,699	0,416
Observations	138	368	377	138	368	377
VI Iris-Courts	Non	Non	Non	Oui	Oui	Oui
VI Arrondissements	Oui	Oui	Oui	Non	Non	Non

Notes : Les résultats représentés dans cette table correspondent à l'estimation de l'équation (2) de notre modèle avec les variables jugées "pertinentes" grâce à la TABLE 3. Les trois premières spécifications incluent des variables indicatrices arrondissements et les trois suivantes des indicatrices Iris-Courts. Nous avons ici trois groupes d'arrondissements (riches, moyens, pauvres).

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.

interpréter avec prudence les résultats issus de ces estimations, sans tirer de conclusion trop générale notamment pour les arrondissements les plus riches qui représentent un nombre d'observations relativement faible. Les spécifications (1) et (4) concernent les arrondissements riches, les spécifications (2) et (5) les arrondissements aux revenus moyens et les spécifications (3) et (6) les arrondissements les plus pauvres.

Les variables qui restent significatives pour les trois groupes d'arrondissements, y compris avec les spécifications les plus contraignantes (c'est à dire les spécifications (4), (5) et (6) qui incluent des variables indicatrices Iris-Courts), sont le nombre de logements (influence positive sur la variable dépendante) et le nombre de personnes par ménage (influence négative). Cependant nous remarquons que ces variables ont un impact plus fort (positif ou négatif) pour les arrondissements relativement aisés que pour les arrondissements les plus pauvres. Il semblerait donc que d'autres considérations puissent avoir une influence plus importante sur les comportements de consommation d'eau des ménages les plus pauvres.

Il est intéressant de remarquer que certaines de nos variables pertinentes n'influencent la consommation d'eau que des ménages riches ou des classes moyennes (la part des propriétaires, le nombre de pièces des résidences principales et la part de ménages composés d'une seule personne) alors que la part de jeunes (0-18 ans) dans la population semble impacter les niveaux de consommation d'eau uniquement pour les ménages vivant dans les arrondissements les plus pauvres.

Il semblerait donc que les variables qui concernent le niveau de richesse ont un impact potentiel sur le comportement des ménages les plus aisés (à la fois en ce qui concerne des consommations plus élevées pour des habitations plus grandes, et des consommations tirées à la baisse par des comportements volontaires d'économies d'eau pour les propriétaires ou pour les ménages composés d'une seule personne). Nous pouvons penser que les personnes aisées qui vivent seules ont un travail très prenant

et sortent beaucoup, ils passent une majeure partie de leur temps en dehors de leur habitation ce qui limite leur consommation d'eau.

Il est important de remarquer ici (dans la Table 5), que bien que nous ayons peu d'observations pour le premier groupe d'arrondissements, il semble que le modèle que nous avons construit s'applique mieux et soit plus pertinent pour expliquer les comportements de consommation d'eau des ménages les plus aisés ou des classes moyennes puisque le coefficient de détermination ajusté (R^2) est plus élevé pour les deux groupes d'arrondissements que pour les arrondissements les plus pauvres. Ceci peut s'expliquer par le fait que les ménages les plus pauvres ont des comportements de consommation similaires entre eux et qui sont déterminés principalement par leur niveau de richesse mais qui varie peu en fonction d'autres critères comme la taille de leur logement ou le nombre de personnes dans le ménage puisqu'ils sont contraints en premier lieu par leur revenu.

Etude économétrique sur deux groupes d'arrondissements en fonction du niveau de revenu

Nous répliquons ici la table précédente mais en regroupant les arrondissements les plus modestes et les classes moyennes car il se peut que les comportements des plus aisés soient vraiment différents alors que ceux des classes moyennes peuvent se rapprocher des comportements des plus modestes. Il faudra à nouveau faire attention au fait que le premier groupe comporte beaucoup moins d'observations que le second (138 contre 745). La table 6 représente donc les résultats de l'estimation de l'équation (2) répliquée sur les deux groupes d'arrondissements formés à partir des trois précédents. Les deux premières spécifications présentent les estimations incluant des variables indicatrices arrondissements et les deux dernières incluent des variables indicatrices iris-courts (spécifications plus contraignantes). Les spécifications (1) et (3) correspondent aux ar-

rondissements les plus riches (ce sont les mêmes estimations que les spécifications (1) et (4) de la Table 5 mais nous les laissons pour une meilleure lecture et une comparaison plus facile) alors que les spécifications (2) et (4) correspondent au groupe formé des arrondissements les plus pauvres et des classes moyennes.

Nous remarquons que même en regroupant les arrondissements les plus pauvres et les classes moyennes nous obtenons une estimation de moins bonne qualité globale (R2 ajusté plus faible) que pour le groupe des arrondissements les plus riches. Cependant notons que ce résultat n'est pas primordial pour nous puisque le R2 n'est pas une mesure "idéale" de la qualité d'un modèle, notamment lorsque nous avons peu d'observations (beaucoup moins que pour l'autre groupe d'arrondissements).

En analysant la spécification (4) qui est la "meilleure" car la plus contraignante et qui correspond aux arrondissements pauvres et classes moyennes, nous remarquons que seules quatre de nos variables pertinentes restent significatives pour expliquer la consommation d'eau des ménages:

- le nombre de logements joue positivement mais moins que pour les arrondissements plus riches
- le nombre de personnes par ménage joue négativement mais moins que pour les plus riches à nouveau
- la part des ménages composés d'une personne seule influence négativement la consommation d'eau mais dans une moindre mesure que pour les ménages des arrondissements riches
- la proportion d'enfants (moins de 18ans) a une influence significative à la hausse sur les consommations moyennes des arrondissements les plus pauvres et ceux caractérisés par des revenus moyens alors que cette variable n'est pas significative dans le cas des plus riches.

Table 6: Résultats sur 2 groupes d'arrondissements (riches et moyens-pauvres)

Variable explicative	(1)	(2)	(3)	(4)
Part population 0-18 ans	0,073 (0,157)	0,238*** (0,077)	0,078 (0,129)	0,217*** (0,074)
Part ménages 1 pers seule	-0,739*** (0,263)	-0,398*** (0,137)	-0,453 (0,275)	-0,437*** (0,137)
Nb personnes par ménage	-2,829*** (0,286)	-1,442*** (0,201)	-1,891*** (0,298)	-1,223*** (0,194)
Part familles avec enfant(s)	0,089 (0,172)	0,006 (0,062)	0,016 (0,175)	-0,073 (0,057)
Nombre logements	1,874*** (0,230)	1,102*** (0,210)	1,255*** (0,235)	1,032*** (0,212)
Nombre pièces des RP	0,913*** (0,220)	0,287*** (0,107)	0,549** (0,224)	0,138 (0,106)
Part propriétaires	-0,159** (0,062)	-0,025** (0,012)	-0,077 (0,058)	-0,017 (0,012)
Constante	4,883*** (0,684)	5,088*** (0,298)	5,403*** (0,698)	5,144*** (0,231)
R2 ajusté	0,590	0,555	0,690	0,652
Observations	138	745	138	745
VI Iris-Courts	Non	Non	Oui	Oui
VI Arrondissements	Oui	Oui	Non	Non

Notes : Les résultats représentés dans cette table correspondent à l'estimation de l'équation (2) de notre modèle avec les variables jugées "pertinentes" grâce à la TABLE 3. Les deux premières spécifications incluent des variables indicatrices arrondissements et les deux suivantes des indicatrices Iris-Courts. Nous avons ici deux groupes d'arrondissements (les plus riches puis les classes moyennes-pauvres).

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.

Nous pouvons donc conclure que les comportements de consommation d'eau des ménages parisiens sont différents en fonction de leurs niveaux de richesse, mais aussi que les variables qui influencent ce comportement ne sont pas les mêmes non plus en fonction des groupes d'arrondissements considérés. Il semblerait que les comportements de consommation des plus aisés dépendent de leur statut (propriétaire ou non), des caractéristiques de leur logement (taille notamment) alors que les comportements de consommation des moins aisés sont marqués par une consommation "incompressible" plus élevée (la constante peut être interprétée comme cette consommation "minimale"), et par les caractéristiques de la famille (nombre/part de jeunes, enfants etc.).

Etude économétrique sur deux groupes d'arrondissements en fonction de leur localisation géographique

Nous allons maintenant présenter les résultats et l'analyse de l'étude économétrique du modèle avec les variables "pertinentes" sur deux échantillons d'arrondissements regroupés selon le critère "arrondissements du centre" ou "arrondissements périphériques". Ainsi le premier groupe correspond aux arrondissements 1 à 10 alors que le second correspond aux arrondissements 11 à 20. Ce choix a été fait car les arrondissements 11 à 20 sont plutôt résidentiels alors que les arrondissements du centre mêlent plus d'activités économiques aux zones résidentielles.

Notons cependant que cette séparation entre les arrondissements du centre de Paris et les arrondissements périphériques n'est pas totalement neutre en termes de niveaux de richesse. En effet, les arrondissements 1 à 10 comportent une grande majorité des arrondissements les plus riches (6ème, 7ème et 8ème) ainsi que 6 des arrondissements appartenant à notre catégorie "revenus moyens" (1er, 2ème, 3ème, 4ème, 5ème et 9ème). Alors que le groupe des arrondissements périphériques contient 5 des 6 arrondissements de la catégorie précédente "arrondissements les plus pauvres" et un seul arrondissement

de la catégorie des plus riches (le 16ème).

Les spécifications (1) et (2) incluent des variables indicatrices arrondissements alors que les spécifications (3) et (4) incluent des variables indicatrices iris-courts. Les spécifications (1) et (3) correspondent aux estimations sur le groupe des arrondissements du centre de Paris alors que les spécifications (2) et (4) concernent les arrondissements périphériques. Comme le montre la Table 7, les résultats se rapprochent des résultats précédents puisque les arrondissements du centre peuvent approximer un groupe d'arrondissements relativement aisés alors que le groupe des arrondissements périphériques est proche du groupe des arrondissements les plus pauvres. Ainsi, la part des 0-18 ans dans la population a un impact (positif) significatif sur la consommation moyenne d'eau potable des ménages que pour les arrondissements périphériques.

De même, la part de propriétaires semble impacter plus significativement la consommation moyenne d'eau des ménages des arrondissements du centre plus que celle des arrondissements périphériques. Les autres variables sont relativement significatives de la même manière pour les deux groupes d'arrondissements. Notons aussi que notre modèle semble à nouveau plus pertinent (R^2 plus élevé) pour expliquer la consommation moyenne d'eau des ménages qui vivent dans les arrondissements du centre (en grande partie des ménages aisés) que celle des habitants des arrondissements "extérieurs".

Cependant ce constat doit être accompagné d'une remarque sur la quantité des observations disponibles pour ces deux groupes d'arrondissements; en effet nous disposons de beaucoup moins d'observations pour le premier groupe des arrondissements que pour le second ce qui peut atténuer la pertinence de comparaisons des valeurs des coefficients de détermination (R^2).

Table 7: Résultats sur 2 groupes d'arrondissements (1 à 10 - 11 à 20)

Variable explicative	(1)	(2)	(3)	(4)
Part population 0-18 ans	-0,074 (0,165)	0,286*** (0,072)	0,042 (0,161)	0,233*** (0,066)
Part ménages 1 pers seule	-0,754*** (0,265)	-0,461*** (0,120)	-0,547** (0,264)	-0,417*** (0,118)
Nb personnes par ménage	-2,436*** (0,388)	-1,643*** (0,173)	-1,907*** (0,413)	-1,259*** (0,160)
Part familles avec enfant(s)	0,252 (0,156)	0,009 (0,057)	0,127 (0,147)	-0,091* (0,053)
Nombre logements	1,909*** (0,320)	1,232*** (0,173)	1,605*** (0,355)	0,950*** (0,168)
Nombre pièces des RP	0,498** (0,237)	0,359*** (0,106)	0,217 (0,212)	0,261** (0,102)
Part propriétaires	-0,254*** (0,0751)	-0,026** (0,011)	-0,160* (0,083)	-0,011 (0,011)
Constante	4,744*** (0,546)	5,148*** (0,317)	5,156*** (0,489)	4,487*** (0,298)
R2 ajusté	0,613	0,528	0,721	0,628
Observations	193	690	193	690
VI Iris-Courts	Non	Non	Oui	Oui
VI Arrondissements	Oui	Oui	Non	Non

Notes : Les résultats représentés dans cette table correspondent à l'estimation de l'équation (2) de notre modèle avec les variables jugées "pertinentes" grâce à la TABLE 3. Les deux premières spécifications incluent des variables indicatrices arrondissements et les deux suivantes des indicatrices Iris-Courts. Nous avons ici deux groupes d'arrondissements (1 à 10 puis 11 à 20).

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.

6 Conclusion

Ce mémoire s'est intéressé à la consommation d'eau de la ville de Paris. Nous avons étudié différents aspects de cette consommation en fonction des données que nous avons pu obtenir auprès d'Eau de Paris. Le premier travail concernant la consommation d'eau totale de la Capitale nous permet de remarquer que les variables climatiques jouent un rôle significatif sur la consommation d'eau. Nous avons aussi pu conclure que les volumes totaux d'eau consommés quotidiennement par tous les usagers de la ville sont fortement marqués par des cycles hebdomadaires (baisse de la consommation le week end) ainsi que des cycles "scolaires" (baisse de la consommation lors des vacances scolaires notamment en Juillet et en Août et pendant les vacances de Noël).

De même les jours fériés sont en général caractérisés par une plus faible consommation d'eau. Cette étude est assez robuste puisqu'elle s'appuie sur des séries temporelles sur une vingtaine d'années ce qui offre un nombre d'observations important. Nous pouvons donc conclure de cette première étude économétrique que la consommation totale d'eau de la ville de Paris est principalement tirée par les activités économiques de ville (commerces bureaux, écoles et universités), mais que les variables climatiques impactent également les volumes d'eau consommés mais dans une moindre mesure.

Le comportement de consommation d'eau des ménages semble donc impacter dans une faible mesure le volume total d'eau consommé chaque jour. Cependant il est intéressant de comprendre et d'étudier les variables qui peuvent influencer ce comportement et expliquer les différences de volumes consommés au niveau des ménages. Ainsi notre deuxième étude s'est attachée à comprendre les déterminants de la consommation moyenne d'eau potable des ménages même si dans la ville de Paris ce comportement n'est pas primordial pour expliquer la consommation globale d'eau, il peut être utile d'une manière générale pour comprendre ce qui motive les consommations d'eau de cette catégorie d'usagers qui au niveau d'un pays représente une part non négligeable

des volumes totaux consommés.

Notre travail a donc révélé que ce comportement pouvait s'expliquer en grande partie par certaines variables sociodémographiques telles que le nombre de logements (proxy de la population) dans une zone, l'âge de cette population (notamment les jeunes qui tirent la consommation vers le haut). De même le niveau de richesse semble impacter les comportements de consommation puisque les consommations d'eau potable des arrondissements les plus riches s'expliquent par des variables différentes que ceux des arrondissements les moins aisés.

Nous avons également remarqué que les propriétaires de leur habitation de la ville de Paris (bien que à priori plus riches que les locataires) semblent faire plus attention aux volumes d'eau qu'ils consomment. La consommation d'eau des ménages aisés semble être impactée par la taille de leur logement alors que la consommation d'eau des ménages les plus pauvres semble être principalement liée à une consommation incompressible.

7 Bibliographie

1. Bernard Barraqué et Alexis Nercessian, « Mieux Comprendre Comment Évolue la Consommation d'Eau à Paris » (Rapport final), Décembre 2008.
2. David L. Chicoine & Ganapathi Ramamurthy, "Evidence on the Specification of Price in the Study of Domestic Water Demand", *Land Economics*, Vol 62, Num 1, February 1986, p26-32
3. Steven C. Deller, David L. Chicoine, Ganapathi Ramamurthy, "Instrumental Variables Approach to Rural Water Service Demand", *Southern Economic Journal*, p333-346.
4. James J. Heckman, "Sample Selection Bias as a Specification Error", *Econometrica*, Vol 47, Num 1, Janvier 1979, pages 153-161.
5. Marielle Montginoul, « La consommation d'eau des ménages en France : Etat des lieux », Juin 2002.
6. Céline Nauges, A. Reynaud, et A. Thomas, « Consommation Domestique D'eau Potable et Tarification », Département d'Economie et de Sociologie Rurales de l'Institut National de la Recherche Agronomique, N° 5 - Décembre 1998.
7. Céline Nauges et A. Reynaud « Estimation de la demande domestique d'eau potable en France », *Revue économique* 1/2001 (Vol. 52), p. 167-185.
8. Céline Nauges et Alban Thomas, « Dynamique de la consommation d'eau potable des ménages : une étude sur un panel de communes françaises », *Économie & prévision*. Numéro 143-144, 2000-2-3. *Economie de l'environnement et des ressources naturelles*, p175-184.
9. Guy Poquet, H. Henriquez, L. Pouquet, et M. Mazars, « Etude de l'évolution des consommations d'eau potable à Paris », CREDOC, Décembre 2005.

10. Guy Poquet, « La baisse de la consommation d'eau dans les grandes villes: moins d'usines et des économies de gestion », CREDOC consommation et modes de vies n°170, Rapport du CREDOC, Novembre 2003.
11. « Rapport de développement soutenable », Eau de Paris, 2007.
12. Julien Souriau, « La consommation d'eau potable à Paris : un état des lieux », Rapport n°1, Eau de Paris, Septembre 2010.
13. Julien Souriau, « Les principaux déterminants de la consommation d'eau à Paris : un état des lieux » Rapport n°2, Eau de Paris, Juillet 2011.

8 Annexes

8.1 Présentation des différentes catégories disponibles dans la base de consommation d'Eau de Paris utilisées pour distinguer les différents usagers des services d'eau

- La première catégorie utilisée correspond au “libellé type abonnement” qui peut prendre les valeurs “Eau potable” ou “Eau non potable”. Nous avons utilisé cette catégorie pour ne conserver que les compteurs d'eau potable.
- La deuxième catégorie utilisée correspond au “libellé état abonnement” qui peut prendre les valeurs “En cours/ Fourniture Service”, “Résilié par abonné”, “Résilié par société” ou “Résilié suite à mutation” : nous n'avons conservé que les compteurs “En cours / Fourniture Service”.
- Ensuite, nous disposons d'une variable “nom du client titulaire” ainsi que “nom du client payeur” : ces deux variables ont été utilisées notamment pour décider au cas par cas le type d'abonné derrière le compteur lorsque ce n'était pas possible de le faire autrement.
- L'information la plus importante contenue dans cette base correspond au “Libellé qualité client” qui peut prendre les valeurs “Habitation Collective”, “Professionnel”, “Administration”, “Habitation”, “Individualisé SRU” ou “Habitation individuelle” : l'utilisation de cette variable a été expliquée dans la partie de présentation des données.
- L'information suivante correspond au “Libellé Code NAF” qui renseigne du secteur d'activité concerné en attribuant un code de 1 à 99 (ou vide si non professionnel). Cette variable a été utilisée pour distinguer certains compteurs de la catégorie “Professionnel” qui auraient été mal classés (notamment des syndicats de copropriété, des logements sociaux etc.).

8.2 Revenu net déclaré par foyer fiscal par arrondissement : données utilisées pour le découpage des groupes d'arrondissements par niveaux de revenus

La Table 8 présente les revenus déclarés des foyers fiscaux des différents arrondissements de la ville de Paris en 2009. La table classe les arrondissements par ordre décroissant des niveaux de revenus et laisse apparaître trois groupes d'arrondissements:

- les arrondissements les plus aisés : 7ème, 8ème, 16ème et 6ème
- les arrondissements caractérisés par des revenus moyens : 1er, 5ème, 17ème, 4ème, 9ème, 3ème, 15ème, 2ème, 14ème et 12ème
- les arrondissements les plus pauvres : 11ème, 10ème, 13ème, 18ème, 20ème et 19ème

Table 8: Revenus nets déclarés des foyers fiscaux en 2009 (en Euros)

Arrondissement	Nb Foyers fiscaux	Revenu ND des foyers fiscaux	Revenu ND par foyer fiscal
7	33 099	2 728 994 342	82 449
8	24 549	1 888 891 597	76 944
16	95 732	6 908 285 579	72 163
6	25 296	1 687 167 646	66 697
1	12 207	581 521 853	47 638
5	34 655	1 611 147 780	46 491
17	106 738	4 544 910 306	42 580
4	20004	821 991 248	41 091
9	40 387	16 32 155 601	40 413
3	24 987	969 885 691	38 816
15	148631	5532907554	37 226
2	16069	579375087	36 055
14	87 929	2 925 181 559	33 268
12	96 371	2 939 609 968	30 503
11	108 647	3 007 878 043	27 685
10	65 155	1 763 629 531	27 068
13	116 379	3 138 108 110	26 965
18	131845	3 144 417 620	23 849
20	134 294	3 029 037 512	22 555
19	121 156	2 676 925 064	22 095

Source : INSEE, 2011

8.3 Comparaison des estimations sur la base domestique et sur la base d'origine

La Table 9 présente les résultats de l'estimation économétrique de l'équation (2) réalisée sur la base des consommations domestiques puis sur la base originale (c'est à dire la base complète comportant tous les compteurs de la ville de Paris sans distinction en fonction des usagers).

Comme nous pouvons le remarquer à travers la significativité des différentes variables (très faible pour les spécifications (3) et (4) correspondant aux estimations sur la base complète), notre modèle semble pertinent pour expliquer la consommation d'eau des ménages parisiens (consommations domestiques d'eau) mais pas pour expliquer la consommation globale. Ce résultat est logique puisque les variables dont nous disposons et que nous avons utilisées ici sont des variables qui nous donnent des informations sur les populations de chaque Iris et sur les caractéristiques des logements. Nous ne disposons d'aucune variable qui puisse se rapprocher de l'activité économique des différents quartiers de la ville de Paris. Les consommations d'eau des administrations, des entreprises, des commerces et des bureaux ne peuvent donc pas être expliquées par le modèle que nous avons détaillé dans ce mémoire.

De plus, comme le montre le coefficient de détermination ajusté (R^2), la qualité globale du modèle est très faible pour l'estimation sur la base complète et très acceptable pour les estimations sur la base domestique.

Ces résultats justifient le choix que nous avons fait et révèlent que la sélection des compteurs domestiques a été relativement pertinente.

Table 9: Résultats pour base domestique et base complète

Variable explicative	(1)	(2)	(3)	(4)
Part population 0-18 ans	0,259*** (0,0674)	0,217*** (0,067)	-0,337*** (0,099)	-0,255** (0,112)
Part ménages 1 pers seule	-0,756*** (0,149)	-0,550*** (0,114)	-0,422*** (0,151)	-0,403*** (0,146)
Nb personnes par ménage	-2,183*** (0,115)	-1,735*** (0,160)	0,220 (0,161)	0,442 (0,334)
Part familles avec enfant(s)	0,181** (0,072)	0,0435 (0,058)	0,232** (0,106)	0,201 (0,130)
Nombre logements	1,895*** (0,150)	1,393*** (0,158)	-0,327* (0,188)	-0,274 (0,220)
Nombre pièces des RP	0,179** (0,083)	0,288*** (0,094)	-0,002 (0,135)	-0,294 (0,237)
Part propriétaires	-0,049*** (0,013)	-0,044*** (0,012)	-0,265*** (0,031)	-0,249*** (0,024)
Constante	5,927*** (0,295)	5,079*** (0,277)	7,404*** (0,339)	7,726*** (0,366)
R2 ajusté	0,467	0,614	0,252	0,314
Observations	883	883	915	915
VI Arrondissements	Non	Oui	Non	Oui

Notes : Les résultats représentés dans cette table correspondent à l'estimation de l'équation (2) de notre modèle avec les variables jugées "pertinentes" grâce à la TABLE 3. Les deux premières spécifications correspondent aux estimations sur la base "domestique" et les deux suivantes sur la base originale (non triée).

* : significatif à 10%, ** : significatif à 5%, *** : significatif à 1%.