A Simple Approach to Choice of a Parameter When Stochastic-Decision-Making Errors have Asymmetric Costs

(preliminary; introduction requires revisions)

Gabriel A. Lozada

Department of Economics, University of Utah,
Salt Lake City, UT 84112
lozada@economics.utah.edu

March 2014

Abstract. Following the lead of authors such as DeGroot (2004), Gigerenzer, Krauss and Vitouch (2004), Leamer (1978), and McCloskey and Ziliak (2009), this paper suggests a simple method for modifying standard parameter estimates when the cost of using an estimate which is smaller than the true value is different from the cost of using an estimate which is larger than the true value. If standard methods suggest a parameter estimate of $\hat{\beta}$ with a variance of $s^2_{\hat{\beta}}$, and if the cost of making one type of mistake is A times the cost of making the other type of mistake, then assuming $\hat{\beta}$ is normally distributed and costs are quadratic, the decision-maker should set the value of the parameter equal to $\hat{\beta} + z(A)\, s^2_{\hat{\beta}}$ where I define $z(A)$ analytically and provide a table of some of its values.

Keywords: (Put keywords here.)

When estimating an equation of the form $y = \beta x$, the standard approach in economics remains to formulate a null hypothesis "$H_0\colon \beta = 0$" and an alternative hypothesis "$H_a\colon \beta \neq 0$," then testing to determine if H_0 can be rejected at a conventional level of significance such as 5% or 1%. The smaller the level of significance the smaller is the chance of Type I error, which is mistakenly rejecting H_0, small, but the larger is the chance of Type II error, which is mistakenly accepting H_0. Ideally, small levels of significance would only be used if the cost (or "loss") in utility or dollar terms to the decision-maker of incorrectly concluding that $\beta \neq 0$ is much greater than the cost of incorrectly concluding that $\beta = 0$. Unfortunately, they are often used without considering the different costs. Past authors decrying this state of affairs or recommending alternative techniques include DeGroot (2004), Gigerenzer, Krauss and Vitouch (2004), Leamer (1978), and McCloskey and Ziliak (2009).

Consider as an example Baicker et al.'s (2013) analysis of the Oregon Medicare experiment, in which H_0 was essentially "giving poor people health care will not make them better off." It is not obvious that, as traditionally implied, the social cost of incorrectly concluding that "Medicare helps poor people" is much greater than the social cost of incorrectly concluding that "Medicare does not help poor people." The first mistake leads society to waste health care on the poor; the second mistake causes the poor to suffer, and causes some of them to die. The conventional position assumes that making the first mistake is to be much more feared than making the second. There exist some members of society who disagree. This makes the conventional position a problematic one for a neutral analyst to take.[1]

There are many other examples of this situation; to name just one more, if x is "air pollution" and y is "human health," the conventional position is that it would be much worse to waste money fruitlessly cleaning up the air than it would be to suffer higher mortality from air pollution we mistakenly thought did not affect mortality.[2]

One solution to this problem would be to flip H_0 and H_a (so that $H_0\colon \beta \neq 0$ and $H_a\colon \beta = 0$);[3] another would be to carefully choose a significance level from the full range of $[0, 1]$, based on a criterion such as maximizing

$$\begin{aligned} &\text{utility[cost of Type I error]} * \text{(Probability of Type I error)} \\ &\quad + \text{utility[cost of Type II error]} * \text{(Probability of Type II error)} \end{aligned}$$

[1]This point is not the same as the important criticism by Frakt, Carroll and Richardson (2013) ("FCR") of Baicker et al. (2013). FCR's criticism is that the medically-relevant sample size of Baicker et al. was so small that their tests had less power than its authors realized. I am criticizing lack of power due to choosing small levels of significance, not due to small sample sizes.

[2]If x is the size of an ocean fish stock and y is its maximum sustainable catch, the uncertainty is not in the relationship between x and y but rather in the estimation of x, a different problem than that posed in this paper. However, it does reflect asymmetric costs, since underestimating x merely leads to poor profits for one year, whereas overestimating x could lead to extinction of the fishery; see Clark (2010 p. 263 ff.) for an analysis.

[3]As Gigerenzer, Krauss, and Vitouch put it (2004 p. 15), "the null need not be a nil hypothesis."

if the decision-maker's preferences obeyed the Expected Utility Hypothesis, or some more general form of "utility[cost of Type I error, Probability of Type I error, cost of Type II error, Probability of Type II error]" otherwise.

However, that leaves unanswered another objection to hypothesis testing, raised by among others McCloskey and Ziliak (2009): how bad it is to make an error usually depends on the *magnitude* of the error. Falsely believing that "air pollution in Salt Lake City does not raise mortality" generates few social costs if the reality is "air pollution in Salt Lake City raises mortality by 1 death per century"; it generates huge social costs if the reality is "air pollution in Salt Lake City raises mortality by 1 death per hour." Ascertaining the costs of erroneous conclusions requires taking into account not just that a conclusion is not true, but how far from the truth it is.

1. Finding the Parameter Analytically

Denote the name of an unknown population parameter by "β" and denote its true value by "β_{true}." The decision-maker, not knowing β_{true}, acts as if the value of β is β_a ("a" for "action"). The decision-maker will choose β_a so as to[4]

$$\min_{\beta_a} \int_{\beta_{\text{true}}} \text{cost}\begin{pmatrix} \text{acting as if } \beta \text{ is } \beta_a \text{ when} \\ \beta \text{ is really } \beta_{\text{true}} \end{pmatrix} \, df(\beta_{\text{true}})\,. \tag{1}$$

The "cost" (or "loss") is measured in utility terms.[5] I wish to consider possibly-asymmetric cost functions, and for simplicity choose

$$\text{cost}\begin{pmatrix} \text{acting as if } \beta \text{ is } \beta_a \text{ when} \\ \beta \text{ is really } \beta_{\text{true}} \end{pmatrix} = \begin{cases} A \cdot (\beta_a - \beta_{\text{true}})^2 & \text{if } \beta_a < \beta_{\text{true}} \text{ and} \\ (\beta_a - \beta_{\text{true}})^2 & \text{otherwise} \end{cases} \tag{2}$$

where A is a positive constant. If $A > 1$, the mistake of setting $\beta_a < \beta_{\text{true}}$ is more costly than the mistake of setting $\beta_a > \beta_{\text{true}}$; the reverse is true if $A < 1$.

Using this cost function, the problem becomes

$$\min_{\beta_a} \left[\int_{-\infty}^{\beta_a} (\beta_a - \beta_{\text{true}})^2 \, df(\beta_{\text{true}}) + \int_{\beta_a}^{\infty} A \cdot (\beta_a - \beta_{\text{true}})^2 \, df(\beta_{\text{true}}) \right]. \tag{3}$$

The distribution function $f(\beta_{\text{true}})$ is unknown; suppose in its place we use $N(\hat{\beta}, s^2_{\hat{\beta}})$ where $\hat{\beta}$ and $s^2_{\hat{\beta}}$ are arrived at by an appropriate standard method, for example Ordinary Least Squares.[6] If $A = 1$, the appropriate β_a is simply $\hat{\beta}$. If $A > 1$, using a β_a which is smaller than β_{true} is more costly than using a β_a which is larger

[4] This closely resembles the approach on p. 122 of DeGroot (2004).

[5] For graphs of loss functions—or equivalently of utility functions—in a context such as this, see for example Fig. 4.4 of Leamer (1978).

[6] One could use either a frequentist or a Baysian approach to find an appropriate estimate for $f(\beta_{\text{true}})$; if that estimate is not normally distributed, changes would be required in the rest of this paper but its basic idea is unaltered. A Baysian analysis would use expert knowledge, if it exists, to inform estimation of $f(\beta_{\text{true}})$. For example, in the Oregon Medicare experiment, economists

than β_{true}, so the appropriate β_a would be larger than $\hat{\beta}$. The reverse is true if $A < 1$; in other words, the sign of $A - 1$ should be the same as the sign of $\beta_a - \hat{\beta}$.

Let $N(x; \hat{\beta}, s^2_{\hat{\beta}})$ denote the value, at x, of the Normal Distribution density function whose mean is $\hat{\beta}$ and whose variance is $s^2_{\hat{\beta}}$. The optimization problem (3) then becomes

$$\min_{\beta_a} \left[\int_{-\infty}^{\beta_a} (\beta_a - x)^2 N(x; \hat{\beta}, s^2_{\hat{\beta}})\, dx + \int_{\beta_a}^{\infty} A \cdot (\beta_a - x)^2 N(x; \hat{\beta}, s^2_{\hat{\beta}})\, dx \right]. \tag{4}$$

Use the abbreviation $CDF[f(x)]$ to denote the "cumulative density function" associated with the probability distribution function $f(x)$ at the point x.

Proposition 1. *The solution to (4) is*

$$\beta_a - \hat{\beta} = \frac{2\,(A-1)\, s^2_{\hat{\beta}}\, N(\beta_a; \hat{\beta}, s^2_{\hat{\beta}})}{A + (1-A)\, CDF\big[N(\beta_a; \hat{\beta}, s^2_{\hat{\beta}})\big]}\,. \tag{5}$$

For the proof, see Section 3. Note that the denominator can be rewritten as $A\,(1 - CDF) + CDF$; since $0 \le CDF \le 1$, this is always positive, so the sign of $\beta_a - \hat{\beta}$ is the same as the sign of $A - 1$, which is the intuitive outcome.

2. Finding the Parameter Numerically

(5) is one equation in the one unknown β_a, and can in principle be solved by root-finding algorithms. It is more convenient to study its solution when rewritten in a standard form:

Proposition 2. *Define*

$$z = \frac{\beta_a - \hat{\beta}}{s_{\hat{\beta}}} \tag{6}$$

---

have expert knowledge that giving people free commodities usually makes them better off, and a Bayesian analysis would use this information to affect the estimation of the effect. Frequentist analysis throws away this "other than the data" information.

In taking the position that one could use either a frequentist or a Baysian approach to finding the estimate for $f(\beta_{\text{true}})$, I am agreeing with Spanos (2012) that a decision-theoretic analysis can be completely free of Baysian influences. In the words of Spanos (p. 9),

> . . . decision-theoretic set up makes perfectly good sense. . . [when]. . . :
>
> [**a**] The primary aim is to use statistical rules to guide actions astutely. . . , and
>
> [**b**] The sagacity of actions is determined by applicable 'losses' based on "relevant information other than the data (Cox and Hinkley, *Theoretical Statistics*, 1974, p. 251).

These conditions are satisfied in the decision-making context of this paper, the first obviously, and the second because, for example, the "social cost of someone dying" is not determined using the data from a study on how air pollution or Medicare affects health.

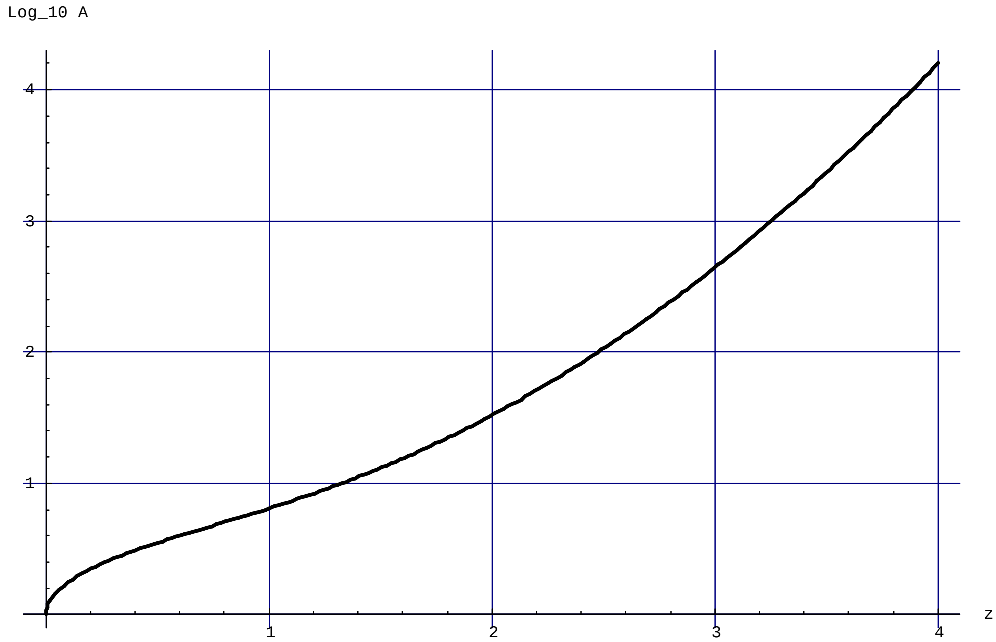

Figure 1. $\log_{10} A$ versus z for $z > 0$ and $A > 1$, that is, for $\log_{10} A > 0$. (The value of z corresponding to a $\log_{10} A_1 < 0$ can be obtained by multiplying by negative one the z corresponding to $-\log_{10} A_1$. The value of $\log_{10} A$ corresponding to a $z_1 < 0$ can be obtained by multiplying by negative one the $\log_{10} A$ corresponding to $-z_1$.)

and for a given A, locate the z which satisfies

$$0 = z \cdot \left\{ A + (1-A)\left(\tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}\left(\tfrac{z}{\sqrt{2}}\right)\right)\right\} - \frac{\sqrt{2}\,(A-1)}{\sqrt{\pi}} \exp(-\tfrac{1}{2} z^2)\,. \tag{7}$$

For example, (7) is satisfied by

A	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*100*	*1000*
z	*0.53*	*0.81*	*0.99*	*1.1*	*1.2*	*1.30*	*1.37*	*1.42*	*1.48*	*2.41*	*3.13*

If for a particular value of A called A_1 the associated z value from (7) is z_1, then for $1/A_1$ the associated z value from (7) is $-z_1$. (7) is equivalent to $A = A(z)$ where

$$A(z) = \frac{\frac{z}{2} + \frac{z}{2}\,\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}}\exp(-\frac{1}{2}z^2)}{-\frac{z}{2} + \frac{z}{2}\,\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right) + \sqrt{\frac{2}{\pi}}\exp(-\frac{1}{2}z^2)} \tag{8}$$

and $A(-z) = 1/A(z)$. A graph of $\log_{10}(A)$ versus z is given in Figure 1.

Once the value of z has been obtained, the optimal β_a is

$$\beta_a = z\, s_{\hat{\beta}} + \hat{\beta}\,. \tag{9}$$

The proof is in Section 3. The z which satisfies (7) has, from (6), the same sign as $A-1$. The interpretation of (9) is that for the values of A given in the table of Proposition 2, the decision-maker should use a parameter value which differs from the traditional estimate by somewhere between a half and three standard deviations.

As an illustration of Proposition 2, suppose that $A = 1/4$, reflecting a judgment that using a β_a which is larger than β_{true} by a certain distance is four times more costly than using a β_a which is smaller than β_{true} by the same distance. The sentence following the table implies that for $A = 1/4$, $z = -0.99$. Hence we would simply set $\beta_a = -0.99 s_{\hat\beta} + \hat\beta$.

Since A is probably not known with much precision, the table only gives a few significant figures for z. Indeed, if the decision-maker has quite a bit of doubt about the value of A—or doubt about "*his*" value of A, depending on whether one thinks the value of A is objectively determined, as for instance through an opinion poll taken by a decision-maker who is the agent for other people, or the value of A is subjectively determined by the decision-maker who owes allegiance only to himself—one conceivable way to use Proposition 2 would be to ask "how reasonable is the value of A which makes β_a equal to zero?" This is equivalent to the question, "how reasonable is the value of A which corresponds to $z = -\hat\beta/s_{\hat\beta}$?" The answer to this question is clearly not to be found in the data, and therefore this question is utterly different from the conventional question "what is the probability of $H_0\colon \hat\beta = 0$?" The conventional question is easier to answer but is less important when an action has to be decided upon.

3. Proofs

I first present two proofs of Proposition 1.

Proof of Proposition 1 (Method 1). The first-order condition is obtained by differentiating (4) with respect to β_a and setting the result equal to zero. Formally, Leibnitz' Rule leads to

$$\begin{aligned}0 = \int_{-\infty}^{\beta_a} & 2(\beta_a - x)\, N(x; \hat\beta, s_{\hat\beta}^2)\, dx + (\beta_a - \beta_a)\, N(\beta_a; \hat\beta, s_{\hat\beta}^2)\, (d\beta_a/d\beta_a) \\ & - (\beta_a - (-\infty))\, N(-\infty; \hat\beta, s_{\hat\beta}^2)\, (d(-\infty)/d\beta_a) \\ + \int_{\beta_a}^{\infty} & 2A(\beta_a - x)\, N(x; \hat\beta, s_{\hat\beta}^2)\, dx + (\beta_a - \infty)\, N(\infty; \hat\beta, s_{\hat\beta}^2)\, (d\infty/d\beta_a) \\ & - (\beta_a - \beta_a)\, N(\beta_a; \hat\beta, s_{\hat\beta}^2)\, (d\beta_a/d\beta_a)\end{aligned}$$

which actually means

$$\begin{aligned}0 &= \int_{-\infty}^{\beta_a} 2(\beta_a - x)\, N(x; \hat\beta, s_{\hat\beta}^2)\, dx + \int_{\beta_a}^{\infty} 2A(\beta_a - x)\, N(x; \hat\beta, s_{\hat\beta}^2)\, dx \iff \\ 0 &= \int_{-\infty}^{\beta_a} (\beta_a - x)\, N(x; \hat\beta, s_{\hat\beta}^2)\, dx + \int_{\beta_a}^{\infty} A(\beta_a - x)\, N(x; \hat\beta, s_{\hat\beta}^2)\, dx\,. \qquad (10)\end{aligned}$$

If $A = 1$ then as expected, $\beta_a = \hat{\beta}$.

To simplify (10), begin by writing A as $1 + (A - 1)$ and expand:

$$
\begin{aligned}
0 &= \int_{-\infty}^{\beta_a} (\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx + \int_{\beta_a}^{\infty} (1)(\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx \\
&\quad + \int_{\beta_a}^{\infty} (A-1)(\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx \\
&= \int_{-\infty}^{\infty} (\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx + \int_{\beta_a}^{\infty} (A-1)(\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx \\
&= \beta_a - \hat{\beta} + \int_{\beta_a}^{\infty} (A-1)(\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx \iff \\
\hat{\beta} - \beta_a &= \int_{\beta_a}^{\infty} (A-1)(\beta_a - x)\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx \qquad (11) \\
&= (A{-}1)\, \beta_a \int_{\beta_a}^{\infty} N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx - (A{-}1) \int_{\beta_a}^{\infty} x\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx \\
&= (A{-}1)\, \beta_a \left\{1{-}CDF[N(\beta_a; \hat{\beta}, s_{\hat{\beta}}^2)]\right\} - (A{-}1) \int_{\beta_a}^{\infty} x\, N(x; \hat{\beta}, s_{\hat{\beta}}^2)\, dx\,. \qquad (12)
\end{aligned}
$$

To simplify the last term on the right-hand side of (12), [?? appeal to a standard result or??] let $n(x; \mu, \sigma)$ denote the value at x of the Normal Distribution function whose mean is μ and whose standard deviation is σ. Then

$$
\begin{aligned}
\int x \cdot n(x; \mu, \sigma)\, dx &= \int x \cdot \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int (x - \mu + \mu)\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\,;
\end{aligned}
$$

changing variables to $w = (x-\mu)^2$, so that $dw = 2\,(x-\mu)\,dx$ and $dx = dw/(2\sqrt{w})$, gives

$$
\begin{aligned}
&= \frac{1}{\sigma\sqrt{2\pi}} \left[\int \sqrt{w}\, e^{-\frac{w}{2\sigma^2}}\, \frac{1}{2\sqrt{w}}\, dw + \mu \int e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\right] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \left[\frac{1}{2} \int e^{-\frac{w}{2\sigma^2}}\, dw + \mu \int e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\right]\,.
\end{aligned}
$$

Hence for an arbitrary constant R,

$$
\begin{aligned}
\int_R^{\infty} x \cdot n(x; \mu, \sigma)\, dx &= \frac{1}{\sigma\sqrt{2\pi}} \left[\frac{1}{2} \int_{(R-\mu)^2}^{\infty} e^{-\frac{w}{2\sigma^2}}\, dw + \mu \int_R^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\right] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \left[(-2\sigma^2)\, e^{-\frac{w}{2\sigma^2}}\Big|_{(R-\mu)^2}^{\infty} + \mu \int_R^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx\right] \\
&= \frac{-2\sigma^2}{\sigma\sqrt{2\pi}} \left(0 - e^{-\frac{(R-\mu)^2}{2\sigma^2}}\right) + \mu \int_R^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx
\end{aligned}
$$

$$= 2\sigma^2 \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(R-\mu)^2}{2\sigma^2}} + \mu \int_R^\infty \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$
$$= 2\sigma^2 n(R;\mu,\sigma) + \mu \left\{1 - CDF[n(R;\mu,\sigma)]\right\}. \tag{13}$$

(12) thus becomes

$$\begin{aligned}\hat\beta - \beta_a &= (A{-}1)\,\beta_a \left\{1{-}CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\right\} \\ &\quad - (A{-}1)\left[2s^2_{\hat\beta}\, N(\beta_a;\hat\beta,s^2_{\hat\beta}) + \hat\beta\left\{1 - CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\right\}\right] \qquad (14) \\ &= (A{-}1)(\beta_a - \hat\beta)\left\{1{-}CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\right\} - (A{-}1)\cdot 2s^2_{\hat\beta}\, N(\beta_a;\hat\beta,s^2_{\hat\beta})\end{aligned}$$

so

$$\begin{aligned}(\hat\beta - \beta_a)\left[1 + (A{-}1)\left\{1{-}CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\right\}\right] &= -(A{-}1)\cdot 2s^2_{\hat\beta}\, N(\beta_a;\hat\beta,s^2_{\hat\beta}) \\ (\hat\beta - \beta_a)\left[1 + (A{-}1)\left\{1\right\} - (A{-}1)\left\{CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\right\}\right] &= -(A{-}1)\cdot 2s^2_{\hat\beta}\, N(\beta_a;\hat\beta,s^2_{\hat\beta}) \\ (\hat\beta - \beta_a)\left[A - (A{-}1)CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\right] &= -(A{-}1)\cdot 2s^2_{\hat\beta}\, N(\beta_a;\hat\beta,s^2_{\hat\beta})\end{aligned}$$

which leads to (5). ∎

Proof of Proposition 1 (Method 2). This proof follows the proof via Method 1 until (10). That equation implies

$$\begin{aligned}0 &= 2\beta_a \int_{-\infty}^{\beta_a} N(x;\hat\beta,s^2_{\hat\beta})\,dx + 2A\beta_a \int_{\beta_a}^{\infty} N(x;\hat\beta,s^2_{\hat\beta})\,dx \\ &\quad - 2\int_{-\infty}^{\beta_a} x\,N(x;\hat\beta,s^2_{\hat\beta})\,dx - 2A\int_{\beta_a}^{\infty} x\,N(x;\hat\beta,s^2_{\hat\beta})\,dx\,. \qquad (15) \\ &= 2\beta_a CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})] + 2A\beta_a\Big(1 - CDF[N(\beta_a;\hat\beta,s^2_{\hat\beta})]\Big) \\ &\quad - 2\int_{-\infty}^{\beta_a} x\,N(x;\hat\beta,s^2_{\hat\beta})\,dx - 2A\int_{\beta_a}^{\infty} x\,N(x;\hat\beta,s^2_{\hat\beta})\,dx\,. \qquad (16)\end{aligned}$$

To simplify the last two terms, [?? appeal to a standard result or??] let $n(x;\mu,\sigma)$ denote the value at x of the Normal Distribution function whose mean is μ and whose standard deviation is σ. Then

$$\begin{aligned}\int x\cdot n(x;\mu,\sigma)\,dx &= \int x\cdot\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int (x-\mu+\mu)\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx\,;\end{aligned}$$

changing variables to $w = (x-\mu)^2$, so that $dw = 2(x-\mu)\,dx$ and $dx = dw/(2\sqrt{w})$, gives

$$= \frac{1}{\sigma\sqrt{2\pi}} \left[\int \sqrt{w}\, e^{-\frac{w}{2\sigma^2}} \frac{1}{2\sqrt{w}}\, dw + \mu \int e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right]$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \left[\frac{1}{2} \int e^{-\frac{w}{2\sigma^2}}\, dw + \mu \int e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right].$$

Hence for an arbitrary constant R,

$$\begin{aligned}
\int_{-\infty}^{R} x \cdot n(x;\mu,\sigma)\, dx &= \frac{1}{\sigma\sqrt{2\pi}} \left[\frac{1}{2} \int_{+\infty}^{(R-\mu)^2} e^{-\frac{w}{2\sigma^2}}\, dw + \mu \int_{-\infty}^{R} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \left[(-2\sigma^2)\, e^{-\frac{w}{2\sigma^2}} \Big|_{+\infty}^{(R-\mu)^2} + \mu \int_{-\infty}^{R} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right] \\
&= \frac{-2\sigma^2}{\sigma\sqrt{2\pi}} \left(e^{-\frac{(R-\mu)^2}{2\sigma^2}} - 0 \right) + \mu \int_{-\infty}^{R} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= -2\sigma^2 \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(R-\mu)^2}{2\sigma^2}} + \mu \int_{-\infty}^{R} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= -2\sigma^2 n(R;\mu,\sigma) + \mu\, CDF[n(R;\mu,\sigma)]\,. \qquad (17)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\int_{R}^{\infty} x \cdot n(x;\mu,\sigma)\, dx &= \frac{1}{\sigma\sqrt{2\pi}} \left[\frac{1}{2} \int_{(R-\mu)^2}^{\infty} e^{-\frac{w}{2\sigma^2}}\, dw + \mu \int_{R}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \left[(-2\sigma^2)\, e^{-\frac{w}{2\sigma^2}} \Big|_{(R-\mu)^2}^{\infty} + \mu \int_{R}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right] \\
&= \frac{-2\sigma^2}{\sigma\sqrt{2\pi}} \left(0 - e^{-\frac{(R-\mu)^2}{2\sigma^2}} \right) + \mu \int_{R}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= 2\sigma^2 \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(R-\mu)^2}{2\sigma^2}} + \mu \int_{R}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= 2\sigma^2 n(R;\mu,\sigma) + \mu \Big(1 - CDF[n(R;\mu,\sigma)] \Big)\,. \qquad (18)
\end{aligned}$$

Substituting (17) and (18) into (16) yields

$$\begin{aligned}
0 &= 2\beta_a CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] + 2A\beta_a \Big(1 - CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] \Big) \\
&\quad - 2 \Big\{ -2s_{\hat{\beta}}^2 N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2) + \hat{\beta}\, CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] \Big\} \\
&\quad - 2A \Big\{ 2s_{\hat{\beta}}^2 N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2) + \hat{\beta} \Big(1 - CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] \Big) \Big\} \\
&= 2\beta_a CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] + 2A\beta_a - 2A\beta_a CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] \\
&\quad + 4s_{\hat{\beta}}^2 N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2) - 2\hat{\beta}\, CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)] \\
&\quad - 4As_{\hat{\beta}}^2 N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2) - 2A\hat{\beta} + 2A\hat{\beta}\, CDF[N(\beta_a;\hat{\beta},s_{\hat{\beta}}^2)]\,. \qquad (19)
\end{aligned}$$

Some algebraic steps lead from (19) to (5). ∎

Proof of Proposition 2. From (5), the optimal β_a solves

$$
\begin{aligned}
0 &= (\beta_a - \hat{\beta}) \cdot \big\{A + (1{-}A)\, \mathit{CDF}[N(\beta_a; \hat{\beta}, s_{\hat{\beta}}^2)]\big\} - 2\,(A{-}1)\, s_{\hat{\beta}}^2\, N(\beta_a; \hat{\beta}, s_{\hat{\beta}}^2) \\
&= (\beta_a - \hat{\beta}) \cdot \big\{A + (1{-}A)\, (\tfrac{1}{2} + \tfrac{1}{2}\operatorname{erf}(\tfrac{\beta_a - \hat{\beta}}{\sqrt{2 s_{\hat{\beta}}^2}}))\big\} - \tfrac{\sqrt{2}\,(A-1)\, s_{\hat{\beta}}}{\sqrt{\pi}} \exp(-\tfrac{1}{2}(\tfrac{\beta_a - \hat{\beta}}{s_{\hat{\beta}}})^2)\,.
\end{aligned}
$$

Using (6), this is

$$
0 = z\, s_{\hat{\beta}} \cdot \big\{A + (1{-}A)\, (\tfrac{1}{2} + \tfrac{1}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}))\big\} - \tfrac{\sqrt{2}\,(A-1) s_{\hat{\beta}}}{\sqrt{\pi}} \exp(-\tfrac{1}{2} z^2)\,.
$$

(7) follows. The table following (7) was generated from (7) using *Mathematica*.

(7) implies

$$
\begin{aligned}
0 &= z \cdot \big\{A + \tfrac{1}{2} + \tfrac{1}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}) - \tfrac{1}{2}A - \tfrac{1}{2}A\operatorname{erf}(\tfrac{z}{\sqrt{2}})\big\} - \sqrt{\tfrac{2}{\pi}}A\exp(-\tfrac{1}{2}z^2) + \sqrt{\tfrac{2}{\pi}}\exp(-\tfrac{1}{2}z^2) \\
&= Az + \tfrac{z}{2} + \tfrac{z}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}) - \tfrac{z}{2}A - \tfrac{z}{2}A\operatorname{erf}(\tfrac{z}{\sqrt{2}}) - \sqrt{\tfrac{2}{\pi}}A\exp(-\tfrac{1}{2}z^2) + \sqrt{\tfrac{2}{\pi}}\exp(-\tfrac{1}{2}z^2) \\
&= \Big(z - \tfrac{z}{2} - \tfrac{z}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}) - \sqrt{\tfrac{2}{\pi}}\exp(-\tfrac{1}{2}z^2)\Big)A + \tfrac{z}{2} + \tfrac{z}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}) + \sqrt{\tfrac{2}{\pi}}\exp(-\tfrac{1}{2}z^2) \iff \\
&\Big(-\tfrac{z}{2} + \tfrac{z}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}) + \sqrt{\tfrac{2}{\pi}}\exp(-\tfrac{1}{2}z^2)\Big)A = \tfrac{z}{2} + \tfrac{z}{2}\operatorname{erf}(\tfrac{z}{\sqrt{2}}) + \sqrt{\tfrac{2}{\pi}}\exp(-\tfrac{1}{2}z^2)
\end{aligned}
$$

from which (8) follows.

To prove the sentence following the table, define the right-hand side of (7) by $g(A, z)$. The claim is that if $g(A, z) = 0$ then $g(1/A, -z) = 0$. This is equivalent to the claim made after (8), namely that $A(-z) = 1/A(z)$. Because erf is an "odd function"—that is, $\operatorname{erf}(-x) = -\operatorname{erf}(x)$ for all x—that claim is true from (8) by inspection. ▮

BIBLIOGRAPHY

Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein (2013), "The Oregon Experiment—Effects of Medicaid on Clinical Outcomes," *New England Journal of Medicine* **368**: 1713—1722. Also at `http://www.nejm.org/doi/full/10.1056/NEJMsa1212321`.

Clark, Colin W. (2010), *Mathematical Bioeconomics, Third Edition*. Hoboken, New Jersey: John Wiley & Sons.

DeGroot, Morris H. (2004), *Optimal Statistical Decisions*. Hoboken, NJ: Wiley-Interscience.

Frakt, Austin, Aaron Carroll, and Sam Richardson (2013), "NEJM Letters on the Oregon Medicaid Study." At `http://theincidentaleconomist.com/wordpress/nejm-letters-on-the-oregon-medicaid-study/`.

Gigerenzer, Gerd, Stefan Krauss, and Oliver Vitouch (2004), "The Null Ritual: What You Always Wanted to Know About Significance Testing but Were Afraid to Ask." Pages 391-408 in D. Kaplan, ed., *The Sage handbook of quantitative methodology for the social sciences*. Thousand Oaks, California: Sage Publications.

Leamer, Edward E. (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

McCloskey, Deirdre N., and Stephen T. Ziliak (2009), "The Cult of Statistical Significance." Presented at the Joint Statistical Meetings, Washington, DC, in a contributed session of the Section on Statistical Education. Available at www.statlit.org/pdf/2009ZiliakMcCloskeyASA.pdf and at www.deirdremccloskey.com/docs/jsm.pdf.

Spanos, Aris (Oct. 2012), "How the Decision Theoretic Perspective Misrepresents Frequentist Inference." Available at www.econ.vt.edu/faculty/2008vitas_research/Spanos/2\%20Spanos-inadmissibility-2012.pdf.