

**L'évaluation des impacts des programmes d'efficacité énergétique :  
principes économiques et applications aux programmes  
de l'Agence d'efficacité énergétique**

**RAPPORT D'EXPERT**

**Pierre Emmanuel Paradis  
Économiste principal  
Groupe d'analyse, Ltée**

**14 novembre 2008**

## TABLE DES MATIÈRES

1. Mandat	3
2. Auteur et qualifications	4
3. Principes économiques d'évaluation quantitative des impacts de programmes et applications	5
3.1 Principes économiques de l'évaluation quantitative des impacts de programme	6
3.2 Applications des méthodes d'évaluation des impacts de programmes	13
4. Revue des évaluations de programmes d'efficacité énergétiques du Québec	18
4.1 Évaluations de programmes récentes	18
4.2 Le PIMVR	24
5. Conclusion et recommandations	30

## 1. MANDAT

Ce rapport présente mon opinion d'économiste concernant l'évaluation des impacts des programmes de l'Agence d'efficacité énergétique (« l'Agence »). Ce rapport est déposé dans le cadre de l'audience R-3671-2008 portant sur la *Demande relative à l'approbation du premier Plan d'ensemble en efficacité énergétique et nouvelles technologies*. En particulier, il aborde principalement trois des sujets cités dans la décision procédurale D-2008-104 (page 5) :

- Rentabilité du PEEÉNT, tests utilisés et résultats;
- Indicateurs de performance du PEEÉNT;
- Évaluation et suivi des programmes et interventions du PEEÉNT.

Selon cette même décision, l'une des conclusions recherchées par la présente demande est la suivante:

*« ...à compter du deuxième Plan d'ensemble débutant le 1er avril 2010, tous les programmes et interventions en efficacité énergétique et nouvelles technologies des distributeurs d'électricité et de gaz naturel à savoir, Hydro-Québec Distribution, Gaz Métro, incluant le FEÉ, et Gazifère, devront, en vertu des articles 22.11 de la Loi sur l'Agence et 85.26 de la Loi sur la Régie, obligatoirement être intégrés au dossier du Plan d'ensemble aux fins de leur approbation commune par la Régie; ... ».* (page 2)

Dans ce contexte d'intégration de tous les programmes et interventions d'Hydro-Québec Distribution (« HQD »), Gaz Métro et Gazifère sous la gouverne de l'Agence, il est important de passer en revue l'évaluation du rendement de ces interventions. Plus spécifiquement, cet examen s'attardera aux impacts des programmes d'efficacité énergétique et aux méthodes employées pour les évaluer, afin de s'assurer qu'ils livrent une information optimale aux fins de prise de décision aux responsables de ces programmes.

Dans le cadre de la rédaction de ce rapport, j'ai conservé le libre choix des moyens d'exécution et il n'existait aucun lien direct ou indirect de subordination entre la FCEI et Groupe d'analyse.

## 2. AUTEUR ET QUALIFICATIONS

Au cours des dix-sept dernières années, Pierre Emmanuel Paradis a réalisé des mandats économiques de haut niveau pour des cabinets d'avocats, des sociétés privées et des organismes gouvernementaux dans de nombreux secteurs, notamment les soins de santé, l'énergie, le transport et l'environnement. Il a produit des rapports et opinions d'expert dans les domaines suivants : étude sectorielle et de marché, analyse avantages-coûts, évaluation financière, impact économique, économétrie, évaluation de programme et développement économique. Dans le cadre de ses travaux, monsieur Paradis a élaboré des modèles quantitatifs complexes et possède une riche expérience dans la planification et la conduite d'enquêtes.

Dans le domaine de l'énergie, il a réalisé plusieurs études de retombées économiques pour les centrales hydroélectriques de Sainte-Marguerite-3, de Toulnostouc, de Gentilly-2 et de Mercier pour Hydro-Québec de 1999 à 2002. En outre, cela lui a valu de participer à la rédaction de la méthodologie de référence d'Hydro-Québec en matière d'analyse de ces impacts économiques. Il a produit une expertise sur les conditions de service d'Hydro-Québec Distribution (R-3535-2004 Phase 2), en plus de participer aux travaux sur le partage des coûts post-patrimoniaux (R-3610-2006). Également, il a réalisé et présenté en commission parlementaire une modélisation d'un Fonds énergétique destiné à rembourser la dette publique du Québec. Il a aussi rédigé des évaluations d'impact économique pour des projets de gazoduc et de parc éolien.

Avant de se joindre à Groupe d'analyse, il était économiste principal à la Fédération canadienne de l'entreprise indépendante (FCEI) où il a travaillé avec les petites et moyennes entreprises sur des dossiers tels que la main-d'œuvre et la rémunération, les finances publiques, la réglementation et le financement. En outre, il agissait également à titre de porte-parole de la Fédération sur ces dossiers, aussi bien auprès de divers ministères et organismes gouvernementaux que dans les médias (télévision, presse écrite, radio).

M. Paradis détient un diplôme d'études supérieures spécialisées (D.E.S.S.) en finance d'entreprise de HEC Montréal et une maîtrise en économie (M.A.) de l'Université de Sherbrooke. Il a publié de nombreux articles et rapports, donne régulièrement des conférences et a témoigné à titre d'expert et d'économiste devant des organismes publics au Québec (Régie de l'énergie, Tribunal administratif du Québec, Autorité des marchés financiers, Régie des rentes du Québec, commissions parlementaires).

### 3. PRINCIPES ÉCONOMIQUES D'ÉVALUATION QUANTITATIVE DES IMPACTS DE PROGRAMMES ET APPLICATIONS

L'évaluation de programme s'intéresse à l'impact d'une foule d'interventions<sup>1</sup> visant à améliorer nos sociétés. Ces interventions couvrent une vaste gamme de domaines d'activité : développement économique, santé, éducation, sécurité publique, énergie, environnement, habitation, aide humanitaire, etc. L'évaluation de ces programmes regroupe de nombreuses approches analytiques, aussi bien quantitatives que qualitatives, qui s'intéressent :

- aux expériences et perceptions des personnes concernées par l'intervention : administrateurs, participants et non-participants, autorités gouvernementales, autres groupes d'intérêt; et
- au processus et aux impacts de ces programmes : organisation logique et livraison du programme, intrants et résultats.

Ce rapport se concentre sur l'évaluation quantitative des impacts des programmes d'efficacité énergétique. Pour ce faire, il comporte les quatre sections suivantes :

- les principes économiques fondamentaux en évaluation quantitative des impacts de programmes seront exposés;
- l'application de ces principes sera décrite dans plusieurs contextes, notamment en développement économique, en éducation, en santé et en efficacité énergétique, à la fois au Canada, aux États-Unis et à l'international;
- l'évaluation des programmes d'efficacité énergétique au Québec sera passée en revue, ce qui permettra d'identifier certaines différences par rapport à l'application usuelle des méthodes économiques en évaluation de programme;
- la conclusion et les recommandations à la Régie, dans le contexte d'unification prochaine des évaluations de programmes d'efficacité énergétique au sein de l'Agence.

---

<sup>1</sup> Les termes « programme » et « intervention » seront utilisés comme synonymes dans ce rapport. Toutefois, ils doivent être compris au sens large, c'est-à-dire tout projet, politique, loi, programme, traitement, etc. ayant pour but d'améliorer une situation.

### 3.1 Principes économiques de l'évaluation quantitative des impacts de programme

À l'origine, un programme est une initiative qui consiste à entreprendre une série d'actions visant à modifier un résultat social observable : la consommation personnelle d'énergie, la création d'emploi, le degré de scolarité atteint, le niveau de salaire, la criminalité, etc.

Dans ce contexte, il peut être utile d'évaluer différents aspects du programme, par exemple sa pertinence, l'efficacité de son mécanisme d'intervention, la satisfaction des bénéficiaires, etc. Toutefois, l'objectif premier d'un programme, aussi bien pour ses bénéficiaires que pour ceux qui le financent et l'administrent, est qu'il se traduise par un **effet concret et observable** sur ces bénéficiaires. En d'autres mots, il faut que le programme change les choses, sans quoi son utilité est limitée.

Pour vérifier si un tel résultat a été atteint, il faut procéder à une évaluation rétrospective et quantitative de l'impact du programme. Si un impact net positif est constaté, l'intervention peut être poursuivie et les leçons de son succès peuvent servir dans d'autres contextes similaires. Dans le cas contraire, le promoteur du programme doit forcément apprendre de ses échecs et revoir la nature de son intervention. Par conséquent, l'évaluation quantitative des impacts du programme demeure le principal exercice à effectuer en vue d'en établir la pertinence et la rentabilité.

L'évaluation quantitative des impacts (« *impact evaluation* ») est un sous-ensemble de l'évaluation basée sur les résultats (« *outcome-based evaluation* »). Cet exercice détermine si le programme a causé un changement comparativement à aucun programme ou à un autre type d'intervention :

*« Impact analysis is not easy, but it is an essential outcome-based analysis to complete if you want to determine whether a given program made a difference compared to either no program or an alternative program. Thus, an absolute requirement in impact analysis is that you have a comparison group or condition against which you compare the significance of your results. »<sup>2</sup> (nos caractères gras)*

L'évaluation des impacts consiste d'abord à choisir une mesure appropriée de l'impact du programme. Ensuite, il faut procéder à l'évaluation de l'impact du programme à proprement parler. Après avoir exposé la mesure théorique optimale de cet impact, les difficultés pratiques et statistiques

---

<sup>2</sup> Robert L. Schalock, *Outcome-Based Evaluation*, 1995, Plenum Press, New York, p. 57.

accompagnant ce type d'estimation seront discutées, de même que les solutions employées à ces fins.

### ***Choix d'une mesure de l'impact du programme***

Si cette étape peut sembler triviale, elle l'est certainement moins en pratique. À preuve, la rentabilité et les indicateurs de performance font partie des sujets à l'étude dans le cadre de la présente demande. Du document AEE-8, Document 1, nous comprenons que les indicateurs de performance seront employés à des fins de balisage des activités de l'Agence (pages 173 et 174), tandis que la rentabilité désigne plus directement l'impact des programmes en termes d'énergie et d'argent épargnés. Ainsi, ce dernier sujet fera l'objet de la discussion économique qui suit.

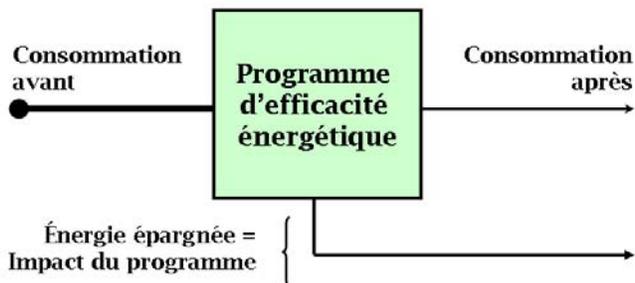
La Figure 1 représente l'intervention d'un programme sur la consommation d'énergie d'un agent économique (individu, entreprise, ou institution), ainsi que son impact. L'évolution de cette consommation dans le temps est représentée par la ligne horizontale. Sans programme, la consommation énergétique de l'agent resterait inchangée. Avec le programme, cette consommation est réduite suite à l'intervention.

**Figure 1**  
**Impact du programme sur la consommation**

#### **SANS PROGRAMME**



#### **AVEC PROGRAMME**



Dans cette figure, l'hypothèse implicite est que, sans le programme, la consommation « après » serait demeurée constante et aussi forte que durant la période « avant ». Ainsi, l'impact observé est causé par le

programme. Nous verrons plus loin quels problèmes d'évaluation cela peut représenter et comment il est possible d'y pallier.

Ainsi, pour un programme d'efficacité énergétique, la **mesure physique** d'impact à privilégier est tout simplement la quantité d'énergie non utilisée. Grâce au programme, une certaine quantité d'énergie est désormais libérée et peut servir à d'autres usages. Plus précisément, cette quantité d'énergie peut être évaluée sur une base unitaire et pour une certaine durée de temps, c'est-à-dire par agent bénéficiaire et par année (par exemple : GJ d'énergie, kWh d'électricité ou m<sup>3</sup> de gaz épargnés par agent par année).

Maintenant, cet impact physique du programme doit aussi être converti en une **valeur économique**. Pour ce faire, il faut estimer la valeur actualisée nette (VAN) du programme, qui est la somme de ses avantages économiques moins le total de ses coûts économiques. Le Tableau 1 précise quels sont ces principaux postes d'avantages et de coûts en question, ainsi que les agents économiques qui les reçoivent et/ou les paient. Les cellules ayant une notation, par exemple 'AC1', illustrent la présence d'un avantage ou d'un coût (lignes) pour un agent donné (colonnes).

**Tableau 1**  
**Avantages et coûts économiques**  
**des programmes d'efficacité énergétique**

	Bénéficiaire/payeur		
	Client	Distributeur et/ou administrateur	Société - autres
<b>Avantages</b>			
Valeur marchande de l'énergie libérée	AC1		
Coûts évités de production, transport et distribution		AD1	
Avantages non-énergétiques: réduction GES, autres			AS1
<b>Coûts</b>			
Administration du programme		CD1	
Incentif payé au client		CD2	
Coût de mise en place pour le client	CC1		

Note : Les avantages non-énergétiques regroupent la réduction des gaz à effet de serre (« GES ») et d'autres avantages, par exemple sur le développement économique.

Les notations de ce tableau comptent trois caractères :

- le premier indique s'il s'agit d'un avantage ('A') ou d'un coût ('C');
- le second indique l'identité du bénéficiaire ou payeur ('C' : client; 'D' : distributeur et/ou administrateur; 'S' : société);

- le troisième numérote l'avantage ou le coût pour chaque bénéficiaire/payeur : '1' pour le premier et '2' pour le deuxième.

Du point de vue économique, l'idéal est de considérer l'ensemble des avantages et coûts associés au programme<sup>3</sup>, soit :

$$\text{VAN} = \text{AC1} + \text{AD1} + \text{AS1} - (\text{CC1} + \text{CD1} + \text{CD2})$$

Cette expression correspond essentiellement à la définition du test du coût social (« TCS »). Toutefois, il est difficile d'estimer précisément ces avantages non-énergétiques, représentés ici par AS1. Notamment, l'Agence propose de retenir 25 % des bénéfices énergétiques, ce qui semble être plutôt arbitraire et peu documenté<sup>4</sup>. En outre, les retombées économiques<sup>5</sup> ne devraient pas être considérées à titre de bénéfices non-énergétiques, car il s'agit en fait d'une analyse de répartition de dépenses et non d'un avantage économique réel. De plus, comme la redevance au Fonds vert est déjà incluse dans le prix des énergies émettant du CO<sub>2</sub>, ainsi que dans l'estimation des coûts évités, il faut éliminer ces sommes de tout calcul d'impact pour éviter la prise en compte de transferts. Par ailleurs, malgré toute l'incertitude associée à la valeur économique réelle des émissions de GES, les 15 \$ par tonne de CO<sub>2</sub>e représentent une valeur unitaire raisonnable, compte tenu des informations disponibles à ce jour sur la valeur des échanges publics et privés de droits d'émission de GES<sup>6</sup>.

Pour contourner cette difficulté de mesurer les avantages non-énergétiques, il serait pertinent d'employer un TCS plus conservateur, qui ne tiendrait compte que de la valeur des GES non émis à 15 \$ la tonne de CO<sub>2</sub>e, en excluant les autres bénéfices non-énergétiques.

Une autre solution est de se concentrer sur les effets directement quantifiables, soit en éliminant la composante sociale AS1 de l'équation:

---

<sup>3</sup> Tous ces avantages et coûts doivent être exempts de toutes formes de taxes et subventions. De plus, il conviendrait en théorie de considérer la valeur marchande de l'énergie libérée, dans le cas où elle différerait du prix au détail de cette énergie.

<sup>4</sup> Défini au AEE-8, Document 1 amendé, page 189, lignes 1-2 et décrit davantage au AEE-9, document 15, Engagement 15.

<sup>5</sup> Plus spécifiquement, on entend ici les « retombées économiques » mesurées à l'aide d'un modèle d'échanges intersectoriels comme celui de l'Institut statistique du Québec.

<sup>6</sup> Cette valeur concorde avec les prévisions de New Carbon Finance, qui estime qu'avec l'expansion prévue des marchés d'échange de crédits CO<sub>2</sub>e (« cap-and-trade »), ce prix s'établirait à 15 \$US la tonne. Plus généralement, en 2007, la valeur moyenne des échanges sur des marchés volontaires (1 % des marchés mondiaux) s'établissait à 5 \$US, tandis qu'elle s'élevait à 22 \$US sur les marchés réglementés. Source : Katherine Hamilton, Milo Sjardin, Thomas Marcello, and Gordon Xu, *Forging a Frontier State of the Voluntary Carbon Markets 2008*, New Carbon Finance, Mai 2008.

$$VAN_{\text{sans social}} = AC1 + AD1 - (CC1 + CD1 + CD2)$$

Cette nouvelle expression correspond alors au test du coût total en ressources (« TCTR »), qui est plus englobant que le test du coût à l'administrateur de programme (« TCAP »), qui ne s'intéresse qu'au bénéfice du participant et aux seuls coûts du programme :  $AC1 - (CD1 + CD2)$ .

En somme, la mesure idéale du rendement économique des programmes serait un TCS complet et quantifié de façon fiable. Toutefois, en raison des difficultés d'estimation des effets non énergétiques, une forme réduite de ce TCS, qui n'inclurait que la valeur des GES non émis, serait un indicateur plus réaliste et conservateur de la rentabilité économique des programmes. De plus, le TCTR, et non le TCAP, apparaît comme étant la mesure la plus appropriée pour juger de la rentabilité économique d'un point de vue « réseau », c'est-à-dire des participants, des fournisseurs d'énergie et des administrateurs de programme.

### ***Évaluation de l'impact - Objectif théorique***

Les principes économiques, méthodes et applications présentées ci-après constituent les fondements économiques modernes de l'évaluation quantitative des impacts de programme. Dans le reste de cette section, nous présentons la théorie et les techniques employées pour réaliser ces évaluations d'impact selon les règles de l'art. La section suivante exposera ensuite leurs applications pratiques dans plusieurs domaines.

Soit les notations suivantes :

- e : quantité d'énergie consommée
- 0 : temps au début du programme
- 1 : temps à la fin du programme
- P : participant au programme
- T : a bénéficié du programme (« traité »)
- ≠T : n'a pas bénéficié du programme (« non traité »)

L'impact du programme sur la consommation d'énergie s'exprime ainsi<sup>7</sup> :

$$(e_1^{PT} - e_0^{PT}) - (e_1^{P\neq T} - e_0^{P\neq T})$$

---

<sup>7</sup> Pour un exposé vulgarisé, voir Ravallion, M., The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation, *The World Bank Economic Review*, Vol. 15, No. 1, p. 115-140. Une présentation plus formelle peut être trouvée dans Cameron, A.C. et Trivedi, P.K., *Microeconometrics Methods and Applications*, Cambridge, 2005, 1034 p.

L'impact du programme est égal à la différence entre la consommation d'énergie des participants suite au traitement (ou à l'intervention) moins la consommation d'énergie qui aurait été constatée si les mêmes participants n'avaient pas été traités. En effet, dans un tel cas, la consommation d'énergie des participants aurait quand même changé durant la période du programme, que ce soit en raison de leur situation financière, de leur conscientisation personnelle vis-à-vis l'efficacité énergétique ou pour toute autre raison. Ainsi, l'impact du programme représente précisément la différence entre la consommation énergétique avec le programme et ce qu'elle aurait été sans traitement pour ces mêmes participants. Avec ce résultat, il serait alors possible de conclure que l'impact estimé a été causé par le programme.

Or, une difficulté majeure se pose pour effectuer cette estimation : il est impossible d'observer simultanément ce que les participants auraient fait avec et sans le programme. En effet, personne ne dispose d'une machine à remonter dans le temps, qui permettrait d'observer de nouveau ce résultat dans les deux scénarios. C'est pourquoi il faut utiliser des techniques statistiques qui permettent de réaliser ce calcul avec le plus de précision possible.

### ***Évaluation de l'impact - Considérations pratiques et statistiques***

Le principal problème de l'évaluation d'impact est l'impossibilité d'observer les participants s'ils n'avaient pas bénéficié du programme et ce, durant la période de déroulement du programme. Pour résoudre ce problème, il faut trouver un groupe contrôle d'individus ou d'entreprises n'ayant pas bénéficié du programme, donc des non-participants, présentant toutefois des caractéristiques aussi proches que possible du groupe des participants. Ainsi, il est alors possible d'observer l'évolution parallèle de ces non-participants durant le programme et de la comparer à l'effet du programme sur les participants.

Naturellement, la constitution de groupes de contrôle pose problème, car les participants et non-participants sont fondamentalement différents et ce, pour deux raisons :

- *Caractéristiques observables* : les individus ou entreprises participant au programme peuvent présenter des caractéristiques individuelles différentes de celles des non-participants. Par exemple, si tous les ménages participants sont à faible revenu, il sera difficile de les comparer à l'ensemble des ménages du Québec sans un ajustement statistique approprié. De la même façon, les disparités géographiques peuvent aussi être source de biais dans l'évaluation de l'impact du programme;

- *Caractéristiques inobservables* : ce biais survient lorsqu'il existe une relation entre la participation au programme et les résultats. Cette source de biais s'appelle le biais de sélection. Par exemple, certains traits de caractère peuvent faire en sorte que les participants soient des individus fondamentalement différents des non-participants, ce qui influence simultanément la consommation d'énergie des deux groupes (avec le programme pour les participants; sans le programme pour les non-participants).

À ces fins, la solution optimale est l'**expérience aléatoire** (« *randomised experiment* »). Cette méthode consiste à choisir au hasard les participants et les non-participants parmi une population donnée. Si ces deux groupes sont bel et bien semblables du point de vue statistique, alors la seule différence pouvant expliquer leur consommation d'énergie est la participation au programme. Ainsi, l'impact du programme dans un tel contexte d'expérience aléatoire est simplement la différence entre l'évolution de la consommation d'énergie des participants traités moins celle des non-participants non traités.

Toutefois, la mise en place de telles expériences peut poser plusieurs problèmes. En effet, de tels programmes sont rares en économie, en plus d'être difficiles et coûteux à réaliser. Ces évaluations peuvent aussi créer des problèmes d'éthique. Le groupe contrôle ne doit pas être indirectement affecté par le programme. Pour contrer ces difficultés, trois autres méthodes dites « quasi-expérimentales » sont généralement utilisées pour comparer des données obtenues dans des contextes non aléatoires.

Premièrement, la méthode des **différences-en-différences** tente de reproduire l'expérience aléatoire, d'abord en sélectionnant un groupe contrôle non traité ('C≠T') et en faisant une double différence avec des participants traités ('PT') : résultat avant et après le programme; participants moins non-participants. L'équation suivante présente cette double différence :

$$(e_1^{PT} - e_0^{PT}) - (e_1^{C\neq T} - e_0^{C\neq T})$$

La seule différence avec l'impact théorique du programme se trouve du côté droit de l'expression, alors que la consommation énergétique est observée pour le groupe contrôle et non pour les participants non traités. En théorie, les deux groupes n'ont pas besoin d'être semblables sur le plan statistique. Il existe diverses façons de raffiner cette technique, notamment en choisissant le groupe de contrôle à l'aide d'une technique de jumelage présentée ci-après. Toutefois, il convient d'être prudent avec

l'autocorrélation, qui peut artificiellement gonfler la significativité des estimés si aucune correction appropriée n'est effectuée.

Deuxièmement, la méthode du **jumelage** (« *matching* ») permet de choisir un groupe contrôle en attribuant aux participants et aux non-participants un score estimant leur « probabilité de participer ». Les non-participants qui constitueront le groupe de contrôle sont ensuite choisis de façon à avoir le score le plus rapproché de celui des participants, selon une règle de décision donnée. Ensuite, il ne reste qu'à calculer la différence de résultats entre ces deux groupes ou, mieux encore, de la combiner à la méthode des différences-en-différences.

Troisièmement, la méthode des **variables instrumentales** permet d'éliminer le biais de sélection. Pour ce faire, il faut choisir des variables statistiquement liées avec la participation au programme, mais indépendantes de la consommation d'énergie, et les intégrer dans l'équation d'évaluation de l'impact. Ce faisant, il ne subsiste en principe que des différences dans les caractéristiques observables entre participants et non-participants. Or, ces différences peuvent être contrôlées à l'aide d'une régression linéaire multiple. Si aucune variable instrumentale de qualité n'est trouvée, il demeure néanmoins possible de recourir à une analyse de régression multiple standard entre participants et non-participants, avec toutefois les limites décrites ci-haut.

Toutes ces méthodes, incluant l'expérience aléatoire, pourraient être employées pour fins d'évaluation des impacts des programmes d'efficacité énergétique de l'Agence. Fait à noter, leur application s'effectue au moyen de calculs microéconométriques, c'est-à-dire réalisés à partir de données sur des individus ou entreprises, et non sur la base de résultats globaux. En fait, nous verrons plus loin qu'au moins un des programmes du PGEÉ a été évalué avec ces méthodes. La section suivante présente des exemples concrets d'application de ces principes et méthodes dans divers domaines. Par la suite, nous examinerons les évaluations récentes conduites en efficacité énergétique, puis passerons en revue le protocole d'évaluation que compte employer l'Agence.

### **3.2 Applications des méthodes d'évaluation des impacts de programmes**

Les méthodes décrites précédemment sont reconnues par la littérature économique comme étant optimales pour l'évaluation quantitative de l'impact de programme, en outre pour établir ce qu'aurait été la situation pour les personnes traitées par le programme en l'absence de ce dernier.

Elles sont largement utilisées à travers divers domaines où il est nécessaire d'évaluer les programmes et politiques économiques et sociaux<sup>8,9,10,11</sup>.

À ce sujet, cette section présente l'application de ces méthodes par divers organismes publics, aussi bien au Canada qu'à l'international, dans les domaines du travail, du développement économique, de l'aide humanitaire, de la santé et de l'efficacité énergétique.

### ***Politiques publiques***

Afin de vérifier si l'argent des contribuables est dépensé de façon efficace et responsable, les gouvernements et organismes parapublics évaluent l'impact de leurs programmes de façon systématique. Plusieurs de ces organismes ont produit des guides d'évaluation de programme afin d'assurer une uniformité et un certain niveau de qualité dans la conduite de ces évaluations.

Au Canada, le Conseil du Trésor a produit un document intitulé *Méthodes d'évaluation des programmes : Mesure et attribution des résultats des programmes*<sup>12</sup> dans lequel sont présentées les bonnes pratiques à suivre pour réaliser une évaluation de programme. Ce document présente la plupart des méthodes reconnues d'évaluation de programme décrites précédemment. Une grande attention est portée au fait qu'il faut choisir un modèle aussi proche que possible du modèle idéal (expérience aléatoire), sous contrainte des données et ressources disponibles.

Les États-Unis ont développé le *Program Assessment Rating Tool* (« PART ») afin d'évaluer la performance des programmes fédéraux. Les résultats de ce processus sont utilisés dans la gestion annuelle de ces programmes. Le document *Guide to the Program Assessment Rating Tool (PART)*<sup>13</sup> présente la méthodologie à suivre pour réaliser l'évaluation des différents programmes fédéraux, de la conception du programme jusqu'à l'atteinte des résultats voulus, en passant par l'application du programme selon les règles. Afin de réaliser une évaluation de programme valide, l'objectif est de déterminer l'impact causal du programme, c'est-à-dire qui

---

<sup>8</sup> Blundell, R., Costa Dias, M. Evaluation Methods for Non-Experimental Data, *Fiscal Studies*(2000) vol. 21, no. 4, p. 427-468.

<sup>9</sup> Ravallion, M. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation, *The World Bank Economic Review*, Vol. 15, No. 1, p. 115-140.

<sup>10</sup> Ezemenari E. et al. *Impact Evaluation: A Note on Concepts and Methods*, The World Bank, May 1999, 33 pages.

<sup>11</sup> Angrist J.D., Krueger A.B. *Empirical Strategies in Labor Economics*, Working Paper #780, Princeton University, Department of Economics, Industrial Relations Section.

<sup>12</sup> Secrétariat du Conseil du Trésor du Canada, *Méthodes d'évaluation des programmes : Mesure et attribution des résultats des programmes*, Troisième édition, 1998, 156 pages.

<sup>13</sup> Office of Management and Budget (OMB). *Guide to the Program Assessment Rating Tool (PART)*, January 2008.

ne se serait pas produit sans la présence de cette intervention. Le détail des techniques privilégiées, très similaires à celles exposées plus haut et utilisées par le Secrétariat du Conseil du Trésor du Canada, est présenté dans une note séparée du document principal<sup>14</sup>.

La Banque mondiale a aussi produit plusieurs guides détaillant ses recommandations relatives à l'évaluation quantitative des impacts de programmes de développement économique, social et humain<sup>15</sup>. Pour décider s'il faut reconduire ou non un programme, la Banque mondiale détermine quel a été l'impact net du projet sur les bénéficiaires, c'est-à-dire quelle part de l'amélioration observée résulte directement du projet et quelle part aurait été obtenue de toute façon. Pour ce faire, elle se base sur les méthodes décrites précédemment. La Banque mondiale accorde aussi une place importante aux aspects qualitatifs dans l'évaluation d'un programme.

### **Santé**

L'évaluation des effets d'un médicament ou d'un traitement médical spécifique se fait aussi à l'aide des méthodes décrites plus haut. Comme l'exactitude des résultats est très importante en santé et qu'il est plus facile de contrôler l'accès à l'intervention dans le cas de traitements médicaux que dans le cas de programmes sociaux, la comparaison des résultats d'un groupe traité à ceux d'un groupe contrôle est la norme à respecter<sup>16</sup>.

L'essai clinique est la méthode principale utilisée pour évaluer la sécurité et l'efficacité d'un traitement spécifique à une maladie. Il s'agit d'une expérience aléatoire qui évalue les impacts d'un traitement dans un environnement contrôlé. Deux groupes de patients sont observés : l'un recevant le traitement à être évalué et l'autre recevant un placebo ou un autre traitement existant. Les patients sont séparés de façon aléatoire dans un groupe et dans l'autre. La différence entre les deux résultats « avant-après » des deux groupes constitue l'impact net du traitement à l'étude.

---

<sup>14</sup> Office of Management and Budget (OMB), *What Constitutes Strong Evidence of a Program's Effectiveness*. Accessible au:

[http://www.whitehouse.gov/omb/part/2004\\_program\\_eval.pdf](http://www.whitehouse.gov/omb/part/2004_program_eval.pdf).

<sup>15</sup> Voir par exemple : Baker, J. *Évaluation de l'impact des projets de Développement sur la pauvreté : Manuel à l'attention des praticiens*, Banque Mondiale, Direction du Développement, 2000, 196 pages. et *OED and Impact Evaluation - A Discussion Note*, Banque Mondiale. Accessible au

[www.worldbank.org/ieg/docs/world\\_bank\\_oed\\_impact\\_evaluations.pdf](http://www.worldbank.org/ieg/docs/world_bank_oed_impact_evaluations.pdf)

<sup>16</sup> Berger M.L. et al. *Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms*, International Society for Pharmacoeconomics and Outcomes Research, 2003, 264 pages.

Un autre type d'étude est l'étude rétrospective de type « *case-control*. » Il s'agit de comparer deux cohortes de patients de façon rétrospective, l'une comprenant des patients ayant reçu le traitement à analyser et l'autre composée de patients semblables au premier groupe, mais ne recevant pas le traitement. Cette technique est beaucoup moins coûteuse que l'essai clinique et permet d'obtenir davantage de patients ayant les caractéristiques désirées, puisqu'ils sont choisis expressément en fonction de ces caractéristiques<sup>17</sup>.

Le choix de la méthode peut aussi dépendre de la nature de l'élément à analyser. Par exemple, si l'objectif est de comparer les coûts associés à la prise de deux médicaments, un essai clinique ne permet pas de connaître le coût dans la population totale, puisque l'environnement est contrôlé. En effet, dans un essai clinique, le choix du traitement n'est pas affecté par des considérations telles que les autres médicaments consommés ou les contraintes économiques, puisque c'est le protocole d'étude qui détermine avec précision les modalités de traitement. Ainsi, pour évaluer l'effet et le coût des traitements au sein d'une vaste population hétérogène, une analyse rétrospective de cohortes est davantage appropriée.

### ***Efficacité énergétique***

Dans le cas plus spécifique de l'efficacité énergétique, les mêmes méthodes d'évaluation expérimentales et quasi-expérimentales sont utilisées au niveau international, notamment par l'*Office of Energy Efficiency and Renewable Energy* (« EERE ») aux États-Unis. Cependant, la nature des interventions en efficacité énergétique rend difficile l'utilisation d'une expérience aléatoire. En effet, bien des programmes d'efficacité énergétique sont de type volontaire, où les participants décident de leur propre chef de participer ou non au programme. Pour ce type d'intervention, il est très difficile de construire des groupes de participants et non-participants de façon aléatoire, puisque les participants disposent d'une importante caractéristique non-observable : un intérêt particulier pour l'efficacité énergétique. Même si un groupe de contrôle est construit, les résultats obtenus pourraient difficilement être généralisés à la population entière sans ajustement statistique approprié, puisque les personnes ne disposant pas d'intérêt pour l'efficacité énergétique ne se retrouvent dans aucun des groupes de comparaison<sup>18,19</sup>.

---

<sup>17</sup> Mann C.J. Observational research methods. Research design II: cohort, cross sectional, and case-control studies, *Emergency Medicine Journal*, 2003, Vol. 20, p. 54-60.

<sup>18</sup> Energy Efficiency and Renewable Energy. *EERE Guide for Managing General Program Evaluation Studies*, February 2006, 154 pages.

<sup>19</sup> National Action Plan for Energy Efficiency Leadership Group. *Model Energy Efficiency Program Impact Evaluation Guide*, November 2007, 152 pages.

L'évaluation de l'impact d'un programme d'efficacité énergétique repose donc habituellement sur des analyses quasi-expérimentales où le groupe de contrôle est défini en jumelant les non-participants aux participants selon certaines caractéristiques importantes. Les deux cohortes peuvent par la suite être comparées avant et après le programme pour déterminer l'impact net du programme.

D'autres méthodes moins coûteuses que les analyses expérimentales et quasi-expérimentales existent, mais celles-ci présentent une validité théorique inférieure, puisqu'elles ne tiennent pas compte des phénomènes externes au programme pouvant avoir eu une influence sur les comportements. Par exemple, voici ce que l'EERE indique concernant les deux méthodes alternatives : « après seulement » et études des participants seulement, donc sans groupe de contrôle :

*« 2. After-Only Comparison Group Design: A less defensible variant of this design (note: comparaison « avant-après » des participants et non-participants, donc la méthode des différences-en-différences) eliminates the before-program measurements and simply compares the two measurements at the same point after the participants participated in the program. (...)*

*3. Before-After Participant Group Time-Series Design: If you do not have a good non-participant comparison group, compare trends in participant behavior before and after participation. This design has less defensibility than the two described above, but costs less to implement and may be all that is feasible with the available data.<sup>20</sup> »*

Ainsi, selon l'EERE, une évaluation sans groupe de non-participants est inférieure aux principales méthodes d'évaluation expérimentales et quasi-expérimentales avec groupe de contrôle. Il s'agit également de l'opinion de Schalock, qui écrit qu'une évaluation de type « avant-après » des participants sans groupe de contrôle représente un « *low level of certainty in one's analysis*<sup>21</sup>. »

---

<sup>20</sup> EERE, page 33.

<sup>21</sup> Schalock, page 46.

## 4. REVUE DES ÉVALUATIONS DE PROGRAMMES D'EFFICACITÉ ÉNERGÉTIQUES DU QUÉBEC

Cette section compare les méthodes d'évaluation des impacts des programmes d'efficacité énergétique au Québec avec celles décrites précédemment. D'abord, les techniques utilisées dans certaines évaluations récentes sont passées en revue. Ensuite, la méthode proposée par l'Agence, soit le Protocole international de mesure et de vérification du rendement (« PIMVR »), sera aussi décortiquée.

### 4.1 Évaluations de programmes récentes

Cette section passe en revue certaines évaluations récentes de programmes d'efficacité énergétiques, en vue de répertorier les méthodes employées et d'examiner leur constance à travers les différents programmes évalués. À ces fins, nous avons examiné quatre évaluations du Plan global en efficacité énergétique (« PGEÉ ») pour Hydro-Québec Distribution (« HQD ») et une évaluation pour l'Agence (Éconologis). Le Tableau 2 présente un résumé des différentes méthodologies utilisées pour l'évaluation des programmes.

**Tableau 2**  
**Méthodes d'évaluation des programmes d'efficacité énergétique**

Évaluation	Présence d'un groupe contrôle	Ajustement sur non-participants	Impact = Impact P + Impact NP	Évaluation des traités si non-traités
1. Thermostats électroniques	Oui	Non	Oui	Non
2. Minuterie filtre de piscine	Oui	Non	Oui	Non
3. Diagnostic résidentiel	Oui	Oui	Non	Oui
4. Moteur et éclairage phase 1	N/A	N/A	N/A	N/A
5. Programme Éconologis	Non	Non	Non	Oui

À prime abord, il ressort que ces cinq évaluations utilisent des méthodes et analyses différentes pour l'évaluation des impacts :

- trois évaluations avaient constitué un groupe contrôle;
- une évaluation avait effectué un ajustement statistique (« *matching* ») sur les non-participants;
- deux évaluations ont conclu que l'impact du programme était égal aux économies des participants **plus** celles des non participants;
- deux évaluations ont tenté de déterminer le résultat des participants s'ils n'avaient pas été traités;

- une évaluation n'a présenté aucun calcul d'impact.

Avant d'entamer la revue de ces évaluations, une remarque mérite d'être formulée sur les non-participants. Par définition, un non-participant n'a pas été influencé par le programme. Par exemple, un individu qui achète un thermostat électronique mais ne reçoit pas de subvention en retour n'est pas un participant. Ainsi, dans la population des non-participants, certains achèteront des thermostats électroniques, même sans subvention. Par conséquent, on ne peut attribuer au programme que les économies découlant des achats réalisés directement en raison du programme, point à la ligne.

Maintenant, il se peut que certains achats aient été influencés par les communications d'HQD, mais cela est autre chose. En effet, si les participants et non-participants au programme des thermostats avaient été choisis de façon aléatoire, ou même appariés à l'aide d'une méthode de jumelage, cette influence des communications sera en principe égale des deux côtés. Pour calculer l'impact propre aux communications, il faudrait faire une évaluation séparée en procédant de la même façon : identifier un groupe de participants et un groupe de contrôle, et calculer l'impact causal des communications sur leurs habitudes de consommation d'énergie.

### ***1. Thermostats électroniques<sup>22</sup>***

L'objectif de ce programme est:

*« d'obtenir des économies d'énergie en stimulant la vente de thermostats électroniques. Le programme vise aussi à sensibiliser la clientèle aux avantages des thermostats électroniques et à favoriser l'installation et l'utilisation efficace des thermostats électroniques. » (p.10)*

L'impact du programme est calculé comme étant la somme des économies d'énergie des participants et des non-participants (P : 87 GWh + NP : 72GWh = 159 GWh). L'évaluation d'Hydro-Québec rapporte aussi des économies annuelles générées par Hydro-Québec dites « globales » (P : 97 GWh + NP : 184 GWh = 281 GWh), qui comprennent en plus des économies liées à l'installation de thermostats, des économies « générées par les communications du tronc commun ». Additionner l'impact des ménages participants et celui des ménages non-participants, puis rajouter à tous cela un impact global est une méthode pour le moins généreuse d'estimer l'impact de ce programme. En fait, tel qu'indiqué plus haut, cet impact en question devrait plutôt correspondre à la différence entre le

---

<sup>22</sup> HQD, *Évaluation du programme des thermostats électroniques - marché existant 2004 à 2006*, Octobre 2007.

résultat des participants moins le résultat qu'ils auraient obtenus sans participer au programme. Quant au tronc commun, son impact devrait être évalué séparément.

L'évaluation rapporte aussi les économies nettes par ménage (P = 670 kWh et NP = 90 kWh). Au niveau du ménage, le programme aurait donc un effet incrémental de 580 kWh, soit la différence entre les économies nettes des ménages participants et non-participants. Cela représenterait des économies totales de  $580 \text{ kWh} * 130\,000 \text{ ménages} = 75,4 \text{ GWh}$  et non 87 GWh d'économies tel qu'indiqué dans le document. Cependant, pour valider ce dernier résultat, il faudrait que les ménages participants présentent des caractéristiques identiques à celles des non-participants, ce qui est loin d'être le cas.

En effet, les ménages participants possèdent en moyenne sept thermostats électroniques alors que les non-participants en ont quatre en moyenne. Ainsi, les économies nettes d'énergie par thermostat des ménages sont  $670/7=95,7 \text{ kWh}$  pour les ménages participants et  $90/4=22,5 \text{ kWh}$  pour les ménages non-participants. La différence entre ces économies d'énergie par thermostat et par ménage pour les groupes P et NP donne  $95,7 - 22,5 = 73,2 \text{ kWh}$ . De plus, il est possible de croire que les ménages participants auraient réalisé des économies d'énergie plus fortes que les non-participants. Par conséquent, ces 22,5 kWh épargnés auraient pu s'élever, pour donner un exemple fictif, à 40 kWh. Cela aurait donné alors un impact net de 55,7 kWh par thermostat, soit environ 390 kWh par ménage participant, donc les deux tiers de l'impact rapporté. Sans affirmer que cela est l'impact effectif du programme, il reste que ce genre de considérations n'a pas été intégré à l'analyse.

## **2. Minuterics pour filtre de piscine<sup>23</sup>**

L'objectif de ce programme est:

*« d'obtenir des économies d'énergie électrique en stimulant la vente de minuterics de filtre de piscine. On visait à doubler le nombre de minuterics vendues avant la mise en place du programme (de 10 000 à 20 000). Le programme vise aussi à sensibiliser la clientèle aux avantages énergétiques des minuterics et à favoriser l'installation et l'utilisation efficace de ces dernières pour réduire le nombre d'heures de fonctionnement du filtre de la piscine. » (p. 10)*

Ici, l'évaluation fait état d'économies nettes de 376 kWh par minuterie subventionnée pour les participants. On y ajoute des économies liées « aux

---

<sup>23</sup> HQD, *Évaluation du programme des minuterics pour filtre de piscine 2004-2006*, Octobre 2007.

communications », ce qui fait passer ce résultat à 404 kWh. En parallèle, on rapporte des économies liées aux communications de 27 kWh par minuterie non-subventionnée (participant) et de 38 kWh par minuterie non-subventionnée (non-participant). Par contre, on ne fait état d'aucune économie nette des non-participants et qui aurait dû servir pour fins de comparaison avec l'impact observé chez les participants.

Encore ici, l'impact du programme est obtenu en additionnant l'impact des ménages participants et celui des ménages non-participants, puis en y rajoutant un impact global. Pour les raisons déjà décrites ci-haut, il s'agit d'une nette surestimation de l'impact réel du programme. Il est vrai que des effets induits ont effectivement été observés en raison de l'achat des minuterie additionnelles; toutefois, il faut séparer cet effet de l'évaluation du programme. Pour ce faire, il aurait fallu évaluer l'impact en constituant un seul groupe de participants, peu importe l'intensité de cette participation, et en le comparant à un groupe jumelé de non-participants, qui peuvent avoir réalisé aussi des économies de leur propre chef.

### ***3. Diagnostic résidentiel « Mieux consommer »<sup>24</sup>***

L'objectif de ce programme est:

*« de générer des économies en modifiant les comportements énergétiques qui ne sont pas optimaux ou en stimulant le remplacement d'équipements moins efficaces ou l'implantation de mesures nécessitant un investissement. De façon secondaire, le rapport permet aussi aux clients de mieux comprendre leur consommation d'électricité. » (p.10)*

Il est ici question d'économies réalisées grâce à un vaste programme de diagnostic résidentiel. Les méthodes employées par l'évaluateur comprennent des entrevues individuelles, des sondages téléphoniques, l'analyse de plus de 600 000 factures et une analyse technique des économies brutes.

Des cinq rapports révisés, il s'agit clairement de l'évaluation la plus conforme aux meilleures pratiques économiques. Deux méthodes d'évaluation sont comparées, soit 1) la méthode de facturation et 2) l'analyse technique. L'évaluation des économies unitaires nettes par ménage selon la méthode de facturation donne des économies de 213 kWh, alors que l'analyse technique rapporte des économies unitaires

---

<sup>24</sup> HQD, *Évaluation du programme du diagnostic résidentiel « Mieux consommer » 2004 à 2006*, Octobre 2007.

nettes par ménage de 212 kWh<sup>25</sup>. Entre autres, dans la méthode de facturation, un groupe témoin a été constitué (deux membres pour chaque participant) en se basant sur la consommation annualisée et le secteur Hydro-Québec. Ensuite, une validation croisée de l'impact énergétique net du programme combine ces deux estimations, livrant un impact final de 213 kWh.

La démarche se poursuit en estimant l'impact d'un effet de chevauchement avec le programme des minuteriers et celui des thermostats. L'économie nette de 213 kWh par ménage est ramenée à 195 kWh par ménage participant après la prise en compte de cet effet. Finalement, en multipliant les économies nettes par ménage par le nombre de ménages, on obtient 157 GWh d'économies nettes totales. Puis, les économies provenant des communications du tronc commun et des programmes spécifiques sont ajoutées, soit 65 GWh, pour des économies annuelles finales de 222 GWh.

Cette dernière évaluation est clairement plus étoffée que les deux premières. On y emploie des méthodes statistiques avancées, semblables à celles décrites à la section 3, à quelques variantes près. Encore ici, on ajoute des montants de tronc commun aux économies nettes estimées plutôt que d'isoler l'impact propre au programme. À ce sujet, nous croyons qu'il est préférable d'être plus conservateurs et de n'isoler que le seul effet de l'intervention. S'il faut faire une évaluation de l'impact du tronc commun, il serait préférable qu'elle soit réalisée séparément.

#### ***4. Moteur et éclairage phase 1***<sup>26</sup>

Dans cette évaluation, on indique que :

*« la présente intervention de l'évaluateur est la première d'une série d'interventions sur trois ans. Il s'agit, pour le moment, de procéder à une évaluation de processus seulement (Théorie du programme). Les évaluations de marché et d'impact énergétique seront réalisées plus tard. »* (p. 1)

Le document porte principalement sur les objectifs fixés, la philosophie de conception, la description des deux volets et un exposé de la Théorie du programme. Il traite aussi du modèle logique du programme et de son organisation générale. Aucune estimation de l'impact n'est donc réalisée.

---

<sup>25</sup> Il s'agit d'économies brutes de 450 kWh contrebalancées par un taux d'opportunisme de 47 %, représentant la proportion des participants ayant reçu chacune des recommandations qui auraient de toute façon implanté la mesure.

<sup>26</sup> HQD, *Évaluation de deux programmes du marché « Affaires », Programme Produits efficaces Volets Moteurs Phase 1 et Éclairage Phase 1*, Janvier 2007.

## **5. Évaluation du programme Éconologis (2004-2006) destiné aux ménages à budget modeste<sup>27</sup>**

Éconologis est un programme destiné aux ménages à budget modeste où les participants doivent payer une facture de chauffage et avoir des revenus en deçà du seuil fixé annuellement par l'AEÉ:

*« Le programme Éconologis se compose de deux volets. Le premier offre des conseils personnalisés aux ménages ainsi que l'installation de produits économiseurs d'énergie pour améliorer l'efficacité énergétique de leurs logements. Le deuxième volet consiste en l'installation gratuite de thermostats électroniques par des électriciens qualifiés dans les logements chauffés par des plinthes électriques munies de contrôles muraux. » (p.1)*

Les impacts énergétiques sont estimés dans ce rapport en comparant les factures des participants seulement, sans groupe contrôle donc, avant et après l'implantation du programme. Cette méthode est donc analogue à celle du PIMVR, qui est examinée à la section suivante.

*« Les impacts énergétiques ont été estimés selon l'approche de l'analyse de la facturation qui utilise les données de plusieurs années de consommation des participants afin de comparer leur consommation énergétique avant et après leur participation au programme et, ainsi, estimer les diminutions de consommation générées par le programme.» (p.1)*

Avec une telle façon d'évaluer le programme, comment savoir si les économies d'énergies sont réellement liées à l'intervention à l'étude? D'autres phénomènes parallèles au programme (lois, changement d'habitude, nouveaux produits sur le marché, récessions, etc.) peuvent être responsables des économies d'énergie puisqu'on n'observe pas les participants durant la même période de temps. Si on choisit de comparer les coûts pré-programme versus post-programme, il faut aussi inclure à l'analyse un groupe contrôle et calculer les mêmes coûts avant et après le programme. Cela reviendrait tout simplement à calculer une « différences-en-différences », tel que présenté à la section précédente.

En somme, les cinq évaluations des impacts des programmes d'efficacité énergétiques (4 PGEÉ, 1 Agence) utilisent des méthodes variées et non uniformes. Certaines ne comprennent pas de groupe de contrôle ou, lorsqu'il y en a un, il n'est pas toujours utilisé correctement. En effet, alors que l'évaluation du programme du diagnostic résidentiel « Mieux

---

<sup>27</sup> AEÉ, HQD, Gaz Métro et Gazifère, *Évaluation du programme Éconologis (2004-2006) destiné aux ménages à budget modeste*, 18 juillet 2008.

consommer » compare le groupe de contrôle au groupe de participants comme il se doit, les évaluations du programme des thermostats électroniques et des minuteries pour filtres de piscine additionnent les économies d'énergies du groupe participant et du groupe contrôle. Quant à Éconologis, l'évaluation est de type avant-après sur les participants, sans groupe de contrôle, ce qui n'est pas idéal en contexte d'évaluation des impacts d'un programme.

## 4.2 Le PIMVR

Dans le cadre de la présente demande, l'Agence n'a pas indiqué de préférence pour l'une ou l'autre des méthodes déjà appliquées dans les évaluations d'HQD, de Gaz Métro ou de Gazifère. En fait, l'Agence compte plutôt appliquer le Protocole international de mesure et de vérification du rendement (« PIMVR ») produit par l'*Efficiency Valuation Organization* (« EVO »)<sup>28</sup>, affirmant qu'il s'agit du « protocole le plus reconnu et le plus utilisé à travers le monde »<sup>29</sup>. En fait, il s'agit certes d'une norme reconnue en « mesure et vérification », ce qui ne veut toutefois pas dire que ces méthodes sont la référence incontestée pour l'évaluation quantitative des impacts de programmes d'efficacité énergétique destinés à des groupes de clients ou d'entreprises.

L'EVO est un regroupement d'acteurs du secteur de l'efficacité énergétique dont la mission est de développer et promouvoir un protocole international de mesure et vérification permettant de mesurer les économies d'énergie résultant de projets d'efficacité énergétique. Cette organisation est née d'un regroupement de bénévoles chapeautés par le *Department of Energy* américain et le *Lawrence Berkeley National Laboratory*. Cette organisation est financée par des compagnies pétrolières et autres entreprises impliquées dans le domaine de l'énergie. Le conseil d'administration et le comité technique du PIMVR sont principalement composés d'acteurs du secteur énergétique.

Le PIMVR présente une méthode qui, sans être incorrecte, n'est clairement pas la meilleure façon d'évaluer l'impact de programmes énergétiques. En effet, la méthode utilisée s'applique davantage à l'évaluation d'une modification appliquée à une machine ou un processus de production dans un endroit connu et contrôlé qu'à l'évaluation d'un programme d'efficacité énergétique touchant des milliers d'agents économiques à l'échelle du Québec.

---

<sup>28</sup> Efficiency Valuation Organization *Protocole International de Mesure et Vérification du Rendement : Concepts et options pour l'évaluation des économies d'énergie et d'eau*, Volume 1, Avril 2007.

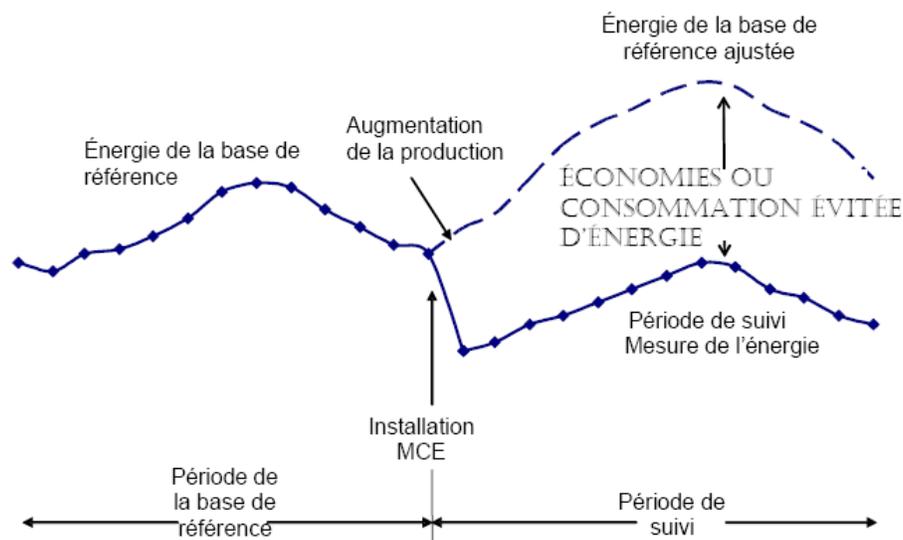
<sup>29</sup> AÉE-8, Document 1, page 212, ligne 7.

La méthode du PIMVR consiste à comparer la consommation d'énergie des participants avant et après l'implantation d'un programme, sans groupe de contrôle, en faisant les ajustements nécessaires pour tout « changement de condition » survenu entre la période de référence (avant) et la période de suivi (après). La méthode tente ainsi d'isoler les effets énergétiques d'un programme d'économie d'énergie des effets d'autres éléments pouvant avoir aussi causé certains changements.

La Figure 2 illustre cette méthode, qui comprend les étapes suivantes :

- un modèle économétrique de la consommation durant la période de référence est estimé;
- le même modèle est utilisé pour « prolonger » le calcul durant la période de suivi, à l'aide des faits physiques identifiables et observés lors de cette période de suivi<sup>30</sup>;
- l'impact est calculé comme étant la différence entre cette consommation projetée et celle observée réellement, donc suite à la participation au programme.

**Figure 2**  
**Méthode d'évaluation du PIMVR**



Source : PIMVR (2007)

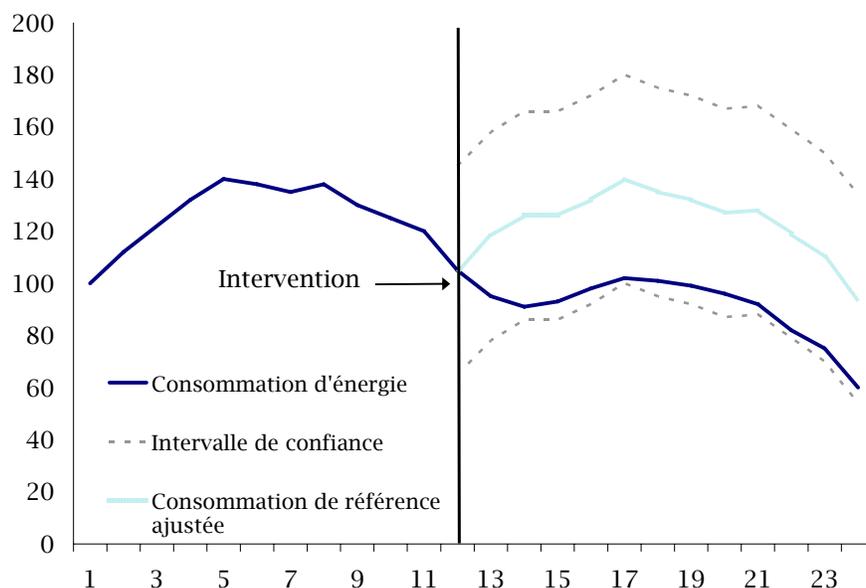
Deux types d'ajustements sont proposés : la consommation d'énergie évitée et les économies normalisées. Après avoir identifié les variables

<sup>30</sup> Ces faits peuvent être périodiques (température au cours d'une année) ou non périodiques (agrandissement d'un bâtiment).

ayant un impact sur la consommation d'énergie, la consommation d'énergie évitée est calculée comme étant la différence entre l'énergie consommée durant la période de suivi et celle projetée à l'aide du modèle estimé durant la période de référence et employant les valeurs des variables indépendantes pendant la période de suivi. La deuxième technique consiste à établir un ensemble de conditions considérées « normales » pour un cycle déterminé. Les économies normalisées sont donc mesurées comme étant la différence entre l'énergie utilisée pendant la période de référence sous des conditions normales et celle observée dans la période de suivi sous les mêmes conditions normalisées.

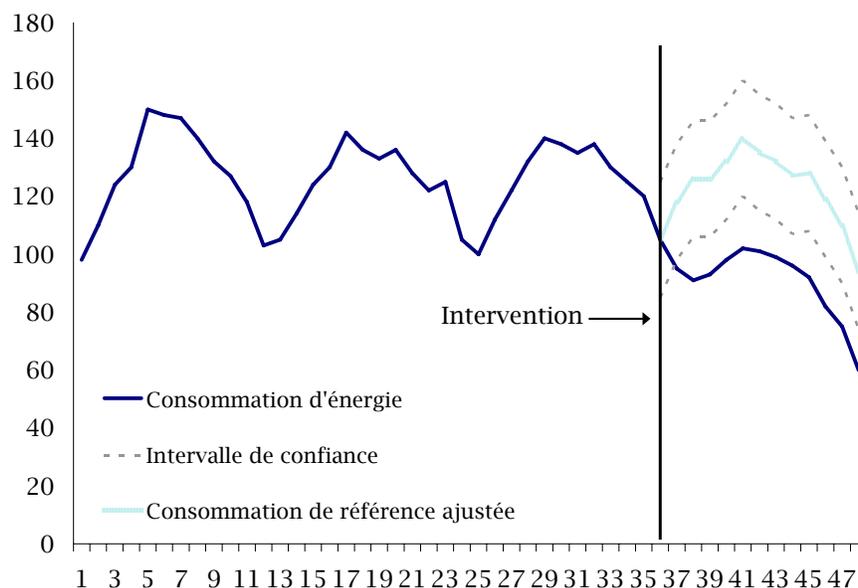
D'un point de vue économétrique, cette méthode présente quelques lacunes lorsque appliquée dans un contexte d'évaluation d'impact d'un programme. Le protocole fait mention de l'incertitude statistique en annexe, mais il n'est fait aucunement mention de l'impact de la longueur de la période de base de référence sur cette incertitude. En effet, il est même précisé qu'un seul cycle est suffisant pour construire l'analyse, même si celui-ci est aussi court qu'une semaine. Or, plus la période d'analyse historique est courte, plus l'incertitude sera grande. Les deux figures suivantes illustrent ce phénomène en montrant l'intervalle de confiance statistique autour de la projection. Plus la période de référence est courte, plus l'intervalle de confiance est large et moins l'impact est précis. En fait, il est possible de concevoir que cet intervalle soit trop large pour qu'il soit possible de tirer une conclusion sur cet impact, comme à la Figure 3.a.

**Figure 3.a**  
**Intervalle de confiance avec peu de données historiques**



L'avantage d'un petit intervalle de confiance (Figure 3.b) est de pouvoir déterminer avec plus de précision s'il y a eu un impact de l'intervention ou non. Les deux figures présentent les mêmes données sur la consommation d'énergie et sur la consommation de référence ajustée. Si on ne dispose que de peu de données historiques, l'intervalle de confiance est beaucoup plus large (Figure 3.a). Par conséquent, il n'est pas possible de conclure que l'intervention a eu un impact significatif, car la consommation de référence ajustée se retrouve dans l'intervalle de confiance de la consommation d'énergie observée. Or, avec une série historique plus longue, il est possible d'obtenir un intervalle de confiance plus étroit, permettant ainsi d'identifier une différence statistique entre la consommation de référence ajustée et la consommation observée.

**Figure 3.b**  
**Intervalle de confiance avec beaucoup de données historiques**



Dans le cadre de l'évaluation de programmes d'efficacité énergétique, il peut être difficile d'obtenir de longues séries temporelles sur la consommation passée des ménages et des entreprises. Entre autres, les déménagements sont un phénomène causant un bris temporel dans les données. De plus, les ménages et entreprises effectuent annuellement des travaux importants de rénovation, ce qui modifie sans cesse l'environnement observé et n'est pas nécessairement pris en compte, à moins de disposer de cette information.

De plus, il est impossible de savoir avec précision les conditions qui peuvent avoir eu un impact sur la consommation d'énergie, étant donné que l'évaluateur (l'Agence) n'est pas le même que celui qui consomme (les participants). En effet, le PIMVR est construit de manière à être appliqué

par une entreprise dans un environnement connu et contrôlé, par exemple lorsqu'elle intègre de nouveaux procédés dans sa chaîne de production ou qu'elle modifie l'espace dans lequel sa production s'effectue (agrandissement, rénovation des fenêtres, etc.). Il est précisé au PIMVR qu'il faut « inclure seulement les périodes de temps pour lesquelles tous les faits, fixes et variables, régissant l'énergie, sont connus »<sup>31</sup>. Or, l'Agence ne dispose pas de telles informations détaillées sur les habitudes, les comportements et les environnements physiques des individus et des entreprises. Ainsi, ce protocole n'est pas destiné à évaluer l'impact causal de programmes visant des agents économiques hétérogènes.

Un autre aspect problématique de la méthode du PIMVR est qu'elle ne permet pas de modification dans la structure de consommation d'énergie entre les périodes étudiées. Essentiellement, la méthode utilise les coefficients estimés durant la période de référence pour les appliquer sur des « conditions » différentes, celles de la période de suivi. Cependant, il est fort possible que face aux mêmes conditions, les mêmes individus réagissent différemment en termes de consommation d'énergie durant une période de temps différente. Par exemple, la structure de décision de consommer de l'essence s'est certainement modifiée entre une période où le prix du litre était à 0,90 \$ et une autre où ce prix était à 1,45 \$. Cette méthode ne permet donc pas de capter des modifications structurelles des habitudes de consommation d'énergie entre la période de référence et la période de suivi. Entre autres, plus la période de suivi est longue par rapport à la période de référence, plus cette lacune est présente.

Finalement, une des caractéristiques associées aux séries temporelles est la présence d'autocorrélation, qui survient lorsque les « résidus » ou « termes d'erreur » de la régression sont corrélés à travers les années. Cela signale un effet de persistance, ce qui ne respecte pas les conditions de base d'utilisation des régressions linéaires. Il a été démontré que l'utilisation de régressions linéaires avec des données en séries temporelles entraîne une surestimation de l'impact d'une intervention, comparativement à des méthodes plus raffinées qui prennent en compte l'autocorrélation<sup>32</sup>. Le PIMVR discute quelque peu des problèmes d'incertitude dans l'estimation de régressions linéaires, mais ne parle aucunement du problème d'autocorrélation associé aux séries temporelles. Si l'Agence se basait sur ce guide, elle risquerait de ne pas traiter ces problèmes, ce qui produirait des résultats non valides statistiquement.

Bien que la méthodologie favorisée par le PIMVR soit économiquement valable dans certains contextes précis, son application n'est pas indiquée

---

<sup>31</sup> PIMVR p. 17

<sup>32</sup> Bertrand M. et al. How much should we trust differences-in-differences estimates?, *The Quarterly Journal of Economics*, February 2004, p. 249-275.

pour l'évaluation de programmes visant des groupes d'individus ou d'entreprises. D'abord, plusieurs éléments sont problématiques quant à l'application de la méthode proposée, par exemple la longueur de la période de référence, l'impossibilité d'ajustement structurel et le traitement de l'autocorrélation. Plus fondamentalement, ce protocole n'est pas conçu pour évaluer l'impact de programmes d'efficacité énergétique. En effet, les techniques « avant-après » sans groupe de contrôle sont parfois utiles, notamment dans un environnement unique et bien contrôlé, mais clairement inférieures aux méthodes expérimentales et quasi-expérimentales décrites à la section 3 pour fins d'évaluation quantitative des impacts de programmes destinés à plusieurs individus ou entreprises à travers le Québec.

## 5. CONCLUSION ET RECOMMANDATIONS

Ce rapport a examiné l'évaluation quantitative des impacts des programmes d'efficacité énergétique.

D'abord, la nature et la valeur de l'impact des programmes d'efficacité énergétique ont été présentées. À ce sujet, la mesure d'impact appropriée des programmes est l'énergie libérée ou non consommée. Quant à la rentabilité, elle correspond à la valeur économique de l'ensemble des actions menant à la libération de cette énergie. Cette valeur économique est le mieux représentée par un TCS réduit, donc ne présentant que la valeur des émissions non émises, évaluées à 15 \$ la tonne CO<sub>2</sub>e.

Ensuite, les méthodes d'évaluation de ces impacts ont été exposées et leur application par les organismes responsables d'évaluation a été décrite. Les méthodes expérimentales, le « *gold standard of evaluation research* » selon l'EERE<sup>33</sup>, et quasi-expérimentales sont les plus reconnues et performantes pour fins d'évaluation quantitative des impacts de programmes. Ces méthodes consistent à comparer l'évolution des résultats avant et après le programme, à la fois pour les participants et un groupe de contrôle de non-participants. Ces méthodes représentent la référence absolue en la matière dans la littérature économique, et sont employées au Canada, aux États-Unis et à l'international pour déterminer précisément l'impact des interventions. À ce sujet, plutôt que le PIMVR, l'Agence devrait élaborer ses propres lignes directrices d'évaluation de programmes, à l'instar du guide de l'EERE.

Enfin, certaines différences entre ces meilleures pratiques et l'évaluation des programmes d'efficacité énergétique au Québec ont été relevées, aussi bien pour des évaluations récentes que pour le protocole envisagé par l'Agence. À ce sujet, les cinq évaluations analysées ont utilisé des méthodes variées et non uniformes. Plusieurs ne comprennent aucun groupe de contrôle ou, lorsqu'il y en a un, celui-ci n'est pas toujours utilisé correctement. Quant au PIMVR, son application n'est pas indiquée pour l'évaluation de programmes visant des groupes d'individus ou d'entreprises. En effet, les techniques « avant-après » sans groupe de contrôle sont parfois utiles, notamment dans un environnement unique et bien contrôlé, mais clairement inférieures aux méthodes expérimentales et quasi-expérimentales pour fins d'évaluation des impacts de programmes destinés à plusieurs individus ou entreprises à travers le Québec.

En somme, voici mes recommandations à la Régie sur ces questions :

---

<sup>33</sup> EERE, page 32.

- **Recommandation 1** : Présenter les économies d'énergie réalisées par les programmes, exprimées en unité comparable d'énergie libérée, tous programmes confondus et par source d'énergie, à la fois globalement et sur une base unitaire, donc par agent et par année
- **Recommandation 2** : Utiliser le TCS réduit (incluant seulement la valeur des réductions d'émissions de CO<sub>2</sub>e) pour mesurer la rentabilité des programmes
- **Recommandation 3** : Élaborer des lignes directrices pour la conduite des évaluations quantitatives des impacts de programmes, privilégiant l'emploi de méthodes expérimentales et quasi-expérimentales à ces fins