Supplementary Materials for "Fixing Boundary Violations: Applying Constrained Optimization to the Truncated Regression Model"

Min-Hua Huang

Department of Political Science

Texas A&M University, College Station, TX 77843

email: admin@mhhspace.org

December 19, 2011

1. INTRODUCTION

This is a supplementary document for the article "Fixing Boundary Violations: Applying Constrained Optimization to the Truncated Regression Model." It is comprised of three parts: model information, technical notes on numerical issues, and instructions for replication. Model information includes the mathematical detail of the constrained optimization problem (COP) for the simulation and replication studies in the main text. Technical notes on numerical issues explains how to handle two numerical problems that result in non-convergence. The instructions for replication section explains how to replicate all of the findings in the article.

2. MODEL INFORMATION

We first present the model specification for the three simulation studies. Set x^* as the covariate matrix after being centered, including the constant.

$$
\begin{aligned}
&\textit{Minimize} \qquad && \ln L = \sum_{i=1}^{n} \ln D_i + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x_i^*} \boldsymbol{\beta} \right)^2 \\
&\textit{Subject to} \qquad && g_1 = \beta_0 + \hat{y}_{\sim 0}^{\max} - b \leq 0 \\
& && g_2 = a - \beta_0 - \hat{y}_{\sim 0}^{\min} \leq 0 \\
& && g_3 = \beta_0 - b \leq 0 \\
& && g_4 = -\beta_0 + a \leq 0 \\
& && g_5 = \beta_1 - min\left(\frac{b - \hat{y}_{\sim 1}^{\max}}{x_1^{*\max}}, -\frac{\hat{y}_{\sim 1}^{\min} - a}{x_1^{*\min}} \right) \leq 0 \\
& && g_6 = -\beta_1 + max\left(\frac{a - \hat{y}_{\sim 1}^{\min}}{x_1^{*\max}}, -\frac{\hat{y}_{\sim 1}^{\max} - b}{x_1^{*\min}} \right) \leq 0 \\
& && \vdots
\end{aligned}
$$

$$g_{2m+3} = \beta_m - min\left(\frac{b-\hat{y}_{\sim m}^{\max}}{x_m^{*\max}}, -\frac{\hat{y}_{\sim m}^{\min}-a}{x_m^{*\min}}\right) \le 0$$
$$g_{2m+4} = -\beta_m + max\left(\frac{a-\hat{y}_{\sim m}^{\min}}{x_m^{*\max}}, -\frac{\hat{y}_{\sim m}^{\max}-b}{x_m^{*\min}}\right) \le 0$$
$$g_{2m+5} = \sigma - b + a \le 0$$
$$g_{2m+6} = -\sigma + 0.001 \le 0,$$

where $D_i = \sqrt{2\pi}\sigma\left\{\Phi\left(\frac{b-\boldsymbol{x_i}^*\boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{a-\boldsymbol{x_i}^*\boldsymbol{\beta}}{\sigma}\right)\right\}$.

We can re-specify the model by setting $\boldsymbol{\gamma} = (\beta_0 \cdots \beta_m\ \sigma)^T$ and $c(\boldsymbol{\gamma}) = (g_1 \cdots g_{2m+6})^T$ and derive

$$\begin{aligned} &Minimize \quad f(\boldsymbol{\gamma}) \\ &Subject\ to \quad c(\boldsymbol{\gamma}) \le 0. \end{aligned}$$

Now, we add a Lagrangian multiplier and form a new objective function

$$l(\boldsymbol{\gamma}; \boldsymbol{\lambda}) = f(\boldsymbol{\gamma}) + \boldsymbol{\lambda}^T c(\boldsymbol{\gamma}),$$

where $\boldsymbol{\lambda} = (\lambda_1 \cdots \lambda_{2m+6})^T$, and $\boldsymbol{d} = (d_{\beta_0} \cdots d_\sigma)^T$. To solve a COP problem, we need to calculate $c(\boldsymbol{\gamma})$, $\partial f(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$, $\partial c(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$, and $\partial^2 l(\boldsymbol{\gamma}; \boldsymbol{\lambda})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T$. We have already determined $c(\boldsymbol{\gamma})$. Regarding $\partial f(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$,

$$\frac{\partial f(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \left(\frac{\partial f}{\partial \beta_0} \cdots \frac{\partial f}{\partial \beta_m} \frac{\partial f}{\partial \sigma}\right)^T.$$

Specifically,

$$\frac{\partial f}{\partial \beta_j} = \sum_{i=1}^{n} \frac{1}{D_i}\frac{\partial D_i}{\partial \beta_j} - \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x_i^*}\boldsymbol{\beta}\right)x_{ij}^*$$
$$\frac{\partial f}{\partial \sigma} = \sum_{i=1}^{n} \frac{1}{D_i}\frac{\partial D_i}{\partial \sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x_i^*}\boldsymbol{\beta}\right)^2,$$

where

$$\frac{\partial D_i}{\partial \beta_j} = -x_{ij}^* \exp\left[\frac{-\left(b - \boldsymbol{x_i^*\beta}\right)^2}{2\sigma^2}\right] + x_{ij}^* \exp\left[\frac{-\left(a - \boldsymbol{x_i^*\beta}\right)^2}{2\sigma^2}\right]$$

$$\frac{\partial D_i}{\partial \sigma} = \frac{-\left(b - \boldsymbol{x_i^*\beta}\right)}{\sigma} \exp\left[\frac{-\left(b - \boldsymbol{x_i^*\beta}\right)^2}{2\sigma^2}\right] + \frac{\left(a - \boldsymbol{x_i^*\beta}\right)}{\sigma} \exp\left[\frac{-\left(a - \boldsymbol{x_i^*\beta}\right)^2}{2\sigma^2}\right] + \frac{D_i}{\sigma}.$$

Regarding $\partial c\left(\boldsymbol{\gamma}\right)/\partial \boldsymbol{\gamma}$,

$$\frac{\partial c\left(\boldsymbol{\gamma}\right)}{\partial \boldsymbol{\gamma}} = \begin{pmatrix} \frac{\partial g_1}{\partial \beta_0} & \cdots & \frac{\partial g_1}{\partial \beta_m} & \frac{\partial g_1}{\partial \sigma} \\ \vdots & \cdots & \vdots & \vdots \\ \frac{\partial g_{2m+6}}{\partial \beta_0} & \cdots & \frac{\partial g_{2m+6}}{\partial \beta_m} & \frac{\partial g_{2m+6}}{\partial \sigma} \end{pmatrix}.$$

Specifically,[1]

$$\begin{aligned}
&\frac{\partial g_1}{\partial \beta_0} = 1, && \frac{\partial g_1}{\partial \beta_j} = v_j^+ x_j^{*\max} + v_j^- x_j^{*\min}, && \frac{\partial g_1}{\partial \sigma} = 0,\\
&\frac{\partial g_2}{\partial \beta_0} = -1, && \frac{\partial g_2}{\partial \beta_j} = -v_j^+ x_j^{*\min} - v_j^- x_j^{*\max}, && \frac{\partial g_2}{\partial \sigma} = 0,\\
&\frac{\partial g_3}{\partial \beta_0} = 1, && \frac{\partial g_3}{\partial \beta_j} = 0, && \frac{\partial g_3}{\partial \sigma} = 0,\\
&\frac{\partial g_4}{\partial \beta_0} = -1, && \frac{\partial g_4}{\partial \beta_j} = 0, && \frac{\partial g_4}{\partial \sigma} = 0.
\end{aligned}$$

For the 5th to $(2m+4)$th rows, if $i = j$, then

$$\frac{\partial g_{2i+3}}{\partial \beta_j} = 1, \frac{\partial g_{2i+4}}{\partial \beta_j} = -1.$$

Otherwise,

$$\frac{\partial g_{2i+3}}{\partial \beta_j} = \begin{cases} \frac{v_j^+ x_j^{*\max} + v_j^- x_j^{*\min}}{x_i^{*\max}} & if \quad \frac{b - \hat{y}_{\sim 1}^{\max}}{x_1^{*\max}} \leq -\frac{\hat{y}_{\sim 1}^{\min} - a}{x_1^{*\min}} \\ \frac{v_j^+ x_j^{*\min} + v_j^- x_j^{*\max}}{x_i^{*\min}} & if \quad \frac{b - \hat{y}_{\sim 1}^{\max}}{x_1^{*\max}} > -\frac{\hat{y}_{\sim 1}^{\min} - a}{x_1^{*\min}} \end{cases}$$

[1] Here we set $v_0^+ = 1$ and $v_0^- = 1$, since every predicted value has the constant.

$$\frac{\partial g_{2i+4}}{\partial \beta_j} = \begin{cases} -\frac{v_j^+ x_j^{*\,\min} + v_j^- x_j^{*\,\max}}{x_i^{*\,\max}} & if \quad \frac{a - \hat{y}_{\sim 1}^{\min}}{x_1^{*\,\max}} \geq -\frac{\hat{y}_{\sim 1}^{\max} - b}{x_1^{*\,\min}} \\ -\frac{v_j^+ x_j^{*\,\max} + v_j^- x_j^{*\,\min}}{x_i^{*\,\min}} & if \quad \frac{a - \hat{y}_{\sim 1}^{\min}}{x_1^{*\,\max}} < -\frac{\hat{y}_{\sim 1}^{\max} - b}{x_1^{*\,\min}}. \end{cases}$$

Also,

$$\frac{\partial g_{2i+3}}{\partial \sigma} = 0, \frac{\partial g_{2i+4}}{\partial \sigma} = 0.$$

For the last two rows,

$$\frac{\partial g_{2i+5}}{\partial \beta_j} = 0, \frac{\partial g_{2i+5}}{\partial \sigma} = 1,$$
$$\frac{\partial g_{2i+6}}{\partial \beta_j} = 0, \frac{\partial g_{2i+6}}{\partial \sigma} = -1.$$

Regarding $\partial^2 l\,(\boldsymbol{\gamma}; \boldsymbol{\lambda})/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T$,

$$\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} \frac{-1}{{D_i}^2} \left(\frac{\partial D_i}{\partial \beta_j}\right)\left(\frac{\partial D_i}{\partial \beta_k}\right) + \sum_{i=1}^{n} \frac{1}{D_i} \frac{\partial^2 D_i}{\partial \beta_j \partial \beta_k} + \frac{1}{\sigma^2} \sum_{i=1}^{n} x_{ik}^* x_{ij}^*$$
$$\frac{\partial^2 l}{\partial \beta_j \partial \sigma} = \sum_{i=1}^{n} \frac{-1}{{D_i}^2} \left(\frac{\partial D_i}{\partial \beta_j}\right)\left(\frac{\partial D_i}{\partial \sigma}\right) + \sum_{i=1}^{n} \frac{1}{D_i} \frac{\partial^2 D_i}{\partial \beta_j \partial \sigma} + \frac{2}{\sigma^3} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x_i^*} \boldsymbol{\beta}\right) x_{ij}^*$$
$$\frac{\partial^2 l}{\partial \sigma \partial \sigma} = \sum_{i=1}^{n} \frac{-1}{{D_i}^2} \left(\frac{\partial D_i}{\partial \sigma}\right)\left(\frac{\partial D_i}{\partial \sigma}\right) + \sum_{i=1}^{n} \frac{1}{D_i} \frac{\partial^2 D_i}{\partial \sigma \partial \sigma} + \frac{3}{\sigma^4} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2,$$

where

$$\frac{\partial^2 D_i}{\partial \beta_j \partial \beta_k} = \frac{-x_{ij}^* x_{ik}^* \left(b - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)}{\sigma^2} \exp\left[\frac{-\left(b - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{2\sigma^2}\right] + \frac{x_{ij}^* x_{ik}^* \left(a - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)}{\sigma^2} \exp\left[\frac{-\left(a - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{2\sigma^2}\right]$$
$$\frac{\partial^2 D_i}{\partial \beta_j \partial \sigma} = \frac{-x_{ij}^* \left(b - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{\sigma^3} \exp\left[\frac{-\left(b - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{2\sigma^2}\right] + \frac{x_{ij}^* \left(a - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{\sigma^3} \exp\left[\frac{-\left(a - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{2\sigma^2}\right]$$
$$\frac{\partial^2 D_i}{\partial \sigma \partial \sigma} = -\frac{\left(b - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^3}{\sigma^4} \exp\left[\frac{-\left(b - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{2\sigma^2}\right] + \frac{\left(a - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^3}{\sigma^4} \exp\left[\frac{-\left(a - \boldsymbol{x_i^*} \boldsymbol{\beta}\right)^2}{2\sigma^2}\right].$$

Given the above information, we can implement the sequential quadratic programming algorithm to solve the truncated regression model as a COP problem. As for the replication of Hellwig and Samuels's models, we fixed the covariate matrix at the minimum, and the interaction terms are recalculated by using the covariate values after being fixed. The new covariate matrix is marked as $x^{\dagger}$. We use the notion as stated in the appendix of the article. The model specification is

$$
\begin{aligned}
&\textit{Minimize} \qquad && \ln L = \sum_{i=1}^{n} \ln D_i + \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_{\boldsymbol{i}}^{\dagger} \boldsymbol{\beta} \right)^2 \\
&\textit{Subject to} \qquad && g_1 = \beta_0 + \hat{y}_{\sim 0}^{\max} - b \leq 0 \\
& && g_2 = a - \beta_0 - \hat{y}_{\sim 0}^{\min} \leq 0 \\
& && g_3 = \beta_0 - b \leq 0 \\
& && g_4 = -\beta_0 + a \leq 0 \\
& && g_5 = \beta_1 - \frac{b - \hat{y}_{\sim 1}^{\max}}{x_1^{\dagger \max}} \\
& && g_6 = -\beta_1 + \frac{a - \hat{y}_{\sim 1}^{\min}}{x_1^{\dagger \max}} \\
& && \vdots \\
& && g_{29} = \beta_{13} - \frac{b - \hat{y}_{\sim 13}^{\max}}{x_1^{\dagger \max}} \\
& && g_{30} = -\beta_{13} + \frac{a - \hat{y}_{\sim 13}^{\min}}{x_1^{\dagger \max}} \\
& && g_{31} = \sigma - b + a \leq 0 \\
& && g_{32} = -\sigma + 0.001 \leq 0.
\end{aligned}
$$

Notice that the definitions of $\hat{y}_{\sim j}^{\max}$ and $\hat{y}_{\sim j}^{\min}$ are different from the simulation model due to the dummy variables (see the article's Appendix section). The rest of the model information is the same except $c(\boldsymbol{\gamma})$ and $\partial c(\boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}$.

The difference of $c(\boldsymbol{\gamma})$ results from different centering methods. Thus, there are some corresponding changes in $\partial c(\boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}$, and they are all associated with the dummy variables.

For the first two rows, these changes are

$$\frac{\partial g_1}{\partial \beta_j} = v_j^+ x_j^{\dagger\max} \qquad , \frac{\partial g_2}{\partial \beta_j} = -v_j^- x_j^{\dagger\max}$$

$$\frac{\partial g_1}{\partial \beta_k} = \begin{cases} 1 & if \quad \beta_k = \max\left(\beta_{10}\cdots\beta_{13}\right) \\ 0 & if \quad otherwise \end{cases} \qquad , \frac{\partial g_2}{\partial \beta_k} = \begin{cases} -1 & if \quad \beta_k = \min\left(\beta_{10}\cdots\beta_{13}\right) \\ 0 & if \quad otherwise \end{cases}$$

where $j = \{0\cdots 9\}$ and $k = \{10\cdots 13\}$.

From the 5th to 22th rows, if $i = j$,

$$\frac{\partial g_{2i+3}}{\partial \beta_j} = 1, \frac{\partial g_{2i+4}}{\partial \beta_j} = -1.$$

Otherwise,

$$\frac{\partial g_{2i+3}}{\partial \beta_j} = \frac{v_j^+ x_j^{\dagger\max}}{x_i^{\dagger\max}} \qquad , \frac{\partial g_{2i+4}}{\partial \beta_j} = -\frac{v_j^- x_j^{\dagger\max}}{x_i^{\dagger\max}}$$

$$\frac{\partial g_{2i+3}}{\partial \beta_k} = \begin{cases} \frac{1}{x_i^{\dagger\max}} & if\, \beta_k = \max\left(\beta_{10}\cdots\beta_{13}\right) \\ 0 & if\, otherwise \end{cases} \qquad , \frac{\partial g_{2i+4}}{\partial \beta_k} = \begin{cases} \frac{-1}{x_i^{\dagger\max}} & if\, \beta_k = \min\left(\beta_{10}\cdots\beta_{13}\right) \\ 0 & if\, otherwise \end{cases}$$

where $i = \{1\cdots 9\}$, $j = \{0\cdots 9\}$ and $k = \{10\cdots 13\}$.

From the 23th to 30th rows

$$\frac{\partial g_{2i+3}}{\partial \beta_j} = \frac{v_j^+ x_j^{\dagger\max}}{x_i^{\dagger\max}} \qquad , \frac{\partial g_{2i+4}}{\partial \beta_j} = -\frac{v_j^- x_j^{\dagger\max}}{x_i^{\dagger\max}}$$

$$\frac{\partial g_{2i+3}}{\partial \beta_k} = \begin{cases} 1 & if \quad i = k \\ 0 & if \quad otherwise \end{cases} \qquad , \frac{\partial g_{2i+4}}{\partial \beta_k} = \begin{cases} -1 & if \quad i = k \\ 0 & if \quad otherwise \end{cases}$$

where $i = \{10\cdots 13\}$, $j = \{0\cdots 9\}$ and $k = \{10\cdots 13\}$.

3. TWO NUMERICAL ISSUES

Two numerical problems sometimes occur and cause non-convergence. The first problem is related to outliers of the predicted value for the location parameter. When the estimated location parameter is far smaller than the lower limit or far greater than the upper limit, the definite Gaussian integral in the denominator of the probability density function will approach zero, and hence, lead to a singularity problem. The second problem is about the magnitude of the step size in each numerical iteration. In some cases, the step size generated by the inverse Hessian is overdriven by the scale parameter, and thus, too large to generate eligible parameter estimates. Through a proper adjustment to the Hessian, the step size can be under control, and thus, reaching an admissible solution becomes possible.

To illustrate the first problem, we refer to the objective function (negative loglikelihood) specified in the constrained optimization problem

$$-\log L = \sum_{i=1}^{n} \ln D_i + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \boldsymbol{x_i \beta})^2.$$

When the following two conditions both exist, $\left|b - \boldsymbol{x_i^* \beta}\right| \gg 5\sigma$ and $\left|a - \boldsymbol{x_i^* \beta}\right| \gg 5\sigma$, D_i will approach zero, and thus, the natural logarithm of D_i will approach negative infinity. The value of the loglikelihood function is, therefore, dominated by a few large contributions from those outliers. The same issue, on the other hand, does not pose a problem to the second term of the loglikelihood function. The contribution of the square standardized deviation is relatively milder so that a few outliers do not seriously distort the result.

To cope with the outlier problem, we can set

$$D_i = \sqrt{2\pi}\sigma \qquad if \quad \left|a - \boldsymbol{x_i^* \hat{\beta}}\right| > 5\hat{\sigma} \quad \& \quad \left|b - \boldsymbol{x_i^* \hat{\beta}}\right| > 5\hat{\sigma},$$

where $\Phi\left(\frac{\boldsymbol{b - x_i^* \beta}}{\sigma}\right) - \Phi\left(\frac{\boldsymbol{a - x_i^* \beta}}{\sigma}\right) = 1$. Doing so is equivalent to setting the outlier's probability density value to almost zero

$$f\left(y_i|\boldsymbol{x_i\beta},\sigma\right)=\frac{\exp\left[-\frac{\left(y_i-\boldsymbol{x_i^*\beta}\right)^2}{2\sigma^2}\right]}{\Phi\left(\frac{b-\boldsymbol{x_i^*\beta}}{\sigma}\right)-\Phi\left(\frac{a-\boldsymbol{x_i^*\beta}}{\sigma}\right)}\rightarrow 0,$$

and thus, outliers do not cause great disturbance in a particular iteration. Since the occurrence of outliers is associated with inadmissible beta estimates, the number of outliers will reduce to zero when later iterations generate admissible solutions or reach convergence. The main effect of setting $D_i=\sqrt{2\pi}\sigma$ is to retain a smooth convergent sequence of parameter estimates in numerical optimization.

The second problem is about the proper control of the step parameter $\boldsymbol{d}$ by adjusting the Hessian. Without loss of generality, the following discussion only assumes one independent variable in the truncated regression model. Both x and β are used as a scalar. The essential rule of Newton's method in achieving sequential convergence to the minimum is

$$\boldsymbol{\gamma_{k+1}}=\boldsymbol{\gamma_k}-\boldsymbol{H}^{-1}\boldsymbol{g},$$

where the negative gradient represents the steepest direction that the parameter estimate should move. And the inverse Hessian indicates how far the move should be by the rate of quadratic convergence. Let the step parameter $\boldsymbol{d}^{(k)}=\left(d_\beta^{(k)},d_\sigma^{(k)}\right)^T$ represent the difference of the parameter estimate from the kth to $(k+1)$th iteration, thus

$$\begin{pmatrix} d_\beta^{(k)} \\ d_\sigma^{(k)} \end{pmatrix}=-\begin{pmatrix} \frac{\partial^2(-\log L)}{\partial\beta^2} & \frac{\partial^2(-\log L)}{\partial\beta\partial\sigma} \\ \frac{\partial^2(-\log L)}{\partial\beta\partial\sigma} & \frac{\partial^2(-\log L)}{\partial\sigma^2} \end{pmatrix}^{-1}\begin{pmatrix} \frac{\partial(-\log L)}{\partial\beta} \\ \frac{\partial(-\log L)}{\partial\sigma} \end{pmatrix}\Bigg|_{\boldsymbol{\gamma}^{(k)}},$$

and specifically,[2]

$$
\begin{aligned}
\frac{\partial\left(-\log L\right)}{\partial\beta} &= \sum_{i=1}^{n}\frac{1}{D_i}\left(\frac{\partial D_i}{\partial\beta}\right) - \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i-\beta x_i^*\right)x_i^* \\
\frac{\partial\left(-\log L\right)}{\partial\sigma} &= \sum_{i=1}^{n}\frac{1}{D_i}\left(\frac{\partial D_i}{\partial\sigma}\right) - \frac{1}{\sigma^3}\sum_{i=1}^{n}\left(y_i-\beta x_i^*\right)^2 \\
\frac{\partial^2\left(-\log L\right)}{\partial\beta^2} &= \sum_{i=1}^{n}\frac{-1}{D_i^2}\left(\frac{\partial D_i}{\partial\beta}\right)^2 + \sum_{i=1}^{n}\frac{1}{D_i}\left(\frac{\partial^2 D_i}{\partial\beta^2}\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i^*\right)^2 \\
\frac{\partial^2\left(-\log L\right)}{\partial\beta\partial\sigma} &= \sum_{i=1}^{n}\frac{-1}{D_i^2}\left(\frac{\partial D_i}{\partial\beta}\right)\left(\frac{\partial D_i}{\partial\sigma}\right) + \sum_{i=1}^{n}\frac{1}{D_i}\left(\frac{\partial^2 D_i}{\partial\beta\partial\sigma}\right) + \frac{2}{\sigma^3}\sum_{i=1}^{n}\left(y_i-\beta x_i^*\right)\left(x_i^*\right) \\
\frac{\partial^2\left(-\log L\right)}{\partial\sigma^2} &= \sum_{i=1}^{n}\frac{-1}{D_i^2}\left(\frac{\partial D_i}{\partial\sigma}\right)^2 + \sum_{i=1}^{n}\frac{1}{D_i}\left(\frac{\partial^2 D_i}{\partial\sigma^2}\right) + \frac{3}{\sigma^4}\sum_{i=1}^{n}\left(y_i-\beta x_i^*\right)^2.
\end{aligned}
$$

By multiplying the last term of $\partial^2\left(-\log L\right)/\partial\beta^2$ with a factor, we can control the step parameter d_σ at the same level while reducing the step parameter d_β. To see why this is the case, we first assume

$$
\begin{aligned}
&\frac{\partial^2\left(-\log L\right)}{\partial\beta^2} = t_1 + t_2, \quad \frac{\partial^2\left(-\log L\right)}{\partial\beta\partial\sigma} = t_3, \quad \frac{\partial^2\left(-\log L\right)}{\partial\sigma^2} = t_4 \\
&\frac{\partial\left(-\log L\right)}{\partial\beta} = g_1, \quad \frac{\partial\left(-\log L\right)}{\partial\sigma} = g_2,
\end{aligned}
$$

where

$$
t_1 = \sum_{i=1}^{n}\frac{-1}{D_i^2}\left(\frac{\partial D_i}{\partial\beta}\right)^2 + \sum_{i=1}^{n}\frac{1}{D_i}\left(\frac{\partial^2 D_i}{\partial\beta^2}\right), \quad t_2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i^*\right)^2.
$$

If we multiply t_2 with a factor τ, the new Hessian becomes

$$
\boldsymbol{H}^* = \begin{pmatrix} t_1 + \tau t_2 & t_3 \\ t_3 & t_4 \end{pmatrix}.
$$

With a few manipulations, we derive the new step parameter d_β^*, which is always smaller

[2] The following items specify the gradient and the Hessian for the simplest truncated regression model without the constant. The specification can be easily extended to the multivariate context by adding variable indicators.

than the original step parameter d_β if $\tau > 1$.

$$d_\beta^* = -\frac{t_4 g_\beta - t_3 g_\sigma}{(t_1 + t_2)\, t_4 - t_3^2 + (\tau - 1)\, t_2 t_4} \leq -\frac{t_4 g_\beta - t_3 g_\sigma}{(t_1 + t_2)\, t_4 - t_3^2} = d_\beta,$$

where $(\tau - 1)\, t_2 t_4 > 0$. Meanwhile, the new step parameter d_σ^* remains close to the original step parameter d_σ, and when $\tau \gg 1$

$$d_\sigma^* = -\frac{(t_1 + t_2)\, g_\sigma - t_3 g_\beta + (\tau - 1)\, t_2 g_\sigma}{(t_1 + t_2)\, t_4 - t_3^2 + (\tau - 1)\, t_2 t_4} \to -\frac{g_\sigma}{t_4}.$$

This indicates that the Hessian will approach a diagonal matrix ($t_3 \to 0$) when τ is far greater than 1

$$\begin{pmatrix} d_\beta^* \\ d_\sigma^* \end{pmatrix} = -\begin{pmatrix} t_1 + \tau t_2 & 0 \\ 0 & t_4 \end{pmatrix}^{-1} \begin{pmatrix} g_\beta \\ g_\sigma \end{pmatrix}.$$

With the above knowledge, we can reduce the step parameter d_β^* arbitrarily while keeping the step parameter d_σ^* at a certain level by increasing the factor τ. This technique of adjusting the Hessian is important in finding an admissible solution during the numerical analysis. In general, the original Hessian tends to generate too large a step and causes non-convergence as the number of covariates increases. The convergent rate would be slower when a smaller step d_β^* is in use. Most importantly, the final results of parameter estimation do not differ much if τ is chosen within a limited range. In the replication studies, we set τ as 14 and 4 for Model I and II, respectively.

4. REPLICATION

There are three stages of statistical analysis in this article. The first is to replicate the three political studies and demonstrate their out-of-bounds violations. The second includes three simulation tests to compare TRM and TRMCO. The third is to apply the TRMCO model to the Hellwig and Samuels's regression analysis.

4.1. *Three Political Studies*

The first part in replicating three political studies is executed in the Stata environment. Save all the data files under your designated directory. The default is set "C:\".

Hellwig and Samuels 2007

OLS's out-of-bounds violation for Model I:

- use "C:\HellwigSamuelsCPS2007.dta", clear
- regress incvotet incvotet1 dgdp tradeshr gdpxtradeshr electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code)
- predict yhat1
- sum yhat1

TRM's out-of-bounds violation for Model I:

- truncreg incvotet incvotet1 dgdp tradeshr gdpxtradeshr electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code) ll(0) ul(100)
- predict yhat2
- sum yhat2

OLS's out-of-bounds violation for Model II:

- regress incvotet incvotet1 dgdp grosscap gdpxgrosscap electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code)
- predict yhat3
- sum yhat3

TRM's out-of-bounds violation for Model II:

- truncreg incvotet incvotet1 dgdp grosscap gdpxgrosscap electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code) ll(0) ul(100)
- predict yhat4
- sum yhat4

Hansford and Gomez 2010

OLS's out-of-bounds violation in the F-test for excluded instruments:

- set mem 500m
- use "C:\HansfordGomez_Data.dta", clear
- xtreg GOPIT DemVoteShare2_3MA Yr52 Yr56 Yr60 Yr64 Yr68 Yr72 Yr76 Yr80 Yr84 Yr88 Yr96 Yr92 Yr2000 DNormPrcp_KRIG RainGOPI Rain_DVS23MA, fe robust
- predict yhat5
- sum yhat5

TRM's out-of-bounds violation in the F-Tests for excluded instruments:

- use "C:\HansfordGomez_MeanbyGroup.dta", clear
- truncreg gopit demvoteshare2_3ma yr52 yr56 yr60 yr64 yr68 yr72 yr76 yr80 yr84 yr88 yr96 yr92 yr2000 dnormprcp_krig raingopi rain_dvs23ma, robust ll(-100) ul(100)
- predict yhat6
- gen origin_gopit=yhat6+mean_gopit
- sum origin_gopit

Acemoglu et al. 2008

Before we start replications, run the following commands:

- set mem 500m
- use "C:\income_demo_5years.dta", clear
- tsset code_numeric year_numeric
- sort code_numeric year_numeric
- tab year, gen (yr)
- tab code, gen(cd)
- set matsize 800

OLS's out-of-bounds violation for the pooled model:

- reg fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, cluster(code)
- predict yhat7
- sum yhat7

TRM's out-of-bounds violation for the pooled model:

- truncreg fhpolrigaug L.(fhpolrigaug lrgdpch) yr* if sample==1, cluster(code) ll(0) ul(1)
- predict yhat8
- sum yhat8

OLS's out-of-bounds violation for the fixed-effect model:

- reg fhpolrigaug L.(fhpolrigaug lrgdpch) yr* cd* if sample==1, cluster(code)
- predict yhat9
- sum yhat9

TRM's out-of-bounds violation for the fixed-effect model:

- truncreg fhpolrigaug L.(fhpolrigaug lrgdpch) yr* cd* if sample==1, cluster(code) ll(0) ul(1)
- predict yhat10
- sum yhat10

4.2. *Simulation Tests*

All three simulations are executed in the Stata environment for OLS and TRM, and in the Matlab environment for TRMCO. All the files are saved in the directory labeled "OLS_simulation1," "TRM_simulation1," "TRMCO_simulation1," $\cdots$. The simulation datasets are "trialdata1.mat," "trialdata2.mat," and "trialdata3.mat" generated by "samlping1.m," "sampling2.m," and "sampling3.m" in Matlab. For OLS and TRM, we separate the dataset into 10 subsamples for each simulation test, and it is labeled as "sample1.dta," "sample2.dta,"$\cdots$,"sample30.dta" for the three tests.

Simulation tests for OLS

- do "C:\OLS_simulation1\olsreg1.do"
- do "C:\OLS_simulation2\olsreg2.do"
- do "C:\OLS_simulation3\olsreg3.do"

Simulation tests for TRM

- do "C:\TRM_simulation1\truncreg1.do"
- do "C:\TRM_simulation2\truncreg2.do"
- do "C:\TRM_simulation3\truncreg3.do"

Note that the result files are "ols_result1.dta," "truncreg_result1.dta,"$\cdots$. The coefficient b_3 is the constant.

Simulation Tests for TRMCO and Table 1

- format shortG
- enter "C:\TRMCO_simulation1\" under the Matlab environment and run "experiment1.m"
- run "getresult1.m"
- display(getstats)
- enter "C:\TRMCO_simulation2\" under the Matlab environment and run "experiment2.m"
- run "getresult2.m"
- display(getstats)
- enter "C:\TRMCO_simulation3\" under the Matlab environment and run "experiment3.m"
- run "getresult3.m"
- display(getstats)

Simulation Tests for Table 2

- format shortG
- enter "C:\TRMCO_simulation1\" under the Matlab environment and run "adt1.m"
- display(adcoeff1)
- enter "C:\TRMCO_simulation2\" under the Matlab environment and run "adt2.m"
- display(adcoeff2)
- enter "C:\TRMCO_simulation3\" under the Matlab environment and run "adt3.m"
- display(adcoeff3)

4.3. *Replications*

We replicate Hellwig and Samuels's (2007) Model I and Model II (p.292). Given that the original model is not strictly linear, the replication result will be slightly different due to the different centering method. OLS and TRM are implemented in the Stata environment, whereas TRMCO is executed in the Matlab environment.

Replication of Hellwig and Samuels's Model I, Table 3

- use "C:\fixmin_Hellwig1.dta", clear
- regress incvotet incvotet1 dgdp tradeshr gdpxtradeshr electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code)
- truncreg incvotet incvotet1 dgdp tradeshr gdpxtradeshr electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code) ll(0) ul(100)
- enter "C:\TRMCO_replication1\" under the Matlab environment and run "model1.m"

Replication of Hellwig and Samuels's Model II, Table 4

- use “C:\fixmin_Hellwig2.dta”, clear
- regress incvotet incvotet1 dgdp grosscap gdpxgrosscap electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code)
- truncreg incvotet incvotet1 dgdp grosscap gdpxgrosscap electype gdpxelect presrun enlp income regafrica regasia regcee reglatam, cluster(code) ll(0) ul(100)
- enter “C:\TRMCO_replication2\” under the Matlab environment and run “model2.m”

Translation of Boundary Violations, Table 5

- enter “C:\TRMCO_replication1\” under the Matlab environment.
- run “load stataresult1.mat”
- run “identify([trmb;trmsig])”
- run “load overall_result1.mat”
- run “load identify(trmcom(1:1:m+2)′)”
- enter “C:\TRMCO_replication2\” under the Matlab environment.
- run “load stataresult2.mat”
- run “identify([trmb;trmsig])”
- run “load overall_result2.mat”
- run “load identify(trmcom(1:1:m+2)′)”