Improving the Numerical Stability of `ivmte`

A. Torgovitsky*

January 4, 2021

1 Motivation

Strange numerical warnings and errors from Gurobi are a recurring problem in `ivmte`. They almost certainly result from the optimization problem being poorly scaled, a point that is supported by the scaling statistics reported by Gurobi. This note contains a proposal that should systematically improve scaling.

I will focus on the "direct" procedure for now, since it is easier. After we check that all of this works for the direct procedure, we can try to apply a similar strategy to the original case with IV–like estimands.

2 Problem Setup

- Assume that the MTRs have the following form:

$$\mathbb{E}[Y(d)|U = u, X = x] \equiv m(d|u, x) = \sum_{k=1}^{K_d} \theta_{dk} b_{dk}(u, x), \tag{1}$$

  where $\theta_d \equiv [\theta_{d1}, \ldots, \theta_{dK_d}]'$ are unknown coefficients and b_{dk} are known basis functions for the MTR with d. (*Note: In contrast to the previous note, I am now indexing $d = 0$ and $d = 1$ separately. This seems to be easier for addressing the collinearity issue that arose in the previous note.*) Let $\theta \equiv [\theta_0', \theta_1']'$.

- Assume that $\theta \in \Theta$ can be represented as $r_{\text{lb}} \leq R\theta \leq r_{\text{ub}}$ for some known constraint matrix R and vector r. (In practice, R and r change on each iteration of the audit procedure, but I will ignore this in the notation.)

*Kenneth C. Griffin Department of Economics, University of Chicago.

- Let $p(x,z) \equiv \mathbb{P}[D=1|X=x, Z=z]$ and $P \equiv p(X,Z)$ as usual.

- As we know, the basis representation implies that

$$\mathbb{E}[Y|D,X,P] = \sum_{k=1}^{K_0} \theta_{0k}(1-D)B_{0k}(P,X) + \sum_{k=1}^{K_1} \theta_{1k} D B_{1k}(P,X) \tag{2}$$

  where

$$B_{0k}(p,x) \equiv \frac{1}{1-p}\int_p^1 b_{0k}(u,x)\,du \quad \text{and} \quad B_{1k}(p,x) \equiv \frac{1}{p}\int_0^p b_{1k}(u,x)\,du. \tag{3}$$

- Denote the least squares criterion by

$$\hat{Q}(\theta) = \sum_{i=1}^{n}\left(Y_i - \sum_{k=1}^{K_0}\theta_{0k}(1-D)B_{0ki} - \sum_{k=1}^{K_1}\theta_{1k} D B_{1ki}\right)^2. \tag{4}$$

3 Rescaling the Least Squares Objective

- The general problem is that $\hat{Q}(\theta)$ is a poorly scaled quadratic form. See previous version of the note for an explanation of why. (It is more clumsy to explain in the new notation.)

- The idea is to rescale each B_{dki} to lie in $[0,1]$ by defining:

$$\tilde{B}_{dki} \equiv \frac{B_{dki} - \text{lb}_{dk}}{(\text{ub}_{dk} - \text{lb}_{dk})}, \tag{5}$$

  where lb_{dk} and ub_{dk} are the minimum and maximum of $\{B_{dki}\}_{i=1}^{n}$.

- If there is a constant in the specification of the d MTR, then we will have $B_{dki} = 1$, so obviously we can't normalize this term. This is a bit annoying, since we ideally don't want to require the MTRs to always have a constant. To handle this, let's assume that $b_{d1}(u,x) = 1$ for both $d = 0, 1$ always. We can conceptualize the no-constant case as one where we know $\theta_{d1} = 0$. So in (5) we assume that $k \geq 2$ since we can't normalize the constant term.

- Substituting (5) we have

$$\begin{aligned}
\sum_{k=1}^{K_0}\theta_{0k}(1-D)B_{0ki} &= (1-D)\theta_{01}+\sum_{k=2}^{K_0}\theta_{0k}(1-D)B_{0ki}\\
&= (1-D)\left(\theta_{01}+\sum_{k=2}^{K_0}\theta_{0k}\left(\mathrm{ub}_{0k}-\mathrm{lb}_{0k}\right)\frac{(B_{0ki}-\mathrm{lb}_{0k})}{(\mathrm{ub}_{0k}-\mathrm{lb}_{0k})}+\theta_{0k}\mathrm{lb}_{0k}\right)\\
&= (1-D)\left(\left(\theta_{01}+\sum_{k=2}^{K_0}\theta_{0k}\mathrm{lb}_{0k}\right)+\sum_{k=2}^{K_0}\theta_{0k}\left(\mathrm{ub}_{0k}-\mathrm{lb}_{0k}\right)\tilde{B}_{0ki}\right)\\
&\equiv (1-D)\left(\xi_{01}+\sum_{k=2}^{K_0}\xi_{0k}\tilde{B}_{0ki}\right)\\
&= \xi_{01}(1-D)+\sum_{k=2}^{K_0}\xi_{0k}(1-D)\tilde{B}_{0ki},
\end{aligned}$$

  where

$$\xi_{01}\equiv\theta_{01}+\sum_{k=2}^{K_0}\theta_{0k}\mathrm{lb}_{0k}\quad\text{and}\quad\xi_{0k}=\theta_{0k}(\mathrm{ub}_{0k}-\mathrm{lb}_{0k})\text{ for all }k=2,\ldots,K.$$

  Similarly,

$$\sum_{k=1}^{K_1}\theta_{1k}DB_{1ki}=\xi_{11}D+\sum_{k=2}^{K_1}\xi_{1k}DB_{1ki},$$

  where

$$\xi_{11}\equiv\theta_{11}+\sum_{k=2}^{K_1}\theta_{1k}\mathrm{lb}_{1k}\quad\text{and}\quad\xi_{1k}=\theta_{1k}(\mathrm{ub}_{1k}-\mathrm{lb}_{1k})\text{ for all }k=2,\ldots,K.$$

- Substituting back into (4), we have a regression of Y onto $(1-D)$, $\{(1-D)\tilde{B}_{0k}\}_{k=2}^{K_0}$, D, and $\{D\tilde{B}_{1k}\}_{k=2}^{K_1}$. This should not be collinear.

- For $k\geq 2$, we can directly back out θ_{dk} from ξ_{dk}. We can back out the intercepts ($k=1$) via

$$\theta_{d1}=\xi_{d1}-\sum_{k=2}^{K_d}\theta_{dk}\mathrm{lb}_{dk}. \tag{6}$$

If there isn't an intercept in the d MTR specification, then instead of (6), we just take $\theta_{d1} = 0$.