Web Appendix to “Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments”

Jörg Stoye
New York University

December 17, 2009

This web appendix provides alternative minimax regret solutions to the decision problems analyzed in section 4. Recall that the minimax regret treatment rules found there coincide with the benchmark solution up to some critical sample size, but do not take into account additional data that would take the sample size beyond this critical size (modulo randomization to overcome an integer problem). It is possible to construct rules that nominally take account of all data. For example, one could base decisions on a randomly selected subsample from the data as long as this subsample is appropriately sized with probability 1. I will here present a very different solution that does not discard any data point either deterministaically or randomly. Sensitivity of the decision rule to the signal is rather toned down by randomizing between δ_1^* and δ_3^* respectively δ_4^*.

I show the result explicitly for the missing data scenario.

Proposition 1 *Consider the missing data model of example 3, thus* $\widetilde{\Delta} \in [(1-\varepsilon)\Delta - \varepsilon, (1-\varepsilon)\Delta + \epsilon]$. *Then minimax regret is achieved by the decision rule*

$$\delta_{MD}^* = \alpha^* \delta_1^* + (1-\alpha^*)\delta_3^*,$$

where

$$\alpha^* = \begin{cases} 0, & \varepsilon \geq \frac{1}{2} \\ \dfrac{2^{N'-1}\frac{1-\varepsilon}{\varepsilon} - \sum_{z \geq \frac{N'(1-2\varepsilon)}{2-2\varepsilon}} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{1-\varepsilon}{\varepsilon}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\}(2z - N')}{\sum_{z > N'/2} \binom{N'}{z}(2z-N') - \sum_{z \geq \frac{N'(1-2\varepsilon)}{2-2\varepsilon}} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{1-\varepsilon}{\varepsilon}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\}(2z-N')}, & \frac{1}{2} < \varepsilon < \varepsilon_N^* \\ 1, & \varepsilon \leq \varepsilon_N^* \end{cases}$$

with $N' = \max\{M \leq N : M \text{ is odd}\}$ *as before and* $\varepsilon_N^* \equiv \left(1 + 2\sum_{z > N'/2} \binom{N'}{n}(2n - N')\right)^{-1}$.

In particular, δ_1^* *is optimal if* $\varepsilon \leq \varepsilon_N^*$ *(as before).*

Proof. The proof will be conducted explicitly for randomized treatment assignment with N odd; the extension to the other scenarios is as before. The least favorable prior is as before.

Step 1: Simplifying Natures best response problem.

This step is essentially as before, and Nature's best-response problem can ultimately be written as

$$\max_{\Delta \in [\max\{-\varepsilon/(1-\varepsilon),-1\},1]} \phi(\Delta;\alpha^*,\varepsilon), \phi(\Delta;\alpha^*,\varepsilon) = ((1-\varepsilon)\Delta + \varepsilon)\left(1 - \alpha^* f_1(\Delta) - (1-\alpha^*) f_3(\Delta)\right), \quad (1)$$

where $f_3(\Delta) = \mathbb{E}\delta_3^*(\omega)$.

Step 2: Introducing a simplified game.

For this and the next few steps, consider a simplified game in which the decision maker's strategy set consists of randomizations over $\{\delta_1, \delta_3\}$. Identify feasible strategies for DM with probabilities α of playing δ_1. Nature, on the other hand, is restricted to choosing Δ (possibly at random); every choice of Δ is identified with even randomization over $(1,(1+\Delta)/2, 0, (1-\Delta)/2), (0,(1-\Delta)/2, 1, (1+\Delta)(/2))$. This game has a Nash equilibrium (by the Glicksberg fixed point theorem). Nature's best-response problem in this game is (1), and the previous step's analysis showed that any best response for Nature in this game is also a best response for her in the original one. On the other hand, it is easy to see the following: If Nature's strategy is supported on $(0,1]$, then $\alpha^* = 1$ is a unique best response for DM in both the simplified and the original game. If Nature plays $\Delta^* = 0$, then the sample data are noise and any choice of α is a best response in both the simplified and the original game. If Nature's strategy is supported on $[-1,0)$, then $\alpha^* = 0$ is a unique best response in the simplified (albeit not the original) game. The latter case (and other, more intricate ones) will not in fact obtain, so that equilibria of the simplified game correspond to equilibria of the original one.

Step 3: Equilibrium when $\varepsilon \geq 1/2$.

As before.

Step 4: Equilibrium when $\alpha^* = 1$.

As before.

Step 5: Analysis of revealing equilibria in general.

Call an equilibrium revealing if Nature's strategy is *not* degenerate at $\Delta^* = 0$. Suppose by contradiction that Nature's strategy is supported on $[-1,0)$, then DM's best response is to set $\alpha^* = 0$, but then the algebra of step 3 can be adapted to show that Nature's best response is some $\Delta^* > 0$, breaking equilibrium (because the planner would now like to play δ_1^*). It follows that in any informative equilibrium, (1) is solved by at least one $\Delta^* > 0$. The value of the equilibrium can therefore be bounded above by

$$\max_{\Delta \in [0,1]} \min_{\alpha \in [0,1]} \phi(\Delta;\alpha,\varepsilon).$$

On this restricted domain,

$$\frac{\partial \phi(\Delta;\alpha^*,\varepsilon)}{\partial \varepsilon} = (1-\Delta)(1-\alpha^* f_1(\Delta) - (1-\alpha^*) f_3(\Delta)) \leq 1/2$$

for any (α^*, ε), hence $\frac{\partial}{\partial \varepsilon} \max_{\Delta \in [0,1]} \min_{\alpha \in [0,1]} \phi(\Delta; \alpha, \varepsilon) \leq 1/2$ by an envelope theorem. In contrast, a pooling equilibrium (if it exists) has value $\phi(0; \alpha^*, \varepsilon) = \varepsilon/2$ (independent of α) with derivative 1/2. It follows that if there exists a pooling equilibrium for some ε, then there also exists a pooling equilibrium for any $\varepsilon' > \varepsilon$. But inspection of algebra in step 4 reveals that $\Delta^* = 0$, thus the equilibrium is pooling, at $\varepsilon = \varepsilon_N^*$. Hence, there exists a pooling equilibrium for every $\varepsilon \geq \varepsilon_N^*$.

Step 6: Characterization of the pooling equilibrium.

A necessary condition for a pooling equilibrium is that $\max_{\Delta \in [-1,1]} \phi(\Delta; \alpha^*, \varepsilon) = \varepsilon/2$. As $\phi(0; \alpha, \varepsilon) = \varepsilon/2$ for every α, this requires a first-order condition to hold at $\Delta^* = 0$, thus

$$\phi'(0; \alpha^*, \varepsilon) = -\varepsilon \left(\alpha^* f_1'(0) + (1-\alpha^*) f_2'(0)\right) + \frac{1-\varepsilon}{2} = 0 \Longrightarrow \alpha^* = \frac{\frac{1-\varepsilon}{2\varepsilon} - f_2'(0)}{f_1'(0) - f_2'(0)}.$$

The closed-form expressions for α^* provided in the proposition follow upon substituting for $f_1'(0)$ from above and using

$$\begin{aligned} f_2(\Delta) &= \sum_{z: \frac{1}{2} + \frac{1-\varepsilon}{\varepsilon}\left(\frac{z}{N} - \frac{1}{2}\right) \geq 0} \binom{N}{z} 2^{-N} (1+\Delta)^z (1-\Delta)^{N-z} \max\left\{\frac{1}{2} + \frac{1-\varepsilon}{\varepsilon}\left(\frac{z}{N} - \frac{1}{2}\right), 1\right\} \\ \Longrightarrow f_2'(0) &= \sum_{z \geq \frac{N(1-2\varepsilon)}{2-2\varepsilon}} \binom{N}{z} 2^{-N} \max\left\{\frac{1}{2} + \frac{1-\varepsilon}{\varepsilon}\left(\frac{z}{N} - \frac{1}{2}\right), 1\right\} (2z - N). \end{aligned}$$

This always leads to a positive expression because $f_1'(0) > f_2'(0)$ and $\frac{1-\varepsilon}{2\varepsilon} \geq f_2'(0)$ can be verified. α^* will equal 0 whenever $\frac{1-\varepsilon}{2\varepsilon} = f_2'(0)$, that is whenever the truncation of δ_2 fails to bind; this is the case iff $\varepsilon \geq 1/2$, confirming algebra from step 3. α^* will equal 1 iff $f_1'(0) = \frac{1-\varepsilon}{2\varepsilon}$, which occurs iff $\varepsilon = \varepsilon_N^*$. $\alpha^* > 1$ if $\varepsilon < \varepsilon_N^*$, in which case the pooling equilibrium does not exist, confirming algebra from step 4. ■

For the other scenarios, similar arguments lead to the following. (The result for hidden covariates requires more substantial adaptation, but the changes are very similar to arguments presented in the paper.)

Proposition 2 *Consider the noncompliance scenario with unrestricted behavior of noncompliers of example 1. Define the decision rule*

$$\delta_5^*(\omega) = \begin{cases} 0, & \frac{1}{2} + \frac{I_N}{2\varepsilon N} < 0 \\ \frac{1}{2} + \frac{I_N}{2\varepsilon N}, & 0 \leq \frac{1}{2} + \frac{I_N}{2\varepsilon N} \leq 1 \\ 1, & \frac{1}{2} + \frac{I_N}{2\varepsilon N} > 1 \end{cases}$$

and the decision rule

$$\delta_{NC1}^* = \alpha^* \delta_1^* + (1-\alpha^*) \delta_5^*,$$

where

$$\alpha^* = \max\left\{0, \frac{\frac{1}{2\varepsilon} - \sum_{z \geq N'(1-\varepsilon)} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{1}{2\varepsilon}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\} (2z - N')}{\sum_{z > N'/2} \binom{N'}{z} (2z - N') - \sum_{z \geq N'(1-\varepsilon)} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{1}{2\varepsilon}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\} (2z - N')}\right\},$$

with $N' = \max\{M \leq N : M$ *is odd*$\}$.

Then proposition 1(i)-(ii) applies with δ_1^* *replaced by* δ_{NC1}^*. *In particular, proposition 1(i)-(ii) applies unchanged if* $\varepsilon \leq \left(2\sum_{z>N'/2}\binom{N'}{n}(2n-N')\right)^{-1}$.

Proposition 3 *Consider the noncompliance scenario and impose monotonicity as in example 2. Define the decision rule*

$$\delta_6^*(\omega) = \begin{cases} 0, & \frac{1}{2} + \frac{I_N}{4\varepsilon N} < 0 \\ \frac{1}{2} + \frac{I_N}{4\varepsilon N}, & 0 \leq \frac{1}{2} + \frac{I_N}{4\varepsilon N} \leq 1 \\ 1, & \frac{1}{2} + \frac{I_N}{4\varepsilon N} > 1 \end{cases}$$

and the decision rule

$$\delta_{NC2}^* = \alpha^* \delta_1^* + (1-\alpha^*)\delta_6^*,$$

where

$$\alpha^* = \begin{cases} 0, & \varepsilon \geq 1/2 \\ \frac{\frac{1}{4\varepsilon} - \sum_{z \geq N'(1-2\varepsilon)} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{1}{4\varepsilon}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\}(2z-N')}{\sum_{z>N'/2}\binom{N'}{z}(2z-N') - \sum_{z \geq N'(1-2\varepsilon)} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{1}{4\varepsilon}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\}(2z-N')}, & \varepsilon_N^* < \varepsilon < 1/2 \\ 1, & \varepsilon \leq \varepsilon_N^* \end{cases},$$

with $N' = \max\{M \leq N : M$ *is odd*$\}$ *and* $\varepsilon_N^* = \left(4\sum_{z>N'/2}\binom{N'}{n}(2n-N')\right)^{-1}$.

Then proposition 1(i)-(ii) applies with δ_1^* *replaced by* δ_{NC1}^*. *In particular, proposition 1(i)-(ii) applies unchanged if* $\varepsilon \leq \varepsilon_N^*$.

Proposition 4 *Consider the hidden covariate setting of example 4. Define the decision rule*

$$\delta_{HC}^* = \alpha^* \delta_1^* + (1-\alpha^*)\delta_4^*,$$

where

$$\alpha^* = \begin{cases} 0, & b \geq \log\left(2+\sqrt{5}\right) \\ \frac{\frac{4\exp(b)}{\exp(2b)-1} - \sum_{z \geq N'(1-\exp(2b)/2)} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{4\exp(b)}{\exp(2b)-1}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\}(2z-N')}{\sum_{z>N'/2}\binom{N'}{z}(2z-N') - \sum_{z \geq N'(1-\exp(2b)/2)} \binom{N'}{z} \min\left\{\frac{1}{2} + \frac{4\exp(b)}{\exp(2b)-1}\left(\frac{z}{N'} - \frac{1}{2}\right), 1\right\}(2z-N')}, & b_N^*(N) < b < \log\left(2+\sqrt{5}\right) \\ 1, & b \leq b_N^* \end{cases}$$

with $N' = \max\{M \leq N : M$ *is odd*$\}$ *as before and*

$$b_N^* = \log\left(\left(\sum_{n<N^*/2}\binom{N'}{n}(2n-N')\right)^{-1} + \left(1 + \left(\sum_{n<N^*/2}\binom{N'}{n}(2n-N')\right)^{-2}\right)^{1/2}\right).$$

Then proposition 1(i)-(ii) applies with δ_1^* *replaced by* δ_{HC}^*. *In particular, proposition 1(i)-(ii) applies unchanged if* $b \leq b_N^*$.