



**HAL**  
open science

# Mesurer le recouvrement du coût financier des services d'eau : exploitation des données SISPEA et modélisation économétrique

A. Vigneron, S. Garcia, L. Guérin-Schneider

## ► To cite this version:

A. Vigneron, S. Garcia, L. Guérin-Schneider. Mesurer le recouvrement du coût financier des services d'eau : exploitation des données SISPEA et modélisation économétrique. [Rapport de recherche] irstea. 2009, pp.51. hal-02592983

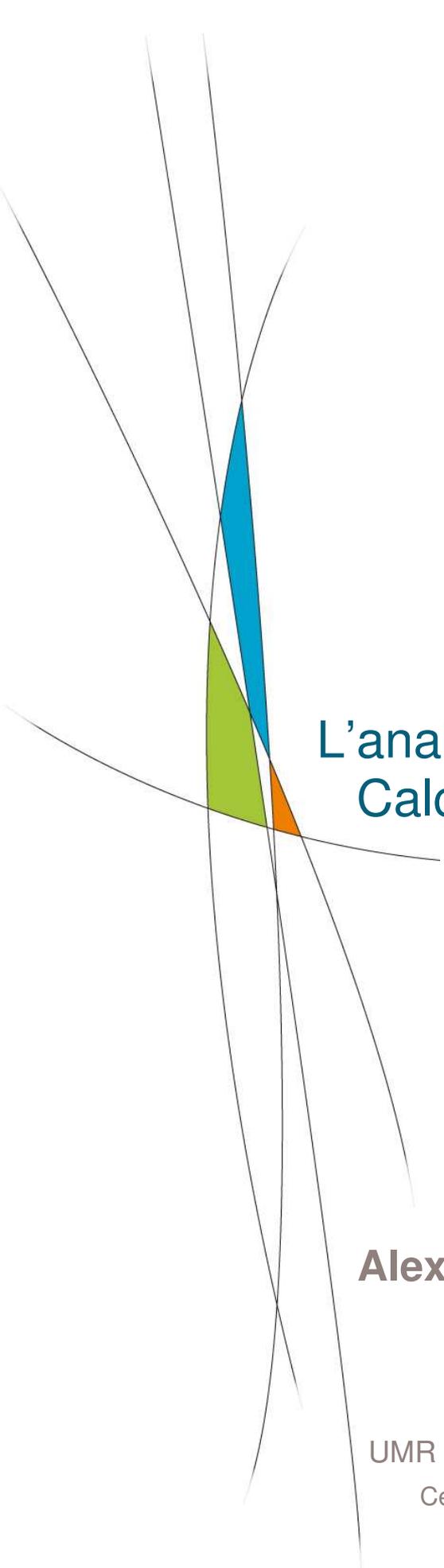
**HAL Id: hal-02592983**

**<https://hal.inrae.fr/hal-02592983>**

Submitted on 15 May 2020

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

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



# Convention Cemagref – ONEMA. Année 2009

Action N°30-1

## L'analyse économique des usages : Calcul de recouvrement des coûts

Rapport final :

Mesurer le recouvrement du coût financier  
des services d'eau : exploitation des  
données SISPEA et modélisation  
économétrique

*Décembre 2009*

**Alexandre Vigneron, Serge Garcia,  
Lætitia Guérin-Schneider**

UMR G-EAU "Gestion de l'Eau, Acteurs et Usages"

Cemagref, 361 rue Jean François Breton, BP 5095

34196 Montpellier Cedex 5 – France



## Table des matières

|   |           |
|---|-----------|
| Liste des figures.....  | 4         |
| Liste des abréviations.....   | 5         |
| Introduction .....  | 6         |
| <b>1. Contexte et objectif de l'étude .....</b>   | <b>8</b>  |
| 1.1. Définition et mesure des taux de recouvrement des coûts.....   | 8         |
| 1.2. Affiner la caractérisation du recouvrement des coûts des services par une<br>approche économétrique sur une population de services ..... | 9         |
| 1.3. Choix et validation d'une méthode économétrique.....   | 9         |
| <b>2. Analyse micro-économétrique du recouvrement des coûts des services d'eau en<br/>France pour l'année 2007 .....</b>                      | <b>11</b> |
| 2.1. Présentation de la base de données SISPEA.....   | 11        |
| 2.2. Première méthode d'estimation des coûts des services d'eau potable .....   | 14        |
| 2.2.1. Présentation générale de la méthode.....   | 14        |
| 2.2.2. Théorie Néoclassique du producteur : Minimisation des coûts des services AEP   | 15        |
| 2.2.3. De la Base Aube-Gironde.....   | 17        |
| 2.2.4. ...à la base SISPEA.....   | 28        |
| 2.2.5. Intérêt théorique de la méthode mais limite des résultats liés au flou dans la<br>définition des données .....                         | 33        |
| 2.3. Deuxième méthode d'estimation des coûts des services AEP .....   | 36        |
| 2.3.1. Présentation générale de la méthode.....   | 36        |
| 2.3.2. Théorie Néoclassique du producteur : Maximisation du profit des services AEP.  | 37        |
| 2.3.3. Estimation des recettes par « <i>switching model</i> ».....  | 37        |
| 2.3.4. Limites et extension de la deuxième méthode.....   | 39        |
| <b>3. Validation de la deuxième méthode et typologie des services recouvrant ou non<br/>leurs coûts .....</b>                                 | <b>41</b> |
| 3.1. Corrélation entre les calculs des taux de recouvrement suivant les deux méthodes<br>41   |           |
| 3.2. Caractéristiques des populations en non recouvrement.....  | 42        |
| <b>Conclusion générale .....</b>  | <b>44</b> |

## Liste des figures

|   |    |
|---|----|
| <b>Figure 1</b> Présentation des différents coûts mentionnés dans la directive cadre, d'après Rogers <i>et al.</i> (1997) cité dans Wateco .....      | 8  |
| <b>Figure 2</b> Comparaison entre les coûts observés et les estimations du modèle de coût construit sur la base AG .....                              | 28 |
| <b>Figure 3</b> Comparaison entre les coûts prédits (méthode 1) et les recettes HT observées sur la base SISPEA.....                                  | 32 |
| <b>Figure 4</b> Taux de recouvrement calculé par différence entre les coûts prédits (méthode 1) et les recettes HT observées sur la base SISPEA ..... | 32 |
| <b>Figure 5</b> Comparaison entre les coûts prédits (méthode 2) et les recettes HT observées sur la base SISPEA.....                                  | 38 |
| <b>Figure 6</b> Taux de recouvrement calculé par différence entre les coûts prédits (méthode 2) et les recettes HT observées sur la base SISPEA ..... | 38 |

## Liste des abréviations

AEP : Alimentation en eau potable

AG: Aube-Gironde (base de données)

CARE : Compte Annuel de Résultat d'Exploitation [d'une délégation]

CRF : Compte Rendu Financier [d'une délégation]

CRT : Compte Rendu Technique [d'une délégation]

DCE : Directive Cadre sur l'Eau

E&Y : Ernst et Young

FP2E : Fédération Professionnelle des Entreprises de l'Eau

LEMA : Loi sur l'Eau et les Milieux Aquatiques

M49 : instruction comptable applicable au budget annexe publique d'eau et d'assainissement

ONEMA : Office Nationale de l'Eau et des Milieux Aquatiques

SISPEA : Système d'Information sur les Services Publics de l'Eau et de l'Assainissement

TRANSLOG : Transcendental logarithmique

## Introduction

Les travaux sur le recouvrement des coûts présentés dans ce rapport ont été réalisés dans le cadre de la Convention de partenariat 2009 entre le Cemagref et l'ONEMA (thème « instruments et évaluation économique »).

L'INRA a contribué activement à ce travail à travers Serge Garcia, économètre au Laboratoire d'Economie Forestière, qui a co-encadré avec Lætitia Guérin-Schneider le stage de Mastère d'Alexandre Vigneron (Mastère 2 Modélisation statistique pour l'économie et la gestion de l'Université Nancy 2).

Ce rapport présente une approche économétrique, complémentaire à l'approche gestionnaire développée dans un premier rapport intitulé :

*"Des approches économiques aux outils comptables et financiers disponibles : quelles mesures possibles du recouvrement des coûts des services d'eau potable et d'assainissement".*

Le premier rapport montrait pourquoi les coûts opérationnels suivis par les gestionnaires, qu'il s'agisse d'informations comptables ou de calculs financiers, ne recouvrent pas exactement la notion des coûts économiques au sens de la Directive Cadre sur l'Eau (DCE).

Nous avons mobilisé les études macroéconomiques déjà conduites pour estimer les coûts financiers *gestionnaires* puis les coûts financiers *économiques* et pour montrer qu'il n'y a pas de lecture univoque du taux de recouvrement.

**Le présent rapport complète l'analyse. Il s'agit ici de tester deux méthodes économétriques différentes permettant de caractériser la situation de recouvrement des coûts pour une population de services d'eau potable, issue des premières données du Système d'Information des Services d'Eau et d'Assainissement (SISPEA) portant sur l'année 2007.**

\*\*\*\*\*

### Avertissement

**Il faut garder à l'esprit durant la lecture de ce rapport que les résultats chiffrés ne sont pas représentatifs de la situation des services d'eau en France.** Par ailleurs, les données financières collectées dans SISPEA portant sur l'exercice 2007 se sont avérées peu fiables et non homogènes.

**L'apport du présent travail est avant tout méthodologique et ne permet pas d'interprétation des résultats chiffrés en tant que tels.**

\*\*\*\*\*

**Pour une vision synthétique le lecteur pourra se reporter aux cadres de synthèse présents :**

- Structuration des résultats présentés p 10,
- Conclusion sur l'estimation de la population en sous-recouvrement en construisant une fonction de coût basée à partir d'une base de données des coûts comptables et financiers de services externes à SISPEA, p35
-

Conclusion sur l'estimation de la population en sous-recouvrement en construisant une fonction de coût à partir des recettes des services de la base SISPEA, p40

- Conclusion sur la possibilité de caractériser la population des services en non recouvrement, p 43

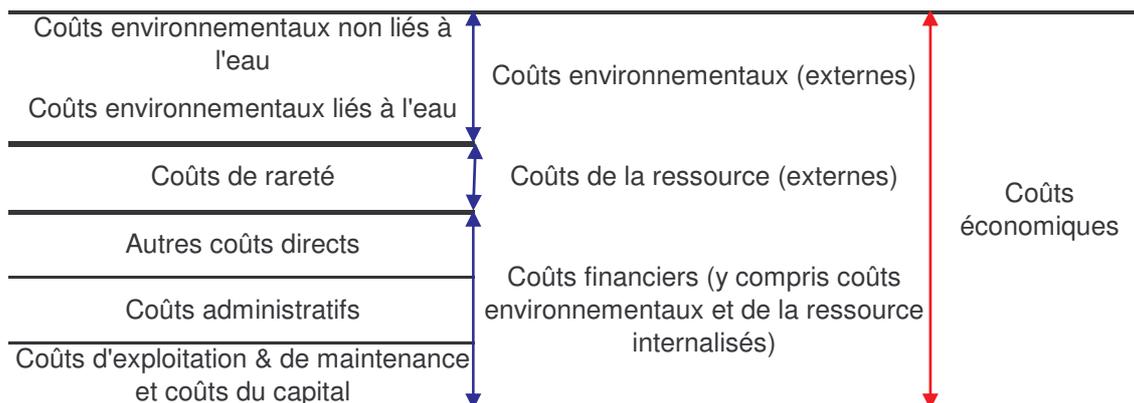
- Conclusion générale, p44

# 1. Contexte et objectif de l'étude

## 1.1. Définition et mesure des taux de recouvrement des coûts

Au sens de la Directive Cadre sur l'Eau (DCE), le coût complet d'un service comprend le coût financier des services, le coût d'opportunité de la ressource et le coût environnemental.

Les deux derniers coûts ne sont actuellement pas couverts par les recettes venant des bénéficiaires des services d'eau et d'assainissement. Par contre, ces derniers assument une partie substantielle des coûts financiers.



**Figure 1** Présentation des différents coûts mentionnés dans la directive cadre, d'après Rogers *et al.* (1997) cité dans Wateco

Le coût financier est sensé être mesuré à partir de la comptabilité de services. Mais la comparaison entre la définition économique (DCE) et la définition gestionnaires (à partir des données financières et comptables réellement disponibles) montre qu'il existe des décalages significatifs<sup>1</sup>.

L'écart entre *coût financier gestionnaire* et *coût financier économique* vient principalement des éléments suivants :

- sous-investissement (qui minore la consommation de capital fixe),
- non prise en compte des coûts d'opportunité sur les capitaux d'origine publique (minore le coût d'opportunité des capitaux),
- à l'inverse, les usagers des services d'eau assument des coûts compensatoires liés à l'utilisation de la ressource en amont (pollution, manque d'eau), ce qui augmente les coûts financiers.

A partir d'une revue des principales études macroéconomiques disponibles sur l'économie et le recouvrement des coûts des services d'eau, il est possible d'estimer, en ordre de grandeur, la différence entre les coûts financiers gestionnaires (c'est-à-dire issus de données comptables) et les coûts financiers économiques (au sens de la DCE).

En nous appuyant sur ces différentes sources (BIPE notamment), nous sommes arrivés au constat suivant : le taux de recouvrement des coûts financiers qui transparait des données comptables macroéconomiques du secteur de l'eau et de l'assainissement serait proche de 95% (le complément venant de subventions). Toutefois, la valeur varie grandement suivant les sources (ainsi l'étude Ernst&Young conduit à estimer un taux de recouvrement plus faible) et suivant les retraitements économiques adoptés (traitement des redevances agences de l'eau notamment).

<sup>1</sup> Cf. rapport intermédiaire de juillet 2009

**Pour la suite du rapport, nous retiendrons conventionnellement comme ordre de grandeur du taux de recouvrement des coûts comptables des services la valeur de 95%, en nous appuyant sur les données BIPE portant sur l'année 2004.**

## **1.2. Affiner la caractérisation du recouvrement des coûts des services par une approche économétrique sur une population de services**

Outre la difficulté d'obtenir des données consolidées au niveau français, les approches macroéconomiques présentent une limite importante : elles n'offrent une vision que de la situation moyenne en France. On aboutit à un taux global national, où éventuellement par bassin versant, sans pouvoir caractériser quels types de services couvrent ou non leurs coûts.

C'est pourquoi **il est apparu utile de travailler par une approche économétrique sur une base de données non consolidée.**

Les données utilisées sont les premières données collectées par le Système d'Information sur les Services d'Eau et d'Assainissement (SISPEA) au sein de l'ONEMA. Il s'agit de données issues des rapports prix qualité du service (RPQS) d'eau et d'assainissement, portant sur 2007 et rassemblées à titre de test avant la mise en œuvre officielle de SISPEA.

*N.B. : La base de données des services d'assainissement n'a pas pu être utilisée car les informations ont été jugées trop peu fiables par l'ONEMA. Seules les données eau potable sont exploitées.*

La finalité d'un modèle économétrique est la suivante : **en identifiant des services qui sont statistiquement en situation de sous-recouvrement de coûts par rapport aux autres, le modèle permet d'analyser si ces services ont des caractéristiques particulières et si oui, lesquelles.** L'intérêt de SISPEA est en effet d'offrir, en parallèle à des éléments financiers (recettes et prix du service), de nombreuses variables descriptives du contexte technique et des performances atteintes (indicateurs obligatoires du Rapport Prix Qualité du Service). A terme, le travail proposé permettrait donc, avec une permanence des méthodes, d'isoler économétriquement une population de services qui semblent en sous-recouvrement et de caractériser et d'analyser cette population. **Cela ouvre alors la voie à une orientation ciblée des politiques publiques visant les services en non recouvrement (ou en sous-recouvrement par rapport aux autres) et cela peut aussi contribuer à une évaluation des politiques publiques sur le financement des services d'eau.**

Les deux approches, macro-économiques et micro-économétriques s'avèrent donc complémentaires :

- la consolidation des données comptables fournit une mesure moyenne, a priori absolue, du taux de recouvrement des coûts comptables à une échelle macro-économique,
- l'exploitation de la base SISPEA, par une méthode économétrique micro-économique, permet d'identifier de manière relative des services en situation de moindre-recouvrement des coûts et permet de caractériser ces services et, à terme, de suivre leur évolution.

## **1.3. Choix et validation d'une méthode économétrique**

La démarche économétrique vise à construire un modèle de coût à partir de variables descriptives pour pouvoir comparer pour chaque service ce coût modélisé aux recettes prélevées sur les usagers. On obtient ainsi un taux de recouvrement par service.

Bien que ce soit techniquement possible, il est contestable d'exploiter le modèle sur un service donné. En effet, on peut toujours avoir un service mal modélisé (du fait de facteurs exceptionnels non pris en compte dans le modèle). Par contre, **l'hypothèse est qu'en travaillant non pas sur un service isolé, mais sur la population de services identifiés**

**comme ne recouvrant pas leurs coûts, on pourra mettre en évidence des tendances statistiquement correctes.**

L'objet de cette première étude est de tester le **principe** d'un modèle économétrique construit à partir des données SISPEA.

Pour ce faire, deux méthodes économétriques ont été choisies pour établir la fonction de coût.

**La première méthode servira pour la validation : elle utilise un modèle de coût établi sur la base d'informations à la fois techniques et comptables issues d'une base de donnée distinctes de SISPEA.** Ces données ont été rassemblées par Serge Garcia à l'occasion d'études antérieures. La base comporte des informations issues des comptes rendus financiers (CRF) des délégataires et des budgets annexes (M49) des collectivités. Elle comporte également les principales variables techniques que l'on retrouve dans SISPEA. Par contre, elle couvre un territoire limité à deux départements.

**La fonction de coût ainsi produite traduit donc une fonction de coût comptable, établie de manière indépendante des données SISPEA.** Cette fonction va pouvoir être appliquée sur les services de SISPEA et comparée aux recettes. On obtient donc ainsi un taux de recouvrement par rapport à une référence externe à SISPEA.

L'avantage de cette méthode est qu'elle permet de reconstituer une véritable fonction de coûts indépendante de l'échantillon SISPEA. Par contre, pour être valable, elle oblige à disposer d'observations récentes sur la comptabilité et sur les conditions techniques d'une population représentative de services. Cette condition est difficile à remplir : même dans notre cas, l'échantillon est limité à 2 départements et date d'environ 10 ans, ce qui n'est pas entièrement satisfaisant.

**La seconde méthode se veut plus simple à mobiliser. Elle repose sur l'exploitation des seules données de SISPEA, couplées à une hypothèse de taux moyen de recouvrement des services** (issue par exemple de l'étude macro-économique BIPE ou Ernst&Young). **La fonction de coût est obtenue en couplant la fonction de recette construite sur les seuls services de SISPEA avec le taux de recouvrement moyen** (coût théorique = recette théorique / taux de recouvrement moyen). A nouveau on peut comparer le coût théorique ainsi obtenu à la recette réelle pour chaque service.

Par construction cette méthode ne va pas permettre de mesurer le taux moyen de recouvrement des coûts (puisque c'est une hypothèse exogène fixée sur la base d'autres études macro-économiques), par contre **elle permet d'isoler une population de service en sous-recouvrement par rapport à la tendance moyenne.**

### **Structuration des résultats présentés**

Le rapport est construit suivant le fil conducteur suivant :

- dans une première partie les deux méthodes sont successivement mises en œuvre et permettent d'isoler une population de services en sous-recouvrement,
- puis on cherche à valider la seconde méthode (plus simple à mettre en œuvre à partir de SISPEA) par la première en voyant dans quelle mesure les résultats sont convergents ; on explore aussi si une typologie des services en sous-recouvrement se dégage,
- sur cette base, le rapport conclut sur les voies d'amélioration à explorer pour permettre l'exploitation de SISPEA afin d'isoler les services en sous-recouvrement par des outils économétriques.

## 2. Analyse micro-économétrique du recouvrement des coûts des services d'eau en France pour l'année 2007

L'objectif principal de cette partie est de tester la faisabilité et la validité d'une approche économétrique mobilisant les données SISPEA pour isoler des services en sous-recouvrement des coûts financiers comptables.

Nous mettons en œuvre deux méthodes différentes afin d'estimer les coûts comptables des services de SISPEA<sup>2</sup> :

- La première méthode d'estimation de ces coûts se fera via les coefficients d'une fonction de coût estimés à partir d'une autre base de données (Aube-Gironde) (la méthode sera décrite par la suite dans la partie 2.2.1). **Cette méthode couple donc l'exploitation de la base SISPEA à une seconde base de données. [Cette méthode servira à valider la pertinence de la seconde méthode].** Elle a l'avantage de s'appuyer sur des données fines (car issues des documents comptables produits par les services de l'Aube et de la Gironde) mais présente le double inconvénient de constituer un échantillon peu représentatif de l'ensemble des services d'eau en France (car elle concerne seulement 2 départements) et de porter sur des données assez anciennes (datant d'environ 10 ans).
- **La deuxième méthode d'estimation de ces coûts n'utilise que la base de données SISPEA en partant des recettes. Elle mobilise néanmoins une donnée macroéconomique externe : le taux de recouvrement global** calculé grâce au rapport BIPE ( $\approx 95\%$ ) (la méthode sera décrite par la suite dans la partie 2.3.1). Elle a le principal avantage de s'appuyer sur des données plus représentatives et récentes (car on utilise le rapport BIPE 2006 qui concerne l'ensemble des services d'eau en France) mais le principal inconvénient de s'appuyer sur un taux de recouvrement des coûts de 95 % fondé sur une analyse macro-économiques peu transparente et non officielle (source des données non détaillées par BIPE).

### 2.1. Présentation de la base de données SISPEA

La loi sur l'eau et les milieux aquatiques (LEMA) a confié à l'ONEMA la mise en place et la coordination technique d'un système d'information l'eau (SIE) dont fait partie le système d'information sur les services publics d'eau et d'assainissement (SISPEA). SISPEA s'appuie en grande partie sur les données réglementaires à intégrer dans le rapport prix qualité du service en application du décret et de l'arrêté du 2 mai 2007, applicable en 2009 sur les données 2008.

**En 2008, seules les collectivités volontaires ont transmis à l'ONEMA les valeurs des indicateurs caractérisant leur service d'eau durant l'exercice 2007. Ces données sont celles utilisées dans la présente étude.**

De ce fait, la première base de données SISPEA ne comporte que 219 services AEP français

En nous appuyant sur une analyse réalisée par Guillem Canneva (Laboratoire Gestion de l'Eau et de l'Assainissement) dans son étude typologique réalisée pour l'ONEMA, on peut conclure que ce premier échantillon n'est pas tout à fait représentatif de la population française des services d'eau. Ainsi, en comparant la base des services SISPEA 2007 aux données de BANATIC, Canneva montre qu'il y a une surreprésentation des services

---

2 On possède les recettes observées des services de SISPEA mais non les coûts : il faut donc les estimer afin d'évaluer le taux de recouvrement des coûts.

intercommunaux, des délégations et des services de grande taille. Toutefois on respecte tout de même une certaine diversité de services.

Les services renseignés dans la base sont caractérisés par les indicateurs suivants :

Tableau 1 Liste des variables décrivant chaque service dans SISPEA

| Indicateurs des services AEP de SISPEA                            | Définitions  |
|---|--|
| Taux de conformité microbiologique (en %)                         | Indicateur de la qualité de l'eau potable. Il représente le pourcentage des prélèvements (aux fins d'analyses microbiologiques) jugés conformes selon la réglementation en vigueur.  |
| Taux de conformité physico-chimique (en %)                        | Indicateur de la qualité de l'eau potable. Il représente le pourcentage des prélèvements (aux fins d'analyses physico-chimiques) jugés conformes   |
| Rendement du réseau de distribution (en %)                        | Traduit la part des volumes introduits dans le réseau de distribution qui est consommée avec autorisation sur le périmètre du service ou vendue en gros à un autre service AEP.<br>Rendement = (volume consommé autorisé + volume vendu en gros) / (volume produit + volume acheté en gros)  |
| Indice linéaire des volumes non comptés (m <sup>3</sup> /km/jour) | Traduit les volumes mis en distribution qui sont perdus ou qui sont consommés sans faire l'objet d'un comptage par km de réseau.<br>Indice = (volume mis en distribution - volume comptabilisé) / longueur du réseau de desserte / 365 (ou 366)<br>Les volumes comptabilisés sont les volumes consommés issus du relevé des compteurs chez les abonnés et autres usagers équipés de compteurs                |
| Indice linéaire de pertes en réseau (m <sup>3</sup> /km/jour)     | Traduit les volumes mis en distribution qui sont perdus ou consommés sans autorisation par km de réseau<br>Indice = (volume mis en distribution - volume consommé autorisé) / longueur du réseau de desserte / 365 (ou 366).<br>Les volumes consommés autorisés sont la somme des volumes comptabilisés, des volumes des consommateurs sans comptage (estimés) et des volumes de service du réseau (estimés) |
| Indice avancement protection de la ressource (En %)               | Niveau d'avancement (exprimé en %) de la démarche administrative et opérationnelle de protection du ou des points de prélèvement dans le milieu naturel d'où provient l'eau potable distribuée.<br>Il donne une information sur la performance atteinte pour assurer une protection effective de la ressource selon la réglementation en vigueur.  |
| Durée d'extinction de la dette (En années)                        | Durée théorique nécessaire pour rembourser la dette du service AEP si la collectivité affecte à ce remboursement la totalité de l'autofinancement dégagé par le service.   |
| Taux de réclamation (Nombre par milliers d'abonnés)               | Nombre de réclamations écrites rapporté au nombre d'abonnés divisé par 1000 du service AEP considéré.<br>Il permet de traduire le niveau d'insatisfaction des abonnés au service d'eau.  |
| Estimation du nombre d'habitants desservis                        | Nombre d'habitants desservis par le service AEP, y compris les résidents saisonniers.<br>Cet indicateur permet d'apprécier la taille du service.   |
| Montant de la facture 120 m <sup>3</sup> (en euros)               | Prix de l'eau potable TTC par m <sup>3</sup> pour une facture type de 120 m <sup>3</sup> .   |

| <b>Indicateurs des services AEP de SISPEA</b>                         | <b>Définitions</b>  |
|---|---|
| Mode de Gestion du service AEP<br>(0 pour régie et 1 pour délégation) | Les services AEP sont exploités en régie ou en délégation (affermage, concession, régie intéressée ...)   |
| Volume prélevé dans les eaux souterraines<br>(m <sup>3</sup> )        | Volume d'eau capté de l'exploitation des nappes souterraines.   |
| Volume prélevé dans les eaux de surface<br>(m <sup>3</sup> )          | Volume d'eau capté de l'exploitation des eaux de surface.   |
| Volume produit<br>(m <sup>3</sup> )                                   | Volume d'eau total produit par le service.  |
| Volume acheté à d'autres services<br>(m <sup>3</sup> )                | Volume d'eau non produit par le service mais acheté en gros à d'autres services.  |
| Volume vendu à d'autres services<br>(m <sup>3</sup> )                 | Volume d'eau produit par le service mais vendu en gros à d'autres services  |
| Volume vendu aux abonnés domestiques<br>(m <sup>3</sup> )             | Volume produit (ou acheté à d'autres services), facturé aux abonnés domestiques.  |
| Volume vendu aux autres abonnés<br>(m <sup>3</sup> )                  | Volume produit (ou acheté à d'autres services), facturé aux autres abonnés.   |
| Volume mis en distribution<br>(m <sup>3</sup> )                       | Volume produit + volume acheté à d'autres services - volume vendu à d'autres services.  |
| Nombre d'abonnements  | Nombre d'abonnés du service AEP.  |
| Linéaire de réseaux de desserte<br>(km)                               | Nombre de kilomètres du réseau de canalisations du service AEP.   |
| <b>Montant des recettes liées à la facturation (euros)</b>            | <i>Montant des recettes provenant de la facture d'eau des abonnés (sont a priori exclues les autres recettes d'exploitation ainsi que des subventions du budget général).</i> |

La plupart de ces indicateurs a fait l'objet d'une définition relativement précise reprise dans la circulaire du 24 novembre 2008.

Toutefois, **certaines informations font exception. C'est notamment le cas de l'information centrale qui nous intéresse : le montant des recettes liées à la facturation n'a pas été clairement défini.**

Après un sondage auprès de plusieurs DDAF ayant rempli les données dans la base SISPEA nous avons fait l'hypothèse que les données saisies sont généralement des montants TTC, correspondant aux recettes mentionnées dans les CRF dans la ligne "recettes d'exploitation" ou mentionné en recettes de vente de produit dans la M49. Pour cette raison, en l'absence d'information supplémentaire, les recettes utilisées dans le modèle ont été obtenues en retranchant la TVA correspondant à 5,5%.

**Toutefois, cette hypothèse n'est pas forcément vérifiée pour tous les services.** D'abord, il conviendrait d'affiner ce calcul en fonction du régime de TVA de chaque service. Surtout un risque existe, suivant la personne qui a renseigné la base, que les données soient parfois hors taxe, parfois TTC, qu'elles intègrent plus que la facturation du service (intégration par exemple des comptes de tiers comme les redevances des agences de l'eau) ou à l'inverse moins (si seules les recettes du délégataire ont été retranscrites, hors recettes de la collectivité).

**La plus grande prudence s'impose donc quant à l'exploitation des données.**

#### Avertissement

En raison de cette représentativité imparfaite de l'échantillon SISPEA 2007 ainsi que de l'absence de définition officielle et univoque du montant des recettes, les résultats chiffrés ne doivent pas être pris comme des références fiables.

**On constate à la lecture du Tableau 1 que les données sur les coûts comptables ou portés aux CRF ne font pas partie de la base SISPEA. On se propose de les estimer économétriquement afin d'estimer un taux de recouvrement des coûts des services de la base SISPEA.**

## **2.2. Première méthode d'estimation des coûts des services d'eau potable**

### **2.2.1. Présentation générale de la méthode**

Le principe de l'estimation des coûts consiste à utiliser une autre base de données de services d'eau (base Aube-Gironde dans notre cas) où les coûts comptables sont effectivement observés. Il s'agit pour les délégations des coûts fournis dans les CRF. Cela concerne donc l'exploitation des services à l'exclusion des coûts d'investissement (pris en charge généralement par la collectivité délégante). En cas de régies, les données sont issues des charges portées dans le budget annexe M49, en excluant à nouveau les investissements.

Ces coûts d'exploitation comptables sont régressés sur des variables caractéristiques des services communes entre notre base d'estimation (Aube-Gironde, notée AG) et notre base d'intérêt (SISPEA).

Les coefficients obtenus sont alors repris d'AG et associés aux variables de SISPEA : on a alors une estimation des coûts<sup>3</sup> et donc, par comparaison aux recettes de facturation, une estimation des taux de recouvrement pour chaque service AEP de SISPEA :

---

<sup>3</sup> Puisque les coûts de la base de données Aube-Gironde représentent les coûts observés dans les comptabilités des délégataires et des collectivités, les coûts estimés des services de SISPEA seront uniquement les coûts financiers comptables.

- Dans Aube-Gironde :

$$\begin{pmatrix} \text{Coût observé}^1 \\ \vdots \\ \text{Coût observé}^m \end{pmatrix}^{AG} = \begin{pmatrix} 1 & X_1^1 & \dots & X_n^1 \\ 1 & X_1^2 & \dots & X_n^2 \\ \dots & \dots & \dots & \dots \\ 1 & X_1^m & \dots & X_n^m \end{pmatrix}^{AG} X \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \mu^1 \\ \mu^2 \\ \dots \\ \mu^m \end{pmatrix}^{AG} \quad (3)$$

Notes : Nombre de services AEP de Aube-Gironde : m

Nombre de variables : n (  $X_1$  à  $X_n$  )

Nombre de coefficients à estimer : n+1 (  $\beta_0$  à  $\beta_n$  )

Résidus : de  $\mu^1$  (pour le 1<sup>er</sup> service) à  $\mu^m$  (pour le m<sup>ième</sup> service)

Les coefficients  $(\bar{\beta}_0 \ \bar{\beta}_1 \ \dots \ \bar{\beta}_n)$  estimés de (3) sont alors reportés dans la base SISPEA et associés aux mêmes variables que celles de SISPEA :

- Dans SISPEA (pour exemple on considère z services AEP) :

$$\begin{pmatrix} 1 & X_1^1 & \dots & X_n^1 \\ 1 & X_1^2 & \dots & X_n^2 \\ \dots & \dots & \dots & \dots \\ 1 & X_1^z & \dots & X_n^z \end{pmatrix}^{SISPEA} X \begin{pmatrix} \bar{\beta}_0 \\ \bar{\beta}_1 \\ \dots \\ \bar{\beta}_n \end{pmatrix} = \begin{pmatrix} \text{Coût estimé}^1 \\ \vdots \\ \text{Coût estimé}^z \end{pmatrix}^{SISPEA} \quad (4)$$

Ainsi le taux de recouvrement des coûts pour chaque service « i » de SISPEA sera égal à :

$$\frac{\text{Recette}^i}{\text{Coût estimé}^i}$$

Et le taux de recouvrement global des coûts des services de SISPEA :

$$\frac{\sum_{i=1}^z \text{Recette}^i}{\sum_{i=1}^z \text{Coût estimé}^i}$$

### 2.2.2. Théorie Néoclassique du producteur : Minimisation des coûts des services AEP

La théorie Néoclassique de la firme montre que la technologie est comme une « boîte noire » où, seuls les « inputs » (facteurs de production, autres caractéristiques de l'entreprise...) entrent dans la fonction de production, sans prise en compte des aspects d'efficacité organisationnelle ou des répartitions des droits de propriétés...

Nous supposons que les services AEP distribuent aux usagers<sup>4</sup> un volume d'eau potable Q qui est caractérisé par un vecteur de caractéristiques techniques CAR (Longueur du réseau,

<sup>4</sup> Pour un service :

*Nombre d'abonnés, topographie de la zone de distribution...*) et, par un vecteur d'indicateurs de qualité  $lq$  (*Indice linéaire des volumes non comptés, Rendement du réseau ...*).

En outre, différents facteurs de production variables  $x$  (travail, électricité...) sont utilisés dans la production et la distribution d'eau potable.

Par conséquent, le processus de production de type Néoclassique peut être représenté par la fonction de transformation suivante :

$$G(X)=Y \quad (5)$$

$$(5) \Leftrightarrow G(lq, x, CAR, u) = Q \quad \text{où } u \text{ représente les aléas de production non observés.}$$

$$(5) \Leftrightarrow F(Q, lq, x, CAR, u) = 0$$

Le principe de dualité en théorie de la production énonce qu'il est équivalent pour le producteur de maximiser son profit ou de minimiser ses coûts de production en ajustant au mieux sa capacité de production en fonction des caractéristiques de sa technologie. Ainsi, on aura une fonction de coût duale à la fonction de production où le producteur cherche à minimiser ses dépenses de production en utilisant de manière optimale ses ressources c'est-à-dire ses facteurs de production entrant dans le processus de production.

A court terme, ce sont les dépenses d'exploitation qui sont minimisées, le facteur de production « capital » représentant l'ensemble des infrastructures étant considéré comme un facteur fixe.

Le programme du gestionnaire du service s'écrit donc :

Min  $\sum (w \cdot x)$  sous contrainte  $F(Q, lq, x, CAR, u) = 0$  où  $w$  représente le vecteur des prix des facteurs de production variables et  $x$ , leurs quantités associées.

La fonction de coût variable est issue du choix optimal des quantités de facteurs de production variables qui minimise les dépenses d'exploitation (variables) pour produire la quantité  $Q$  au niveau de qualité  $lq$  étant donné les prix des facteurs de production  $w$  et les caractéristiques du service  $CAR$ , et sous contrainte technologique (définie par la fonction de transformation  $F$ )

Ainsi, la fonction de coût adoptée pour le modèle dépendra donc de la quantité d'eau mis en distribution ( $Q$ ), du vecteur des prix des facteurs de production ( $w$ ), d'indicateurs de qualité ( $lq$ ), des caractéristiques du service ( $CAR$ ) et du terme d'erreur habituel  $\varepsilon$  :

$$C(Q, CAR, lq, w, \varepsilon) \quad (6)$$

Etant donné que l'eau est un produit lourd dont le transport est coûteux et le stockage difficile, plus le volume d'eau  $Q$  mis en distribution est important, plus les coûts engendrés par son stockage et son transport seront importants ce qui entraînera naturellement des coûts plus élevés.

De plus, l'exigence que cette eau distribuée soit de bonne qualité se traduira par des coûts de traitement de l'eau brute et par des coûts d'exploitation supplémentaires afin de préserver cette bonne qualité de l'eau dans les réseaux AEP.

D'autres caractéristiques des réseaux comme le nombre d'abonnés ou la longueur du réseau auront un également un impact sur les coûts variables : en effet, plus le service AEP est « de taille importante » c'est-à-dire compte beaucoup d'abonnés et un grand réseau, plus le nombre de branchements à effectuer sera important ce qui induira des coûts plus importants.

---

Volume mis en distribution «  $Q$  » = volume produit + (volume acheté en gros - volume vendu en gros).

La topographie des zones desservies est aussi un aspect des services important à prendre en compte dans les coûts des services AEP : en effet, selon le relief de la zone de distribution d'eau potable, les coûts variables pourront être différents d'un service AEP à l'autre. L'environnement dans lequel le service évolue explique également les coûts des services dans le sens où certains types de coûts ne seront pas les mêmes selon que l'eau potable soit produite à partir des eaux souterraines ou à partir des eaux de surface<sup>5</sup>.

En ce qui concerne les indicateurs de qualité des services AEP, on peut citer le rendement du réseau de distribution *RDT* ou encore l'indice linéaire des volumes non comptés *ILVNC* qui, au travers des fuites dans les réseaux de distribution d'eau potable (dus à des ruptures de conduites ou à des joints fuyants), permettent de décrire l'état du réseau : celui-ci induira des coûts de production et de distribution plus ou moins élevé.

Enfin, il est évident qu'il faut également payer les facteurs de production (travail, électricité, capital) tout au long de la chaîne de production-distribution de l'eau potable : les prix des facteurs de production sont des déterminants du coût.

### 2.2.3. De la Base Aube-Gironde...

#### ❖ Présentation de la base de données AG

Nous avons à notre disposition 2 bases de données construites à partir des CRT (comptes rendus techniques) et CRF (comptes rendu financiers) pour les services AEP délégués et à partir des comptabilités des collectivités pour les services AEP gérés en régie (budget annexe M49) :

La première base concerne 47 services AEP de Gironde, observés sur 4 années (1995 à 1998) et, délégués à des entreprises privées sous contrat d'affermage.

La seconde base concerne 52 services AEP en régie situés dans le département de l'Aube et observés sur 3 ans (1995 à 1997).

Afin d'avoir un échantillon plus grand, nous avons compilé ces deux bases de données et empilées leurs données : les 344 observations qui en résultent ( $47 \times 4 + 52 \times 3$ ) seront donc traitées en coupe transversale<sup>6</sup>.

Le fait de traiter en coupe instantanée, des données de panel, entraînera un problème d'hétéroscédasticité de la matrice de variance-covariance des erreurs<sup>7</sup> : les estimateurs des MCO seront sans biais mais non efficaces (les coefficients seront centrés mais leurs écarts-types, donc leurs significativités seront biaisés). Cependant, puisque l'objectif principal est de faire de la prédiction de coût dans la base SISPEA, ce problème d'hétéroscédasticité ne sera pas dommageable.

**La description de la technologie des services AEP faite dans la section précédente nous a permis d'identifier plusieurs déterminants importants des coûts :**

- Le volume d'eau mis en distribution **Q**.

---

<sup>5</sup> Une eau potable produite à partir des eaux souterraines engendrera des coûts de pompage et de forage plus importants (car il faut creuser plus profondément) mais des coûts de potabilisation moindre par rapport à une eau produite à partir des eaux de surface (car l'eau souterraine est plus « pure »)

<sup>6</sup> On fait comme si on avait 344 services AEP indépendants.

<sup>7</sup> L'hétéroscédasticité pourrait éventuellement être corrigée par l'intégration d'une matrice de White robuste à l'hétéroscédasticité.

- Le vecteur des caractéristiques techniques **CAR** :
  - Le type de traitement de potabilisation (*Trait*)
  - Le nombre d'abonnés desservis (*AB*)
  - La longueur du réseau (*Longres*)
  - La topographie de la zone de distribution (*Topo*)
  - L'origine des eaux brutes (*EB*)
- Le vecteur des indicateurs de qualité **Iq** :
  - Le rendement du réseau de distribution (*RDT*)
  - L'indice linéaire de pertes en distribution (*ILVNC*)
- Les prix des facteurs de production **w**.

La fonction de coût variable peut donc s'écrire :  
 $C ( Q, ( Trait, AB, Longres, Topo, EB), (RDT, ILVNC), w) \quad (6)$

Nous n'utiliserons pas toutes les variables spécifiées dans cette fonction de coût (6) pour plusieurs raisons :

- Les données ne sont pas disponibles (*Topo*) ou les observations ne présentent que peu de variations pour avoir un effet significatif (*EB, Trait*).
- Il faut que les variables explicatives utilisées dans la modélisation de la fonction de coût pour la base Aube-Gironde soient communes à celles observées dans la base SISPEA (voir section 3.2.1 pour la description de la méthode) : Ainsi, « w » n'étant pas présente dans la base SISPEA, elle ne sera pas utilisée non plus<sup>8</sup>.

Les statistiques descriptives des variables finalement utilisées pour le modèle<sup>9</sup> sont reportées dans le tableau suivant :

---

<sup>8</sup> Puisque les variables concernant le prix des facteurs de production sont exogènes (définies par le marché), elles ne varient que très peu. Le fait de ne pas avoir ces données dans SISPEA ne sera donc pas préjudiciable.

<sup>9</sup> C'est-à-dire les variables expliquant les coûts totaux ou le mode de gestion (voir la section « switching modèle » p20), qui sont communes à SISPEA et à AG.

**Statistiques descriptives de l'échantillon AG**

|                |                         | Gestion publique<br>(156 observations) |         |        |           | Gestion déléguée<br>(188 observations) |         |        |           |
|----------------|-------------------------|--|---------|--------|-----------|--|---------|--------|-----------|
| Variable       | Unité                   | Moy                                    | Std     | Min    | Max       | Moy                                    | Std     | Min    | Max       |
| <i>C</i>       | Millier de Franc        | 502                                    | 699     | 59     | 4 448     | 1 688                                  | 2 069   | 184    | 12 736    |
| <i>Q</i>       | m <sup>3</sup>          | 152 843                                | 228 348 | 13 540 | 1 512 776 | 541 700                                | 722 188 | 38 824 | 4 278 094 |
| <i>AB</i>      |                         | 710                                    | 992     | 81     | 6 028     | 2 876                                  | 1 898   | 257    | 17 210    |
| <i>ILVNC</i>   | m <sup>3</sup> /km/jour | 4,371                                  | 4,662   | 0,222  | 35,294    | 2,962                                  | 2,159   | 0,499  | 13,222    |
| <i>Longres</i> | Km                      | 40                                     | 74      | 4      | 500       | 143                                    | 160     | 9      | 890       |
| <i>RDT</i>     | %                       | 69,406                                 | 13,478  | 24,744 | 96,894    | 74,4                                   | 7,858   | 64,113 | 90,134    |

Notes : Moy = moyenne arithmétique, Std = écart-type.

❖ **Spécification du modèle : la fonction « transcendantal logarithmique »**

Comme dans beaucoup d'études antérieures sur les services d'eau, nous avons choisi de spécifier la fonction de coût selon la forme flexible « translog » pour ses bonnes propriétés. En particulier, l'avantage de cette forme fonctionnelle est qu'elle permet de décrire n'importe quelle technologie car elle n'impose a priori aucune restriction sur la structure de celle-ci.

C'est une approximation locale à la moyenne géométrique des variables (en log) de la base de données SISPEA<sup>10</sup> ; de ce fait, elle ne décrira la « vraie » technologie qu'au point d'approximation (ici la moyenne géométrique des variables de SISPEA) et à son voisinage.

Les estimations des coûts seront donc moins fiables pour des services « éloignés » de ce point d'approximation.

**L'approximation « translog » (Christensen *et al.*,1971) évaluée à la moyenne des variables de SISPEA s'écrit :**

Equation (7)

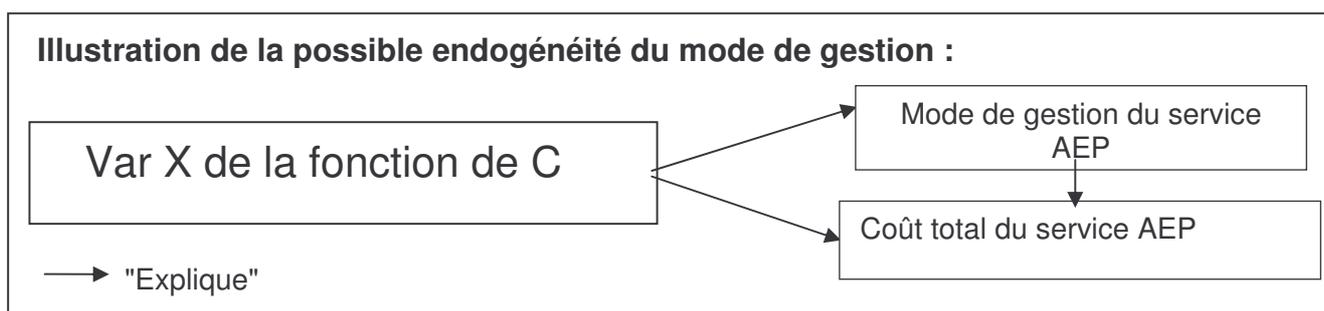
$$\begin{aligned}
 \ln(CT) = & \beta_0 + \beta_1 * (\ln Q - \overline{\ln Q}) + \beta_2 * (\ln AB - \overline{\ln AB}) + \beta_3 * (\ln ILVNC - \overline{\ln ILVNC}) \\
 & + \beta_4 * 0,5 * (\ln Q - \overline{\ln Q})^2 + \beta_5 * 0,5 * (\ln AB - \overline{\ln AB})^2 + \beta_6 * 0,5 * (\ln ILVNC - \overline{\ln ILVNC})^2 \\
 & + \beta_7 * (\ln Q - \overline{\ln Q}) * (\ln AB - \overline{\ln AB}) + \beta_8 * (\ln Q - \overline{\ln Q}) * (\ln ILVNC - \overline{\ln ILVNC}) \\
 & + \beta_9 * (\ln AB - \overline{\ln AB}) * (\ln ILVNC - \overline{\ln ILVNC})
 \end{aligned}$$

<sup>10</sup> Les coefficients sont estimés sur la base Aube-Gironde mais puisque notre base d'intérêt est la base SISPEA, il paraît plus judicieux de choisir les moyennes des variables de SISPEA comme point de référence.

❖ **Le modèle à régimes « *switching model* »**

La collectivité locale peut gérer elle-même le service d'eau ou confier son exploitation à une entreprise privée. Si elle souhaite déléguer le service AEP, son appel d'offres lui permettra, à partir des comptes d'exploitation prévisionnels des différentes entreprises candidates, de déterminer celle qui sera la plus à même de gérer son service AEP.

Le choix du mode de gestion du service se fait en fonction de variables qui déterminent les coûts du service (cf. équation 7) mais aussi de variables qui n'entrent pas dans la fonction de coût comme la longueur et le rendement du réseau<sup>11</sup> : le mode de gestion des services AEP peut donc s'avérer endogène et son inclusion (sous la forme d'une *dummy* 0 et 1 respectivement pour gestion en régie et gestion déléguée) dans le modèle de coût<sup>12</sup> pourrait biaiser les estimations des coefficients (estimés par MCO).



Pour remédier à ce problème, nous définissons un modèle à régimes (*switching model*) à la Lee (1978) sur la base des fonctions de coût de chaque mode de gestion et, d'une fonction critère (ou équation de sélection) décrivant le choix (programme sas)

Cette équation de sélection  $I^*$  (variable latente non observée) représente la prédisposition de la collectivité «  $i$  » à déléguer son service AEP :

$$\begin{aligned}
 I^* = & \gamma_0 + \gamma_1 * (\ln Q - \overline{\ln Q}) + \gamma_2 * (\ln AB - \overline{\ln AB}) + \gamma_3 * (\ln ILVNC - \overline{\ln ILVNC}) \\
 & + \gamma_4 * 0,5 * (\ln Q - \overline{\ln Q})^2 + \gamma_5 * 0,5 * (\ln AB - \overline{\ln AB})^2 + \gamma_6 * 0,5 * (\ln ILVNC - \overline{\ln ILVNC})^2 \\
 & + \gamma_7 * (\ln Q - \overline{\ln Q}) * (\ln AB - \overline{\ln AB}) + \gamma_8 * (\ln Q - \overline{\ln Q}) * (\ln ILVNC - \overline{\ln ILVNC}) \\
 & + \gamma_9 * (\ln AB - \overline{\ln AB}) * (\ln ILVNC - \overline{\ln ILVNC}) + \gamma_{10} * (\ln Longres - \overline{\ln Longres}) \\
 & + \gamma_{11} * (\ln RDT - \overline{\ln RDT}) + \mu
 \end{aligned}$$

On réécrit cette fonction critère sous forme matricielle :

$$I^* = \mathbf{Z} \cdot \boldsymbol{\gamma} + \mu \quad (8)$$

où  $\mathbf{Z}$  est la matrice des variables explicatives,  $\boldsymbol{\gamma}$  est la matrice des coefficients et  $\mu$ , la matrice de termes d'erreur avec  $\mu \sim N(0, \sigma_\mu^2)$

<sup>11</sup> En théorie la longueur du réseau et le rendement du réseau ont un impact sur les coûts totaux des services mais ces variables ne sont pas incluses dans la fonction de coût total (7).

<sup>12</sup> La variable « gestion AEP » est une caractéristique du service pouvant expliquer son coût total.

En normalisant  $\sigma_{\mu}^2 = 1$ , l'équation (8) a la forme d'un modèle Probit :

$$\left\{ \begin{array}{l} \text{Gestion AEP} = 1 \text{ si } \mathbf{I}^* > \mathbf{0} \text{ (la collectivité est disposée à déléguer son service)} \\ \text{Gestion AEP} = 0 \text{ si } \mathbf{I}^* \leq \mathbf{0} \text{ (la collectivité n'est pas disposée à déléguer son service)} \end{array} \right.$$

Si l'on note  $\Psi = \cdot \mathbf{Y}$ , la règle de décision probabiliste du modèle s'écrit :

$$\left\{ \begin{array}{l} \text{Prob (Gestion AEP} = 1) = \text{Prob}(\mathbf{I}^* > \mathbf{0}) = \text{Prob}(\Psi > -\mu) = \Phi(\Psi) \\ \text{Prob (Gestion AEP} = 0) = \text{Prob}(\mathbf{I}^* \leq \mathbf{0}) = \text{Prob}(\Psi \leq -\mu) = 1 - \Phi(\Psi) \end{array} \right.$$

Où  $\Phi$  est la fonction de distribution cumulative de la loi normale centrée réduite.

L'équation d'intérêt est, pour chaque mode de gestion (i=0 pour régie et i=1 pour délégation) :

$$\begin{aligned} \text{Ln}(CT)^i &= \beta_0 + \beta_1 * (\text{Ln}Q^i - \overline{\text{Ln}Q}) + \beta_2 * (\text{Ln}AB^i - \overline{\text{Ln}AB}) + \beta_3 * (\text{Ln}ILVNC^i - \overline{\text{Ln}ILVNC}) \\ &+ \beta_4 * 0,5 * (\text{Ln}Q^i - \text{Ln}Q)^2 + \beta_5 * 0,5 * (\text{Ln}AB^i - \text{Ln}AB)^2 + \beta_6 * 0,5 * (\text{Ln}ILVNC^i - \text{Ln}ILVNC)^2 \\ &+ \beta_7 * (\text{Ln}Q^i - \overline{\text{Ln}Q}) * (\text{Ln}AB^i - \overline{\text{Ln}AB}) + \beta_8 * (\text{Ln}Q^i - \overline{\text{Ln}Q}) * (\text{Ln}ILVNC^i - \overline{\text{Ln}ILVNC}) \\ &+ \beta_9 * (\text{Ln}AB^i - \overline{\text{Ln}AB}) * (\text{Ln}ILVNC^i - \overline{\text{Ln}ILVNC}) + \varepsilon \end{aligned}$$

On réécrit cette fonction d'intérêt sous forme matricielle :

$$\text{Ln}(C)^i = \mathbf{X}^i \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}^i \quad (9)$$

où  $\mathbf{X}$  est la matrice des variables explicatives,  $\boldsymbol{\beta}$  est la matrice des coefficients et  $\boldsymbol{\varepsilon}$ , la matrice des termes d'erreurs avec  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_{\varepsilon}^2)$ .

Nous allons alors utiliser une méthode d'estimation en 2 étapes développée par Heckman<sup>13</sup> en 1976.

Dans un premier temps, nous estimons par maximum de vraisemblance, le modèle probit (8) puis, dans un second temps, nous estimerons les 2 équations d'intérêts séparément en corrigeant l'éventuel biais de sélection dû au fait que l'on estime le coût sur une partie seulement de l'échantillon (pour les services en régie et pour les services délégués) dont la sélection n'est pas aléatoire<sup>14</sup>.

Cette correction du biais de sélection repose sur le calcul de l'inverse du ratio de Mills simulé sur la base du vecteur des caractéristiques du groupe supposé discriminé incluses dans la sélection considérée et du vecteur des paramètres issu de l'estimation de l'équation de

<sup>13</sup> James Heckman est le premier à avoir introduit le concept de biais de sélection en économétrie ; il fut récompensé du prix Nobel d'économie en 2000.

<sup>14</sup> La sélection est faite par les coefficients estimés du modèle probit ; si on ne corrige pas le biais de sélection, les coefficients estimés par MCO seront biaisés.

sélection pour l'ensemble de la population : l'hypothèse sous-jacente est que les régies ont des coûts souvent moindre par rapport à l'ensemble de la population des services AEP ( et inversement les services délégués ont des coûts souvent plus important que l'ensemble de la population).

Si aucun biais de sélection n'est avéré, on pourra alors inclure, si elle n'est pas endogène, la variable « gestion aep » sous forme d'une *dummy* (0 pour régie et 1 pour délégation) directement dans notre modèle MCO simple<sup>15</sup>.

Concrètement, le biais de sélection est détecté lorsque les termes d'erreurs  $\varepsilon^i$  et  $\mu$  des équations (9) et (8) sont corrélés c'est-à-dire lorsque le coefficient de corrélation  $\rho^i$ , entre  $\varepsilon^i$  et  $\mu$  est significativement différent de zéro.

Si tel est le cas, la covariance entre les termes d'erreurs  $\varepsilon^i$  et  $\mu$  des modèles (9) et (8) n'est pas nulle soit :

$$\rho^i = \frac{Cov(\varepsilon^i, \mu)}{\sigma_\varepsilon} \neq 0 \Leftrightarrow Cov(\varepsilon^i, \mu) \neq 0$$

$$\Leftrightarrow E(\varepsilon^i \cdot \mu) \neq 0$$

La correction de ce biais passe par l'inclusion<sup>16</sup>, dans le modèle (9) des services en délégation (resp. en régie) estimé par MCO, de la probabilité estimée par le modèle probit (8) que la gestion soit déléguée (resp. en régie). L'ajout des ratios de Mills inverses dans chaque équation de régime a pour but d'ajuster les termes d'erreur des équations de coût afin que leur moyenne soit égale à 0.

En effet, le biais de sélection provoque la violation d'une des hypothèses fondamentales de la régression par MCO à savoir que l'espérance conditionnelle des aléas pour chaque équation de régime devient non nulle :

- Si l'on considère les services AEP délégués ( $i=D$ ), nous avons :

$$E(\varepsilon^D \mid X^1, \mathbf{I}^* > 0) = E(\varepsilon^D \mid \mathbf{I}^* > 0) \quad (\text{car } X \text{ est exogène : } X^D \perp \varepsilon^D)$$

$$= E(\varepsilon^D \mid \mu > -Z \cdot \gamma) \quad (\neq 0 \text{ car } Z \text{ contient } X) \quad (10)$$

Avec :  $\rho^D = \frac{Cov(\varepsilon^D, \mu)}{\sigma_\varepsilon} \Leftrightarrow Cov(\varepsilon^D, \mu) = \rho^D \cdot \sigma_\varepsilon$

Et :  $Cov(\rho^D \cdot \sigma_\varepsilon \cdot \mu, \mu) = \rho^D \cdot \sigma_\varepsilon \cdot Cov(\mu, \mu) \quad \text{où } Cov(\mu, \mu) = 1$

On peut déduire :  $\varepsilon^D = \rho^D \cdot \sigma_\varepsilon \cdot \mu + \text{résidu (10')}$

<sup>15</sup> Si en revanche, il s'avérait que la variable « gestion aep » était endogène, il faudrait alors la remplacer par une matrice d'instrument corrélée à elle mais indépendante du terme d'erreur (exogène).

<sup>16</sup> Au travers des ratios de Mills inverses.

A partir de (10) et de (10'), on a :

$$E(\varepsilon^D \mid \mu > -Z \cdot \gamma) = E(\rho^D \cdot \sigma_\varepsilon \cdot \mu \mid \mu > -Z \cdot \gamma) \\ = \rho^D \cdot \sigma_\varepsilon \cdot E(\mu \mid \mu > -Z \cdot \gamma)$$

$$\text{Où } E(\mu \mid \mu > -Z \cdot \gamma) = \frac{\varphi(-Z\gamma)}{1-\Phi(-Z\gamma)} = \frac{\varphi(Z\gamma)}{\Phi(Z\gamma)} \quad (\text{car } \varphi(-Z\gamma) = \varphi(Z\gamma))$$

$$\text{On a alors : } E(\varepsilon^D \mid X^D, I^* > 0) = \rho^D \cdot \sigma_\varepsilon \cdot \frac{\varphi(Z\gamma)}{\Phi(Z\gamma)} \neq 0$$

La nouvelle variable  $\frac{\varphi(Z\hat{\gamma})}{\Phi(Z\hat{\gamma})}$  appelée inverse du ratio de Mills, est estimée par le probit puis réinjectée dans l'équation de coût des services délégués afin d'annuler l'espérance du terme d'erreur.

On considérera donc l'équation de coût suivante (pour les services délégués) :

$$\ln(CT)^D = X^D \cdot \beta + \rho^D \cdot \sigma_\varepsilon \frac{\varphi(Z\hat{\gamma})}{\Phi(Z\hat{\gamma})} + \eta^D \quad (\eta^D \sim N(0, \sigma_\eta^2)) \quad (11)$$

$$\text{où : } E(\eta^D \mid \text{Gestion AEP} = 1) = E(\ln(CT)^D) - E(X^D \cdot \beta) - E(\rho^D \cdot \sigma_\varepsilon \cdot \frac{\varphi(Z\hat{\gamma})}{\Phi(Z\hat{\gamma})}) \\ = X^D \cdot \beta + E(\rho^D \cdot \sigma_\varepsilon \cdot \frac{\varphi(Z\hat{\gamma})}{\Phi(Z\hat{\gamma})}) - X^D \cdot \beta - E(\rho^D \cdot \sigma_\varepsilon \cdot \frac{\varphi(Z\hat{\gamma})}{\Phi(Z\hat{\gamma})}) \\ = 0$$

- Si l'on considère les services AEP gérés en régie (i=R), un calcul similaire montre que :

$$E(\varepsilon^R \mid X^R, I^* \leq 0) = \rho^R \cdot \sigma_\varepsilon \cdot \frac{-\varphi(Z\gamma)}{1-\Phi(Z\gamma)} \neq 0$$

La nouvelle variable  $\frac{-\varphi(Z\hat{\gamma})}{1-\Phi(Z\hat{\gamma})}$  appelée inverse du ratio de Mills est estimée par le probit puis réinjectée dans l'équation du coût des services délégués afin d'annuler l'espérance du terme d'erreur.

On considérera donc l'équation de coût suivante (pour les services en régie) :

$$\ln(CT)^R = X^R \cdot \beta - \rho^R \cdot \sigma_\varepsilon \frac{\varphi(Z\hat{\gamma})}{1-\Phi(Z\hat{\gamma})} + \eta^R \quad (\eta^R \sim N(0, \sigma_\eta^2)) \quad (12)$$

$$\text{où : } E(\eta^R \mid \text{Gestion AEP} = 0) = 0$$

Par construction, la correction du biais de sélection provoque l'hétéroscédasticité des termes d'erreurs  $\eta^i$  (la variance de  $\eta^i$  varie pour tous les services) qui devra donc être corrigée par une matrice de variance-covariance robuste afin de rendre l'estimation MCO efficace<sup>17</sup> et sans biais.

La première étape de la méthode d'Heckman nous donne l'estimation du probit :

#### Estimation du modèle probit (8) :

|                          | Estimation | Ecart-Type |
|--------------------------|------------|------------|
| <b>Constante</b>         | 2,0757***  | 0,2617     |
| <b>Q</b>                 | -7,2799*** | 1,6592     |
| <b>AB</b>                | 5,3801***  | 1,4071     |
| <b>ILVNC</b>             | 2,3394***  | 0,8332     |
| <b>Q<sup>2</sup></b>     | -0,9959    | 4,3834     |
| <b>AB<sup>2</sup></b>    | 1,8985     | 4,7814     |
| <b>ILVNC<sup>2</sup></b> | 0,6036     | 0,3979     |
| <b>Q * AB</b>            | -0,5945    | 4,5257     |
| <b>Q * ILVNC</b>         | 1,0287     | 0,8978     |
| <b>AB * ILVNC</b>        | -1,1741    | 0,9679     |
| <b>RDT</b>               | 8,4494***  | 2,3971     |
| <b>Longres</b>           | 3,1293***  | 0,8971     |

Notes :

- Par soucis de clarté  $(\ln Q - \overline{\ln Q})$  est représenté par « Q »...
- Niveau de significativité : \*\*\* : 1% \*\* : 5% \* : 10%

Le modèle probit (8) du choix de mode de gestion est estimé de manière très satisfaisante avec un pseudo R<sup>2</sup> de Mac Fadden égal à 0,5095 et 83,72% des observations correctement prédites comme le montre le tableau suivant :

<sup>17</sup> Autrement dit, les écarts-types des coefficients estimés ne seront plus biaisés.

| FREQUENCE<br>Pourcentage<br>Pct en ligne<br>Pct en col. |              |              | Total         |
|---|--------------|--------------|---------------|
|   | 0            | 1            |               |
| 0   | 124          | 32           | 156           |
|   | 36.05        | 9.30         | 45.35         |
|   | 79.49        | 20.51        |               |
|   | 83.78        | 16.33        |               |
| 1   | 24           | 164          | 188           |
|   | 6.98         | 47.67        | 54.65         |
|   | 12.77        | 87.23        |               |
|   | 16.22        | 83.67        |               |
| <b>Total</b>  | <b>148</b>   | <b>196</b>   | <b>344</b>    |
|   | <b>43.02</b> | <b>56.98</b> | <b>100.00</b> |

Il est difficilement possible d'interpréter « l'impact » des variables explicatives sur la probabilité que le service soit en délégation (délégation=1) puisque celles-ci sont contenues également dans d'autres variables du modèle : ainsi pour interpréter l'influence du volume mis en distribution sur la probabilité de la gestion déléguée du service d'eau par exemple, il faudrait isoler les « Q » et voir le signe du coefficient associé mais, non seulement la significativité du coefficient changerait et de plus, il est impossible d'isoler tous les « Q » :

On le constate en reprenant uniquement les variables contenant « Q » et leurs coefficients associés<sup>18</sup> :

$$\begin{aligned}
 & -5,8781 * (\ln Q - \overline{\ln Q}) + 2,6981 * \left( \ln \left( \frac{Q - Q_c}{365} \right) - \overline{\ln ILVNC} \right) - 0,9959 * 0,5 * (\ln Q - \overline{\ln Q})^2 \\
 & + 0,6036 * 0,5 * \left( \ln \left( \frac{Q - Q_c}{365} \right) - \overline{\ln ILVNC} \right)^2 - 0,5945 * (\ln Q - \overline{\ln Q}) * (\ln AB - \overline{\ln AB}) \\
 & + 1,0287 * (\ln Q - \overline{\ln Q}) * \left( \ln \left( \frac{Q - Q_c}{365} \right) - \overline{\ln ILVNC} \right) - 1,1741 * (\ln AB - \overline{\ln AB}) * \left( \ln \left( \frac{Q - Q_c}{365} \right) - \overline{\ln ILVNC} \right) \\
 & + 8,4494 * \left( \ln \left( \frac{Q}{Q_c} \right) - \overline{\ln RDT} \right)
 \end{aligned}$$

Si l'on avait voulu interpréter les signes des coefficients, il faudrait changer les variables explicatives du modèle probit en s'appuyant sur un modèle théorique et en évitant la redondance entre les variables mais ce n'est pas l'objectif de ce travail.

La deuxième étape de la méthode d'Heckman estime les 2 équations de coûts suivantes sur la base de la sélection faite par l'estimation du probit de la 1<sup>ère</sup> étape :

<sup>18</sup> Q<sub>c</sub> représente le volume d'eau consommé (facturé).

Estimation du modèle de coût sur le sous-échantillon des services délégués (11) :

|                                 | Estimation | Ecart-type |
|---------------------------------|------------|------------|
| <b>Constante</b>                | 13,9315*** | 0,0205     |
| Q                               | 0,2676**   | 0,1058     |
| AB                              | 0,6705***  | 0,1110     |
| ILVNC                           | 0,0044     | 0,0278     |
| Q <sup>2</sup>                  | 1,6528**   | 0,8379     |
| AB <sup>2</sup>                 | 1,1211     | 0,8004     |
| ILVNC <sup>2</sup>              | 0,1811***  | 0,0540     |
| Q * AB                          | -1,4109*   | 0,8124     |
| Q * ILVNC                       | -0,4742*** | 0,1681     |
| AB * ILVNC                      | -0,5077*** | 0,1658     |
| Cov ( $\varepsilon^D$ , $\mu$ ) | 0,1874***  | 0,0171     |

Notes :

- Par soucis de clarté  $(\ln Q - \overline{\ln Q})$  est représenté par « Q »...
- Niveau de significativité : \*\*\* : 1% \*\* : 5% \* : 10%

L'ajustement du modèle est très bon puisque le R<sup>2</sup> ajusté est de 0,9434 (94,34% de la variance du coût est expliqué par le modèle) : par conséquent le modèle donne des bonnes prédictions de coûts pour les services en délégation.

Par ailleurs, le biais de sélection est avéré puisque le coefficient associé au ratio de Mills inverse est significatif à 1 % : par conséquent il faudra bien inclure le ratio de Mills inverse sous forme de variable supplémentaire dans la régression.

Estimation du modèle de coût sur le sous-échantillon des services en régie (12) :

|   | Estimation | Ecart-type |
|---|------------|------------|
| <i>Constante</i>  | 13,3751*** | 0,1225     |
| <i>Q</i>  | 0,6389**   | 0,2583     |
| <i>AB</i>   | 0,3327     | 0,3274     |
| <i>ILVNC</i>  | -0,2001*** | 0,0455     |
| <i>Q<sup>2</sup></i>  | 0,0874     | 0,6249     |
| <i>AB<sup>2</sup></i>                                       | 0,1865     | 0,7055     |
| <i>ILVNC<sup>2</sup></i>                                    | -0,0575    | 0,0588     |
| <i>Q * AB</i>   | 0,1209     | 0,6357     |
| <i>Q * ILVNC</i>  | 0,0777     | 0,1583     |
| <i>AB * ILVNC</i>   | -0,1372    | 0,1584     |
| <i>Cov ( <math>\varepsilon^R</math>, <math>\mu</math> )</i> | 0,0167     | 0,1243     |

Notes :

- Par soucis de clarté  $(\ln Q - \overline{\ln Q})$  est représenté par « Q »...
- Niveau de significativité : \*\*\* : 1% \*\* : 5% \* : 10%

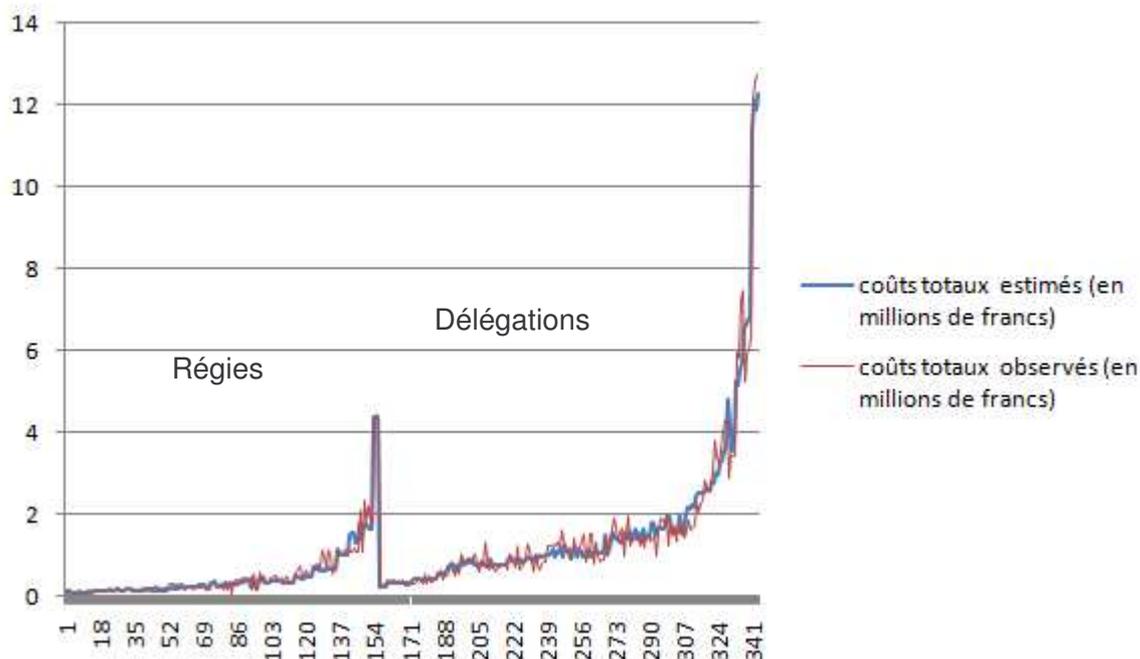
L'ajustement du modèle est très bon puisque le R<sup>2</sup> ajusté est de 0,9134 : par conséquent le modèle donne des bonnes prédictions de coûts pour les services en régie.

Le coefficient associé au ratio de Mills inverse est non significatif mais il conviendra tout de même de l'inclure dans la régression.

L'interprétation des coefficients associés aux variables explicatives pour les deux modes de gestion est impossible (comme pour le probit estimé précédemment) mais l'objectif de ce travail préalable était d'avoir un bon ajustement du modèle aux données de Aube – Gironde afin de pouvoir prédire au mieux (via les coefficients estimés sur AG) les coûts de la base d'intérêt SISPEA.

**Nous reportons sur la figure suivante, les coûts observés et nos estimations pour avoir une vue d'ensemble de l'excellent ajustement de notre modèle de coût aux données<sup>19</sup> :**

<sup>19</sup> L'axe des ordonnées représente les coûts totaux estimés et observés des services AEP ; l'axe des abscisses représente les services AEP (de 1 à 345) triés par leur mode de gestion (de 1 à 154 : régie



*N.B. L'abscisse traduit le numéro de rang des services de la base, classés par mode de gestion et par taille (abonnés)*

**Figure 2** Comparaison entre les coûts observés et les estimations du modèle de coût construit sur la base AG

### 2.2.4. ...à la base SISPEA

Après suppression des services AEP ayant des données aberrantes, manquantes ou influentes pour les variables que l'on va utiliser dans la prédiction de leurs taux de recouvrement<sup>20</sup>, nous retenons une base de données SISPEA retravaillée contenant 184 services AEP dont les statistiques descriptives des variables utilisées pour les deux types de gestion figurent dans le tableau suivant :

---

/ de 154 à 345 : délégation) et par leurs nombres d'abonnés (on constate de ce fait une tendance à ce que les coûts totaux estimés des services augmentent avec le nombre d'abonnés).

<sup>20</sup> On a supprimé 35 services : 4 à cause des recettes, 24 à cause des volumes mis en distribution, 4 à cause de l'indice linéaire des volumes non comptés, 2 à cause du mode de gestion et 1 qui était en double dans la base.

**Statistiques descriptives de l'échantillon SISPEA (retravaillé)**

|                       |                         | Gestion publique<br>(38 observations) |         |        |           | Gestion déléguée<br>(146 observations) |         |       |           |
|-----------------------|-------------------------|---------------------------------------|---------|--------|-----------|--|---------|-------|-----------|
| Variable              | Unité                   | Moy                                   | Std     | Min    | Max       | Moy                                    | Std     | Min   | Max       |
| <i>Recette totale</i> | Milliers d'Euros        | 467                                   | 514     | 16     | 2 341     | 663                                    | 738     | 2     | 3 886     |
| <i>Q</i>              | m <sup>3</sup>          | 613 678                               | 894 466 | 35 000 | 4 514 000 | 708 845                                | 847 229 | 8 475 | 4 731 000 |
| <i>AB</i>             |                         | 2 672                                 | 2 212   | 167    | 9 477     | 3 833                                  | 3 892   | 67    | 21 928    |
| <i>ILVNC</i>          | m <sup>3</sup> /km/jour | 4,703                                 | 9,366   | 0,453  | 53,18     | 3,413                                  | 4,8     | 0,257 | 36,738    |
| <i>Longres</i>        | Km                      | 178                                   | 195     | 10     | 981       | 220                                    | 229     | 2     | 1357      |
| <i>RDT</i>            | %                       | 76,4                                  | 8,8     | 44     | 89,2      | 77,9                                   | 9,7     | 51,5  | 98,7      |

Notes : Moy = moyenne arithmétique, Std = écart-type.

Les services en régie sont dans l'ensemble plus petits puisqu'ils desservent moins d'abonnés, distribuent moins de m<sup>3</sup> d'eau potable et présentent une longueur totale de canalisations moindre. Les valeurs moyennes de l'indice linéaire des volumes non comptés et du rendement du réseau de distribution indiquent un état du réseau meilleur pour les services en gestion déléguée.

Afin de prédire les coûts des services AEP de SISPEA et comme expliqué dans la section 3.2.1, nous reportons le vecteur des coefficients estimés des 2 équations d'intérêts d'AG (régie et délégation) que l'on associe aux variables de SISPEA<sup>21</sup>.

La spécification adoptée pour prédire les coûts de SISPEA est donc la fonction « translog » (équation 7) évaluée à la moyenne des variables de SISPEA en y ajoutant la variable supplémentaire correspondant aux ratios de Mills pour les deux modes de gestion.

<sup>21</sup> Les variables utilisées dans SISPEA sont bien entendu les mêmes que celles d'AG.

L'équation suivante résume les prédictions des coûts comptables des services de SISPEA pour les services en régie (38 services) :

$$\underbrace{\begin{pmatrix} 1 & Q^1 & \dots & ILVNC^1 * AB^1 & \frac{\varphi(Z^1 \gamma)}{1-\varphi(Z^1 \gamma)} \\ 1 & Q^2 & \dots & ILVNC^2 * AB^2 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Q^{38} & \dots & ILVNC^{38} * AB^{38} & \frac{\varphi(Z^{38} \gamma)}{1-\varphi(Z^{38} \gamma)} \end{pmatrix}}_{(A)} * \underbrace{\begin{pmatrix} 14,1563 \\ 0,5856 \\ \vdots \\ -0,1372 \\ -0,0167 \end{pmatrix}}_{(B)} = \underbrace{\begin{pmatrix} (LnCP)^1 \\ (LnCP)^2 \\ \vdots \\ \vdots \\ (LnCP)^{38} \end{pmatrix}}_{(C)}$$

Notes :

La matrice (A) représente les variables explicatives des coûts de SISPEA pour l'ensemble des services en régie (y compris le ratio de Mills inverse pour cet échantillon)

- $Q^i = ((LnQ)^i - \overline{Ln(Q)})$  où i représente le i-ème service en régie de SISPEA.

Il en va de même pour les autres variables.

- $Z^i$  représente le vecteur des variables explicatives du modèle probit (8) et  $\gamma$  le vecteur des coefficients associés :

$$Z^i \gamma = (1 \quad Q^i \quad \dots \quad Longres^i) * \begin{pmatrix} 2,0757 \\ -7,2799 \\ \vdots \\ 3,1293 \end{pmatrix}$$

- où i représente le i-ème service.

La matrice (B) représente les coefficients estimés de l'équation d'intérêt des services en régie.

La matrice (C) = (A)\*(B) nous donne les logarithmes des coûts prédits pour chaque service en régie de SISPEA.

De manière analogue, l'équation suivante résume les prédictions des coûts des services de SISPEA pour les services en délégation (146 services) :

$$\begin{pmatrix} 1 & Q^1 & \dots & ILVNC^1 * AB^1 & \frac{\varphi(Z^1, \gamma)}{\Phi(Z^1, \gamma)} \\ 1 & Q^2 & \dots & ILVNC^2 * AB^2 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & Q^{146} & \dots & ILVNC^{146} * AB^{146} & \frac{\varphi(Z^{146}, \gamma)}{\Phi(Z^{146}, \gamma)} \end{pmatrix} * \begin{pmatrix} 13,9315 \\ 0,2676 \\ \vdots \\ -0,5077 \\ 0,1874 \end{pmatrix} = \begin{pmatrix} (LnCP)^1 \\ (LnCP)^2 \\ \vdots \\ (LnCP)^{146} \end{pmatrix}$$

Puisque l'on dispose des recettes observées (en euros) et des coûts comptables prédits pour chaque service de SISPEA<sup>22</sup>, on peut en déduire un taux de recouvrement des coûts comptables pour chaque service, un taux moyen arithmétique de recouvrement des coûts<sup>23</sup>, un taux moyen pondéré<sup>24</sup> ou un taux moyen géométrique<sup>25</sup>.

**Il est pertinent de retenir ici la moyenne géométrique pour obtenir un taux global de recouvrement** : La moyenne géométrique est utilisée lorsque l'on considère des nombres qui se multiplient plus qu'ils ne s'additionnent. C'est le cas de tous les taux : taux d'intérêt, taux d'inflation, taux d'accroissement...

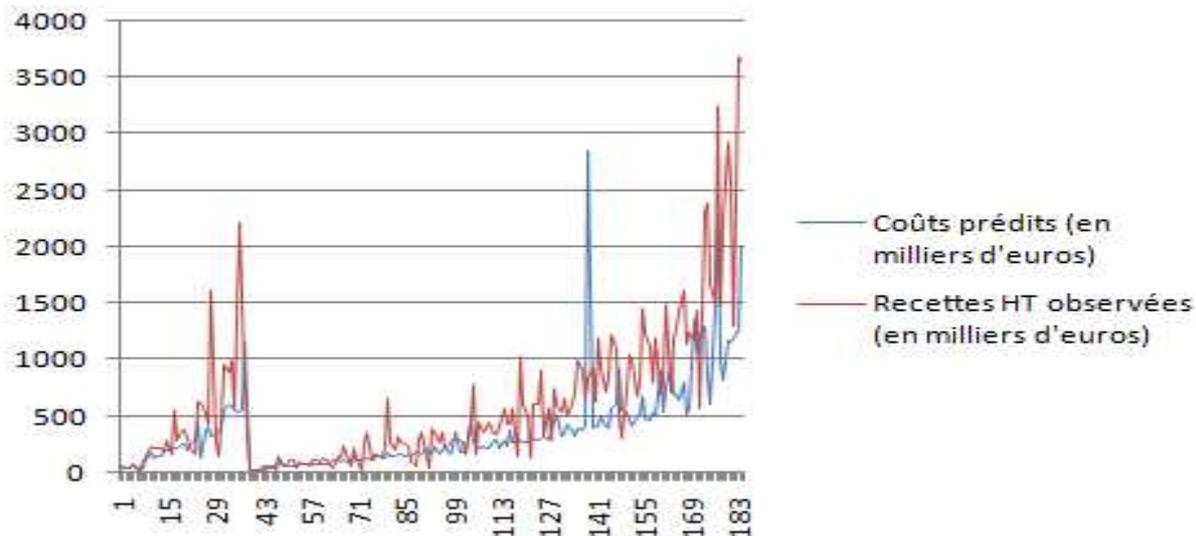
Les taux de recouvrement des coûts comptables pour chaque service AEP de SISPEA (triés par mode de gestion et par nombre d'abonnés) sont retranscrits par les 2 graphiques suivants :

<sup>22</sup> Les coûts prédits sont convertis en euros et tiennent compte de l'inflation entre 1997 (date « moyenne » des services AEP de AG) et 2008 (date d'observation des services de SISPEA).

<sup>23</sup> On ne tient pas compte de la pondération : c'est la somme des taux de recouvrement de chaque service divisée par le nombre de services.

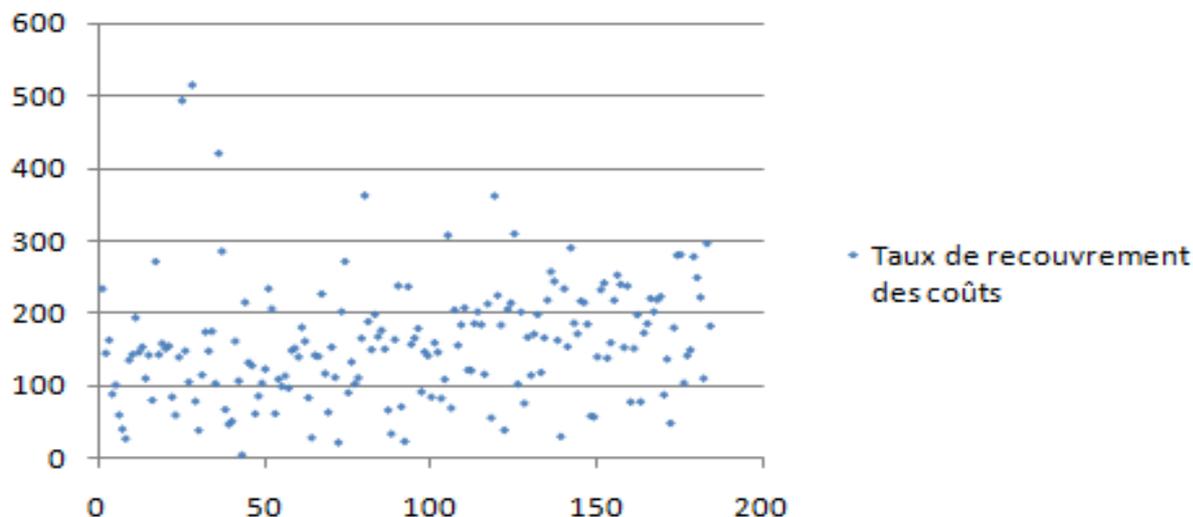
<sup>24</sup> Ici, on tient compte de la pondération : c'est la somme de toutes les recettes des services divisée par la somme de tous les coûts prédits des services.

<sup>25</sup> La moyenne géométrique est la racine nième du produit de n nombres ou la moyenne arithmétique des nombres transformés en logarithme.



*N.B L'abscisse traduit le numéro de rang des services de la base, classés par mod de gestion et par taille (abonnés).*

**Figure 3** Comparaison entre les coûts prédits (méthode 1) et les recettes HT observées sur la base SISPEA



**Figure 4** Taux de recouvrement calculé par différence entre les coûts prédits (méthode 1) et les recettes HT observées sur la base SISPEA

Le premier graphique montre que, pour la plupart des services, les recettes HT sont supérieures aux coûts prédits : la plupart des services de la base SISPEA paraissent donc recouvrer leurs coûts (le deuxième graphique le confirme). Plus précisément, seulement 10 services en régie et 31 services en délégation ne paraissent pas recouvrer leurs coûts comptables.

Les résultats des statistiques descriptives des taux de recouvrement pour les deux modes de gestion sont résumés dans le tableau suivant :

| Statistiques<br><b>Taux de Recouvrement</b> | Moyenne géométrique <sup>26</sup> | Minimum | Maximum |
|---|-----------------------------------|---------|---------|
| Pour les services en régie                  | 129,00%                           | 26%     | 514%    |
| Pour les services en délégation             | 136,00%                           | 4%      | 362%    |

**Lorsque l'on considère l'ensemble de l'échantillon, le taux de couverture des coûts global est d'environ 135%, résultat significativement supérieur à celui calculé par le rapport BIPE (95% pour les services d'eau).**

**Ces résultats contre-intuitifs doivent donc être discutés car ils paraissent aberrants.**

### 2.2.5. Intérêt théorique de la méthode mais limite des résultats liés au flou dans la définition des données

La justesse de ce taux est sujette à caution pour plusieurs raisons :

- La fonction « translog » utilisée est une approximation locale (série de Taylor de second ordre) : les estimations sont fiables à la moyenne de l'échantillon SISPEA et à son voisinage. Par conséquent, les services qui s'en éloignent ont des coûts prédits moins précis.
- Comme on l'a signalé, la représentativité de la base AG sur laquelle on estime les coefficients peut être sujette à caution : les coefficients sont estimés à partir des coûts observés sur 2 départements (Aube et Gironde) et reportés dans la base SISPEA qui regroupe un ensemble hétérogène de services AEP français. De plus, il y a un décalage d'environ 10 ans avec la base SISPEA. Même si l'on a tenu compte de cela en appliquant un coefficient d'inflation aux coûts prédits, il se peut que celui-ci soit approximatif.
- Un biais significatif vient aussi certainement du fait que la base AG modélise les coûts d'exploitation, hors investissement.
- Le plus gros problème selon nous tient au fait que nous ne savons pas exactement comment la variable « recette » a été renseignée. Il y a peu de doute sur le fait que cette donnée n'a pas été calculée de la même façon par tous les services. Ces recettes comprennent-ils uniquement les revenus de la facture d'eau (prix au m<sup>3</sup> consommé et abonnement) ? Ou bien incluent-elles aussi d'autres revenus provenant des travaux, voire dans certains cas les comptes de tiers (subvention des agences de l'eau ?) Nous ne sommes pas certains que les valeurs données des recettes soient

<sup>26</sup> La moyenne géométrique est la racine nième du produit de n nombres ou la moyenne arithmétique des nombres transformés en logarithme. La moyenne géométrique est utilisée lorsque l'on considère des nombres qui se multiplient plus qu'ils ne s'additionnent. C'est le cas de tous les taux : taux d'intérêt, taux d'inflation, taux d'accroissement...

toutes calculées hors taxes, ni même si les différentes subventions ne sont pas parfois comptabilisées dans les recettes.

- Par ailleurs, les fonctions de coût variable sont estimées à partir des coûts d'exploitation (et administratifs) et n'incluent pas les coûts de capital (ou dépenses d'investissement). Il est bien entendu primordial que soient associées aux recettes les dépenses qui leur correspondent.

En conclusion, cette méthode économétrique basée sur l'estimation de fonctions de coût à partir de données comptables (M49) et financière (CRF/CARE) peut se révéler un outil puissant et très informatif à condition que les bases de données à partir desquelles sont effectuées les estimations reposent sur des informations riches, précises et cohérentes pour la totalité des services étudiés.

### **Conclusion sur l'estimation de la population en sous-recouvrement en construisant une fonction de coût basée à partir d'une base de données des coûts comptables et financiers de services externes à SISPEA**

La première méthode testée visait à s'appuyer sur des données de coûts comptables et financières (M49 et CRF) pour établir une fonction de coûts de référence à partir de variables techniques et de performance présentes également dans SISPEA.

Si la démarche est techniquement faisable, comme on l'a montré en utilisant la base AG, elle s'avère difficile à exploiter tant qu'on n'a pas une maîtrise complète de la définition et du contenu :

- d'une part sur les informations de coûts (fiabilité de la source des données ? inclusion des coûts du délégataire et/ou de la collectivité ? inclusion des coûts d'exploitation et/ou d'investissement ?)
- d'autre part sur les informations de recettes (HT ou TTC, avec ou sans les comptes de tiers, avec ou sans les produits autres que la facturation ?).

**Ceci conduit à deux recommandations pour que cette première approche puisse être consolidée :**

- idéalement introduire dans SISPEA des données sur les coûts comptables et financiers des services. Cette optique n'étant pas d'actualité, car les données comptables ne font pas partie des informations du décret et de l'arrêté du 2 mai 2007, **la question se pose d'une étude ad hoc permettant de valider un modèle de coût utilisable ensuite comme référence pendant plusieurs années.**

- **préciser et enrichir les informations concernant les recettes suivies dans SISPEA.** Il conviendrait de clairement définir et détailler les recettes, ainsi que la source d'information à utiliser (concrètement rédiger une fiche du type de celle existant pour les indicateurs de performance). **Le décret du 2 mai prévoit aussi de préciser le montant des subventions du budget général, information non reprise par SISPEA mais qui si elle l'était contribuerait à éclairer la question de la couverture des coûts.**

## 2.3. Deuxième méthode d'estimation des coûts des services AEP

La seconde méthode tente d'isoler les services en situation de non recouvrement en utilisant uniquement les données de la base SISPEA, couplées à une information macro-économique sur le taux de recouvrement moyen des services en France.

### 2.3.1. Présentation générale de la méthode

On va estimer les coûts des services AEP de SISPEA à partir de leurs recettes et en tenant compte d'un taux de recouvrement global de 95% calculé par le rapport BIPE<sup>27</sup>.

En effet le BIPE nous donne la relation suivante (avec z représentant le nombre total des services AEP en France) :

$$\frac{\sum_{i=1}^z \text{Recette}^i}{\sum_{i=1}^z \text{Coût Comptable}^i} \simeq 95\% \quad (13)$$

$$(13) \Rightarrow \sum_{i=1}^z \text{Coût Comptable}^i \simeq \frac{\sum_{i=1}^z \text{Recette}^i}{0,95}$$

Nous allons donc régresser les recettes réellement observées sur les caractéristiques des services de SISPEA. Les recettes estimées pour chaque service AEP de SISPEA seront alors divisées par 0,95 afin d'obtenir les coûts comptables estimés de ces mêmes services.

Pour un service « i »<sup>28</sup>:

$$\text{Coût Comptable estimé}^i = \frac{\text{Recette estimée}^i}{0,95}$$

On divisera alors la recette réellement observée pour ce service « i » par son coût estimé afin d'avoir son taux de recouvrement des coûts :

$$\text{Taux de recouvrement}^i = \frac{\text{Recette observée}^i}{\text{Coût Comptable estimé}^i}$$

Bien évidemment, le taux moyen de recouvrement ainsi obtenu sera par construction proche des 95%.

<sup>27</sup> Avec prise en compte d'une partie des coûts environnementaux internalisés (redevances des agences de l'eau) et en considérant un budget équilibré sur l'année 2006, le taux de recouvrement des coûts est de 97% mais si l'on se ramène aux coûts strictement comptables ce taux de recouvrement est de 95%.

<sup>28</sup> On considère un taux moyen et un taux moyen pondéré de recouvrement des coûts environ égaux.

En effet :

$$\text{Recette observée} = X \cdot \beta + \varepsilon \quad \text{où } \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

d'où :

$$\text{Recette estimée} = X \cdot \hat{\beta}$$

Par conséquent pour un service « i » :

$$\text{Taux de recouvrement}^i = \frac{X^i \cdot \beta + \varepsilon}{\left( \frac{X^i \cdot \hat{\beta}}{0,95} \right)} \quad (14)$$

Et donc pour l'ensemble de l'échantillon SISPEA :

$$\begin{aligned} E(\text{Taux de recouvrement}) &= 0,95 * E\left(\frac{X \cdot \beta + \varepsilon}{X \cdot \hat{\beta}}\right) \\ &= 0,95 * \left( E\left(\frac{X \cdot \beta}{X \cdot \hat{\beta}}\right) + E\left(\frac{\varepsilon}{X \cdot \hat{\beta}}\right) \right) \simeq 0,95 \end{aligned}$$

Cette méthode nous permettra d'identifier, pour chaque mode de gestion, la population de services qui paraissent recouvrir leurs coûts et ceux qui ne les recouvrent pas : on pourra alors en théorie comparer avec les résultats de la 1<sup>ère</sup> méthode<sup>29</sup>.

En pratique la comparaison devra être maniée avec quelques précautions, compte tenu des réserves sur la cohérence des données de coûts utilisées dans la première méthode (cf. §2.2.5).

### 2.3.2. Théorie Néoclassique du producteur : Maximisation du profit des services AEP

Comme relaté dans la section 2.2.2, et d'après la théorie de la dualité, nous pouvons modéliser la technologie des services AEP par une maximisation du profit de l'exploitant.

Max  $\sum (p \cdot Q)$  sous contrainte  $F(Q, Iq, x, CAR, \varepsilon) = 0$  où  $p$  désigne le prix de l'output (ici l'eau) et  $Q$ , la quantité d'eau mis en distribution.

Ainsi, de manière analogue à la fonction de coût total comptable, la fonction de recette totale adoptée pour le modèle dépendra de la quantité d'eau mis en distribution  $Q$ , de  $ILVNC$  et du nombre d'abonnés  $AB$  :

$$RT(Q, ILVNC, AB)$$

### 2.3.3. Estimation des recettes par « switching model »

Comme pour les coûts totaux, nous adoptons, une fonction de recette totale « translog » évaluée à la moyenne de SISPEA.

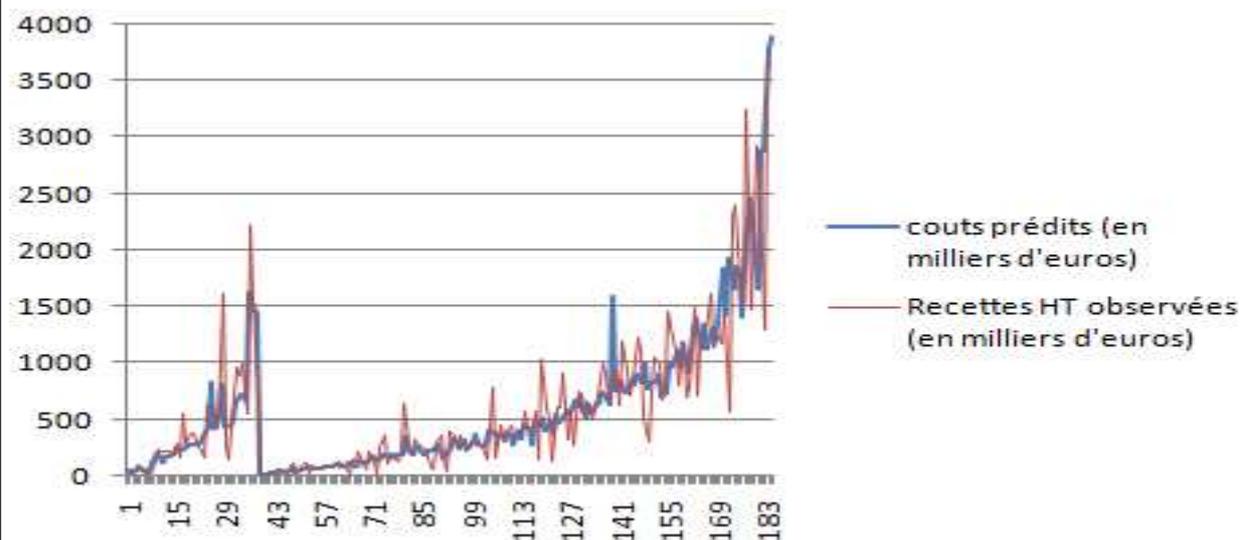
L'estimation se fera également en deux étapes à la Heckman, où notre équation de sélection représentant le mode de gestion (0 pour régie et 1 pour délégation) sera estimée par un modèle probit ; les ratios de Mills inverses corrigeant le biais de sélection seront alors

<sup>29</sup> 10/38 services en régie et 31/146 services en délégation ne recouvrent pas leurs coûts comptables.

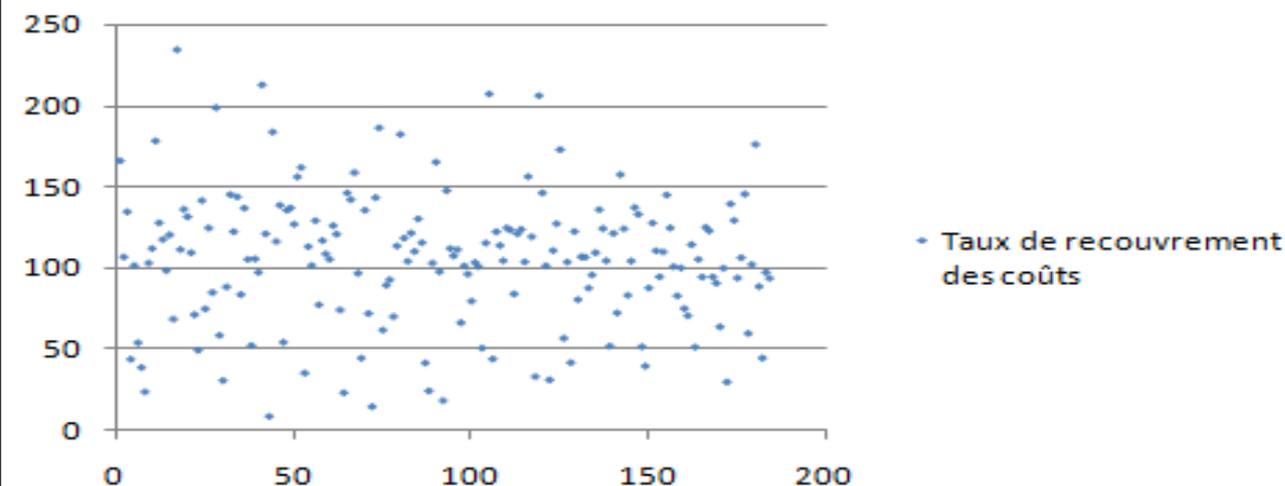
réinjectées dans nos 2 équations d'intérêts représentant les recettes pour chaque mode de gestion, estimées par MCO.

Les coefficients estimés du modèle probit ainsi que les coefficients estimés par MCO de nos 2 équations d'intérêts sont répertoriés dans l'ANNEXE II.

Les principaux résultats sont relatés pas les figures suivantes :



**Figure 5** Comparaison entre les coûts prédits (méthode 2) et les recettes HT observées sur la base SISPEA



**Figure 6** Taux de recouvrement calculé par différence entre les coûts prédits (méthode 2) et les recettes HT observées sur la base SISPEA

On constate par le premier graphique, que **l'allure de la courbe des coûts prédits par la deuxième méthode est relativement similaire à celle des coûts prédits par la première méthode.**

Le deuxième graphique montre que, comme pour la première méthode d'estimation des taux de recouvrement, la plupart des services de la base SISPEA semblent recouvrir leurs coûts comptables.

En particulier, seulement 15/38 services en régie et 55/146 services en délégation ne semblent pas recouvrir pas leurs coûts comptables. Ce constat est assez proche de celui fait par la première méthode d'estimation des taux de recouvrement, surtout pour les régies.

De plus, on a seulement 35 services sur 184 dont les résultats en termes de couverture ou non des coûts ne convergent pas entre les deux méthodes.

#### **2.3.4. Limites et extension de la deuxième méthode**

Cette méthode repose sur une technique de régression des recettes par rapport à différentes variables explicatives, et sur l'hypothèse que certains services ne recouvrent pas exactement la totalité des coûts (la différence entre recettes et coûts est supérieure ou inférieure à zéro). L'écart entre la recette estimée et la recette observée peut donc s'expliquer par cette différence. Par ailleurs, de façon générale, la différence entre les valeurs observées d'une variable et ses valeurs prédites est aussi liée à l'erreur d'estimation due à des imprécisions dans les données ou plus généralement parce qu'il n'est pas possible pour l'économètre d'inclure tous les facteurs qui déterminent les variations d'une variable observée.

**Le terme d'erreur de la régression comprend donc à la fois l'erreur classique économétrique et une partie expliquant la différence entre recette et coût.** L'espérance de ce terme d'erreur total est donc nulle. En conséquence, il n'est pas possible d'estimer un taux moyen (global) de recouvrement des coûts non nul. En revanche, pour chaque service, les résidus (ou erreurs estimées) issus de la différence entre observations et prédictions sont différents de zéro. **Il n'est pas possible d'identifier quelle est la part d'erreur due à l'erreur économétrique et celle liée au non recouvrement des coûts.**

**Une solution envisagée dans nos recherches futures est d'utiliser la technique des frontières stochastiques. Cela consisterait à définir des lois de distribution différentes pour chacun des deux termes d'erreur afin d'identifier précisément la part due à la différence entre recettes et coûts.**

L'idée serait de partir de l'hypothèse qu'un service parmi l'ensemble des services observés peut évidemment avoir des recettes inférieures aux coûts engendrés par l'exploitation, mais pas inférieures à un certain seuil (que nous fixerons en fonction des chiffres donnés par les études nationales faites sur le recouvrement des coûts), ce qui donne un taux de recouvrement (inférieur à un) minimal. Tous les autres services de l'échantillon ont un taux de recouvrement supérieur à celui de ce service, toutes choses égales par ailleurs. Le taux de recouvrement suit donc une distribution (normale) tronquée positive différente de la distribution normale classique suivie par le terme d'erreur « économétrique ». Il est alors possible d'estimer séparément le taux de recouvrement et le résidu classique du modèle économétrique.

**Conclusion sur l'estimation de la population en sous-recouvrement en construisant une fonction de coût à partir des recettes des services de la base SISPEA**

La seconde méthode testée visait à isoler les services en situation de moindre recouvrement en travaillant à partir d'une fonction théorique de recette issue de la base SISPEA et en couplant avec un taux moyen de recouvrement.

La méthode utilisée présente des résultats relativement convergents avec la première méthode.

Toutefois, l'interprétation des non recouvrement n'est pas totalement satisfaisante car la méthode ne permet pas de savoir, pour un service donné, quelle part est imputable à l'erreur classique économétrique et quelle part est imputable au non recouvrement des coûts.

Pour une prochaine étape il serait donc utile de préciser l'approche économétrique en utilisant la technique des frontières stochastiques.

### 3. Validation de la deuxième méthode et typologie des services recouvrant ou non leurs coûts

Les objectifs principaux de cette partie sont :

- de juger la validité la seconde approche économétrique mobilisant les seules données SISPEA,
- de tester s'il est possible, sur la base de la méthode économétrique de dégager une typologie des services recouvrant ou non leurs coûts.

Malgré les réserves sur la fiabilité des données et sur l'amélioration possible des méthodes, on constate donc une certaine convergence entre les deux approches. Même si, en valeur absolue la première méthode prête à discussion, on voit qu'en valeur relative, elle conduit à isoler une population de service en moindre recouvrement proche de celle isolée par la seconde méthode.

On va donc essayer d'aller plus loin, en testant si une corrélation existe entre les taux de recouvrement estimés pour les services de SISPEA par chacune des méthodes. Si la corrélation est positive, importante et significative cela va dans le sens de valider la pertinence de la deuxième méthode d'estimation des coûts, théoriquement moins précise, mais plus facile à mettre en œuvre pour de futurs travaux portant sur ce même thème.

On souhaiterait également arriver à faire une typologie des services en sous-recouvrement en analysant les liens qui existent entre les taux de recouvrements estimés par les deux méthodes et les caractéristiques des services.

#### 3.1. Corrélation entre les calculs des taux de recouvrement suivant les deux méthodes

Comme dit dans la section précédente, les résultats en termes de couverture des coûts convergent entre les deux méthodes pour 149 services sur 184 (soit environ 81%).

Le coefficient de corrélation linéaire entre les taux de recouvrement estimés par les deux méthodes a été calculé :  $r = \frac{\text{Cov}(tx\ 1\text{ère}\ méth, tx\ 2\text{ème}\ méth)}{\sigma(tx\ 1\text{ère}\ méth) * \sigma(tx\ 2\text{ème}\ méth)}$

On obtient :

|  |  |
|--|--|
|  | Taux de couverture des coûts estimés par la première méthode |
| Taux de couverture des coûts estimés par la deuxième méthode | 0,68037<br>(< 0,0001)  |

On constate une forte corrélation positive et significative à un niveau de 1% entre les taux de recouvrement estimés par la première méthode et les taux de recouvrement estimés par la deuxième méthode.

**Cela conforte l'idée suivant laquelle on pourra utiliser la seconde, plus facile à mettre en œuvre, dans les années à venir, dans les limites exprimées dans la section précédente.**

### 3.2. Caractéristiques des populations en non recouvrement

Dans le but de dégager une éventuelle typologie des services recouvrant ou non leurs coûts, nous avons calculé les coefficients de corrélation entre les taux de recouvrement des coûts estimés par la première méthode et les caractéristiques des services de la base SISPEA :

|  | Volume mis en distribution | Nombre d'abonnés    | Longueur du réseau  | Indice linéaire des volumes non comptés |
|--|----------------------------|---------------------|---------------------|---|
| Taux de couverture des coûts estimés par la première méthode | 0,24359<br>(0,0009)        | 0,24446<br>(0,0008) | 0,22053<br>(0,0026) | 0,23683<br>(0,0012)                     |

**Le volume d'eau distribué, le nombre d'abonnés, la longueur du réseau et l'indice linéaire des volumes non comptés sont corrélés positivement au taux de couverture des coûts calculés par la première méthode (coefficient de Pearson significatif à 1%).**

**Les services de grande taille paraissent donc recouvrer plus facilement leurs coûts puisque le taux de recouvrement augmente avec la quantité d'eau mis en distribution, le nombre d'abonnés et la longueur du réseau de distribution.** Ceci semble d'ailleurs assez logique puisque les services d'eau étant des monopoles naturels, ils réalisent des économies d'échelles (leurs coûts moyens diminuent avec l'augmentation du volume d'eau qu'il produise) : en conséquence la croissance du nombre d'abonnés (et donc du volume distribué et de la longueur du réseau) fera augmenter les recettes plus vite que les coûts et donc, le taux de recouvrement augmentera.

**En revanche, l'indice linéaire des volumes non comptés influe également positivement sur le taux de recouvrement des coûts ce qui signifierait de prime abord qu'une perte d'eau plus importante par jour et par kilomètre de réseau entraînerait un taux de couverture plus important.** Mais, cet indice est fortement corrélé positivement au volume mis en distribution qui lui joue positivement sur le taux de recouvrement des coûts : les « gros » services qui recouvrent plus aisément leurs coûts auront également un indice linéaire des volumes non comptés plus important.

**On peut également analyser l'impact du mode de gestion (délégué ou régie) sur le fait que les services recouvrent ou non leurs coûts.** Pour ce faire, on transforme les taux de recouvrement estimés pour chaque service en variable binaire où « 1 » représente les services recouvrant leurs coûts et « 0 » représente les services ne recouvrant pas leurs coûts. Nos deux variables représentant le type de gestion et la couverture des coûts sont donc qualitatives binaires. Nous effectuons alors un test d'indépendance du khi-deux pour juger de la dépendance statistique entre nos deux variables :

gestion(gestion)  
couverture1(couverture1)

| FREQUENCE<br>Pourcentage<br>Pct en ligne<br>Pct en col. | 0                             | 1                              | Total         |
|---|-------------------------------|--------------------------------|---------------|
| 0   | 10<br>5.43<br>26.32<br>24.39  | 28<br>15.22<br>73.68<br>19.58  | 38<br>20.65   |
| 1   | 31<br>16.85<br>21.23<br>75.61 | 115<br>62.50<br>78.77<br>80.42 | 146<br>79.35  |
| Total   | 41<br>22.28                   | 143<br>77.72                   | 184<br>100.00 |

Fréquence manquante = 26

La p-value associée au test du khi-deux est de 0,5024. On a donc 50,24% de chance de se tromper en rejetant l'hypothèse nulle d'indépendance entre les deux variables : en conséquence, **le mode de gestion d'un service AEP n'a pas de corrélation avec le fait qu'il recouvre ou non ses coûts comptables.**

**Des études de corrélations similaires ont été faites sur la seconde méthode. Malheureusement, cette fois-ci, les corrélations se sont révélées non significatives. Il conviendrait de confirmer ce résultat sur une prochaine étude.**

#### **Conclusion sur la possibilité de caractériser la population des services en non recouvrement**

Des méthodes statistiques assez simples (corrélation) ont été mobilisées pour tenter de caractériser la population des services en situation de non recouvrement.

Certains résultats ressortent de l'application de la première méthode : les services de petite taille semblent recouvrer moins facilement leurs coûts. Les résultats sont moins concluants pour la seconde méthode.

Avant tout développement supplémentaire, il conviendra d'affiner la méthode économétrique permettant d'isoler la population des services en non recouvrement et d'améliorer la qualité des données financières de SISPEA.

## Conclusion générale

Les études macroéconomiques déjà réalisées sur les taux de recouvrement des coûts comptables des services d'eau en France donnent des résultats très différents : selon que l'on travaille à partir des comptes des collectivités et des délégataires ou à partir des flux financiers globaux des services, on obtient des taux de recouvrement global des coûts comptables respectivement de 71% (Etude Ernst&Young) ou de 95% (BIPE).

Notre travail visait quant à lui non pas à estimer plus précisément le taux de recouvrement à partir de SISPEA (qui ne contient pas les données nécessaires pour établir directement le niveau de coût complet, mais simplement des informations sur les recettes du service) mais à isoler statistiquement une population de services en situation de moindre recouvrement afin de pouvoir mieux la caractériser.

Pour cela deux méthodes micro-économétriques ont été proposées et testées.

**La première en croisant la base SIPSEA avec une base comportant des informations comptables et techniques a permis d'établir une fonction de coût théorique.** Il s'agit du coût financier comptable et non du coût complet. Appliquée sur les données SISPEA cette fonction conduit à estimer un taux de recouvrement global des coûts comptables de l'ordre de 135%. Ce chiffre élevé paraît pour le moins étonnant et s'explique en partie par un manque d'informations précises et fiables concernant les données comptables et techniques sur les services d'eau :

### *1) La qualité des données dans SISPEA*

- Avant tout, il semble que l'information sur les recettes saisie dans SISPEA pour la première année de collecte des données ne soit pas suffisamment bien définie et que dans de nombreux cas, les recettes des tiers, voire les produits annexes aient été inclus. Le périmètre des recettes saisies dans SISPEA est donc sans doute plus large que celui des coûts calculés dans la base AG.

- Les grosses collectivités<sup>30</sup> pour lesquelles la règle de recouvrement des coûts est en générale appliquée de manière plus stricte (l'interdiction des subventions par le budget annexe ne s'applique pas pour les communes de moins de 3000 habitants) sont surreprésentées dans l'échantillon SISPEA.

- De même, sont surreprésentés les services délégués : étant sensés respecter l'impératif d'une marge bénéficiaire pour les actionnaires, on peut penser qu'ils seront rarement en non recouvrement des coûts.

### *2) La qualité des données dans la base de référence Adour-Gironde*

Les coûts de SISPEA sont prédits via les coefficients estimés sur la base de données Aube-Gironde, peu représentative (car ne comprenant que des services AEP observés sur deux départements) et assez ancienne (environ 10 ans).

### *3) Les limites de la méthode économétrique retenue pour les services extrêmes*

Nous avons adopté pour la fonction de coût une forme "translog". Cette forme fonctionnelle présente de nombreux avantages (flexibilité, facilité de mise en œuvre, lecture directe des paramètres estimés, etc.). Cependant, l'estimation des coûts des services d'eau est d'autant plus précise qu'un service est au voisinage de la moyenne de l'échantillon. L'erreur pour les services extrêmes risque donc d'être non négligeable. Or la variabilité de l'échantillon est assez importante et risque d'être encore plus forte lorsque l'échantillon de SISPEA s'étoffera.

---

<sup>30</sup> Puisque en 2008, seules les collectivités volontaires ont fourni les valeurs des indicateurs caractérisant leur service AEP.

**Malgré les limites soulignées, une population de services identifiés comme ne recouvrant pas leurs coûts a été isolée. Compte tenu de ce qui précède, on peut penser qu'elle est sous-estimée.**

**La seconde méthode utilise un taux de recouvrement exogène (issue de l'analyse macro-économique BIPE) qui permet de calibrer une fonction de coût à partir d'une fonction théorique des recettes, directement produite à partir des données SISPEA.**

Cette méthode ne permet bien-sûr pas de mesurer le taux de recouvrement absolu, puisque c'est une donnée introduite dans le modèle : par construction, on restera proche de la valeur de 95% issue de l'étude BIPE.

On peut toutefois utiliser cette méthode, comme précédemment, pour isoler une population de services en moindre recouvrement des coûts par rapport aux autres.

Une limite de la méthode économétrique proposée est qu'elle ne permet pas de faire la part dans cette population apparaissant en sous-recouvrement de ce qui est lié à l'erreur du modèle (facteurs non pris en compte dans le modèle, qui faussent le coût théorique par rapport au coût réel) et ce qui est réellement lié à un non recouvrement (niveau de recette inférieur aux coûts réels).

La question de la qualité des données de SISPEA sur la recette se pose à nouveau.

Néanmoins, on peut observer que la convergence entre les deux méthodes en termes de services en moindre recouvrement est grande : une large majorité des services AEP non recouvrant par la seconde méthode le sont aussi par la première. Par ailleurs, il existe une corrélation assez importante entre les taux de recouvrement des coûts estimés selon les deux méthodes.

Ces premiers résultats sont préliminaires et nécessitent d'être consolidés à la fois en terme de fiabilité des données et en terme d'affinement de la méthode économétrique adoptée. Néanmoins, ces résultats semblent indiquer que la seconde méthode, la plus simple à mobiliser à partir de la seule base SISPEA serait satisfaisante, puisque validée par la première méthode, la plus précise.

Un dernier résultat ressort de l'analyse de la population des services en sous-recouvrement isolés par la première méthode : une étude de corrélation montre le lien entre la taille des services et le niveau de recouvrement. En tendance, plus les services sont de taille importante, meilleur est le recouvrement.

**En tout état de cause, pour prolonger les travaux commencés et pouvoir valider effectivement une méthode économétrique permettant d'isoler des services en non recouvrement, il s'avère nécessaire de travailler sur des données plus fiables. C'est la raison pour laquelle le projet qui devait initialement se poursuivre en 2010 sera interrompu jusqu'à ce que la qualité des données permette à nouveau d'avancer.**

## Bibliographie

BENJAMIN C., CADORET I., HERRARD N., MARTIN F., TANGUY S. (2004), « Econométrie Appliquée » 1<sup>ère</sup> Edition.

BIPE/FP2E. (2008), « Les services collectifs d'eau et d'assainissement en France - Données économiques, sociales et environnementales ».

BOYER M., GARCIA S. (2008), « Régulation et mode de gestion : une étude économétrique sur le prix et la performance dans le secteur de l'eau potable », *Annales d'Economie et de Statistique*.

BOUSCASSE H., DESTANDAU F, GARCIA S. (2008), « Analyse économétrique des coûts des services d'eau potable et qualité des prestations offertes aux usagers » *Revue d'Economie Industrielle*, 122, pp 7-26.

CANNEVA G., GARCIA S., GUERIN-SCHNEIDER L. (2004), « La production des comptes dans les contrats d'eau et d'assainissement : Outil de régulation pour la collectivité ou de négociation pour le délégataire », Cahiers de recherche du GEA, Cahier No 2-Septembre.

CANNEVA G. (2008), « Les déterminants du prix des services d'eau et d'assainissement : synthèse bibliographique pour la définition d'une typologie en France », *Rapport provisoire*, ONEMA, Paris.

CHRISTENSEN L.R., JORGENSON D.W., LAU L.J. (1971), « Conjugate duality and the transcendental logarithmic production function », *Econometrica*, vol. 39, n° 4, pp. 255-256.

DESTANDAU F. (2004), « Méthodes d'évaluation des coûts d'exploitation des services d'alimentation en eau potable et d'assainissement », Rapport pour le Ministère de l'Ecologie et du Développement et de l'Aménagement Durables.

EL YOUSFI H., LOUBIER S. (2009), « L'analyse économique dans la DCE : 1- Harmonisation des méthodes de calcul du recouvrement des coûts 2- Pratiques et besoins des Agences », *Convention Cemagref - ONEMA Année 2008*.

ERNST & YOUNG. (2007), « Etude relative au calcul de la récupération des coûts des services liés à l'utilisation de l'eau pour les districts français », Rapport pour le Ministère de l'Ecologie et du Développement et de l'Aménagement Durables.

FAUQUERT G. (2007), « Les déterminants du prix des services d'eau potable en délégation », Thèse de doctorat, Engref.

GARCIA S. (2001), « Analyse économique des coûts d'alimentation en eau potable », Thèse de doctorat, Engref, Toulouse.

GUERIN-SCHNEIDER L. (2001), « Introduire la mesure de performance des services d'eau et d'assainissement en France, instrumentation et organisation », Thèse de doctorat, Engref, Paris.

GUERIN-SCHNEIDER L., NAKHLA M. (2003), « Les indicateurs de performance : une évolution clef dans la gestion et la régulation des services d'eau et d'assainissement », *Flux*, 52/53, pp. 55-58.

GUERIN-SCHNEIDER L. (2009), « L'analyse économique des usages : Calcul du recouvrement des coûts », *Convention Cemagref - ONEMA Année 2008*.

GURGAND D, « Econométrie et biais de sélection », Cours.

HAYES K. (1987), « Cost structure of the water industry », *Applied Economics*, vol. 19, n° 3, pp. 417-425.

HECKMAN J. (1979), « Sample selection bias as a specification error », *Econometrica* Vol. 47(1), pp. 153-161.

LEE L.-F. (1978), « Unionism and wage rates : A simultaneous equations models with qualitative and limited dependent variables », *International Economic Review*,19(2), pp. 415-433.

LEE L.-F., MADDALA G.S, TROST R.P (1980), « Asymptotic covariance matrices of two-stages probit and two-stages tobit methods for simultaneous equations models of selectivity » , *Econometrica*, 48(2), pp 491-503.

ROGERS, BHATIA, HUBER, 1997. *Water as social and economic good: How to put the principle into practice?*, *Readings of WRM Course*, World Bank.

**ANNEXE****Estimation du modèle probit :**

|                          | <b>Estimation</b> | <b>Ecart-Type</b> |
|--------------------------|-------------------|-------------------|
| <b>Constante</b>         | 0,7442***         | 0,1764            |
| <b>Q</b>                 | -0,3740           | 0,7326            |
| <b>AB</b>                | 0,6504            | 0,6160            |
| <b>ILVNC</b>             | 0,3049            | 0,3801            |
| <b>Q<sup>2</sup></b>     | 3,1792            | 2,3647            |
| <b>AB<sup>2</sup></b>    | 1,8025            | 2,3447            |
| <b>ILVNC<sup>2</sup></b> | 0,8025            | 0,4998            |
| <b>Q * AB</b>            | -2,3816           | 2,3059            |
| <b>Q * ILVNC</b>         | -2,3276**         | 0,9739            |
| <b>AB * ILVNC</b>        | 2,34**            | 0,9355            |
| <b>RDT</b>               | 1,7425            | 1,3169            |
| <b>Longres</b>           | 0,0353            | 0,3582            |

**Estimation du modèle de recette sur le sous-échantillon des services délégués (11) :**

|   | <b>Estimation</b> | <b>Ecart-type</b> |
|---|-------------------|-------------------|
| <b>Constante</b>  | 12,30735***       | 0,20792           |
| <b>Q</b>  | 0,39805*          | 0,22649           |
| <b>AB</b>   | 0,70697***        | 0,25154           |
| <b>ILVNC</b>  | -0,10582          | 0,06752           |
| <b>Q<sup>2</sup></b>  | 0,01184           | 0,82489           |
| <b>AB<sup>2</sup></b>                                       | 0,11753           | 0,68350           |
| <b>ILVNC<sup>2</sup></b>                                    | 0,17517           | 0,17088           |
| <b>Q * AB</b>   | -0,01009          | 0,72245           |
| <b>Q * ILVNC</b>  | -0,25364          | 0,43850           |
| <b>AB * ILVNC</b>   | 0,21684           | 0,43891           |
| <b>Cov ( <math>\varepsilon^D</math>, <math>\mu</math> )</b> | 0,80076*          | 0,472             |

**Estimation du modèle de recette sur le sous-échantillon des services en régie (12) :**

|   | <b>Estimation</b> | <b>Ecart-type</b> |
|---|-------------------|-------------------|
| <b>Constante</b>  | 13,9497***        | 2,3553            |
| <b>Q</b>  | 1,0390            | 1,3383            |
| <b>AB</b>   | 0,1065            | 1,5282            |
| <b>ILVNC</b>  | -0,0159           | 0,5171            |
| <b>Q<sup>2</sup></b>  | 1,0244            | 7,2672            |
| <b>AB<sup>2</sup></b>                                       | -0,3546           | 6,3141            |
| <b>ILVNC<sup>2</sup></b>                                    | 0,1073            | 1,3599            |
| <b>Q * AB</b>   | -0,3347           | 6,6545            |
| <b>Q * ILVNC</b>  | -0,8528           | 3,4842            |
| <b>AB * ILVNC</b>   | 0,6709            | 3,4596            |
| <b>Cov ( <math>\varepsilon^D</math>, <math>\mu</math> )</b> | -0,9766           | 1,7961            |

N.B. : Programme SAS utilisé pour les modèles est disponible sur demande aux auteurs

---

*Au sens de la DCE, le coût complet comprend le coût financier des services, le coût d'opportunité de la ressource et le coût environnemental.*

*Les deux derniers coûts ne sont actuellement pas couverts par les recettes venant des bénéficiaires des services d'eau et d'assainissement. Par contre, ces derniers assument une partie substantielle des coûts financiers.*

*Est-il possible d'isoler les services qui ne recouvrent pas ce premier niveau des coûts financiers ? C'est la question à laquelle ce rapport contribue.*

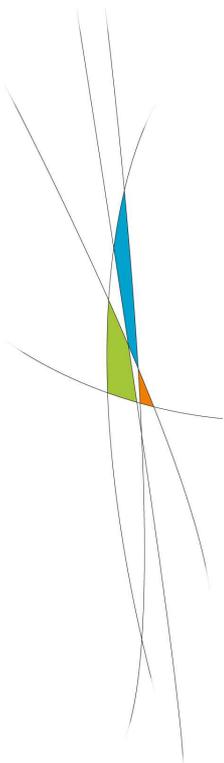
*Il ne s'agit pas de calculer un taux de recouvrement moyen mais d'étudier les caractéristiques des services en sous-recouvrement, pour pouvoir éclairer les décisions publiques notamment en termes de financement des services d'eau.*

*Pour isoler statistiquement une population de service en sous-recouvrement, différentes méthodes économétriques sont testées et appliquées sur la première série de données récoltées par le Système d'Information sur les Services d'Eau et d'Assainissement (SISPEA).*

*Les premiers résultats sont encourageants en termes de faisabilité, mais nécessitent une amélioration substantielle de la qualité des données sur les recettes et un affinement des méthodes économétriques pour donner lieu à des résultats valides et interprétables.*

\*\*\*\*\*

*Ce rapport présente les résultats d'une action entreprise dans le cadre de la convention Cemagref ONEMA 2009, portant sur l'analyse économique au sein de la DCE. Il complète un premier rapport qui a montré la difficulté d'évaluer le taux de recouvrement économique par les outils comptables et gestionnaires disponibles.*



Direction générale  
Parc de Tourvoie  
BP 44 - 92163 Antony cedex  
Tél. 01 40 96 61 21 - Fax 01 40 96 62 25  
[www.cemagref.fr](http://www.cemagref.fr)