Why equivalent MTRs generate different bounds

January 11, 2023

1 Summary

Let X be a binary random variable. Consider the models

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{1}$$

and

$$Y = \alpha_0 \mathbb{1}\{X = 0\} + \alpha_1 \mathbb{1}\{X = 1\} + \eta, \tag{2}$$

where ε and η are the error terms. Since the regressors in (2) can be constructed from linear combinations of the regressors in (1) and vice versa, we could say these models are equivalent. We would expect these models to fit the data equally well, which they do when using the standard linear regression framework.

Given the result above, we may expect `ivmte` to return the same bounds for equivalent models specified using different MTRs. However, in Issue #229 on GitHub, we find this not to be the case. I believe this has to do with the criterion used to fit the data. Specifically, the least squares criterion ensures that equivalent models fit the data equally well and generate equivalent bounds. In contrast, the other criteria supported by `ivmte` do not.

2 Why I think the bounds differ

2.1 Setup

Let B denote the $n \times k$ design matrix used in the direct regression approach.[1] The columns in B are the regressors in the model chosen by the user. Let Y denote the $n \times 1$ vector of outcomes.

Let P denote a $k \times k$ full-rank matrix. Using P, we can transform B to construct design matrices for equivalent models. For example, suppose B is the $n \times 2$ design matrix of (1), with the first column being a vector of 1's for the intercept, and the second column being a vector of realized X's. If we define

$$P \equiv \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix},$$

then the design matrix of (2) is BP.

2.2 Obtaining bounds under the least squares criterion

Define the criterion functions

$$Q(\theta) \equiv (Y - B\theta)'(Y - B\theta),$$
$$\widetilde{Q}(\theta) \equiv (Y - BP\theta)'(Y - BP\theta).$$

These criterion functions correspond to the OLS problems for two equivalent models.

[1]See the file `direct-mtr-procedure.pdf` posted in Issue #194 on GitHub on how B is constructed.

Lemma 1. *The minimum criteria*

$$Q^\star \equiv \min_\theta Q(\theta) \tag{3}$$

and

$$\widetilde{Q}^\star \equiv \min_\theta \widetilde{Q}(\theta) \tag{4}$$

are equal.

Proof. Let $\theta^\star$ denote a solution to (3) and $\widetilde{\theta}^\star$ denote a solution to (4). Then

$$\begin{aligned} Q^\star &= (Y - B\theta^\star)'(Y - B\theta^\star) \\ &\leq (Y - B[P\widetilde{\theta}^\star])'(Y - B[P\widetilde{\theta}^\star]) \\ &= \widetilde{Q}^\star. \end{aligned} \tag{5}$$

Similarly,

$$\begin{aligned} \widetilde{Q}^\star &= (Y - BP\widetilde{\theta}^\star)'(Y - BP\widetilde{\theta}^\star) \\ &\leq (Y - BP[P^{-1}\theta^\star])'(Y - BP[P^{-1}\theta^\star]) \\ &= (Y - B\theta^\star)'(Y - B\theta^\star) \\ &= Q^\star. \end{aligned} \tag{6}$$

Since $Q^\star \leq \widetilde{Q}^\star$ and $\widetilde{Q}^\star \leq Q^\star$, it follows that $Q^\star = \widetilde{Q}^\star$. ■

Corollary 1. *If $\theta^\star$ is a solution to (3), then $P^{-1}\theta^\star$ is a solution to (4). Likewise, if $\widetilde{\theta}^\star$ is a solution to (4), then $P\widetilde{\theta}^\star$ is a solution to (3).*

Proof. From (5), it follows that $Q^\star = Q(P\widetilde{\theta}^\star)$. Likewise, from (6), it follows that

$\widetilde{Q}^\star = \widetilde{Q}(P^{-1}\theta^\star)$. ■

Now consider the constrained optimization problems

$$\begin{aligned} \max_{\theta}/\min \quad & \tau'\theta \\ \text{s.t.} \quad & Q(\theta) \leq Q^\star(1+\kappa) \end{aligned} \tag{7}$$

and

$$\begin{aligned} \max_{\widetilde{\theta}}/\min \quad & (P'\tau)'\widetilde{\theta} \\ \text{s.t.} \quad & \widetilde{Q}(\widetilde{\theta}) \leq \widetilde{Q}^\star(1+\kappa), \end{aligned} \tag{8}$$

where κ is a tuning parameter; the vector τ determines the target parameter, and its entries correspond to the same variables as the columns in B. `ivmte` solves these types of problems when the user selects the least squares criterion (`direct = ''ls''`) and a hard constraint on the criterion (`soft = FALSE`). Shape constraints are omitted for simplicity. Problems (7) and (8) are for equivalent models.

Lemma 2. *The bounds obtained from* (7) *and* (8) *are equivalent.*

Proof. Let Θ denote the feasible set in (7) and $\widetilde{\Theta}$ denote the feasible set in (8).

Take any $\widetilde{\theta} \in \widetilde{\Theta}$. From Corollary 1, we know that $P\widetilde{\theta} \in \Theta$. If we set $\theta = P\widetilde{\theta}$ in (7), then our objective value is equal to $\tau' P\widetilde{\theta}$. It follows that the bounds in (7) must contain the bounds in (8).

Similarly, take any $\theta \in \Theta$. From Corollary 1, we know that $P^{-1}\theta \in \widetilde{\Theta}$. If we set $\widetilde{\theta} = P^{-1}\theta$ in (8), then our objective value is equal to $\tau' P(P^{-1}\theta) = \tau'\theta$. It follows that the bounds in (8) must contain the bounds in (7).

If the bounds in (7) and (8) contain each other, then the bounds must be equal. ■

2.3 Obtaining bounds under the ℓ_1 criterion

Lemma 1 showed that least squares problems (3) and (4) are isomorphic to each other. From this, we were able to show that the bounds in (7) and (8) are also equivalent. However, under the ℓ_1 criterion (`direct = ``l1''`), the criterion minimization problems for equivalent models do not seem to be isomorphic to each other. To see this, consider minimizing the criterion under the ℓ_1 norm when the design matrix is B. This means solving

$$\min_{\theta} \|B'Y - B'B\theta\|_1,$$

which can be expressed as the following linear programming problem,

$$\begin{aligned} Q^\star = \min_{\sigma,\theta} \quad & \sigma'\mathbf{1} \\ \text{s.t.} \quad & A\sigma + B'B\theta = B'Y \\ & \sigma \geq 0, \end{aligned} \tag{9}$$

where σ is a $2k \times 1$ vector of non-negative slack variables; $\mathbf{1}$ is a $2k \times 1$ vector of 1's; and A is a $k \times 2k$ matrix containing 0's, 1's, and -1's.[2]

Now consider minimizing the criterion when the design matrix is BP,

$$\min_{\widetilde{\theta}} \|P'B'Y - P'B'BP\widetilde{\theta}\|_1,$$

[2] A has the form

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

This means solving

$$\widetilde{Q}^{\star} = \min_{\widetilde{\sigma},\widetilde{\theta}} \quad \widetilde{\sigma}'\mathbf{1}$$
$$\text{s.t.} \quad A\widetilde{\sigma} + P'B'BP\widetilde{\theta} = P'B'Y \tag{10}$$
$$\widetilde{\sigma} \geq 0,$$

Lemma 3. *The minimum criterion in* (9) *is 0 if and only if the minimum criterion in* (10) *is 0.*

Proof. Rewrite the constraints in (9) and (10) as

$$A\sigma = B'Y - B'B\theta \tag{11}$$

$$A\widetilde{\sigma} = P'(B'Y - B'BP\widetilde{\theta}). \tag{12}$$

Let $\Sigma \times \Theta$ denote the feasible set in (9), and $\widetilde{\Sigma} \times \widetilde{\Theta}$ denote the feasible set in (10).

Suppose the objective in (9) is 0. That is, suppose there exists a $(\sigma^{\star}, \theta^{\star}) \in \Sigma \times \Theta$ such that $\sigma^{\star\prime}\mathbf{1} = 0$. Since $\sigma^{\star} \geq 0$, it must be that all elements of $\sigma^{\star}$ are 0. It follows from (11) that $B'Y - B'B\theta^{\star} = \mathbf{0}$. Let $\widetilde{\theta}^{\star} = P^{-1}\theta^{\star}$. Then from (12), it follows that

$$\begin{aligned} A\widetilde{\sigma} &= P'(B'Y - B'BP\underbrace{\widetilde{\theta}^{\star}}_{P^{-1}\theta^{\star}}) \\ &= P'(\underbrace{B'Y - B'B\theta^{\star}}_{\mathbf{0}}) \\ &= \mathbf{0}. \end{aligned}$$

It follows that $(\mathbf{0}, P^{-1}\theta^{\star}) \in \widetilde{\Sigma} \times \widetilde{\Theta}$, and the objective in (10) is 0.

The converse can be proven the same way. Suppose the objective in (10) is 0. This implies that $\mathbf{0} \in \widetilde{\Sigma}$. From (12), it also means there exists some $\widetilde{\theta}^{\star} \in \widetilde{\Theta}$ such that

$B'Y - B'BP\widetilde{\theta}^\star = 0$. If we let $\theta^\star = P\widetilde{\theta}^\star$, then it follows from (11) that $B'Y - B'B\theta^\star = 0$ and $\mathbf{0} \in \Sigma$. The objective in (9) must then be 0. ■

Now consider obtaining the bounds under the ℓ_1 criterion for equivalent models. When the design matrix is B, this means solving

$$\begin{aligned} \max_{\sigma,\theta}/\min \quad & \tau'\theta \\ \text{s.t.} \quad & \sigma'\mathbf{1} \leq Q^\star(1+\kappa) \\ & A\sigma + B'B\theta = B'Y \\ & \sigma \geq 0 \end{aligned} \tag{13}$$

When the design matrix is BP, this means solving

$$\begin{aligned} \max_{\widetilde{\sigma},\widetilde{\theta}}/\min \quad & (P'\tau)'\widetilde{\theta} \\ \text{s.t.} \quad & \widetilde{\sigma}'\mathbf{1} \leq \widetilde{Q}^\star(1+\kappa) \\ & A\widetilde{\sigma} + P'B'BP\widetilde{\theta} = P'B'Y \\ & \widetilde{\sigma} \geq 0. \end{aligned} \tag{14}$$

Corollary 2. *(to Lemma 3) The bounds in* (13) *and* (14) *are equal if the minimum criteria are equal to 0.*

Proof. Let $\Sigma \times \Theta$ denote the feasible set in (13) when $Q^\star = 0$. Let $\widetilde{\Sigma} \times \widetilde{\Theta}$ denote the feasible set in (14) when $\widetilde{Q}^\star = 0$.

Take any $\theta \in \Theta$. In the proof of Lemma 3, we saw that $(\mathbf{0}, P^{-1}\theta) \in \widetilde{\Sigma} \times \widetilde{\Theta}$. If we set $\widetilde{\theta} = P^{-1}\theta$ in (14), then the objective value is $(P'\tau)'P^{-1}\theta = \tau'\theta$. So the bounds (14) must contain the bounds in (13).

Likewise, take any $\widetilde{\theta} \in \widetilde{\Theta}$. In the proof of Lemma 3, we saw that $(\mathbf{0}, P\widetilde{\theta}) \in \Sigma \times \Theta$.

If we set $\theta = P\widetilde{\theta}$ in (13), then the objective value is $\tau' P\widetilde{\theta}$. So the bounds (13) must contain the bounds in (14).

Since the bounds in (13) and (14) contain each other, they must be equal. ■

However, if the criteria in (9) and (10) are not zero, there is no guarantee that the criteria are equal to each other. This suggests the constraints in (13) may not be as restrictive as those of (14), and vice versa. As a result, the bounds in (13) and (14) may differ.

The example posted by John Bonney here is an example where two equivalent models resulted in different criteria and different bounds. The different criteria potentially stems from $B'B$ and $P'B'BP$ having different ranks.