# L'(ABSENCE D') IMPACT DE L'IMPACT : POURQUOI LES ÉVALUATIONS D'IMPACT CONDUISENT RAREMENT À UNE PRISE DE DÉCISION POLITIQUE FONDÉE SUR LES FAITS

Jean-Louis Arcand

De Boeck Supérieur | « Revue d'économie du développement »

2013/4 Vol. 21 | pages 193 à 218

ISSN 1245-4060
ISBN 9782804185619
DOI 10.3917/edd.274.0193

# L' (absence d') impact de l'impact : pourquoi les évaluations d'impact conduisent rarement à une prise de décision politique fondée sur les faits

*The (Lack of) Impact of Impact: Why Impact Evaluations Seldom Lead to Evidence-based Policymaking*

Jean-Louis Arcand*

Une question récurrente qui déroute plusieurs chercheurs et certains responsables politiques est celle de savoir pourquoi les évaluations d'impact, qui sont devenues monnaie courante dans le domaine du développement, ont si peu d'impact sur la prise de décision à proprement parler. Dans cet article, j'étudie l'impact des évaluations d'impact. Je fais appel à un cadre bayésien simple emboîté dans un modèle standard reposant sur une fonction de « contest success ». Avec ce modèle de concurrence entre des décideurs anti-évaluation, des décideurs bayésiens et des évaluateurs fréquentistes, je montre que la probabilité d'annulation d'un programme est une fonction décroissante de l'impact estimé par l'évaluation et de la croyance *a priori* sur la base de laquelle le programme a été initialement approuvé. En outre, la probabilité d'annulation est une fonction décroissante de l'efficacité de l'influence exercée par les évaluateurs fréquentistes. Dans la mesure où il est fort probable que cette efficacité en termes de lobbying des évaluateurs fréquentistes soit proche de zéro dans la vraie vie, la probabilité d'annulation d'un programme qui avait été approuvé au départ, bien qu'il soit entaché d'une évaluation très négative, est extrêmement faible. Le modèle fournit ainsi une explication possible de la raison pour laquelle les évaluations d'impact ont si peu d'impact sur la prise de décision, et pourquoi elles ont si peu contribué à la prise de décision fondée sur les faits.

Mots-clés : Évaluation d'impact, analyse bayésienne, fonction de « contest success ».

* Directeur du Centre Finance et Développement, Professeur au département d'économie internationale, Institut de hautes études internationales et du développement, Genève, Suisse et Senior Fellow à la FERDI, Clermont-Ferrand, France.
Email : jean-louis.arcand@graduateinstitute.ch
http://graduateinstitute.ch/economics/faculty/arcand

Classification JEL : O12, D04, D72, C11, C21, C72.

A recurring puzzle to many academics and some policymakers is why impact evaluations, which have become something of a cottage industry in the development field, have so little impact on actual policymaking. In this paper, I study the impact of impact evaluations. I show, in a simple Bayesian framework embedded within a standard contest success function-based model of competition amongst anti-evaluation policymakers, Bayesian policymakers, and frequentist evaluators, that the likelihood of a program being cancelled is a decreasing function both of the impact estimated by the evaluation and of the prior on whose basis the program was approved to begin with. Moreover, the probability of cancellation is a decreasing function of the effectiveness of the influence exerted by frequentist evaluators. Since the latter's effectiveness in terms of lobbying in favor of their findings in the real world is likely to be close to zero, the likelihood of cancelling a program that was approved in the first place, despite its suffering a highly negative evaluation, is extremely low. The model thus provides one possible explanation for why impact evaluations have so little impact in the realm of decisionmaking, and why they have contributed so little to evidence-based policymaking.

Key words: Impact evaluation, Bayesian analysis, contest success functions.

Tout chercheur universitaire qui a été impliqué dans des évaluations d'impact sur le terrain peut raconter des histoires d'horreur à propos du fait que ses résultats n'ont pas été pris en compte par les décideurs pour décider s'il faut maintenir ou pas un programme en cours. Ma propre expérience personnelle peut en être une bonne illustration. Il y a plusieurs années, j'ai effectué une évaluation d'impact pour un grand programme financé par un important donateur multilatéral. Après trois ans de travail, nous avons trouvé des effets négatifs statistiquement significatifs du programme sur notre variable principale. La simple logique budgétaire aurait voulu que le donateur prenne en compte ces résultats issus d'une évaluation d'impact rigoureuse, mais cela n'a pas été le cas. À la fin du cycle de projet et lorsque vint le temps pour le donateur de décider s'il fallait maintenir ou annuler (comme je le recommandais vivement) le programme, il décida de ne pas annuler. Pourquoi ?

Dans cet article, j'utilise des arguments bayésiens de base et un modèle simple de concurrence entre trois types de décideurs pour montrer qu'il est fortement probable qu'un programme qu'un évaluateur fréquentiste aurait recommandé d'annuler, mais qui avait été approuvé au départ, sera maintenu. Ce résultat s'explique par les interactions stratégiques entre trois types de décideurs qui font pression en faveur de leurs préférences en termes de résultats : un évaluateur fréquentiste, un décideur politique qui, sur la base de ses croyances, a initié le programme au départ, et un décideur bayésien qui tente de combiner les croyances et les résultats de l'évaluation d'impact de façon statistiquement rationnelle.

Les acteurs du développement ont toujours été conscients du manque de prise de décision fondée sur des faits. Par exemple, l'article de 2006 du

« Center for Global Development Evaluation Gap Working Group » intitulé *When Will We Ever Learn?*, qui a conduit à la mise en place de l'Initiative Internationale pour l'Evaluation d'Impact (3ie, voir CGD (2006))

> « a présenté l'absence d'évaluations d'impact rigoureuses comme l'élément manquant dans l'apprentissage des efforts de développement social. Il a préconisé une nouvelle approche pour évaluer l'aide, approche qui apporterait de plus grandes précision et crédibilité aux évaluations d'impact, et, par extension, aux politiques et à la pratique du développement. Il a clairement souligné le défi : d'ici 10 ans à compter de la publication du rapport, il est nécessaire que de plus nombreuses et de plus rigoureuses évaluations d'impact soient mises en place si nous souhaitons de solides faits à l'appui de la prise de décision ». (ODI 2011).

Malheureusement pour la prise de décision fondée sur les faits, si plus d'évaluations d'impact rigoureuses sont disponibles aujourd'hui par rapport à 2006, la qualité de la prise de décision a néanmoins connu très peu, voire pas, d'amélioration.

L'intuition sous-jacente au modèle présenté dans cet article est extrêmement simple. Soit $\widehat{\beta}$ l'impact du programme sur la variable d'intérêt mis en évidence par l'évaluation d'impact et, sans perte de généralité, normalisons le coût du programme à zéro. Afin d'être mis en œuvre, un programme doit être approuvé ; il doit donc y avoir de bonnes raisons *a priori* de croire que son impact l'emportera sur son coût. En termes statistiques, cela signifie que les décideurs ont une croyance a priori (notée $\mu_\beta$) concernant le programme, et cette croyance doit être telle que *ex ante*, $\beta$ est supposé être statistiquement plus grand que zéro ($\mu_\beta > 0$)[1]. La théorie bayésienne nous apprend que les croyances révisées (notées $\mu_\beta^*$) d'un décideur bayésien constituent une combinaison convexe de $\widehat{\beta}$ et $\mu_\beta$. Si la croyance a priori est élevée et positive, le résultat $\widehat{\beta}$ de l'évaluation d'impact doit être suffisamment négatif pour que la croyance a posteriori $\mu_\beta^*$ soit négative. Ainsi si la croyance a priori est suffisamment positive, même une évaluation d'impact significativement négative peut ne pas être suffisante pour générer une croyance a posteriori négative et statistiquement significative. Dans ce cas, il serait rationnel pour le décideur

1 D'un point de vue opérationnel, si l'on considère l'exemple de la Banque Mondiale, c'est ce qui se fait en pratique dans les documents d'évaluation de projet (PAD) qui exposent les bénéfices probables du programme (et le fait que ceux-ci excèdent les coûts) afin d'obtenir l'approbation du Conseil d'administration de la Banque pour l'instrument de prêt associé.

bayésien de s'opposer à l'annulation du programme, même si l'évaluateur fréquentiste est en faveur de l'annulation. Si nous combinons le besoin d'un résultat d'évaluation d'impact très négatif avec l'inertie générée par le fait que la plupart des décideurs qui ont été à l'initiative du programme et qui fondent souvent entièrement leur prise de décision sur leurs croyances a priori ont un pouvoir de décision important, il n'est pas surprenant que l'annulation soit relativement peu probable.

L'article s'organise comme suit : dans la section 1 je pose le cadre de l'analyse en recourant à des termes bayésiens simples. J'établis les conditions sous lesquelles le programme sera approuvé à la base et étudie les conditions correspondantes sous lesquelles un chercheur-évaluateur fréquentiste et un décideur bayésien seraient en faveur de maintenir ou de mettre fin au programme. En partant des principes de base, je montre que ces conditions sont données par les résultats de tests de significativité standards, à savoir les tests de Student. Une fois que l'on a établi les préférences de chaque agent comme une fonction des résultats de l'évaluation d'impact, la section 2 ancre le cadre initial dans un modèle de concurrence simple exprimé en termes de « contest success function ». Cette méthode constitue l'approche qui est typiquement adoptée dans la littérature des sciences politiques. J'établis plusieurs résultats. Premièrement, la probabilité d'annulation est une fonction décroissante de $\widehat{\beta}$ et $\mu_\beta$. Deuxièmement, la probabilité d'annulation est une fonction croissante des variances de $\mu_\beta$ et de la muette indiquant l'état du traitement. Troisièmement, la probabilité d'annulation est une fonction décroissante de l'efficacité de l'influence exercée par les évaluateurs fréquentistes sur la prise de décision. La section 3 conclut en faisant part de quelques réflexions en ce qui concerne les limites inhérentes au fait de baser la prise de décision sur l'effet moyen de traitement (EMT), ce qui revient à supposer une fonction de bien-être social avec neutralité vis-à-vis du risque.

# 1 UN MODÈLE BAYÉSIEN SIMPLE

Lorsque nous effectuons une évaluation d'impact, nous obtenons le coefficient estimé de l'impact noté $\widehat{\beta}$. Par souci de simplicité, supposons que ce coefficient découle d'un essai randomisé contrôlé (ERC) où l'effet du traitement est estimé à l'aide d'une simple régression par les moindres carrés ordinaires. Cette régression prend la forme suivante :

$$Y_i = D_i\beta + U_i \tag{1}$$

où $Y_i$ est la variable d'intérêt, $D_i$ est la muette indiquant l'état du traitement (elle vaut 1 lorsque l'individu $i$ est traité par le programme, et 0 sinon), et $U_i$ est le terme d'erreurs qui satisfait les conditions de Gauss-Markov usuelles, avec $U_i \sim N(0,\sigma^2)$[2]. La taille de l'échantillon est notée $n$.

## 1.1 Croyances conjuguées

Dans le modèle bayésien linéaire standard décrit par exemple dans O'Hagan (1994), la croyance conjuguée $p(\beta,\sigma^2)$ pour $\beta$ et $\sigma^2$ est spécifiée par la loi normale gamma inverse ($N\Gamma^{-1}$) :

$$p(\beta,\sigma^2) = p(\beta \mid \sigma^2)\, p(\sigma^2) = N(\mu_\beta,\sigma^2 V_\beta)\,\Gamma^{-1}(a,b) = N\Gamma^{-1}(\mu_\beta,V_\beta,a,b)$$

Ou, de manière plus explicite :

$$N(\mu_\beta,\sigma^2 V_\beta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 V_\beta}}\, exp\left\{-\frac{(\beta-\mu_\beta)^2}{2\sigma^2 V_\beta}\right\}$$

$$\Gamma^{-1}(a,b) = \frac{b^a exp\left\{-\dfrac{b}{\sigma^2}\right\}}{\sigma^{2(1+\alpha)}\Gamma(a)}$$

Ceci donne :

$$p(\beta,\sigma^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 V_\beta}}\, exp\left\{-\frac{(\beta-\mu_\beta)^2}{2\sigma^2 V_\beta}\right\} \frac{b^a exp\left\{-\dfrac{b}{\sigma^2}\right\}}{\sigma^{2(1+\alpha)}\Gamma(a)}$$

En intégrant $\sigma^2$, l'expression devient :

$$p(\beta) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 V_\beta}}\, exp\left\{-\frac{(\beta-\mu_\beta)^2}{2\sigma^2 V_\beta}\right\} \frac{b^a exp\left\{-\dfrac{b}{\sigma^2}\right\}}{\sigma^{2(1+\alpha)}\Gamma(a)}\, d\sigma^2$$

$$= \frac{2^a (bV_\beta)^a \left[2bV_\beta + (\beta-\mu_\beta)^2\right]^{-\frac{1}{2}-a} \Gamma\left(\frac{1}{2}+a\right)}{\sqrt{\pi}\,\Gamma(a)} \qquad (2)$$

[2] Aucun des résultats qui suivent ne dépend du fait que $D_i$ soit une variable muette : les résultats restent valables dans le cas où $D_i$ est une mesure continue de l'intensité du traitement.

Nous pouvons réécrire l'expression (2) de la façon suivante :

$$\frac{2^{a}(bV_{\beta})^{a}\left[2bV_{\beta}+\left(\beta-\mu_{\beta}\right)^{2}\right]^{-\frac{1}{2}-a}\Gamma\left(\frac{1}{2}+a\right)}{\sqrt{\pi}\Gamma(a)}$$

$$=\frac{2^{-\frac{1}{2}+\frac{1}{2}(1+2a)}\left(\frac{a}{2a+\frac{a\left(\beta-\mu_{\beta}\right)^{2}}{bV_{\beta}}}\right)^{\frac{1}{2}(1+2a)}}{\sqrt{a}\sqrt{\frac{b}{a}V_{\beta}}B(a,\frac{1}{2})} \tag{3}$$

À droite de l'égalité dans l'expression (3), nous avons une distribution du $t$ de Student avec une moyenne de $\mu_{\beta}$, un paramètre d'échelle $\frac{b}{a}V_{\beta}$ et $2^{a}$ degrés de liberté[3]. La distribution pour la croyance inconditionnelle de $\beta$, qui correspond à l'information détenue par les décideurs quand ils décident de financer le programme, implique que la moyenne et la variance que les décideurs ont en tête *ex ante facto* sont données par :

$$E[\beta]=\mu_{\beta}, Var[\beta]=\frac{b}{a-1}V_{\beta} \tag{4}$$

## 1.2 Approbation préalable du programme

Pour que le programme soit approuvé, il faut que les décideurs en question soient convaincus que son impact sera plus grand que son coût d'un point de vue statistique. Ceci peut être formalisé à l'aide de l'inégalité suivante :

$$E[\beta]-t_{\alpha,n-1}\sqrt{Var[\beta]}=\mu_{\beta}-t_{\alpha,n-1}\sqrt{\frac{b}{a-1}Var[\beta]}>0$$

où le terme $t_{\alpha,n-1}$ représente la valeur critique du test de Student avec un intervalle de confiance à $1-\alpha$ et $n-1$ degrés de liberté[4]. Intuitivement, le décideur approuve le programme si le zéro est strictement situé *en dessous* de la borne inférieure de l'intervalle de confiance à $1-\alpha$.

3 $B(.,.)$ est la fonction bêta d'Euler.

4 Par exemple, pour un tel test unilatéral sur échantillon infini et un intervalle de confiance à 97.5 %, $t_{a,n-1}\approx 1.96$.

## 1.3 Croyances a posteriori suite à l'évaluation d'impact

Par suite, d'après le théorème de Bayes, la distribution a posteriori jointe $p\left(\beta,\sigma^2 \mid Y\right)=\dfrac{p\left(\beta,\, \sigma^2\right)p\left(Y \mid \beta,\sigma^2\right)}{\iint p\left(\beta,\, \sigma^2\right)p\left(Y \mid \beta,\sigma^2\right)d\beta d\sigma^2}$ pouvant être calculée une fois que l'évaluation d'impact a été effectuée est donnée par une distribution $N\Gamma^{-1}$ avec des paramètres actualisés :

$$\mu_\beta^* = \frac{V_\beta^{-1}\mu_\beta + D'Y}{V_\beta^{-1}+D'D} = \frac{V_\beta^{-1}}{V_\beta^{-1}+D'D}\mu_\beta + \frac{D'D}{V_\beta^{-1}+D'D}\widehat{\beta}, V_\beta^* = \frac{1}{V_\beta^{-1}+D'D} \tag{6}$$

$$a^* = a+\frac{n}{2},\ \ b^* = b+\frac{1}{2}(\mu_\beta^2 V_\beta^{-1} + Y'Y - \mu_\beta^{*2}V_\beta^{*-1}) \tag{7}$$

En procédant comme pour l'expression (2), la distribution a posteriori marginale de $\beta$ est donnée par un $t$ de Student de la même forme que celui présenté à droite de l'égalité dans l'expression (3) :

$$E[\beta]=\mu_\beta^*, Var[\beta]=\frac{b^*}{a^*-1}V_\beta^*$$

## 1.4 L'impact de l'évaluation d'impact

Quel est l'impact de l'évaluation d'impact ?

### *Les évaluateurs fréquentistes*

Pour la plupart des chercheurs qui ne prennent pas en compte le fait que le programme a été approuvé à la base, ce qui importe est le coefficient $\widehat{\beta}=\dfrac{D'Y}{D'D}$ estimé par les moindres carrés à partir de la régression (1). Si $\widehat{\beta}>0$ et

$$\widehat{\beta}-t_{\alpha,n-1}\widehat{\sigma}\sqrt{D'D}^{-1}>0$$

où $\widehat{\sigma}^2 = \dfrac{1}{n-1}\left(Y-D\widehat{\beta}\right)'\left(Y-D\widehat{\beta}\right)=\dfrac{1}{n-1}\left(YY'-D'D\widehat{\beta}^2\right)$ est l'estimation de $\sigma^2$ par les moindres carrés, le chercheur conclura que le programme « fonctionne » dans la mesure où son impact est positif, d'un point de vue statistiquement significatif. En revanche, si $\widehat{\beta}<0$ et

$$\widehat{\beta}+t_{\alpha,n-1}\widehat{\sigma}\sqrt{D'D}^{-1}<0$$

(c'est-à-dire si le zéro est situé strictement *au-dessus* de la borne supérieure de l'intervalle de confiance à $1-\alpha$), le chercheur conclura que le programme est un échec et qu'il faudra y mettre fin. Ainsi, en termes de valeurs critiques de $\widehat{\beta}$, l'évaluateur fréquentiste préconisera l'annulation du programme lorsque :

$$\widehat{\beta} < \beta_* = -\frac{t_{\alpha,n-1} Y'Y}{\sqrt{\left(n-1+t^2_{\alpha,n-1}\right)\left(D'D\right)\left(Y'Y\right)}} < 0$$

Mais recommandera son maintien lorsque :

$$\widehat{\beta} > \beta_* = -\frac{t_{\alpha,n-1} Y'Y}{\sqrt{\left(n-1+t^2_{\alpha,n-1}\right)\left(D'D\right)\left(Y'Y\right)}} > 0$$

Lorsque $\beta_* < \widehat{\beta} < \beta^*$, les résultats de l'évaluation d'impact seront considérés comme peu concluants. Bien sûr, ceci n'a pas de sens du point de vue du décideur car il est raisonnable de supposer qu'au moins certains des décideurs en question révisent leurs croyances a priori en se fondant sur les résultats de l'évaluation d'impact[5].

## *Les décideurs bayésiens*

En révisant leurs croyances suite à l'évaluation d'impact, quand les décideurs choisiront-ils de mettre fin au programme ? Formellement, ceci se produira lorsque $\mu^*_\beta < 0$ et :

$$\mu^*_\beta + t_{\alpha,n-1}\sqrt{\frac{b^*}{a^*-1}V^*_\beta} < 0 \tag{10}$$

où nous travaillons désormais avec les croyances a posteriori. Bien évidemment, si $\mu^*_\beta > 0$ et :

$$\mu^*_\beta - t_{\alpha,n-1}\sqrt{\frac{b^*}{a^*-1}V^*_\beta} > 0 \tag{11}$$

alors le décideur bayésien sera en faveur du maintien du programme.

La substitution des équations (6) et (7) implique que l'inégalité (10) peut être réécrite de la façon suivante :

$$\frac{\mu_\beta + V_\beta D'D\widehat{\beta} + t_{\alpha,n-1}\sqrt{\dfrac{V_\beta \Omega}{n-2a-2}}}{1+V_\beta D'D} < 0 \tag{12}$$

où :

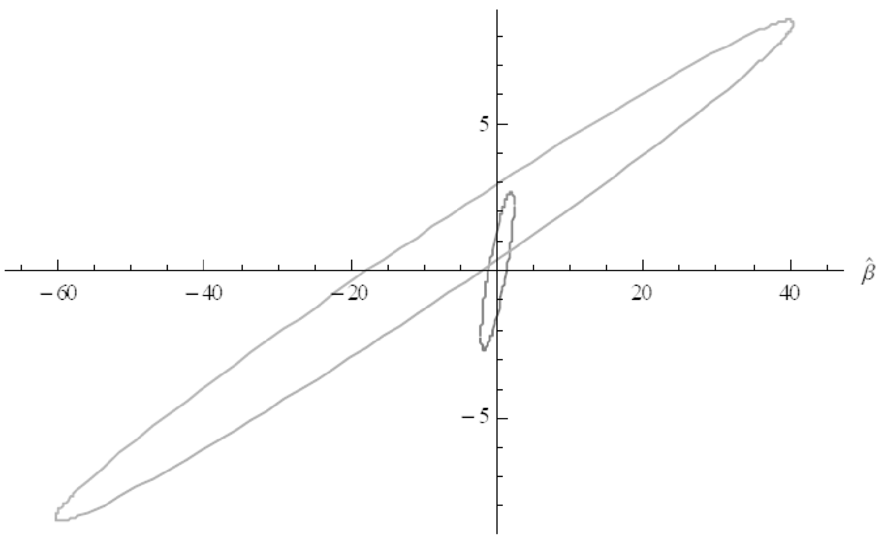
$$\Omega = 2b\left(1+V_\beta D'D\right) + Y'Y + D'D\left[V_\beta\left(Y'Y - D'D\widehat{\beta}^2\right) + \mu_\beta(\mu_\beta - 2\widehat{\beta})\right]$$

---

5 Les problèmes liés à l'incohérence temporelle dans la prise de décision vont au-delà de la portée de cet article.

avec une expression similaire pour (11). La Figure 1 constitue une représentation graphique standard des inégalités (8), (9), (10) et (11). Je configure les paramètres de façon à ce que le programme soit approuvé à la base. Pour cela et à titre purement illustratif, je pose $\mu_\beta = 2, V_\beta = 1.00, a = 75, b = 40, D^{'}D = 0.2, Y^{'}Y = 1$ et $n = 10$. L'aire située au sein de la grande ellipse correspond à des valeurs de $\hat{\beta}$ telles que $\mu_\beta^* - t_{\alpha,n-1}\sqrt{\frac{b^*}{a^*-1}V_\beta^*} < 0 < \mu_\beta^* + t_{\alpha,n-1}\sqrt{\frac{b^*}{a^*-1}V_\beta^*}$ et pour lesquelles le décideur bayésien ne peut donc pas prendre de décision concernant la significativité statistique du coefficient estimé. L'aire située à l'intérieur de la petite ellipse renvoie à la même situation d'absence de significativité statistique pour l'évaluateur fréquentiste.

Figure 1 : *Une illustration de* $\hat{\beta} - t_{\alpha,n-1}\hat{\sigma}\sqrt{D^{'}D}^{-1}$, $\hat{\beta} + t_{\alpha,n-1}\hat{\sigma}\sqrt{D^{'}D}^{-1}$, $\frac{\mu_\beta + V_\beta D^{'}D\hat{\beta} - t_{\alpha,n-1}\sqrt{\frac{V_\beta \Omega}{n-2a-2}}}{1+V_\beta D'D}$, *et* $\frac{\mu_\beta + V_\beta D^{'}D\hat{\beta} + t_{\alpha,n-1}\sqrt{\frac{V_\beta \Omega}{n-2a-2}}}{1+V_\beta D'D}$



Après quelques calculs fastidieux, il peut être démontré que les racines des équations quadratiques en $\beta$ que l'on obtient en égalisant à zéro la partie à gauche de l'inégalité dans les expressions (11) ou (12) sont données par :

$$\beta_+, \beta_- = \frac{-\mu_\beta \pm t_{\alpha,n-1}\sqrt{\frac{\left(1+V_\beta D^{'}D\right)(2bV_\beta + V_\beta Y^{'}Y + \mu_\beta^2)}{n\left(n-2a-2\right)+t^2_{\alpha,n-1}}}}{V_\beta D'D} < 0 \tag{13}$$

Ceci conduit à la proposition suivante :

**Proposition 1** *Il est rationnel pour un décideur bayésien d'annuler le programme lorsque* $\beta < \beta_-$ *et de le maintenir lorsque* $\widehat{\beta} > \beta_+$. *Lorsque* $\beta_- < \widehat{\beta} < \beta_+$ *le décideur bayésien est indifférent entre le maintien et l'annulation du programme.*

**Démonstration.** Il est aisé de démontrer que la partie à gauche de l'inégalité dans l'expression (12) est égale à zéro pour :

$$\widehat{\beta} = \beta_- = \frac{-\mu_\beta - t_{\alpha,n-1}\sqrt{\frac{\left(1+V_\beta D^{'}D\right)(2bV_\beta + V_\beta Y^{'}Y + \mu_\beta^2)}{n\left(n-2a-2\right)+t^2_{\alpha,n-1}}}}{V_\beta D'D}$$

En outre, une simple différenciation de la partie à gauche de l'inégalité dans (12) donne :

$$\frac{\partial}{\partial\widehat{\beta}}\left(\frac{\mu_\beta + V_\beta D^{'}D\widehat{\beta} + t_{\alpha,n-1}\sqrt{\frac{V_\beta\Omega}{n-2a-2}}}{1+V_\beta D^{'}D}\right) = \frac{V_\beta D^{'}D\widehat{\beta} - \frac{t_{\alpha,n-1}V_\beta D^{'}D(\mu_\beta + V_\beta D^{'}D\widehat{\beta}}{(n-2a-2)\sqrt{\frac{V_\beta\Omega}{n-2a-2}}}}{1+V_\beta D^{'}D} > 0$$

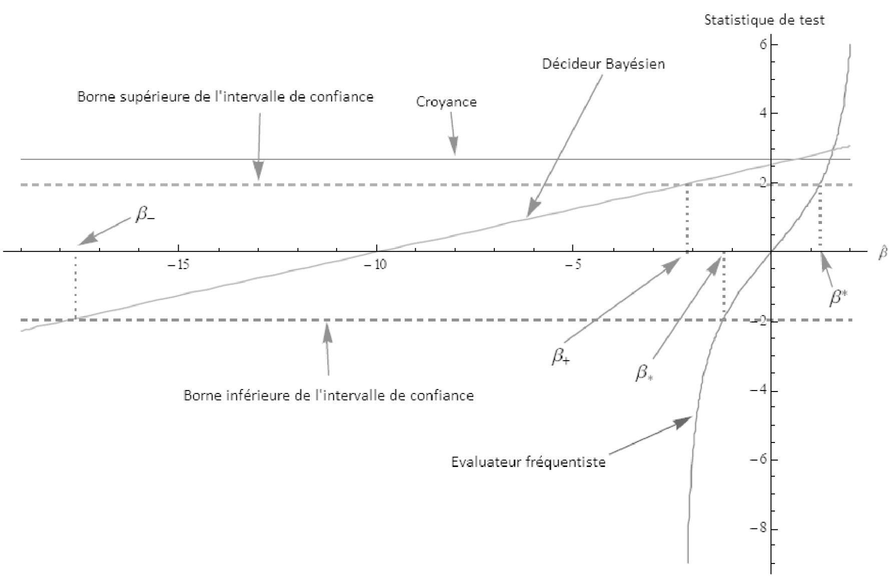
où le signe de l'inégalité découle du fait que dans cette configuration $\mu_\beta + V_\beta D^{'}D\widehat{\beta} < 0$. Ainsi l'expression (12) tient pour $\widehat{\beta} < \beta_-$. Des arguments similaires permettent de démontrer que l'expression (11) tient lorsque $\widehat{\beta} > \beta_+$. Par conséquent aucune des deux inégalités ne tient lorsque $\beta_- < \widehat{\beta} < \beta_+$ et le décideur bayésien sera indifférent entre l'annulation et le maintien du programme en présence de cette configuration.

La Proposition 1 démontre qu'il est rationnel pour un décideur bayésien d'annuler le programme lorsque $\widehat{\beta}$ est « suffisamment petit », alors qu'il sera en faveur du maintien du programme lorsque $\widehat{\beta}$ est « suffisamment grand ».

La Figure 2 présente une illustration de la Proposition 1. Avec la même configuration des paramètres utilisée dans la Figure 1, $\frac{\mu_\beta}{\sqrt{\frac{b}{a-1}V_\beta}} \approx 2.72$ afin que, suivant la croyance, le programme soit approuvé sans problème aux intervalles de confiance conventionnels.

Dans la Figure, je représente les statistiques de Student obtenues en faisant varier le résultat $\hat{\beta}$ de l’évaluation d’impact, tout en maintenant constants les autres paramètres. Pour la croyance, la statistique de test associée est constante et égale à 2.72 pour toutes les valeurs de $\hat{\beta}$. La courbe en « S inversé » dépeint la statistique de test $\dfrac{\hat{\beta}}{\hat{\sigma}\sqrt{D'D}^{-1}}$ associée au résultat de l’évaluation d’impact vu par l’évaluateur fréquentiste. Lorsque la valeur de la statistique de test dépasse la borne supérieure de l’intervalle de confiance et se situe au-dessus d’elle ($\hat{\beta} > \beta^* > 0$), l’évaluateur fréquentiste annoncera le succès du programme et sera en faveur de son maintien. Lorsque la valeur de $\hat{\beta}$ est telle que la statistique de test se retrouve en-dessous de la borne inférieure de l’intervalle de confiance ($\hat{\beta} < \beta_* < 0$), l’évaluateur fréquentiste sera en faveur de l’annulation du programme. Pour toutes les valeurs intermédiaires, il sera indifférent entre l’annulation et le maintien du programme. Le même raisonnement tient pour la statistique de test $\dfrac{\mu^*_\beta}{\sqrt{\dfrac{b^*}{a^*-1}V^*_\beta}}$ de l’évaluateur bayésien. Notons que pour cette configuration du paramètre, et étant donné l’intensité de la croyance, l’évaluateur bayésien sera toujours en faveur du maintien du programme même lorsque le coefficient estimé $\hat{\beta}$ est juste au-dessus de $\beta_+ \approx -2.2$. En revanche, il ne sera en faveur de l’annulation du programme qu’une fois que l’impact estimé du programme est en-dessous de $\beta_- \approx -17.6$.

Figure 2 : *Une illustration de la Proposition 1*



Notons que dans la Figure 2, $\beta_{-} < \beta_{+} < \beta_{*}$. Dans ce cas, pour $\beta_{+} < \widehat{\beta} < \beta_{*}$, l'évaluateur fréquentiste est en faveur de l'annulation du programme et le décideur bayésien en faveur de son maintien. Ceci ne constitue pas une propriété générale et il est parfaitement possible d'obtenir $\beta_{-} < \beta_{*} < \beta_{+}$ avec une configuration différente des valeurs des paramètres. Dans ce cas, lorsque $\beta_{*} < \widehat{\beta} < \beta_{+}$, l'évaluateur fréquentiste sera indifférent entre l'annulation et le maintien du programme, alors que le décideur bayésien sera en faveur de son annulation.

En passant, (comme $D^{'}D = (n-1)\sigma_D^2$, $Y^{'}Y = (n-1)\sigma_Y^2$, et $Y^{'}Y - D^{'}D\widehat{\beta}^2 = (n-1)\widehat{\sigma}^2$ ) il convient de noter que :

$$\lim_{n\to\infty} \frac{\mu_\beta + V_\beta D^{'}D\widehat{\beta} - t_{\alpha,n-1}\sqrt{\dfrac{V_\beta \Omega}{n-2a-2}}}{1+V_\beta D'D} = \lim_{n\to\infty} \frac{\mu_\beta + V_\beta D^{'}D\widehat{\beta} + t_{\alpha,n-1}\sqrt{\dfrac{V_\beta \Omega}{n-2a-2}}}{1+V_\beta D'D} = \widehat{\beta}$$

Ainsi,

$$\lim_{n\to\infty} \beta_{-} = \lim_{n\to\infty} \beta_{+} = \lim_{n\to\infty} \beta_{*} = \lim_{n\to\infty} \beta^{*} = 0$$

Par conséquent, si l'évaluation d'impact reposait sur un échantillon de taille infinie, on s'attendrait intuitivement à ce que son résultat détermine entièrement la décision d'annuler le programme.

## 2 UNE PRISE DE DÉCISION FONDÉE SUR LES FAITS ?

Je considère une situation où, suite à l'évaluation d'impact, trois types de décideurs sont en concurrence pour déterminer le maintien ou l'annulation du programme. Les trois décideurs sont :

- Les décideurs anti-évaluation : ils ont peur de l'évaluation et préfèrent fonder leur décision uniquement sur la croyance. Ils bénéficient du maintien d'un programme qui a été approuvé par le passé[6].
- Les décideurs bayésiens : ils fondent leur décision sur la croyance a posteriori, combinant ainsi leur croyance aux résultats de l'évaluation d'impact.

[6] Je cite ici un haut fonctionnaire de Scandinavie travaillant dans le domaine de l'aide au développement, et que j'eus à interroger sur la raison pour laquelle son agence n'évaluait pas ses programmes à travers des évaluations d'impact : « Nous mettons en œuvre les mêmes programmes depuis 20 ans, ils doivent donc bien fonctionner ».

- Les (chercheurs)-évaluateurs d'impact fréquentistes : ils fondent leur décision uniquement sur les résultats de l'évaluation d'impact.

## 2.1 Concurrence entre acteurs

Je suppose que le processus de prise de décision qui conduit au maintien ou à l'annulation du programme prend la forme d'une concurrence entre les trois décideurs. Chacun d'entre eux investit des ressources pour obtenir le résultat qu'il préfère[7]. À titre illustratif, considérons des fonctions de contest success simples et à la « forme de ratio », dont la probabilité de gagner (d'obtenir son résultat préféré) est notée par[8] :

$$p_i\left(e_A, e_B, e_F\right) = \frac{\alpha_i e_i^{\gamma}}{\alpha_A e_A^{\gamma} + \alpha_B e_B^{\gamma} + \alpha_F e_F^{\gamma}}, \gamma \geq 0, i = A, B, F \qquad (14)$$

où $\alpha_i$ est l'efficacité relative de l'effort exercé par l'agent $i$, $p_i$ donne ainsi la probabilité que l'agent $i$ s'impose, et elle est fonction de l'effort exercé $(e_A, e_B, e_F)$ par les trois parties impliquées.

[7] Lors de la planification des évaluations d'impact, un paramètre clé à établir est la taille de l'échantillon $n$. Comme il a été souligné dans l'article influent de Bloom (1995), un grand $n$ réduit l'effet minimum détectable (EMD), ce qui signifie que la puissance statistique de l'échantillon est plus grande. Si l'on suppose, en guise de première approximation, que la taille de l'échantillon est déterminée de sorte à ce que l'EMD soit égal à la croyance sur l'impact du programme et que l'on est dans le cadre d'un ERC, la formule standard de l'EMD pour un échantillon équilibré ($P = \frac{1}{2}$) est donnée par $EMD = \frac{M_{n-2}}{\sqrt{n}}\sqrt{\frac{1}{P(1-P)}} = \frac{2M_{n-2}}{\sqrt{n}}$ ; $M_{n-2}$ sera à peu près égal à 2.5 pour un test unilatéral avec une puissance du test de 0.80 et un niveau de significativité de 0.05 (fixés conventionnellement). Étant donné que la croyance porte sur le fait que la taille de l'effet est $\frac{\mu_\beta}{\sqrt{\frac{b}{a-1}V_\beta}}$, cela implique que la taille de l'échantillon sera donnée par $n = \frac{4b}{a-1}\mu_\beta^{-2}V_\beta M_{n-2}^2$ sur la base de cette croyance. En revanche, un évaluateur fréquentiste préfèrerait augmenter la taille de l'échantillon afin d'augmenter le poids de l'évaluation d'impact en termes de prise de décision politique. La prise en compte de la taille de l'échantillon comme une variable faisant l'objet de lobbying dépasse le cadre de cet article.

[8] Il existe une vaste littérature sur les fonctions de « contest success ». Dans une étude récente, Jia, Skaperdas et Vaidya (2011) font un excellent tour d'horizon des différentes formes fonctionnelles ainsi que des différentes manières de les dériver axiomatiquement à partir de spécifications stochastiques ou de principes de conception de mécanismes.

En faisant l'hypothèse de neutralité vis-à-vis du risque, le gain découlant du maintien du programme pour l'agent $A$ peut s'écrire :

$$B_A = \mu_\beta \tag{15}$$

où $B_A > 0$ puisque le programme a été approuvé à la base. Lorsque l'évaluation d'impact implique que le programme doit être annulé pour l'agent $F$ ($\widehat{\beta} < \beta_* < 0$) :

$$B_F = -\widehat{\beta} > 0 \tag{16}$$

Notons que la fonction d'objectif de l'agent $F$ fait intervenir $-\widehat{\beta}$ dans la mesure où il s'agit du *gain* obtenu de l'annulation du programme sans infliger une *perte* moyenne de $\widehat{\beta}$ aux bénéficiaires (rappelons au passage que le coût du programme a été normalisé à zéro). Une autre façon de le voir consiste à le considérer comme le gain d'opportunité découlant de l'annulation du programme.

Dans la suite, je considère quatre configurations possibles des résultats de l'évaluation d'impact[9]. Dans tous les cas, le décideur anti-évaluation est en faveur du maintien du programme :

- $\widehat{\beta} < \beta_- < \beta_+ < \beta_* < \beta^*$ : l'agent $B$ est en faveur de l'annulation, l'agent $F$ est en faveur de l'annulation.
- $\beta_- < \widehat{\beta} < \beta_+ < \beta_* < \beta^*$ : l'agent $B$ est indifférent entre l'annulation et le maintien du programme, l'agent $F$ est en faveur de l'annulation.
- $\beta_- < \beta_+ < \widehat{\beta} < \beta_* < \beta^*$ : l'agent $B$ est en faveur du maintien, l'agent $F$ est en faveur de l'annulation.
- $\beta_- < \beta_+ < \beta_* < \widehat{\beta} < \beta^*$ : l'agent $B$ est en faveur du maintien, l'agent $F$ est indifférent entre l'annulation et le maintien, le programme est alors maintenu avec une probabilité de 1.

[9] Dans l'autre configuration où $\beta_* < \beta_+$, seuls deux cas sont intéressants à considérer, à savoir : $\widehat{\beta} < \beta_- < \beta_* < \beta_+ < \beta^*$ : les agents $B$ et $F$ sont en faveur de l'annulation ; $\beta_- < \widehat{\beta} < \beta_* < \beta_+ < \beta^*$ : l'agent $B$ est indifférent entre l'annulation et le maintien du programme, l'agent $F$ est en faveur de l'annulation ; $\beta_- < \beta_* < \widehat{\beta} < \beta_+ < \beta^*$ : les agents $B$ et $F$ sont indifférents entre l'annulation et le maintien, et le programme est maintenu avec une probabilité de 1 ; $\beta_- < \beta_* < \beta_+ < \widehat{\beta} < \beta^*$ : l'agent $B$ est en faveur du maintien, l'agent $F$ est indifférent entre l'annulation et le maintien, et le programme est maintenu avec une probabilité de 1.

## 2.2 Une croyance a posteriori significativement négative : $\widehat{\beta} < \beta_- < \beta_+ < \beta_* < \beta^*$

Lorsque, à travers la croyance a posteriori, l'évaluation d'impact implique l'annulation du programme pour l'agent $B$, on a :

$$B_B = -\mu_\beta^* > 0 \tag{17}$$

En supposant la configuration de résultats précédente, et comme les agents $B$ et $F$ souhaitent tous deux l'annulation du programme (ce qui implique $p_B = p_F = 1 - p_A$), les gains espérés des agents sont donnés par :

$$U_A = p_A\mu_\beta - ce_A, U_B = (1-p_A)(-\mu_\beta^*) - ce_B, U_F = (1-p_A)(-\widehat{\beta}) - ce_F \tag{18}$$

où, par souci de transparence des résultats qui suivent, je suppose que le coût marginal $c$ de l'effort exercé pour atteindre son résultat préféré est constant et identique pour toutes les parties[10]. On fait l'hypothèse que chaque agent résout un problème de maximisation standard qui conduit au niveau d'effort optimal suivant :

$$e_j^* = arg\max_{\{e_j\}} U_j \; s.t. e_j \geq 0, j = A, B, F$$

Là encore, par soucis de simplicité et pour s'assurer de l'unicité de l'équilibre qui en découle, supposons que $\gamma = 2$[11]. Dans ce cas, et après quelques calculs fastidieux, l'équilibre de Nash des niveaux d'effort fourni par chaque partie est donné par :

$$e_A^* = \frac{2\alpha_A\alpha_B\alpha_F\mu_\beta^3\mu_\beta^{*2}\widehat{\beta}^2(\alpha_B\mu_\beta^{*2} + \alpha_F\widehat{\beta}^2)}{c\left[\alpha_B\alpha_F\mu_\beta^{*2}\widehat{\beta}^2 + \alpha_A\mu_\beta^2(\alpha_B\mu_\beta^{*2} + \alpha_F\widehat{\beta}^2\right]^2} > 0 \tag{19}$$

10 À l'équilibre, des résultats plus complexes sont obtenus lorsque la fonction de coût diffère d'un joueur à l'autre, mais je préfère éviter ces complications et mettre l'accent sur l'impact sur l'équilibre des caractéristiques de la croyance, les résultats de l'évaluation d'impact, les variances associées à la croyance et à l'évaluation d'impact, et l'influence relative de chaque partie.

11 Le paramètre $\gamma$ représente le caractère "informationnel" du concours. Lorsque $\gamma \to 0$, le concours tend vers une randomisation où les actions des agents n'ont aucun effet sur le résultat, lorsque $\gamma \to \infty$, le concours tend vers une enchère « all pay » où le concurrent qui fournit un niveau infinitésimalement plus grand d'effort emporte l'intégralité de la récompense. En termes d'unicité de l'équilibre, notons que le fait de poser $\gamma \to 1/2$ conduit à une paire d'équilibres (dont l'un s'obtient sous certaines conditions sur les paramètres) mais la qualité des résultats qui suivent reste inchangée.

$$e_B^* = -\frac{2\alpha_A\alpha_B\alpha_F^2\mu_\beta^2\mu_\beta^{*3}\widehat{\beta}^4}{c\left[\alpha_B\alpha_F\mu_\beta^{*2}\widehat{\beta}^2+\alpha_A\mu_\beta^2(\alpha_B\mu_\beta^{*2}+\alpha_F\widehat{\beta}^2\right]^2} > 0 \tag{20}$$

$$e_F^* = -\frac{2\alpha_A\alpha_B\alpha_F\mu_\beta^2\mu_\beta^{*4}\widehat{\beta}^3}{c\left[\alpha_B\alpha_F\mu_\beta^{*2}\widehat{\beta}^2+\alpha_A\mu_\beta^2(\alpha_B\mu_\beta^{*2}+\alpha_F\widehat{\beta}^2\right]^2} > 0 \tag{21}$$

La substitution des expressions (19), (20) et (21) dans la fonction « contest success » permet d'obtenir, à l'équilibre, la probabilité de maintien du programme. Elle est fonction de l'impact moyen ex ante, l'impact moyen ex post, le coefficient estimé de l'évaluation d'impact, et de l'influence exercé par chaque agent sur le résultat :

$$p_{A-}^* = \frac{\alpha_A\mu_\beta^2(\alpha_F\widehat{\beta}^2+\alpha_B\mu_\beta^{*2})}{\alpha_B\alpha_F\mu_\beta^2\mu_\beta^{*2}+\alpha_A\mu_\beta^2(\alpha_F\widehat{\beta}^2+\alpha_B\mu_\beta^{*2})}$$

En y substituant l'expression pour les résultats ex post :

$$p_{A-}^* = \frac{\alpha_A\mu_\beta^2\left[\alpha_F\left(1+V_\beta D^{'}D\right)^2\widehat{\beta}^2+\alpha_B(V_\beta D^{'}D\widehat{\beta}+\mu_\beta)^2\right]}{\Psi} \tag{22}$$

où

$$\Psi = \alpha_A\alpha_F\left(1+V_\beta D^{'}D\right)^2\widehat{\beta}^2\mu_\beta^2+\alpha_B(V_\beta D^{'}D\widehat{\beta}+\mu_\beta)^2(\alpha_F\widehat{\beta}^2+\alpha_A\mu_\beta^2)$$

D'où la Proposition suivante :

**Proposition 2** *La probabilité d'annulation du programme, sachant qu'il a été approuvé à la base et que les évaluateurs bayésien et fréquentiste sont tous deux en faveur de l'annulation ($\widehat{\beta}<\beta_-<\beta_+<\beta_*<\beta^*$), admet les statiques comparées suivantes :*

$$\frac{\partial(1-p_{A-}^*)}{\partial\mu_\beta} = -\left(\frac{\mu_\beta}{\widehat{\beta}}\right)\frac{\partial\left(1-p_{A-}^*\right)}{\partial\widehat{\beta}} < 0, \frac{\partial\left(1-p_{A-}^*\right)}{\partial V_\beta} = \left(\frac{V_\beta}{D^{'}D}\right)\frac{\partial\left(1-p_{A-}^*\right)}{\partial\left(D^{'}D\right)} > 0$$

**Démonstration**. Une simple mais lourde différenciation de (22) par rapport à $\mu_\beta$ et $\widehat{\beta}$ conduit à :

$$\frac{\partial\left(1-p^*_{A-}\right)}{\partial\mu_\beta} = -2\Psi^{-2}\alpha_A\alpha_B\alpha_F\widehat{\beta}^2\mu_\beta\left(V_\beta D^{'}D\widehat{\beta}+\mu_\beta\right)\times\left[\alpha_F V_\beta D^{'}D(1+V_\beta D^{'}D)^2\widehat{\beta}^3 + \right.$$

$$\left.\alpha_B\left(V_\beta D^{'}D\widehat{\beta}+\mu_\beta\right)^3\right] = -\left(\frac{\mu_\beta}{\widehat{\beta}}\right)\frac{\partial\left(1-p^*_{A-}\right)}{\partial\widehat{\beta}}$$

Dans la configuration que j'ai retenue, $\widehat{\beta}<0$, $V_\beta D^{'}D\widehat{\beta}+\mu_\beta < 0$ et $\widehat{\beta}-\mu_\beta<0$. Il s'ensuit que $-\frac{\partial\left(1-p^*_{A-}\right)}{\partial\mu_\beta} = \frac{\partial\left(1-p^*_{A-}\right)}{\partial\widehat{\beta}}<0$. De même,

$$\frac{\partial\left(1-p^*_{A-}\right)}{\partial V_\beta} = 2\Psi^{-2}\alpha_A\alpha_B\alpha_F^2 V_\beta\left(1+V_\beta D^{'}D\right)\widehat{\beta}^4\left(\widehat{\beta}-\mu_\beta\right)\mu_\beta^2\left(V_\beta D^{'}D\widehat{\beta}+\mu_\beta\right) =$$

$$\left(\frac{V_\beta}{D^{'}D}\right)\frac{\partial\left(1-p^*_{A-}\right)}{\partial\left(D^{'}D\right)}>0$$

Comme à ce qu'on pourrait s'attendre intuitivement, la Proposition 2 indique que, *ceteris paribus*, plus le coefficient $\widehat{\beta}$ associé à l'évaluation d'impact est élevé, moins la probabilité d'annulation du programme est importante. Un résultat similaire et intuitivement séduisant est obtenu par rapport à la croyance $\mu_\beta$. En outre, une hausse de la variance de la muette indiquant l'état du traitement ($D^{'}D$) augmente la probabilité d'annulation du programme, tout comme une hausse de la variance associée à la croyance.

Un autre résultat intéressant est obtenu en considérant l'effet d'un changement de l'efficacité du lobbying par l'évaluateur bayésien ou fréquentiste. J'exprime cela à travers la Proposition évidente suivante :

**Proposition 3** *Lorsque* $\widehat{\beta}<\beta_-<\beta_+<\beta_*<\beta^*$, *la probabilité d'annulation du programme est une fonction croissante de l'influence des évaluateurs bayésien et fréquentiste, et une fonction décroissante de l'influence du décideur anti-évaluation. En outre, lorsque l'influence* $\alpha_F$ ($\alpha_B$) *de l'évaluateur fréquentiste (bayésien) est nulle, la probabilité d'annulation du programme est nulle.*

**Démonstration.** Une différenciation triviale de (22) conduit à :

$$\frac{\partial\left(1-p^*_{A-}\right)}{\partial\alpha_A} = -\Psi^{-2}\alpha_B\alpha_F\widehat{\beta}^2\mu_\beta\left(V_\beta D^{'}D\widehat{\beta}+\mu_\beta\right)^2\times\left[\alpha_F\widehat{\beta}^2\left(1+V_\beta D^{'}D\right)^2 + \right.$$

$$\left.\alpha_B\left(V_\beta D^{'}D\widehat{\beta}+\mu_\beta\right)^2\right]\leq 0$$

$$\frac{\partial\left(1-p_{A-}^{*}\right)}{\partial\alpha_{B}}=\Psi^{-2}\alpha_{A}\alpha^{2}{}_{F}\left(1+V_{\beta}D^{'}D\right)^{2}\widehat{\beta}^{4}\mu_{\beta}^{2}\left(V_{\beta}D^{'}D\widehat{\beta}+\mu_{\beta}\right)^{2}\geq 0$$

$$\frac{\partial\left(1-p_{A-}^{*}\right)}{\partial\alpha_{F}}=\Psi^{-2}\alpha_{A}\alpha^{2}{}_{B}\widehat{\beta}^{2}\mu_{\beta}^{2}\left(V_{\beta}D^{'}D\widehat{\beta}+\mu_{\beta}\right)^{4}\geq 0$$

En outre, il est évident que $1-p_{A-}^{*}\,|_{\alpha_B=0}=1-p_{A-}^{*}\,|_{\alpha_F=0}=0$

La Proposition 3 est frappante : il suffit que l'influence du lobbying par le décideur fréquentiste ($\alpha_F$) ou bayésien ($\alpha_B$) devienne nulle pour que la probabilité d'annulation du programme soit égale à zéro. Dans la vraie vie, et malgré tous ces beaux discours, il est probable que l'influence réelle du lobbying par les chercheurs-évaluateurs afin d'obtenir l'annulation de programmes dont on a « démontré » l'échec à travers des évaluations d'impact, soit minime dans le meilleur des cas. Le modèle montre (à travers l'expression de $e_F^*$ dans l'équation (21)) que lorsque l'influence $\alpha_F$ du lobbying entrepris par le chercheur-évaluateur fréquentiste tend vers zéro, ce qui est probablement proche de la situation qui prévaut dans la vraie vie, son effort en faveur de l'annulation du programme tend vers zéro également. Ainsi, l'absence d'influence alimente l'absence d'effort, et l'absence d'effort conduit au maintien du programme. Il n'est peut-être pas surprenant en tant que tel que l'influence des évaluations d'impact dans le domaine de la prise de décision soit, dans le meilleur des cas, limitée.

## 2.3 Une croyance a posteriori peu concluante : $\beta_- < \widehat{\beta} < \beta_+ < \beta_* < \beta^*$

Lorsque l'évaluation d'impact débouche sur une croyance a posteriori qui est peu concluante, l'évaluateur bayésien sera indifférent entre l'annulation et le maintien du programme, par conséquent il ne fournira aucun effort. Le jeu se réduit alors à sa variante à deux joueurs où le décideur anti-évaluation et l'évaluateur fréquentiste sont en concurrence. Les gains sont donnés par $U_A = p_A\mu_\beta - ce_A$, $U_F = (1-p_A)\left(-\widehat{\beta}\right) - ce_F$, alors que la fonction de « contest success » se simplifie à $p_A = \dfrac{e_A^\gamma}{e_A^\gamma + e_F^\gamma}$. Il est alors aisé d'établir la Proposition suivante :

**Proposition 4** *La probabilité que le programme soit annulé, alors qu'il avait été approuvé à la base, lorsque* *l'évaluateur bayésien est indifférent entre le maintien et l'annulation et que l'évaluateur fréquentiste est en faveur de l'annulation* ($\beta_- < \widehat{\beta} < \beta_+ < \beta_* < \beta^*$), *est égale à* $p^*_{A0} = \dfrac{\alpha_A \mu^2_\beta}{\alpha_F \widehat{\beta}^2 \alpha_A \mu^2_\beta}$*, et admet les statiques comparées suivantes :*

$$\frac{\partial(1-p^*_{A0})}{\partial \mu_\beta} = -\left(\frac{\mu_\beta}{\beta}\right)\frac{\partial\left(1-p^*_{A0}\right)}{\partial \widehat{\beta}} < 0, \frac{\partial\left(1-p^*_{A0}\right)}{\partial V_\beta} = \frac{\partial\left(1-p^*_{A0}\right)}{\partial\left(D'D\right)} = 0.$$

**Démonstration.** En procédant comme dans la démonstration de la Proposition 2 : $e^*_A = \dfrac{2\alpha_A \alpha_F \mu^3_\beta \widehat{\beta}^2}{c(\alpha_F \widehat{\beta}^2 + \alpha_A \mu^2_\beta)^2} > 0, e^*_F = -\dfrac{2\alpha_A \alpha_B \alpha_F \mu^2_\beta \mu^{*2}_\beta \widehat{\beta}^3}{c(\alpha_F \widehat{\beta}^2 + \alpha_A \mu^2_\beta)^2} > 0$ ce qui implique $p^*_{A0} = \dfrac{\alpha_A \mu^2_\beta}{\alpha_F \widehat{\beta}^2 \alpha_A \mu^2_\beta}$. En différenciant :

$\dfrac{\partial(1-p^*_{A0})}{\partial \mu_\beta} = -\dfrac{2\alpha_A \alpha_F \widehat{\beta}^2 \mu_\beta}{\left(\alpha_F \widehat{\beta}^2 + \alpha_A \mu^2_\beta\right)^2} = -\left(\dfrac{\mu_\beta}{\beta}\right)\dfrac{\partial\left(1-p^*_{A0}\right)}{\partial \widehat{\beta}} < 0$. Il est évident que $\dfrac{\partial\left(1-p^*_{A0}\right)}{\partial V_\beta} = \dfrac{\partial\left(1-p^*_{A0}\right)}{\partial\left(D'D\right)} = 0$.

Comme dans le cas d'une croyance a posteriori significativement négative, la probabilité d'annulation baisse à zéro lorsque l'influence $\alpha_F$ du chercheur-évaluateur fréquentiste baisse à zéro.

## 2.4 Une croyance a posteriori significativement positive : $\beta_- < \beta_+ < \widehat{\beta} < \beta_* < \beta^*$

Dans ce cas :

$$B_B = \mu^*_\beta > 0$$

Les agents $A$ et $B$ souhaitent tous deux maintenir le programme et les gains espérés sont donnés par :

$$U_A = p_A \mu_\beta - ce_A, U_B = p_A \mu^*_\beta - ce_B, U_F = \left(1-p_A\right)\left(-\widehat{\beta}\right) - ce_F$$

où $p_A = \dfrac{e^\gamma_A + e^\gamma_B}{e^\gamma_A + e^\gamma_B + e^\gamma_F}$ . D'où la Proposition suivante :

**Proposition 5** *La probabilité que le programme soit annulé, alors qu'il avait été approuvé à la base, lorsque l'évaluateur bayésien est en faveur du maintien et l'évaluateur fréquentiste en faveur de l'annulation* $(\beta_{-} < \beta_{+} < \widehat{\beta} < \beta_{*} < \beta^{*})$ *est égale à* $p_{A+}^{*} = \frac{\alpha_A \alpha_B \mu_\beta{}^2 (V_\beta D^{'} D \widehat{\beta} + \mu_\beta)^2}{\Psi}$*, et admet les statiques comparées suivantes :*

$$\frac{\partial(1-p_{A+}^{*})}{\partial \mu_\beta} = -\left(\frac{\mu_\beta}{\beta}\right)\frac{\partial\left(1-p_{A+}^{*}\right)}{\partial \widehat{\beta}} < 0, \frac{\partial\left(1-p_{A+}^{*}\right)}{\partial V_\beta} = \left(\frac{V_\beta}{D^{'}D}\right)\frac{\partial\left(1-p_{A+}^{*}\right)}{\partial\left(D^{'}D\right)} > 0.$$

**Démonstration.** En procédant comme dans la démonstration de la Proposition 2, l'équilibre de Nash des niveaux d'effort fourni par chaque partie est donné par : $e_A^{*} = \frac{2\alpha_A \alpha_B^{2} \alpha_F \mu_\beta^3 \mu_\beta^{*4} \widehat{\beta}^2}{c\left[\alpha_B \alpha_F \mu_\beta^{*2} \widehat{\beta}^2 + \alpha_A \mu_\beta^2 (\alpha_B \mu_\beta^{*2} + \alpha_F \widehat{\beta}^2)\right]^2} > 0$,

$$e_B^{*} = \frac{2\alpha_A \alpha_B \alpha_F \mu_\beta^4 \mu_\beta^{*3} \widehat{\beta}^2}{c\left[\alpha_B \alpha_F \mu_\beta^{*2} \widehat{\beta}^2 + \alpha_A \mu_\beta^2 (\alpha_B \mu_\beta^{*2} + \alpha_F \widehat{\beta}^2)\right]^2} > 0,$$

$e_F^{*} = -\frac{2\alpha_A \alpha_B \alpha_F \mu_\beta^2 \mu_\beta^{*3} \widehat{\beta}^3}{c\left[\alpha_B \alpha_F \mu_\beta^{*2} \widehat{\beta}^2 + \alpha_A \mu_\beta^2 (\alpha_B \mu_\beta^{*2} + \alpha_F \widehat{\beta}^2)\right]^2} > 0$, ce qui implique

$p_{A+}^{*} = \frac{\alpha_A \alpha_B \mu_\beta^2 \mu_\beta^{*2}}{\alpha_B \alpha_F \widehat{\beta}^2 \mu_\beta^{*2} + \alpha_A \mu_\beta^2 (\alpha_F \widehat{\beta}^2 + \alpha_B \mu_\beta^{*2})}$ et donc

$p_{A+}^{*} = \frac{\alpha_A \alpha_B \mu_\beta^2 (V_\beta D^{'} D \widehat{\beta} + \mu_\beta)^2}{\Psi}$. En différenciant, on obtient :

$$\frac{\partial(1-p_{A+}^{*})}{\partial \mu_\beta} = -2\Psi^{-2} \alpha_A \alpha_B \alpha_F \widehat{\beta}^2 \mu_\beta (V_\beta D^{'} D \widehat{\beta} + \mu_\beta) \left[\alpha_A (1 + V_\beta D^{'} D)^2 \, \mu_\beta^3 + \alpha_B (V_\beta D^{'} D \widehat{\beta} + \mu_\beta)^3\right] = -\left(\frac{\mu_\beta}{\beta}\right)\frac{\partial\left(1-p_{A+}^{*}\right)}{\partial \widehat{\beta}} < 0$$

Comme $\widehat{\beta} < 0$, $V_\beta D^{'} D \widehat{\beta} + \mu_\beta > 0$ et $\widehat{\beta} - \mu_\beta < 0$ dans la configuration retenue, il s'ensuit que $\frac{\partial(1-p_{A+}^{*})}{\partial\mu_{\beta}} = -\left(\frac{\mu_{\beta}}{\widehat{\beta}}\right)\frac{\partial\left(1-p_{A+}^{*}\right)}{\partial\widehat{\beta}} < 0$. De même,

$$\frac{\partial(1-p_{A+}^{*})}{\partial V_{\beta}} = -2\Psi^{-2}\alpha_{A}^{2}\alpha_{B}\alpha_{F}D^{'}D\left(1+V_{\beta}D^{'}D\right)\widehat{\beta}^{2}\left(\widehat{\beta}-\mu_{\beta}\right)'$$

$$\mu_{\beta}^{4}\left(V_{\beta}D^{'}D\widehat{\beta}+\mu_{\beta}\right) = \left(\frac{V_{\beta}}{D^{'}D}\right)\frac{\partial\left(1-p_{A+}^{*}\right)}{\partial\left(D^{'}D\right)} > 0.$$

Au final, les Propositions 2, 4 et 5 indiquent que la probabilité d'annulation d'un programme qui avait été initialement approuvé est, comme pourrait le suggérer le bon sens, une fonction décroissante du résultat $\widehat{\beta}$ de l'évaluation d'impact, *ceteris paribus*. En outre, la Proposition 3 (et les résultats correspondants qui peuvent être établis de façon triviale pour les configurations retenues dans les Propositions 4 et 5) montre que la probabilité d'annulation du programme tend vers zéro lorsque l'influence exercée par les chercheurs-évaluateurs fréquentistes sur les décisions politiques tend vers zéro, une hypothèse qui est probablement satisfaite en pratique.

## 2.5 Faire le lien entre le tout

Afin de faire le lien entre les résultats des Propositions précédentes, une dernière étape est requise : il faut voir ce qui se passe au niveau de la probabilité d'annulation lorsqu'on considère les trois seuils critiques $\beta_{-}, \beta_{+}$ et $\beta_{*}$. Je m'attache à regarder cela dans la Proposition suivante :

**Proposition 6** *Aux valeurs limites* $\beta_{-}, \beta_{+}$ *et* $\beta_{*}$, $1-p_{A-}^{*}\left(\beta_{-}\right) < 1-p_{A0}^{*}\left(\beta_{-}\right)$, $1-p_{A0}^{*}\left(\beta_{+}\right) < 1-p_{A+}^{*}\left(\beta_{+}\right)$ *et* $1-p_{A+}^{*}\left(\beta_{*}\right) > 0$.

**Démonstration.** Au regard des résultats présentés dans les Propositions 2, 4 et 5, il est évident que :

$$\left[1-p_{A-}^{*}\left(\widehat{\beta}\right)\right]-\left[1-p_{A0}^{*}\left(\widehat{\beta}\right)\right] = -\frac{\alpha_{A}\alpha_{F}^{2}\left(1+V_{\beta}D^{'}D\right)^{2}\widehat{\beta}^{4}\mu_{\beta}^{2}}{\Lambda} < 0 \ \forall\widehat{\beta}$$

et

$$\left[1-p_{A0}^{*}\left(\widehat{\beta}\right)\right]-\left[1-p_{A+}^{*}\left(\widehat{\beta}\right)\right] = -\frac{\alpha_{A}{}^{2}\alpha_{F}\left(1+V_{\beta}D^{'}D\right)^{2}\widehat{\beta}^{2}\mu_{\beta}^{4}}{\Lambda} < 0 \ \forall\widehat{\beta}$$

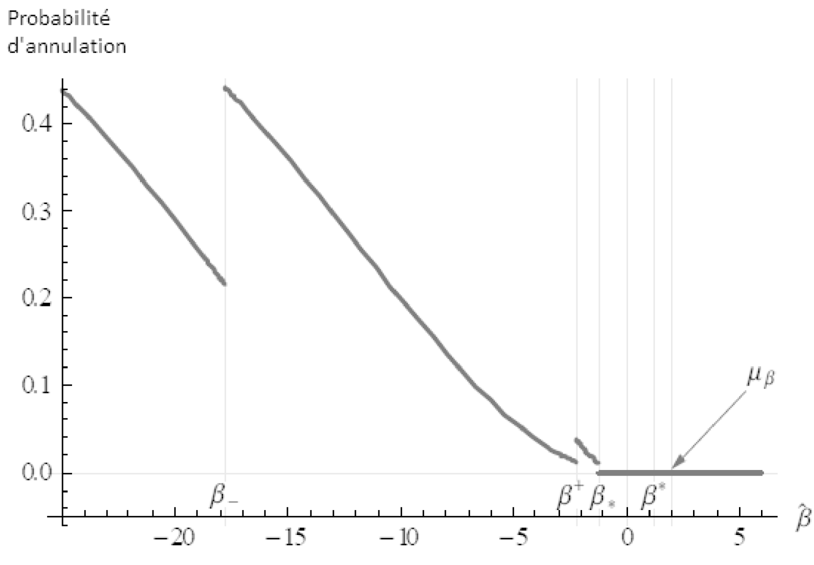
Où :

$$\Lambda = \left(\alpha_F \widehat{\beta}^2 + \alpha_A \mu_\beta^2\right)\left[\alpha_A \alpha_F \left(1 + V_\beta D' D\right)^2 \widehat{\beta}^2 \mu_\beta^2 + \alpha_B \left(V_\beta D' D \widehat{\beta} + \mu_\beta\right)^2 \left(\alpha_F \widehat{\beta}^2 + \alpha_A \mu_\beta^2\right)\right]$$

Il s'ensuit que $1 - p^*_{A-}(\beta_-) < 1 - p^*_{A0}(\beta_-)$ et $1 - p^*_{A0}(\beta_+) < 1 - p^*_{A+}(\beta_+)$. Finalement, il découle naturellement de la Proposition 4 que $1 - p^*_{A+}(\beta_*) > 0$.

La Proposition 6 montre qu'au fur et à mesure que la valeur de $\widehat{\beta}$ diminue, on constate : (i) un saut vers le haut de la probabilité inconditionnelle d'annulation $p^*_A$ en $\beta_*$ et (ii) un saut vers le bas en $\beta_+$ et $\beta_-$.

Figure 3 : *Représentation à l'équilibre de la probabilité d'annulation du programme comme fonction du résultat $\widehat{\beta}$ de l'évaluation d'impact.*

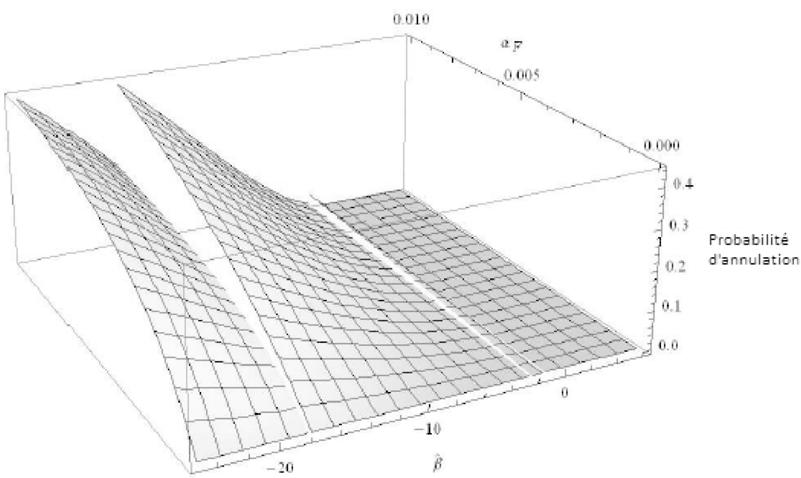



La Figure 3 illustre les Propositions 2, 4, 5 et 6 pour les mêmes valeurs des paramètres utilisées dans les Figures 1 et 2. Je pose $\alpha_A = \alpha_B = 1$ et $\alpha_F = 0.01$ : je surestime probablement le poids placé sur les résultats de l'évaluation d'impact dans les cercles politiques, dans le sens où je donne le même niveau d'influence au décideur anti-évaluation et à son homologue bayésien. Le poids de 0.01 assigné au chercheur-évaluateur fréquentiste est défini à titre d'hypothèse : en réalité son poids dans la prise de décision concernant l'annulation du programme est probablement plus faible.

Dans la Figure 3, la probabilité d'annulation, comme il a été souligné dans les Propositions 2, 4 et 5, est toujours une fonction strictement décroissante de $\widehat{\beta}$, $\forall\widehat{\beta} < \beta_*$ et égale à zéro $\forall\widehat{\beta} \geq \beta_*$. Le saut établi dans la Proposition 6 se produit aux valeurs critiques $\beta_-, \beta_+$ et $\beta_*$, il est plus évident en $\beta_-$. L'aspect le plus frappant de la Figure 3 est la faiblesse relative de la probabilité d'annulation du programme, même lorsque le résultat $\widehat{\beta}$ de l'évaluation d'impact devient excessivement négatif. Par exemple, lorsque $\widehat{\beta} \approx -5$ (cas où il sera considéré comme étant très significatif par l'évaluateur fréquentiste), la probabilité d'annulation est toujours située en-dessous de 5 %. Pour que la probabilité d'annulation s'élève à 50 %, le résultat de l'évaluation d'impact devrait atteindre une valeur bien en-dessous de -25. Tel est le pouvoir de l'inertie bureaucratique dans le monde de la prise de décision politique.

Une dernière représentation graphique des résultats présentés ci-dessus est fournie par la Figure 4, laquelle illustre la Proposition 3 en représentant la probabilité d'annulation en fonction du résultat $\widehat{\beta}$ de l'évaluation d'impact et de l'influence $\alpha_F$ du chercheur-évaluateur fréquentiste. Il est clair que la probabilité d'annulation tend vers zéro pour toutes les valeurs de $\widehat{\beta}$ lorsque l'influence $\alpha_F$ du chercheur-évaluateur fréquentiste tend vers zéro.

Figure 4 : *Une illustration de la Proposition 3 : au fur et à mesure que l'influence $\alpha_F$ du chercheur-évaluateur fréquentiste tend vers 0, la probabilité d'annulation du programme tend vers 0 et ce pour toutes les valeurs de $\widehat{\beta}$.*



## 3 CONCLUSION

Un dernier point reste à aborder avant de tirer une conclusion. La neutralité vis-à-vis du risque qui caractérise fondamentalement les préférences dans toute évaluation d'impact constitue-t-elle une hypothèse raisonnable ? Ne faudrait-il pas tenir compte de la possibilité d'aversion au risque pour un sujet se rapportant à un problème de choix social ?

Au regard de notre méthodologie statistique, lorsque nous estimons l'impact d'un programme, nos estimateurs visent à minimiser une *fonction de risque* fondée sur une fonction de perte quadratique :

$$\arg\min_{\{\hat{\beta}\}} \int \left(\hat{\beta}-\beta\right)^2 f(\beta)\, d\beta = \int \beta f(\beta)\, d\beta = \mu_\beta \tag{23}$$

Ceci revient à faire une hypothèse de neutralité vis-à-vis du risque en termes d'utilité. Du point de vue de la terminologie de l'évaluation d'impact, ceci est la raison pour laquelle on se focalise sur l'effet moyen du traitement (EMT).

Mais du point de vue du choix social, nous devons choisir un estimateur qui corresponde à notre critère de bien-être social, et l'aversion au risque devrait être prise en compte. Considérons une fonction d'utilité de classe CARA : $W(\beta,\theta) = 1 - \exp\{-\theta\beta\}$ où $\theta$ est le coefficient d'Arrow-Pratt d'aversion absolue au risque. Nous devons ainsi choisir un estimateur $\hat{\beta}$ qui correspond à :

$$\arg\min_{\{\hat{\beta}\}} \int \left[W\left(\hat{\beta},\theta\right) - W(\beta,\theta)\right]^2 f(\beta)\, d\beta \tag{24}$$

Comme dans (23), la dérivation de cet estimateur est rendue possible par l'existence de la quantité suivante :

$$\int_{-\infty}^{+\infty} W\left(\hat{\beta},\theta\right) f(\beta)\, d\beta$$

En outre, comme il a été démontré dans cet article, la distribution appropriée qui doit être utilisée pour $f(\beta)$ afin de décider s'il faut approuver ou non le programme à la base, est donnée par la distribution d'un t de Student. En appliquant le résultat bien connu de Geweke (2001) ou en faisant des calculs simples, il est aisé de démontrer que l'intégrale $\int_{-\infty}^{+\infty} W\left(\hat{\beta},\theta\right) t(\beta)\, d\beta$ diverge, et que le décideur adverse au risque ne peut donc même pas calculer l'utilité espérée associée à l'approbation d'un programme au départ[12]. Le même raisonnement s'applique au décideur bayésien suite à l'évaluation d'impact, dans la mesure où la croyance a posteriori est également donnée par un t de Student. Les implications de cette absence d'une fonction de bien-être social calculable, lorsqu'on relâche l'hypothèse de neutralité vis-à-vis du risque, et lorsqu'il s'agit d'évaluer un programme soit *ex ante* en termes d'approbation, soit *ex post* en termes bayésiens, doivent faire l'objet d'une future recherche.

Bien sûr, il existe plusieurs arguments économétriques valides qui peuvent conduire à recommander le maintien d'un programme malgré une évaluation d'impact débouchant sur un EMT fréquentiste négatif et statistiquement significatif. Par exemple, si l'effet marginal de traitement formalisé par Heckman et Vytlacil (1999) augmente avec les caractéristiques inobservées qui déterminent l'état du traitement, et si l'effet du traitement sur les traités est très négatif et l'effet du traitement sur les non-traités est très positif (avec un EMT négatif et statistiquement significatif), alors il existe d'excellentes raisons pour ne pas annuler le programme si les individus non-traités peuvent être traités lorsqu'il est maintenu[13].

Néanmoins, un tel niveau de sophistication dans le processus de prise de décision reste encore à voir. Malheureusement, il est fort probable que les arguments d'inertie bureaucratique développés dans cet article sur la base d'un modèle bayésien soient bien proches de la triste réalité. Ils expliquent pourquoi nous « n'apprenons presque jamais rien » des évaluations d'impact et pourquoi elles conduisent rarement à une prise de décision fondée sur les faits.

## RÉFÉRENCES

BLOOM, H. S. (1995): "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs, " *Evaluation Review*, 19(5), 547—556.

CGD (2006): "When Will We Ever Learn? Improving Lives through Impact Evaluation," Report of the Evaluation Gap Working Group. Washington, DC: CGD (www.cgdev.org/content/publications/detail/7973).

GEWEKE, J. (2001): "A Note on Some Limitations of CRRA Utility," *Economics Letters,* 71, 341—345.

[12] Formellement, le résultat de Geweke s'applique aux fonctions d'utilité de classe CRRA, mais il tient également pour celle de classe CARA.

[13] Voir Heckman et Vytlacil (2005) pour une analyse exhaustive de l'ETM et de ses liens avec l'EMT, l'effet du traitement sur les traités et l'effet du traitement sur les non-traités.

HECKMAN, J. J., E. J. VYTLACIL (1999): “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects,” *Proceedings of the National Academy of Sciences*, 96(8), 4730—4734.

HECKMAN, J. J., E. J. VYTLACIL (2005): “Structural Equations, Treatment Effects and Econometric Policy Evaluation,” *Econometrica*, 73(3), 669—738.

JIA, H., S. SKAPERDAS, S. VAIDYA (2011): “Contest Functions: Theoretical Foundations and Issues in Estimation,” Deakin University and UC Irvine.

ODI (2011): “Learning How to Learn: Eight Lessons for Impact Evaluations that Make a Difference,” Overseas Development Institute Background Note, April 2011.

O’HAGAN, A. (1994): Bayesian Inference, vol. 2B of Kendall’s Advanced Theory of Statistics. Arnold, London, UK.