

# The (lack of) impact of impact: Why impact evaluations seldom lead to evidence-based policymaking

Jean-Louis Arcand

Director, Centre for Finance and Development  
Professor, Department of International Economics,  
The Graduate Institute, Geneva, Switzerland  
and

Senior Fellow, FERDI, Clermont Ferrand, France

email: [jean-louis.arcand@graduateinstitute.ch](mailto:jean-louis.arcand@graduateinstitute.ch)  
<http://graduateinstitute.ch/economics/faculty/arcand>

June 15, 2013

## Abstract

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.

Keywords: impact evaluation, Bayesian analysis, contest success functions  
JEL: O12, D04, D72, C11, C21, C72

Any academic researcher who has been associated with impact evaluations in the field can tell horror stories of how his or her findings were not subsequently taken into account by decisionmakers, in terms of deciding whether or not to continue with an existing program. My own personal experience is a case in point. Several years ago I carried out an impact evaluation of a major program funded by an important multilateral donor. After three years of work, we found statistically significant negative effects of the program on our main response variable. Simple budgetary logic would suggest that the donor should have taken the results of what was a rigorously conducted impact evaluation into account. But it did not. At the end of the project cycle, and when the time came for the donor to decide whether or not to cancel the program (as I was urging), they chose not to cancel it. Why?

In this paper, I show, using elementary Bayesian arguments and a barebones model of competition among three types of decisionmakers, that it is highly likely that a program that a frequentist evaluator would recommend cancelling, but that was approved to begin with, will be continued. The result obtains because of the strategic interaction, in terms of pushing for their preferred outcomes, amongst a frequentist evaluator, a policymaker who initiated the program in the first place on the basis of her priors, and a Bayesian decisionmaker who attempts to combine the priors and the results of the impact evaluation in a statistically rational manner.

The development community has long been aware of the dearth of evidence-based policymaking. For example, the Center for Global Development Evaluation Gap Working Group's 2006 paper, *When Will We Ever Learn?*, which led to the establishment of the International Initiative for Impact Evaluation (3ie, see CGD (2006))

presented the lack of rigorous impact evaluations as the missing piece in learning about social development efforts. It advocated a renewed approach to aid evaluation that would bring greater accuracy and credibility to assessments of impact and, by extension, to development policy and practice.... it set out the challenge in clear terms: that 10 years on from the publication of the report, more and better rigorous impact evaluations would need to be in place

if we wanted a stronger evidence-base for making decisions.(ODI 2011)

Unfortunately for evidence-based policymaking, there may be many more rigorous impact evaluations available today than in 2006, but the quality of policymaking has seen little if any improvement.

The intuition behind the model presented in this paper is extremely simple. Suppose that  $\hat{\beta}$  is the impact of the program on the response variable of interest uncovered by the impact evaluation and, without loss of generality, normalize the cost of the program to zero. In order to be deployed, a program must be approved, and it must therefore be the case that there are good *a priori* reasons to believe that its impact will outweigh its cost. In statistical terms, this means that decisionmakers have a prior (which I shall denote by  $\mu_\beta$ ) concerning the program, and this prior must be such that  $\beta$  is "believed" *ex ante* to be greater than zero ( $\mu_\beta > 0$ ) in a statistically significant sense.<sup>1</sup> Bayesian theory tell us that the updated beliefs (which I shall denote by  $\mu_\beta^*$ ) of a Bayesian decisionmaker will be a convex combination of  $\hat{\beta}$  and  $\mu_\beta$ . If the prior is large and positive, the result  $\hat{\beta}$  of the impact evaluation must be sufficiently negative for the posterior  $\mu_\beta^*$  to be negative. Thus, if the prior is sufficiently positive, even a negative and statistically significant impact evaluation result may not be sufficient to generate a negative and statistically significant posterior, in which case it will be rational for a Bayesian decisionmaker to be against cancelling the program, although the frequentist evaluator is in favor of cancellation. If we combine the need for a strongly negative impact evaluation result with the inertia generated by the fact that many of the decisionmakers who initiated the program in the first place, and who often base their decisionmaking entirely upon their prior, are likely to possess a high degree of decisionmaking power, it should come as no surprise that cancellation will be relatively unlikely.

The paper is organized as follows. In section 1 I set the scene in simple Bayesian terms. I establish the conditions under which the program will be approved to begin with, and study the corresponding conditions under which a frequentist academic evaluator

---

<sup>1</sup>In operational terms, and to take the example of World Bank practice, this is what is done in Project Appraisal Documents (PADs), which set out the likely benefits of the program (and that fact that they exceed costs) in order to secure approval by the Bank's Board of Directors for the associated loan instrument.

and a Bayesian decisionmaker will be in favor either of continuing or of terminating the program. Starting from first principles, these conditions are shown to be given by the results of standard Student  $t$ -tests of statistical significance. Having established each agent's preferences as a function of the results of the impact evaluation, section 2 embeds the initial setup within a simple model of competition phrased in terms of the contest success function approach that constitutes a workhorse of the political science literature. I establish several results. First, the probability of cancellation is a decreasing function of  $\hat{\beta}$  and  $\mu_\beta$ . Second, the probability of cancellation is an increasing function of the variance of  $\mu_\beta$  and of the variance of the treatment status dummy. Third, the probability of cancellation is a decreasing function of the effectiveness of the influence exerted by frequentist evaluators on policymaking. Section 3 concludes by offering some thoughts on the limitations inherent in basing policymaking on the Average Treatment Effect (ATE), which is akin to assuming a risk neutral social welfare function.

## 1 A simple Bayesian model

When we carry out an impact evaluation, we come up with an estimate of its impact,  $\hat{\beta}$ . For simplicity, assume that this estimate comes from a Randomized Control Trial (RCT) in which the treatment effect is estimated through a simple least squares regression of the form:

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

where  $Y_i$  is the outcome of interest,  $D_i$  is the treatment dummy (equal to 1 when individual  $i$  is treated by the program and zero otherwise) and  $U_i$  is a disturbance term that satisfies the usual Gauss-Markov assumptions, with  $U_i \sim N(0, \sigma^2)$ .<sup>2</sup> Sample size will be denoted by  $n$ .

---

<sup>2</sup>None of the results that follow are dependent on  $D_i$  being a dummy variable: they are also valid were  $D_i$  to be a continuous measure of the intensity of treatment.

## 1.1 Conjugate priors

In the standard Bayesian linear model, described, for example, in O’Hagan (1994), one specifies a conjugate prior  $p(\beta, \sigma^2)$  for  $\beta$  and  $\sigma^2$  of the Normal-Inverse-Gamma ( $NI\Gamma^{-1}$ ) form:

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

or more explicitly:

$$\begin{aligned} N(\mu_\beta, \sigma^2 V_\beta) &= \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2 V_\beta}} \exp \left\{ -\frac{(\beta - \mu_\beta)^2}{2\sigma^2 V_\beta} \right\}, \\ \Gamma^{-1}(a, b) &= \frac{b^a \exp \left\{ -\frac{b}{\sigma^2} \right\}}{\sigma^{2(1+a)} \Gamma(a)}, \end{aligned}$$

yielding:

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

Integrating  $\sigma^2$  out of this expression yields:

$$\begin{aligned} p(\beta) &= \int_0^{+\infty} \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2 V_\beta}} \exp \left\{ -\frac{(\beta - \mu_\beta)^2}{2\sigma^2 V_\beta} \right\} \frac{b^a \exp \left\{ -\frac{b}{\sigma^2} \right\}}{\sigma^{2(1+a)} \Gamma(a)} d\sigma^2 \\ &= \frac{2^a (bV_\beta)^a [2bV_\beta + (\beta - \mu_\beta)^2]^{-\frac{1}{2}-a} \Gamma(\frac{1}{2} + a)}{\sqrt{\pi} \Gamma(a)}. \end{aligned} \quad (2)$$

Inspection of (2) implies that it can be rewritten as:

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

where the Right-Hand-Side (RHS) of (3) is a Student’s  $t$  distribution with mean  $\mu_\beta$ , scale parameter  $\frac{b}{a} V_\beta$  and degrees of freedom  $2a$ .<sup>3</sup> This distribution for the unconditional prior of  $\beta$ , which is the information that decisionmakers have when they decide to approve funding for the program, implies that the mean and variance that decisionmakers have

---

<sup>3</sup> $B(\cdot, \cdot)$  is Euler’s Beta function.

in mind *ex ante facto* are given by:

$$E[\beta] = \mu_\beta, \text{Var}[\beta] = \frac{b}{a-1}V_\beta. \quad (4)$$

## 1.2 Prior approval of the program

For the program to have been approved, it must be the case that the decisionmakers in question believed its impact to be greater than its cost in a statistical sense, which we can write as the usual inequality:

$$E[\beta] - t_{\alpha, n-1} \sqrt{\text{Var}[\beta]} = \mu_\beta - t_{\alpha, n-1} \sqrt{\frac{b}{a-1}V_\beta} > 0, \quad (5)$$

where the  $t_{\alpha, n-1}$  term represents the critical value for a standard Student's  $t$ -test with  $1-\alpha$  level of confidence and  $n-1$  degree of freedom.<sup>4</sup> In intuitive terms, the decisionmaker approves the program if zero lies strictly *below* the lower bound of the  $1-\alpha$  level confidence interval.

## 1.3 Posteriors following the impact evaluation

In turn, by Bayes Rule, received theory tells us that the joint posterior distribution  $p(\beta, \sigma^2|Y) = \frac{p(\beta, \sigma^2)p(Y|\beta, \sigma^2)}{\int \int p(\beta, \sigma^2)p(Y|\beta, \sigma^2)d\beta d\sigma^2}$  that can be computed once the impact evaluation is carried out is given by a  $N\Gamma^{-1}$  distribution with updated parameters:

$$\mu_\beta^* = \frac{V_\beta^{-1}\mu_\beta + D'Y}{V_\beta^{-1} + D'D} = \frac{V_\beta^{-1}}{V_\beta^{-1} + D'D}\mu_\beta + \frac{D'D}{V_\beta^{-1} + D'D}\hat{\beta}, V_\beta^* = \frac{1}{V_\beta^{-1} + D'D}, \quad (6)$$

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

In turn, proceeding as in (2), the marginal posterior distribution of  $\beta$  is given by a Student's  $t$  of the same form as that given on the RHS of (3), with:

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

---

<sup>4</sup>For example, for such a one-sided test with infinite sample size and a confidence level of 97.5%,  $t_{\alpha, n-1} \approx 1.96$ .

## 1.4 The impact of the impact evaluation

What is the impact of the impact evaluation?

### 1.4.1 Frequentist evaluators

For many researchers, who do not take into account the fact that the program was approved in the first place, the only quantity of interest is the least squares estimate  $\hat{\beta} = \frac{D'Y}{D'D}$  from the regression given by (1). If  $\hat{\beta} > 0$  and:

$$\hat{\beta} - t_{\alpha, n-1} \hat{\sigma} \sqrt{D'D}^{-1} > 0, \quad (8)$$

where  $\hat{\sigma}^2 = \frac{1}{n-1} (Y - D\hat{\beta})'(Y - D\hat{\beta}) = \frac{1}{n-1} (Y'Y - D'D\hat{\beta}^2)$  is the least squares estimate of  $\sigma^2$ , the researcher will declare that the program "works" in the sense that its impact is positive, in a statistically significant sense. But if  $\hat{\beta} < 0$  and:

$$\hat{\beta} + t_{\alpha, n-1} \hat{\sigma} \sqrt{D'D}^{-1} < 0, \quad (9)$$

(i.e. if zero lies strictly *above* the upper bound of the  $1 - \alpha$  level confidence interval) the researcher will declare that the program is a failure, and should be terminated. In terms of threshold values of  $\hat{\beta}$ , the frequentist evaluator will therefore recommend cancellation of the program when:

$$\hat{\beta} < \beta_* = -\frac{t_{\alpha, n-1} Y'Y}{\sqrt{(n-1 + t_{\alpha, n-1}^2)(D'D)(Y'Y)}} < 0,$$

whereas she will recommend continuation when:

$$\hat{\beta} > \beta^* = \frac{t_{\alpha, n-1} Y'Y}{\sqrt{(n-1 + t_{\alpha, n-1}^2)(D'D)(Y'Y)}} > 0.$$

When  $\beta_* < \hat{\beta} < \beta^*$ , the results of the impact evaluation will be deemed to be inconclusive. Of course, this makes no sense from the decisionmaker's perspective, because it is reasonable to assume that at least some of the decisionmakers in question update their

prior beliefs based on the results of the impact evaluation.<sup>5</sup>

### 1.4.2 Bayesian decisionmakers

When would the decisionmakers, upon updating their beliefs following the impact evaluation, decide that the program should be terminated? In formal terms, this will be the case when  $\mu_\beta^* < 0$  and:

$$\mu_\beta^* + t_{\alpha, n-1} \sqrt{\frac{b^*}{a^* - 1} V_\beta^*} < 0, \quad (10)$$

where we now of course work with the posteriors. Of course, if  $\mu_\beta^* > 0$  and:

$$\mu_\beta^* - t_{\alpha, n-1} \sqrt{\frac{b^*}{a^* - 1} V_\beta^*} > 0, \quad (11)$$

a Bayesian decisionmaker will be in favor of continuing the program.

Substituting from (6) and (7) implies that inequality (10) can be rewritten as:

$$\frac{\mu_\beta + V_\beta D' D \hat{\beta} + t_{\alpha, n-1} \sqrt{\frac{V_\beta \Omega}{n-2a-2}}}{1 + V_\beta D' D} < 0, \quad (12)$$

where:

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

with a similar expression for (11). Figure 1 provides a standard graphical representation of inequalities (8), (9), (10), and (11). I set parameters such that the program is approved in the first place. To do so, and for purely illustrative purposes, I pose  $\mu_\beta = 2$ ,  $V_\beta = 1.00$ ,  $a = 75$ ,  $b = 40$ ,  $D' D = 0.2$ ,  $Y'Y = 1$  and  $n = 10$ . The area within the large ellipse corresponds to values of  $\hat{\beta}$  such that  $\mu_\beta^* - t_{\alpha, n-1} \sqrt{\frac{b^*}{a^* - 1} V_\beta^*} < 0 < \mu_\beta^* + t_{\alpha, n-1} \sqrt{\frac{b^*}{a^* - 1} V_\beta^*}$  and for which the Bayesian decisionmaker can therefore not take a decision concerning the statistical significance of the estimate. The area within the small ellipse corresponds to the same situation of statistical insignificance for the frequentist evaluator.

After some tedious algebra, it can be shown that the roots of the quadratic equations

---

<sup>5</sup>Time inconsistency issues in terms of decisionmaking are beyond the scope of this paper.



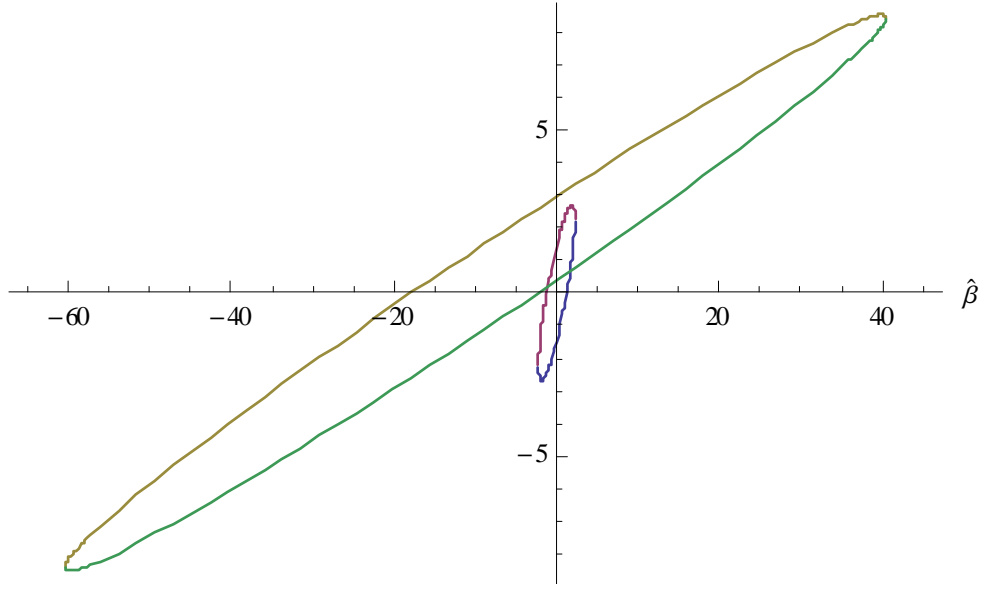


Figure 1: An illustration of  $\hat{\beta} - t_{\alpha, n-1} \hat{\sigma} \sqrt{D'D}^{-1}$ ,  $\hat{\beta} + t_{\alpha, n-1} \hat{\sigma} \sqrt{D'D}^{-1}$ ,  $\frac{\mu_{\beta} + V_{\beta} D' D \hat{\beta} - t_{\alpha, n-1} \sqrt{\frac{V_{\beta} \Omega}{n-2a-2}}}{1+V_{\beta} D' D}$ , and  $\frac{\mu_{\beta} + V_{\beta} D' D \hat{\beta} + t_{\alpha, n-1} \sqrt{\frac{V_{\beta} \Omega}{n-2a-2}}}{1+V_{\beta} D' D}$ .

in  $\beta$  obtained by setting the LHS of (11) or (12) equal to zero are given by:

$$\beta_{+}, \beta_{-} = \frac{-\mu_{\beta} \pm t_{\alpha, n-1} \sqrt{\frac{(1+V_{\beta} D' D)(2bV_{\beta} + V_{\beta} Y' Y + \mu_{\beta}^2)}{n(n-2a-2) + t_{\alpha, n-1}^2}}}{V_{\beta} D' D} < 0. \quad (13)$$

This leads to the following Proposition:

**Proposition 1** *It will be rational for a Bayesian decisionmaker to cancel the program when  $\hat{\beta} < \beta_{-}$  and to continue the program when  $\hat{\beta} > \beta_{+}$ . When  $\beta_{-} < \hat{\beta} < \beta_{+}$  the Bayesian decisionmaker is indifferent between continuing and cancelling the program.*

**Proof.** It is easy to show that the LHS of (12) is equal to zero for

$$\hat{\beta} = \beta_{-} = \frac{-\mu_{\beta} - t_{\alpha, n-1} \sqrt{\frac{(1+V_{\beta} D' D)(2bV_{\beta} + V_{\beta} Y' Y + \mu_{\beta}^2)}{n(n-2a-2) + t_{\alpha, n-1}^2}}}{V_{\beta} D' D}.$$

Moreover, straightforward differentiation of the LHS of (12) yields:

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

where the sign of the inequality follows from the fact that in this configuration  $\mu_{\beta} + V_{\beta} D' D \widehat{\beta} < 0$ . Thus (12) will hold for  $\widehat{\beta} < \beta_-$ . Similar arguments show that (11) will hold when  $\widehat{\beta} > \beta_+$ . As a result, neither inequality holds when  $\beta_- < \widehat{\beta} < \beta_+$ , and the Bayesian decisionmaker will be indifferent between cancelling and continuing the program when this configuration obtains. ■

Proposition 1 shows that it will be rational for a Bayesian decisionmaker to cancel the program when  $\widehat{\beta}$  is "sufficiently small", whereas she will be in favor of continuing the program when it is "sufficient large".

Figure 2 presents an illustration of Proposition 1. For the parameter configuration used in Figure 1,  $\frac{\mu_{\beta}}{\sqrt{\frac{b}{a-1} V_{\beta}}} \approx 2.72$  so that, based on the prior, the program would be approved without trouble at conventional levels of confidence.

In the Figure, I plot the Student  $t$ -statistics associated with varying the result  $\widehat{\beta}$  of the impact evaluation, holding all other parameter values constant. For the prior, the associated  $t$ -statistic is constant and equal to 2.72 for all values of  $\widehat{\beta}$ . The inverse  $S$ -shaped curve plots the  $t$ -statistic  $\frac{\widehat{\beta}}{\widehat{\sigma} \sqrt{D' D}^{-1}}$  associated with the result of the impact evaluation, as seen by a frequentist evaluator. When the value of the associated  $t$ -statistic crosses the upper confidence band and lies above it ( $\widehat{\beta} > \beta^* > 0$ ), the frequentist evaluator will declare the program to be a success, and will be in favor of its continuation. When the value of  $\widehat{\beta}$  is such that the  $t$ -statistic lies below the lower confidence band ( $\widehat{\beta} < \beta_* < 0$ ), the frequentist evaluator will be in favor of cancellation. For all intermediate values, she will be indifferent between cancelling and continuing the program. Similar arguments can be made with respect to the Bayesian evaluator's  $t$ -statistic  $\frac{\mu_{\beta}^*}{\sqrt{\frac{b^*}{a^*-1} V_{\beta}^*}}$ . Notice, for this parameter configuration, and given the strength of the prior, that the Bayesian evaluator will still be in favor of continuing the program even when the estimated  $\widehat{\beta}$  is just

above  $\beta_+ \approx -2.2$ , whereas she will only be in favor of cancellation once the estimated impact of the program is below  $\beta_- \approx -17.6$ .

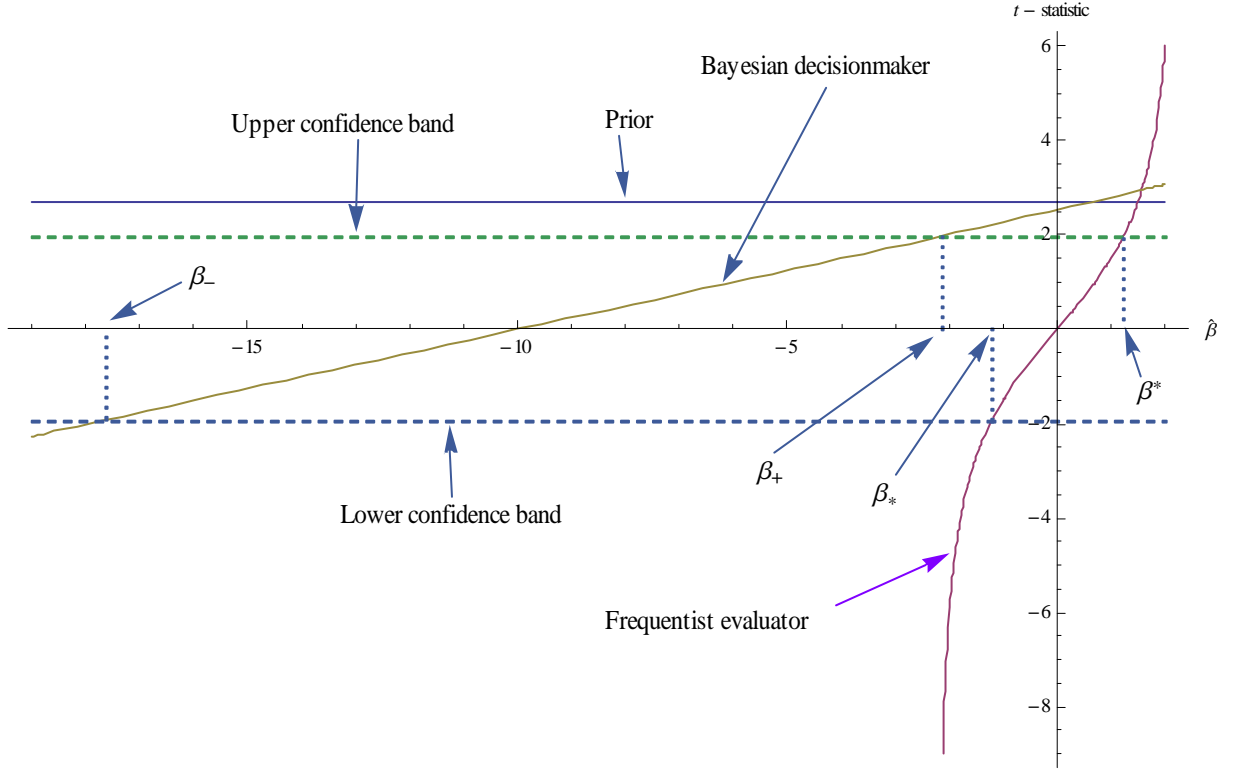


Figure 2: An illustration of Proposition 1.

Note in Figure 2 that  $\beta_- < \beta_+ < \beta_*$ . In this case, for  $\beta_+ < \hat{\beta} < \beta_*$ , the frequentist evaluator is for cancellation and the Bayesian decisionmaker is for continuation. This is not a general property and it is entirely possible, with a different configuration of parameter values, that one would obtain  $\beta_- < \beta_* < \beta_+$ . In this case, when  $\beta_* < \hat{\beta} < \beta_+$ , the frequentist evaluator will be indifferent between cancellation and continuation, whereas the Bayesian decisionmaker will be in favor of cancellation.

In passing, it is worth noting (since  $D'D = (n-1)\sigma_D^2$ ,  $Y'Y = (n-1)\sigma_Y^2$  and  $Y'Y - D'D\hat{\beta}^2 = (n-1)\hat{\sigma}^2$ ), that:

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

and thus that:

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

Therefore, if the impact evaluation were based on a sample of infinite size, its result would entirely determine the decision to cancel the program, as one would intuitively expect.

## 2 Evidence-based policymaking?

I consider a situation in which, following the impact evaluation, there are three decision-makers who compete to determine whether the program is continued or whether it is terminated. The three decisionmakers are:

- **Anti-evaluation decisionmakers:** they are afraid of evaluation and prefer to base their decision solely on the prior; they get benefits from continuing a program that has been approved in the past;<sup>6</sup>
- **Bayesian decisionmakers:** they will base their decision on the posterior, thereby combining their prior with the results of the impact evaluation;
- **Frequentist (academic) impact evaluators:** they will base their decisions solely on the results of the impact evaluation.

### 2.1 Competition amongst actors

I assume that the decisionmaking process that leads to the program being continued or terminated takes the form of a competition amongst the three decisionmakers, who each invest resources to have their preferred outcome obtain.<sup>7</sup> For illustrative purposes,

---

<sup>6</sup>To quote a senior Scandinavian development aid official whom I once questioned about why her agency was not evaluating its programs through impact evaluations: "We have been carrying out the same programs for 20 years, so they must be working."

<sup>7</sup>When most impact evaluations are being planned, a key parameter that must be established is sample size  $n$ . As pointed out in the seminal article by Bloom (1995), a larger  $n$  reduces the Minimal Detectable Effect Size (MDES), meaning that the statistical power of the sample design is greater. If one assumes as a first approximation that sample size is determined so that the MDES is equal to the prior concerning the impact of the program, and if one takes the case of an RCT, the standard formula for the MDES for a balanced sample ( $P = \frac{1}{2}$ ) is given by  $MDES = \frac{M_{n-2}}{\sqrt{n}} \sqrt{\frac{1}{P(1-P)}} = \frac{2M_{n-2}}{\sqrt{n}}$ ;  $M_{n-2}$  will be roughly equal to 2.5 for a one-sided statistical test with conventional power of 0.80 and significance level of 0.05. Since the prior is that the effect size is  $\frac{\mu_\beta}{\sqrt{\frac{b}{a-1}V_\beta}}$ , this implies that sample size would be given by  $n = \frac{4b}{a-1}\mu_\beta^{-2}V_\beta M_{n-2}^2$  on the basis of the prior. A frequentist evaluator, on the other hand, might wish to increase sample size so as to increase the weight of the impact evaluation in terms of policymaking. Including sample size as a variable over which lobbying takes place is beyond the scope of this paper.

consider the following standard simple "ratio form" *contest success functions*, where the probability of winning (having one's preferred outcome obtain) is given by:<sup>8</sup>

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

where  $\alpha_i$  is the relative effectiveness of effort expended by agent  $i$ , and  $p_i$  therefore gives the probability that agent  $i$  prevails, as a function of the effort expended  $(e_A, e_B, e_F)$  by all three parties involved.

Assuming risk neutrality, notice that the benefits to agent  $A$  from continuing the program can be written as:

$$B_A = \mu_\beta, \quad (15)$$

where  $B_A > 0$  since the program was approved in the first place. When the impact evaluation implies that the program should be cancelled for agent  $F$  ( $\hat{\beta} < \beta_* < 0$ ),

$$B_F = -\hat{\beta} > 0. \quad (16)$$

Note that the objective function for agent  $F$  involves  $-\hat{\beta}$  since that is the *gain* obtained from cancelling the program and not inflicting an average *loss* of  $\hat{\beta}$  on beneficiaries (recall in passing that we have normalized the cost of the program to zero). Another way of seeing this is that it represents the opportunity *gain* from cancelling the program.

In what follows, I consider the four possible configurations of the impact evaluation results.<sup>9</sup> In all cases, the anti-evaluation decisionmaker is in favor of continuation of the program:

- $\hat{\beta} < \beta_- < \beta_+ < \beta_* < \beta^*$  : agent  $B$  is in favor of cancellation, agent  $F$  is in favor of

---

<sup>8</sup>A vast literature exists on contest success functions. An excellent recent survey on the various functional forms and the manner in which they may be derived axiomatically, from stochastic specifications, or from mechanism design principles is provided by Jia, Skaperdas, and Vaidya (2011).

<sup>9</sup>In the alternative configuration in which  $\beta_* < \beta_+$ , there are only two interesting cases to consider. To wit:  $\hat{\beta} < \beta_- < \beta_* < \beta_+ < \beta^*$  : agent  $B$  is in favor of cancellation, agent  $F$  is in favor of cancellation;  $\beta_- < \hat{\beta} < \beta_* < \beta_+ < \beta^*$  : agent  $B$  is indifferent between cancellation and continuation, agent  $F$  is in favor of cancellation;  $\beta_- < \beta_* < \hat{\beta} < \beta_+ < \beta^*$  : agent  $B$  is indifferent between cancellation and continuation, agent  $F$  is indifferent between cancellation and continuation –the program is continued with probability 1;  $\beta_- < \beta_* < \beta_+ < \hat{\beta} < \beta^*$  : agent  $B$  is in favor of continuation, agent  $F$  is indifferent between cancellation and continuation –the program is continued with probability 1.

cancellation;

- $\beta_- < \widehat{\beta} < \beta_+ < \beta_* < \beta^*$  : agent  $B$  is indifferent between cancellation and continuation, agent  $F$  is in favor of cancellation;
- $\beta_- < \beta_+ < \widehat{\beta} < \beta_* < \beta^*$  : agent  $B$  is in favor of continuation, agent  $F$  is in favor of cancellation;
- $\beta_- < \beta_+ < \beta_* < \widehat{\beta} < \beta^*$  : agent  $B$  is in favor of continuation, agent  $F$  is indifferent between cancellation and continuation –the program is therefore continued with probability 1.

## 2.2 A significantly negative posterior: $\widehat{\beta} < \beta_- < \beta_+ < \beta_* < \beta^*$

When the impact evaluation, through the posterior, implies that the program should be cancelled for agent  $B$ , we have:

$$B_B = -\mu_\beta^* > 0. \quad (17)$$

Assuming the preceding configuration of outcomes, and since agents  $B$  and  $F$  both want the program to be cancelled (which implies that  $p_B = p_F = 1 - p_A$ ) the expected payoffs of the agents are given by:

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

where, for the sake of the transparency of the results that follow, I assume that the marginal cost  $c$  of expending effort on achieving one's preferred outcome is constant and identical for each of the parties.<sup>10</sup> Each agent is assumed to solve a standard maximization problem which yields her optimal level of effort:

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

---

<sup>10</sup>More complicated equilibrium outcomes will obtain when the cost functions differ by player, but I prefer to eschew these complications and focus on the impact on the ensuing equilibrium of the characteristics of the prior, the results of the impact evaluation, the variances associated with the prior and the impact evaluation, and the relative influence of each party.

Again, both for simplicity and to ensure the uniqueness of the ensuing equilibrium, assume that  $\gamma = 2$ .<sup>11</sup> In this case, and after some tedious algebra, the Nash equilibrium levels of effort furnished by each party are given by:

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

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

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

Substituting from (19), (20) and (21) into the contest success function then yields the equilibrium probability that the program will be continued as a function of the prior mean impact, the posterior mean impact, the estimate stemming from the impact evaluation, and the influence that each agent has on the outcome:

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

Substituting from the expression for the posterior yields:

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

where

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

One then has the following Proposition:

---

<sup>11</sup>The parameter  $\gamma$  represents the "informativeness" of the contest. As  $\gamma \rightarrow 0$ , the contest tends towards a randomization in which the actions of the agents have no effect whatsoever on the outcome, whereas as  $\gamma \rightarrow \infty$  the contest tends towards an "all pay auction" in which the contestant who furnishes an infinitesimally greater level of effort reaps the entirety of the reward. In terms of uniqueness of the equilibrium, note that setting  $\gamma = 1/2$  yields a pair of equilibria (one of which obtains under certain conditions on the parameters) but that the qualitative flavor of the results that follow remain unchanged.

**Proposition 2** *The likelihood of the program being cancelled, when it was approved in the first place and when both Bayesian and frequentist evaluators are in favor of cancellation ( $\widehat{\beta} < \beta_- < \beta_+ < \beta_* < \beta^*$ ) admits the following comparative statics:*

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

**Proof.** Straightforward, though cumbersome, differentiation of (22) with respect to  $\mu_\beta$  and  $\widehat{\beta}$  yields:

$$\begin{aligned} \frac{\partial(1-p_{A-}^*)}{\partial\mu_\beta} &= -2\Psi^{-2}\alpha_A\alpha_B\alpha_F\widehat{\beta}^2\mu_\beta(V_\beta D'D\widehat{\beta} + \mu_\beta) \\ &\quad \times \left[\alpha_F V_\beta D'D(1+V_\beta D'D)^2\widehat{\beta}^3 + \alpha_B(V_\beta D'D\widehat{\beta} + \mu_\beta)^3\right] \\ &= -\left(\frac{\mu_\beta}{\beta}\right) \frac{\partial(1-p_{A-}^*)}{\partial\widehat{\beta}}. \end{aligned}$$

Since, under the configuration I have assumed,  $\widehat{\beta} < 0$ ,  $V_\beta D'D\widehat{\beta} + \mu_\beta < 0$  and  $\widehat{\beta} - \mu_\beta < 0$ , it follows that  $-\frac{\partial(1-p_{A-}^*)}{\partial\mu_\beta} = \frac{\partial(1-p_{A-}^*)}{\partial\beta} < 0$ . Similarly,  $\frac{\partial(1-p_{A-}^*)}{\partial V_\beta} = 2\Psi^{-2}\alpha_A\alpha_B\alpha_F^2 V_\beta(1+V_\beta D'D)\widehat{\beta}^4(\widehat{\beta} - \mu_\beta)\mu_\beta^2(V_\beta D'D\widehat{\beta} + \mu_\beta) = \left(\frac{V_\beta}{D'D}\right) \frac{\partial(1-p_{A-}^*)}{\partial(D'D)} > 0$ . ■

As one would intuitively expect, Proposition 2 shows that, *ceteris paribus*, the higher the coefficient  $\widehat{\beta}$  associated with the impact evaluation, the lower the likelihood that the program is cancelled. A similar, intuitively appealing, result obtains with respect to the prior  $\mu_\beta$ . Moreover, an increase in the variance of the treatment dummy ( $D'D$ ) increases the likelihood of the program being cancelled, as does an increase in the variance associated with the prior.

Another interesting result obtains when one considers the effect of a change in the effectiveness of lobbying by either the Bayesian or the frequentist evaluator. I express this as the following obvious Proposition:

**Proposition 3** *When  $\widehat{\beta} < \beta_- < \beta_+ < \beta_* < \beta^*$ , the likelihood of cancellation of the program is an increasing function of the influence of the Bayesian and the frequentist evaluators, and a decreasing function of the influence of the anti-evaluation decision-maker. Moreover, when the influence  $\alpha_F$  ( $\alpha_B$ ) of the frequentist (Bayesian) evaluator is zero, the probability of cancellation of the program is zero.*



**Proof.** Trivial differentiation of (22) yields:

$$\begin{aligned} \frac{\partial(1 - p_{A-}^*)}{\partial\alpha_A} &= -\Psi^{-2}\alpha_B\alpha_F\widehat{\beta}^2\mu_\beta^2(V_\beta D' D\widehat{\beta} + \mu_\beta)^2 \\ &\times [\alpha_F\widehat{\beta}^2(1 + V_\beta D' D)^2 + \alpha_B(V_\beta D' D\widehat{\beta} + \mu_\beta)^2] \leq 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial(1 - p_{A-}^*)}{\partial\alpha_B} &= \Psi^{-2}\alpha_A\alpha_F^2(1 + V_\beta D' D)^2\widehat{\beta}^4\mu_\beta^2(V_\beta D' D\widehat{\beta} + \mu_\beta)^2 \geq 0, \\ \frac{\partial(1 - p_{A-}^*)}{\partial\alpha_F} &= \Psi^{-2}\alpha_A\alpha_B^2\widehat{\beta}^2\mu_\beta^2(V_\beta D' D\widehat{\beta} + \mu_\beta)^4 \geq 0. \end{aligned}$$

Moreover, it is immediate that  $1 - p_{A-}^*|_{\alpha_B=0} = 1 - p_{A-}^*|_{\alpha_F=0} = 0$ . ■

Proposition 3 is striking: it suffices for the influence of lobbying by either the frequentist ( $\alpha_F$ ) or the Bayesian decisionmaker ( $\alpha_B$ ) to become zero for the likelihood of the program being cancelled going to zero. In the real world, and despite much rhetoric, it is likely that the actual influence of lobbying by academic evaluators in terms of obtaining the cancellation of programs that have been "proven" through the results of impact evaluations not to work is minimal at best. The model shows (through the expression for  $e_F^*$  in equation (21)) that as the influence  $\alpha_F$  of lobbying carried out by the frequentist academic evaluator tends towards zero—which is likely to be close to the situation that obtains in the real world—her effort in favor of cancellation of the program tends towards zero as well. Thus, lack of influence breeds lack of effort, and lack of effort leads to the continuation of the program. As such, it is perhaps not surprising that the influence of impact evaluations in the realm of policymaking is limited, at best.

### 2.3 An inconclusive posterior: $\beta_- < \widehat{\beta} < \beta_+ < \beta_* < \beta^*$

When the impact evaluation yields a posterior that is inconclusive, the Bayesian evaluator will be indifferent between cancellation and continuation of the program and will therefore furnish no effort. The game then reduces to its two-player variant, in which the anti-evaluation decisionmaker and the frequentist evaluator compete. The payoffs are given by  $U_A = p_A\mu_\beta - ce_A$ ,  $U_F = (1 - p_A)(-\widehat{\beta}) - ce_F$ , whereas the contest success function

simplifies to  $p_A = \frac{e_A^\gamma}{e_A^\gamma + e_F^\gamma}$ . It is then easy to establish the following Proposition:

**Proposition 4** *The likelihood of the program being cancelled, when it was approved in the first place, when the Bayesian evaluator is indifferent between continuation and cancellation, and the frequentist evaluator is in favor of cancellation ( $\beta_- < \hat{\beta} < \beta_+ < \beta_* < \beta^*$ ) is equal to  $p_{A0}^* = \frac{\alpha_A \mu_\beta^2}{\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2}$ , and admits the following comparative statics:  $\frac{\partial(1-p_{A0}^*)}{\partial \mu_\beta} = -\left(\frac{\mu_\beta}{\hat{\beta}}\right) \frac{\partial(1-p_{A0}^*)}{\partial \hat{\beta}} < 0$ ,  $\frac{\partial(1-p_{A0}^*)}{\partial V_\beta} = \frac{\partial(1-p_{A0}^*)}{\partial(D'D)} = 0$ .*

**Proof.** Proceeding as in the proof of Proposition 2 yields:  $e_A^* = \frac{2\alpha_A \alpha_F \mu_\beta^3 \hat{\beta}^2}{c(\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2)^2} > 0$ ,  $e_F^* = -\frac{2\alpha_A \alpha_B \alpha_F \mu_\beta^2 \mu_\beta^* \hat{\beta}^3}{c(\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2)^2} > 0$ , which implies that  $p_{A0}^* = \frac{\alpha_A \mu_\beta^2}{\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2}$ . Differentiation then yields:  $\frac{\partial(1-p_{A0}^*)}{\partial \mu_\beta} = -\frac{2\alpha_A \alpha_F \hat{\beta}^2 \mu_\beta}{(\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2)^2} = -\left(\frac{\mu_\beta}{\hat{\beta}}\right) \frac{\partial(1-p_{A0}^*)}{\partial \hat{\beta}} < 0$ . It is immediate that  $\frac{\partial(1-p_{A0}^*)}{\partial V_\beta} = \frac{\partial(1-p_{A0}^*)}{\partial(D'D)} = 0$ . ■

As with the case of a significantly negative posterior, the likelihood of cancellation falls to zero when the influence  $\alpha_F$  of the frequentist academic evaluator falls to zero.

## 2.4 A significantly positive posterior: $\beta_- < \beta_+ < \hat{\beta} < \beta_* < \beta^*$

In this case, we have:

$$B_B = \mu_\beta^* > 0,$$

both agent  $A$  and agent  $B$  wish to continue the program, and the expected payoffs are given by:

$$U_A = p_A \mu_\beta - c e_A, U_B = p_A \mu_\beta^* - c e_B, U_F = (1 - p_A)(-\hat{\beta}) - c e_F,$$

where  $p_A = \frac{e_A^\gamma + e_B^\gamma}{e_A^\gamma + e_B^\gamma + e_F^\gamma}$ . We then have the following Proposition:

**Proposition 5** *The likelihood of the program being cancelled, when it was approved in the first place, when the Bayesian evaluator is in favor of continuation and the frequentist evaluator is in favor of cancellation ( $\beta_- < \beta_+ < \hat{\beta} < \beta_* < \beta^*$ ) is equal to  $p_{A+}^* = \frac{\alpha_A \alpha_B \mu_\beta^2 (V_\beta D' D \hat{\beta} + \mu_\beta)^2}{\Psi}$ , and admits the following comparative statics:  $\frac{\partial(1-p_{A+}^*)}{\partial \mu_\beta} = -\left(\frac{\mu_\beta}{\hat{\beta}}\right) \frac{\partial(1-p_{A+}^*)}{\partial \hat{\beta}} < 0$ ,  $\frac{\partial(1-p_{A+}^*)}{\partial V_\beta} = \left(\frac{V_\beta}{D'D}\right) \frac{\partial(1-p_{A+}^*)}{\partial(D'D)} > 0$ .*

**Proof.** Proceeding as in the proof of Proposition 2, the Nash equilibrium levels of effort

furnished by each of the parties can be shown to be given by  $e_A^* = \frac{2\alpha_A\alpha_B^2\alpha_F\mu_\beta^3\mu_\beta^{*4}\widehat{\beta}^2}{c[\alpha_B\alpha_F\mu_\beta^{*2}\widehat{\beta}^2 + \alpha_A\mu_\beta^2(\alpha_B\mu_\beta^{*2} + \alpha_F\widehat{\beta}^2)]^2} >$

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

which implies that  $p_{A+}^* = \frac{\alpha_A\alpha_B\mu_\beta^2\mu_\beta^{*2}}{\alpha_B\alpha_F\widehat{\beta}^2\mu_\beta^{*2} + \alpha_A\mu_\beta^2(\alpha_F\widehat{\beta}^2 + \alpha_B\mu_\beta^{*2})}$  and thus  $p_{A+}^* = \frac{\alpha_A\alpha_B\mu_\beta^2(V_\beta D' D \widehat{\beta} + \mu_\beta)^2}{\Psi}$ .

Differentiation then yields:  $\frac{\partial(1-p_{A+}^*)}{\partial\mu_\beta} = -2\Psi^{-2}\alpha_A\alpha_B\alpha_F\widehat{\beta}^2\mu_\beta(V_\beta D' D \widehat{\beta} + \mu_\beta)[\alpha_A(1 + V_\beta D' D)^2\mu_\beta^3 +$

$\alpha_B(V_\beta D' D \widehat{\beta} + \mu_\beta)^3] = -\left(\frac{\mu_\beta}{\widehat{\beta}}\right)\frac{\partial(1-p_{A+}^*)}{\partial\widehat{\beta}} < 0$ . Since, under the configuration we have as-

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

$0$ . Similarly,  $\frac{\partial(1-p_{A+}^*)}{\partial V_\beta} = -2\Psi^{-2}\alpha_A^2\alpha_B\alpha_F D' D(1 + V_\beta D' D)\widehat{\beta}^2(\widehat{\beta} - \mu_\beta)\mu_\beta^4(V_\beta D' D \widehat{\beta} + \mu_\beta) =$

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

The upshot of Propositions 2, 4 and 5 is that the likelihood of a program that was initially approved being cancelled is, as common sense would suggest, a decreasing function of the result  $\widehat{\beta}$  of the impact evaluation, *ceteris paribus*. Moreover, Proposition 3 (and the corresponding results that can be trivially established for the configurations considered in Propositions 4 and 5) shows that, as the influence exerted on policy decisions by frequentist academic evaluators tends towards zero –an assumption that is likely to be approximately satisfied in practice– the likelihood of the program being cancelled tends towards zero.

## 2.5 Tying it all together

In order to tie the results of the preceding Propositions together, one last step is needed: to consider what happens to the probability of cancellation at the three critical threshold values  $\beta_-$ ,  $\beta_+$  and  $\beta_*$ . I do this in the following Proposition:

**Proposition 6** *A the limit values  $\beta_-$ ,  $\beta_+$  and  $\beta_*$ ,  $1 - p_{A-}^*(\beta_-) < 1 - p_{A0}^*(\beta_-)$ ,  $1 - p_{A0}^*(\beta_+) < 1 - p_{A+}^*(\beta_+)$ , and  $1 - p_{A+}^*(\beta_*) > 0$ .*

**Proof.** By inspection of the results presented in Propositions 2, 4 and 5, it is immediate that:

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

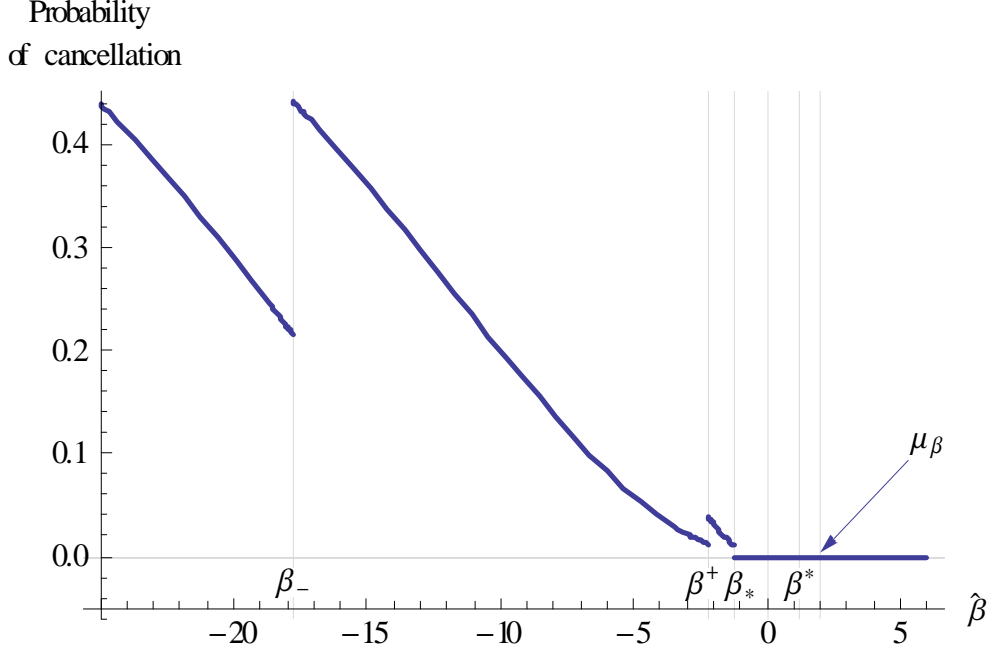


Figure 3: The equilibrium probability of cancellation of the program as a function of the result  $\hat{\beta}$  of the impact evaluation.

and

$$[1 - p_{A0}^*(\hat{\beta})] - [1 - p_{A+}^*(\hat{\beta})] = -\frac{\alpha_A^2 \alpha_F (1 + V_\beta D' D)^2 \hat{\beta}^2 \mu_\beta^4}{\Lambda} < 0, \forall \hat{\beta},$$

where:  $\Lambda = (\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2)[\alpha_A \alpha_F (1 + V_\beta D' D)^2 \hat{\beta}^2 \mu_\beta^2 + \alpha_B (V_\beta D' D \hat{\beta} + \mu_\beta)^2 (\alpha_F \hat{\beta}^2 + \alpha_A \mu_\beta^2)]$ .

It follows that  $1 - p_{A-}^*(\beta_-) < 1 - p_{A0}^*(\beta_-)$ , and  $1 - p_{A0}^*(\beta_+) < 1 - p_{A+}^*(\beta_+)$ . Finally, it is obvious from Proposition 4 that  $1 - p_{A+}^*(\beta_*) > 0$ . ■

Proposition 6 shows, as the value of  $\hat{\beta}$  decreases, that there will be: (i) an upward jump in the unconditional probability of cancellation  $p_A^*$  at  $\beta_*$ , and (ii) downward jumps at  $\beta_+$  and  $\beta_-$ .

Figure 3 provides an illustration of Propositions 2, 4, 5 and 6 in action, for the same parameter values used in Figures 1 and 2. I set  $\alpha_A = \alpha_B = 1$  and  $\alpha_F = 0.01$ : this probably overstates the weight that is placed on the results of the impact evaluation in policy circles, in that it gives equal influence to the anti-evaluation decisionmaker and her Bayesian counterpart. The weight of 0.01 assigned to the frequentist academic evaluator is made for argument's sake: the weight they carry in policy decisions concerning program cancellation is probably even smaller.

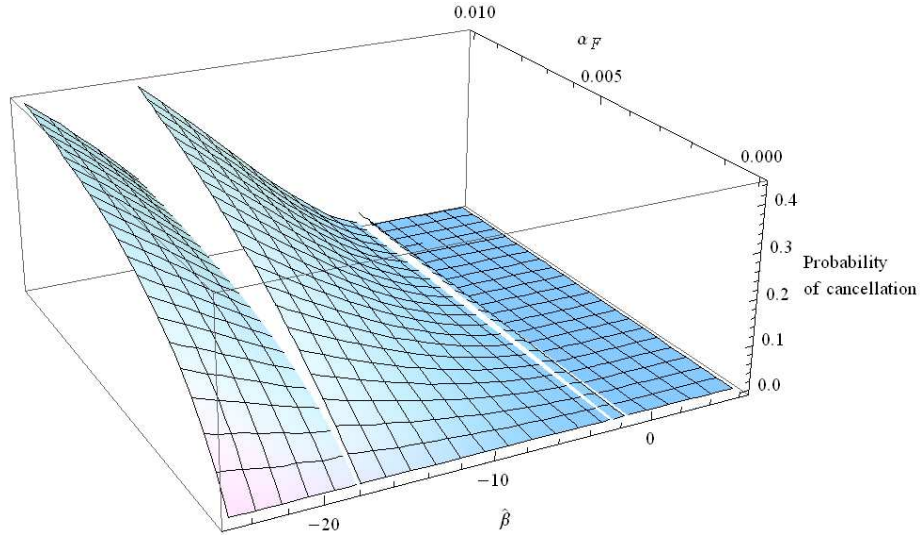


Figure 4: An illustration of Proposition 3: as the influence  $\alpha_F$  of the frequentist academic evaluator tends towards 0, the probability of program cancellation tends toward 0, for all values of  $\hat{\beta}$ .

In Figure 3, the probability of cancellation, as noted in Propositions 2, 4 and 5, is everywhere a strictly decreasing function of  $\hat{\beta}$ ,  $\forall \hat{\beta} < \beta_*$  and equal to zero  $\forall \hat{\beta} \geq \beta_*$ . The jumps established in Proposition 6 obtain at the threshold values  $\beta_-$ ,  $\beta_+$  and  $\beta_*$ , most obviously so at  $\beta_-$ . The most striking aspect of Figure 3 is the relatively low probability of cancellation of the program, even when the result  $\hat{\beta}$  of the impact evaluation becomes exceedingly negative: for example, when  $\hat{\beta} \approx -5$  (in which case it would be seen as being highly significant by the frequentist evaluator), the probability of cancellation is still below 5%. For the likelihood of cancellation to rise to 50%, the result of the impact evaluation would have to reach a value well below  $-25$ . Such is the power of bureaucratic inertia in the world of policymaking.

A final graphical representation of the results presented above is provided in Figure 4, which illustrates Proposition 3 by plotting the probability of cancellation against both the outcome  $\hat{\beta}$  of the impact evaluation and the influence  $\alpha_F$  of the academic frequentist evaluator. As should be clear, as the influence  $\alpha_F$  of the frequentist academic evaluator tends towards zero, the probability of cancellation tends towards zero, for all values of  $\hat{\beta}$ .

### 3 Concluding remarks

One final issue remains to be addressed before drawing to a close. Is the fundamentally risk neutral nature of the preferences being assumed in all impact evaluation reasonable? Should one not be allowing for risk aversion in terms of what is essentially a problem of social choice?

In terms of our statistical methodology, when we estimate the impact of a program, our estimators seek to minimize a *risk function* that is based on a quadratic loss:

$$\arg \min_{\{\hat{\beta}\}} \int (\hat{\beta} - \beta)^2 f(\beta) d\beta = \int \beta f(\beta) d\beta = \mu_{\beta}. \quad (23)$$

This is akin to an assumption of risk-neutrality in utility terms. From the standpoint of impact evaluation terminology, this is why one often focuses on the Average Treatment Effect (ATE).

But from the social choice point of view, we should be choosing an estimator that corresponds to our social welfare criterion, and risk aversion should enter the picture. Consider a standard utility function of the CARA class:  $W(\beta, \theta) = 1 - \exp\{-\theta\beta\}$ , where  $\theta$  is the usual Arrow-Pratt coefficient of absolute risk aversion. Then we should be picking an estimator  $\hat{\beta}$  that corresponds to:

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

As in (23), the key to being able to compute this estimator is that the quantity

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

exists. Moreover, as has been shown in this paper, the appropriate distribution that should be used for  $f(\beta)$  in terms of deciding whether or not to approve the program in the first place is given by a Student's  $t$  distribution. Either by applying the well-known result of Geweke (2001) or by straightforward calculations, it is easy to show that

the integral  $\int_{-\infty}^{+\infty} W(\hat{\beta}, \theta)t(\beta)d\beta$  diverges, and that a risk averse decisionmaker therefore *cannot* even compute the expected utility associated with the approval of a program in the first place.<sup>12</sup> The same reasoning holds for the Bayesian decisionmaker following the impact evaluation, since the posterior is also given by a Student's  $t$ . The implications of this lack of a workable social welfare function, when one relaxes the untenable assumption of risk neutrality, when it comes to evaluating a program either *ex ante* in terms of approval, or *ex post* in Bayesian terms, is left for further research.

Of course, there are many valid econometric reasons that could lead one to recommend the continuation of a program despite an impact evaluation that yields a frequentist ATE that is negative and statistically significant. For example, if the marginal treatment effect (MTE), formalized by Heckman and Vytlacil (1999), is increasing in the unobservables that determine treatment status, and the treatment on the treated (TT) is highly negative and the treatment on the untreated (TUT) highly positive (with a negative and statistically significant ATE), there are excellent reasons for not cancelling the program if the untreated individuals can be brought into the fold when it is continued.<sup>13</sup>

This observer has yet to see this level of sophistication in policymaking circles, and it is highly likely that the bureaucratic inertia arguments developed in this paper within the straightjacket of a simple Bayesian framework are, unfortunately, much closer to the ugly truth, and explain why we almost "never learn" from impact evaluations, and why they seldom lead to evidence-based policymaking.

## References

- 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](http://www.cgdev.org/content/publications/detail/7973)).
- GEWEKE, J. (2001): "A Note on Some Limitations of CRRA Utility," *Economics Letters*, 71, 341–345.

---

<sup>12</sup>The Geweke result applies formally to utility functions of the CRRA class, but also holds for those of the CARA class.

<sup>13</sup>See Heckman and Vytlacil (2005) for a through treatment of the MTE and its link to the ATE, the TT, and the TUT.

- HECKMAN, J. J., AND 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.
- (2005): “Structural Equations, Treatment Effects and Econometric Policy Evaluation,” *Econometrica*, 73(3), 669–738.
- JIA, H., S. SKAPERDAS, AND 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.